The Accuracy of Fast Phylogenetic Methods for Large Datasets (PSB), 2002

Reviewed by Kajori Banerjee
Motivation

• Compares 4 methods
  1) Neighbor-joining
  2) Neighbor
  3) Greedy parsimony and
  4) DCM-NJ+MP

Parameters for comparison:
  1) Model of evolution - Jukes-Cantor and Kimura2-Parameter+Gamma
  2) Tree diameter
  3) Sequence length requirement
  4) Taxon sampling
Dataset Generation

1) Generate Model tree : model true tree using Random birth-death process

2) Make model trees non-ultrametric - the edges are multiplied with a random number from the interval $[1/c, c]$ is

3) Evolve sequences : Jukes-Cantor or Kimura2-Parameter+Gamma model.
Results

Figure 5: Accuracy as a function of the diameter under the K2P+Gamma model for fixed sequence length (500) and two numbers of taxa
Conclusion

1) **Sequence length**:
   - Weighbor - better performance for small sequence lengths
   - DCM-NJ+MP - more appropriate for data with longer sequences.
   - NJ require sequences length to be exponential with respect to the evolutionary diameter of the true tree.

2) **Speed**:
   - Both Neighbor-joining and greedy parsimony are generally faster than Weighbor and DCMNJ+MP.
   - Greedy Parsimony is very fast but low topological accuracy.

3) **Tree diameter**:
   DCM-NJ+MP has better topological accuracy then NJ as the evolutionary distance between the taxa in the dataset increases.

4) **Model of evolution**

5) **Taxon sampling**
   - DCM-NJ+MP has better topological accuracy then NJ with respect to as the number of taxa in the dataset increases.
Comments

• Have not considered boot-strapping.

• Model tree used in these experiments are free from horizontal gene transfer
References


