INTRODUCTION

Phylogenetic analyses of large and diverse data sets generally result in large sets of competing phylogenetic trees. Consensus tree methods used to summarize sets of competing trees discard important information regarding the similarity and distribution of competing trees. A more fine grain approach is to use a dimensionality reduction method to project tree-to-tree distances in 2D or 3D space [1]. In this study, we systematically evaluate the performance of several nonlinear dimensionality reduction (NLD R) methods on tree-to-tree distances obtained from independent nonparametric bootstrap analyses of genes from three mid- to large-sized mitochondrial genome alignments.

Study Goals

1. Evaluate the performance and goodness of fit of several popular NLD R methods
2. Estimate the intrinsic dimensionality of tree-to-tree distances
3. Evaluate 2D and 3D projections
4. Compare the tree projects of different mtDNA data sets

Methods

Data

A tree-to-tree distance matrix was created for the Fish, Mammal, and Salamander data set by concatenating the bootstrap trees found for gene and calculated the unweighted Robinson-Foulds (RF) distance [5].

Compare NLD R Methods

Visual Inspection

Intrinsic Dimensionality

Visual Inspection

Analytical Measures

2D Versus 3D Projections

Landscapes of mtDNA Gene Trees

Efficiency of Cost Function and Optimization

References

Acknowledgements

This work is supported in part by a grant (EF-0849861) from the National Science Foundation. Discussions with Geoffrey Fox regarding the effects of different cost functions and optimization algorithms provided some of the

Goodness of Fit Measures

TABLE 2. Phylogenetic trees were obtained for each of the three mtDNA data sets (GTR+I+G) using nonparametric bootstrap analysis (100 replicates) on each of the 15 mtDNA genes.

TABLE 2. Phylogenetic trees were obtained for each of the three mtDNA data sets (GTR+I+G) using nonparametric bootstrap analysis (100 replicates) on each of the 15 mtDNA genes.

TABLE 3. Several analytical measures of intrinsic dimensionality of the three tree-to-tree distance matrices, where NN = Nearest Neighbour estimator [8,9], COR = Continuity Dimension [10,11], ML = Maximum Likelihood estimator [12], and VIS result from figure 3 [13].

2. Setiamarga et al., 2008

3. Zhang et al., 2008


