

Computational methods in phylogenetic analysis

Tutorial at CSB 2004

Tandy Warnow

Reconstructing the “Tree” of Life



Handling large datasets:
millions of species

Phylogenetic Inference

- Hard optimization problems (e.g. MP, ML)
 - Better heuristics
 - Better approximations/lower boundsRelationship between quality of optimization criterion and topological accuracy

Phylogenetic Inference, cont.

- Bayesian inference
- Whole Genome Rearrangements
- Reticulate evolution
- Processing sets of trees: compact representations and consensus methods
- Supertree methods
- Statistical issues with respect to stochastic models of evolution (e.g., “fast converging methods”)
- Multiple sequence alignment

Major challenge: MP and ML

- Maximum Parsimony (MP) and Maximum Likelihood (ML) remain the methods of choice for most systematists
- The main challenge here is to make it possible to obtain good solutions to MP or ML in reasonable time periods on large datasets

Outline

- Part I (Basics): 40 minutes
- Part II (Models of evolution): 20 min.
- Part III (Distance-based methods): 30 min.
- Part IV (Maximum Parsimony): 30 min.
- Part V (Maximum Likelihood): 15 minutes
- Part VI (Open problems/research directions): 30 minutes