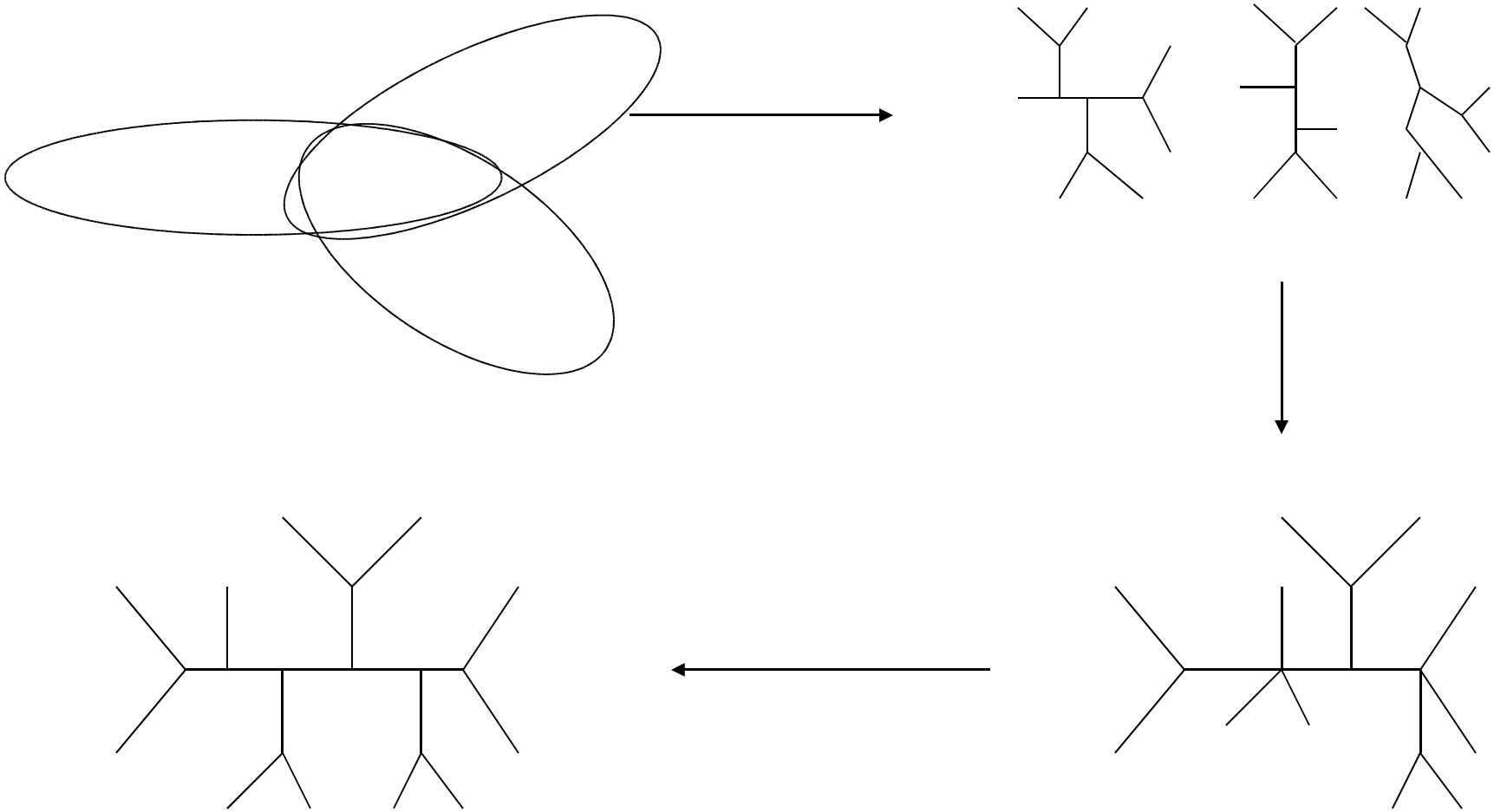
Algorithms group

- Tandy Warnow, coordinator
- Berkeley: Dick Karp, Elchanan Mossel, Gene Myers, Christos Papadimitriou, Satish Rao, and Stuart Russell
- UT-Austin: Warren Hunt and Tandy Warnow
- Others: John Huelsenbeck (UCSD), Bernard Moret (UNM), Luay Nakhleh (Rice University), Usman Roshan (NJIT), Li-San Wang (Penn), and Tiffani Williams (Radcliffe/Texas A&M)

Research problems in phylogenetic reconstruction

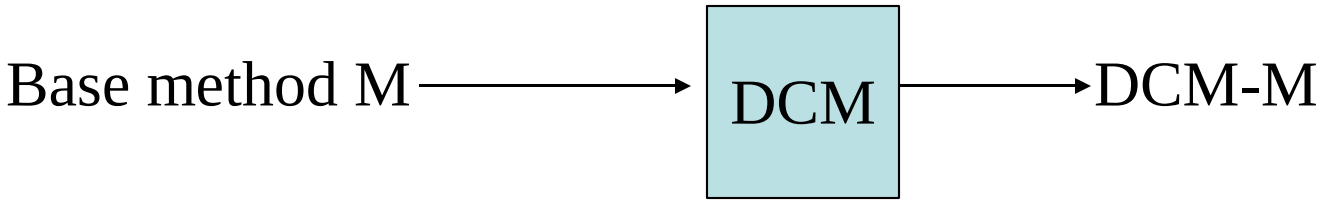
- Hard optimization problems
- Statistical issues with respect to stochastic models of evolution (e.g., “fast converging methods”)
- Bayesian inference
- Whole Genome Phylogeny (e.g., gene order/content)
- Reticulate evolution
- Gene Tree/Species Tree
- Processing sets of trees: compact representations and consensus methods
- Supertree methods

DCMs: Divide-and-conquer for improving phylogeny reconstruction



“Boosting” phylogeny reconstruction methods

- DCMs “boost” the performance of phylogeny reconstruction methods.

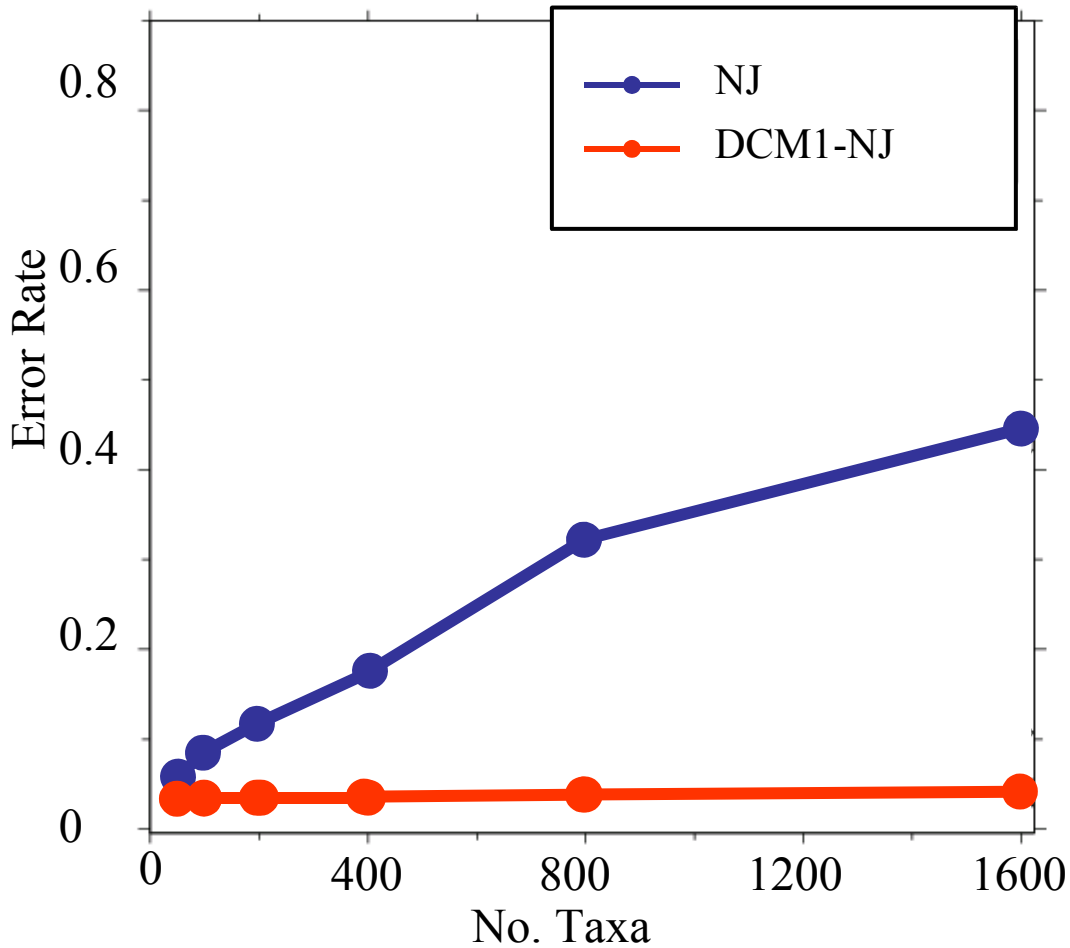


DCMs (Disk-Covering Methods)

- DCMs for polynomial time methods *improve topological accuracy* (empirical observation), and have provable theoretical guarantees under Markov models of evolution
- DCMs for hard optimization problems *reduce running time* needed to achieve good levels of accuracy (empirically observation)

DCM1-boosting distance-based methods

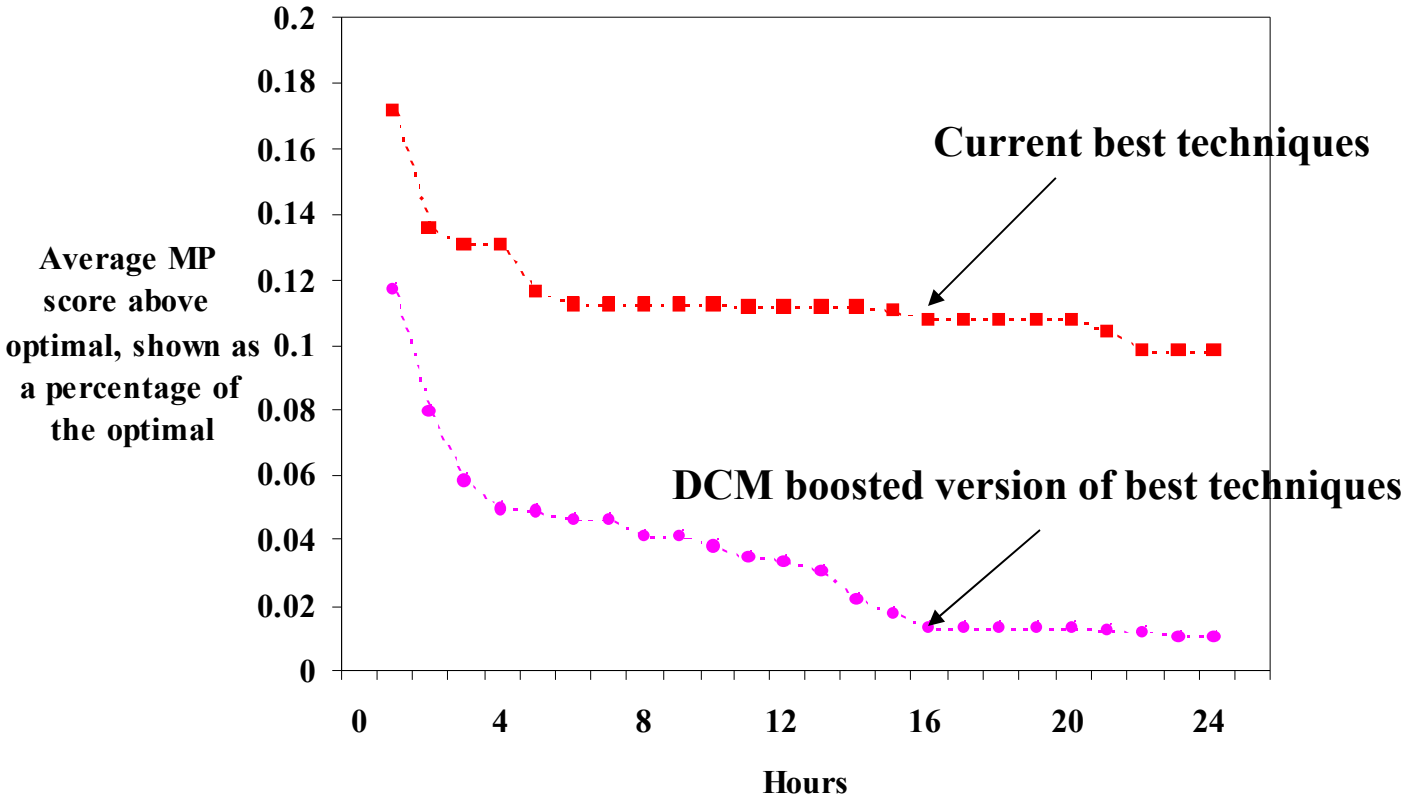
[Nakhleh et al. ISMB 2001]



DCM1-boosting makes distance-based methods more accurate

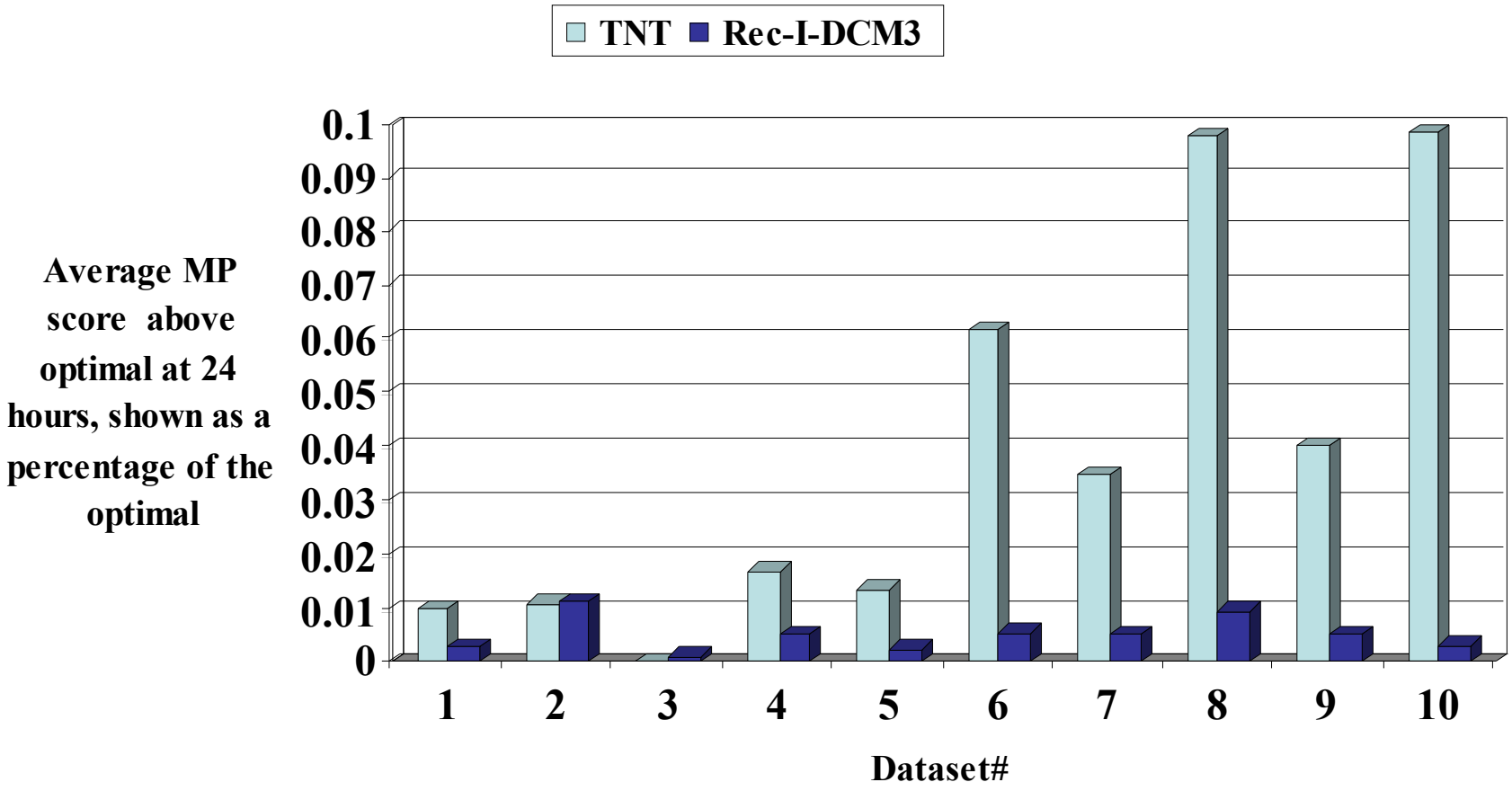
Theoretical guarantees that DCM1-NJ converges to the true tree from polynomial length sequences

Rec-I-DCM3 significantly improves performance



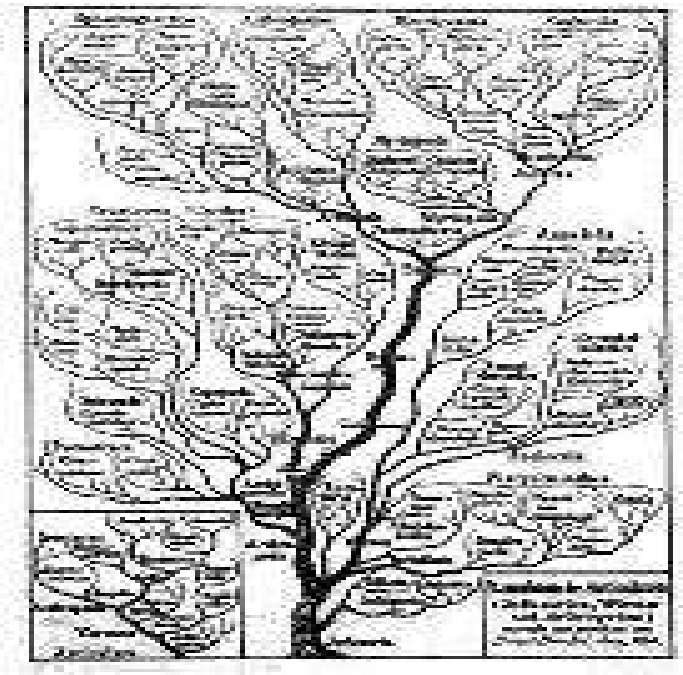
Comparison of TNT to Rec-I-DCM3(TNT) on one large dataset

Rec-I-DCM3(TNT) vs. TNT (Comparison of scores at 24 hours)



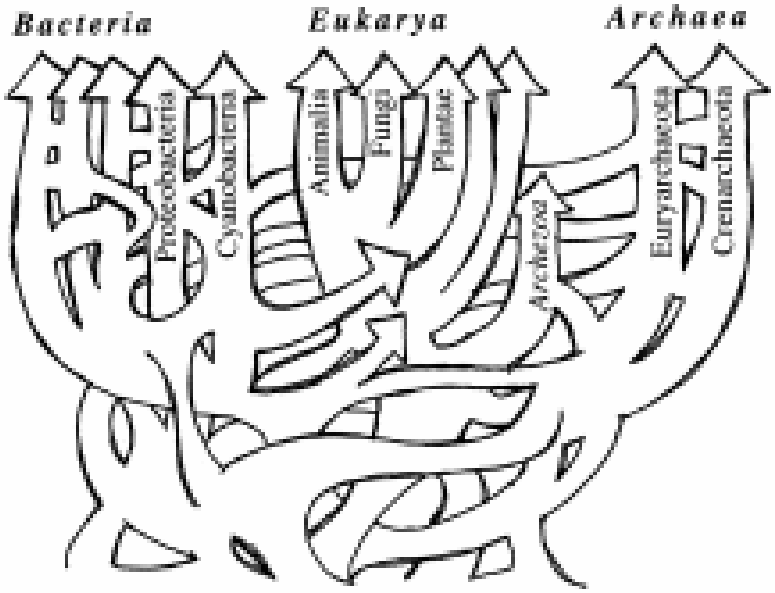
Base method is the default TNT technique, the current best method for MP. Rec-I-DCM3 significantly improves upon the unboosted TNT by returning trees which are at most 0.01% above optimal on most datasets.

Reconstructing the “Tree” of Life



Handling large datasets:

millions of species



The “Tree of Life” is not really a tree:

reticulate evolution