

# PartFastTree: Constructing Large Phylogenetic Trees & Estimating Their Reliability

Zhen Meng, Yanping Gao, Yanfei Hou, Jianhui Li, Yunchun Zhou  
Scientific Data Center, Computer Network Information Center, Chinese Academy of Sciences  
4 Zhongguancun Nansijie, Haidian District, Beijing 100190, China  
{zhenm99, gaoyanping, afeisafei, lijh, zyc}@cnic.cn

**Abstract**— Inferring phylogenies from alignments of thousands of sequences is becoming a known computational problem as DNA sequencing accelerates and gene families are growing rapidly. We present a method named PartFastTree to construct large phylogenetic trees and estimate their reliability. It is improved from FastTree, an approximate Maximum-Likelihood method for constructing phylogenetic trees. Instead of using improved Neighbor-Joining method, PartFastTree adopts PartTree method in the phase of constructing an initial tree. It reduces the memory required from  $O(nsa+n1.25)$  to  $O(ns)$  and at the same time reduces the computation time from  $O(n1.25sa)$  to  $O(n\log(n)s)$ , where  $n$  is the number of sequences,  $s$  is the width of the alignment, and  $a$  is the size of the alphabet. PartFastTree and FastTree are implemented and the evaluation on them is also presented, while PartFastTree is faster than FastTree with a little reduced accuracy when running on the datasets of from 250 to 237,882 sequences.

**Keywords**— *Large Phylogenetic Tree; PartFastTree; Time Complexity; Space Complexity*