

NJML+P: A Hybrid Algorithm of the Maximum Likelihood and Neighbor-Joining Methods Using Parallel Computing

Satoshi Oota^{1,2} Naruya Saitou^{1,2}
soota@lab.nig.ac.jp nsaitou@genes.nig.ac.jp

¹ Japan Science and Technology Corporation

² National Institute of genetics, Mishima 411-8540, Japan

Keywords: NJML method, phylogenetic tree, NJ method, ML method, parallel computing

1 Introduction

The NJML method [2, 3] is a hybrid algorithm of the two well-known methods to reconstruct molecular phylogenetic trees: the neighbor-joining (NJ) method [4] and the maximum likelihood (ML) method [1]. The NJML method is considerably efficient both in reliability and speed comparing with the other existing ML-based methods. By giving appropriate parameters, the NJML method gradually approaches the exhaustive topology search, that is supposed to be the most accurate way to find the true phylogenetic tree. However, it is obvious that more exhaustive searches require more computational time. We have implemented the NJML method by using parallel computing to reduce the computational time, in which the bootstrap trials and the maximum likelihood estimation were parallelized.

2 Method and Results

The NJML method can be divided into two parts: (1) reconstruction of an initial bootstrap NJ tree; (2) search of the topology space only around unreliable parts of the initial NJ tree in terms of bootstrap values. In each part, NJML+ program [3] was parallelized as shown in Fig. 1. We measured efficiencies of the parallelized program (NJML+P) under various conditions (see Table 1 and 2). A parallel computer (PRIME POWER 2000, SPARC64GP 563MHz) of National Institute of Genetics was used for the experiments.

Table 1: Efficiencies of parallelism for bootstrap computation.

	Sequential (sec.)	16 CPU (sec.)	Speed up
ClustalW	13.7	1.4	~9
Modified Phylip	~9,000	~610	~15

Note. The number of bootstrap trials was 2,000 for each run.

Table 2: Efficiencies of parallelism for the maximum likelihood estimation.

# of CPUs	Sequential	1	2	4	8	12	16
Time (sec.)	66.0	67.3	34.2	17.6	9.5	7.4	7.2
Speed up	-	~ 1	~ 2	~ 4	~ 6	~ 10	~ 10

Note. The number of topologies was 105 for each run. Amino acid sequences were used. “Sequential” means results by a program without parallel computing algorithm.

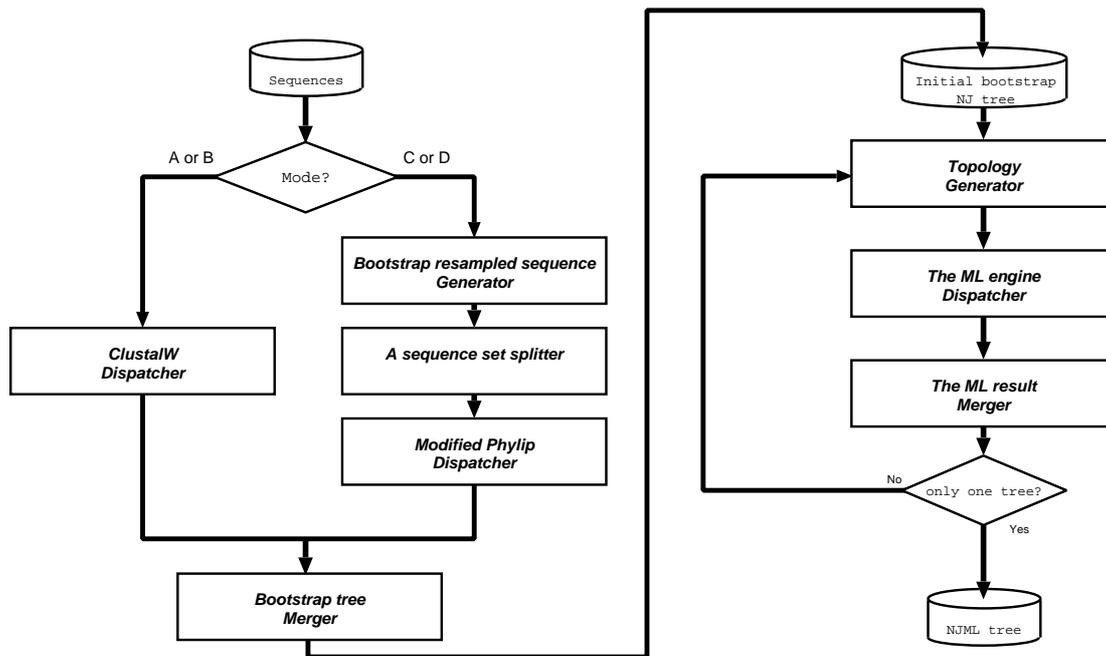


Figure 1: Parallel implementation of NJML+P.

3 Discussion

The NJML method with parallel computing is significantly useful to perform extensive bootstrap resampling and topology search.

References

- [1] Felsenstein, J., Evolutionary trees from DNA sequences: A maximum likelihood approach, *J. Mol. Evol.*, 17(6):368–376, 1981.
- [2] Ota, S. and W-H. Li, NJML: A hybrid algorithm for the neighbor-joining and maximum likelihood methods, *Mol. Biol. Evol.*, 17(9):1401–1409, 2000.
- [3] Ota, S. and W-H. Li, NJML+: An extension of the NJML method to handle protein sequence data and computer software implementation, *Mol. Biol. Evol.*, 18(11):1983–1992, 2001.
- [4] Saitou, N. and Nei, M., The neighbor-joining method: A new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.*, 4(4):406–425, 1987.