Ant Algorithm for Construction of Evlutionary Tree

Shin Ando, Hitoshi Iba
School of Engineering, University of Tokyo, Tokyo, Japan

1 Introduction

This paper proposes an implemetation of ant algorithms for constructing evolutionary tree.

An evolutionary tree analyzes the nucleotide and amino acid sequencesis to infer phylogenetic relationships and evolutionary hypotheses. It is an unrooted binary tree. The sequence of the leaves are not given on the internal nodes. An evolutionary tree with n leaves has $n - 2$ internal nodes and $m = 2n - 3$ edges. The number of possible trees with n leaves is $\prod_{i-3}^{n}(2i - 5)$. Most version of evolutionary trees problems are NP complete and exploration of such tree structure-space is both computationally demanding and time-consuming.

2 Suffix Representation

An evolutionary tree is represented in following suffix representation:
$Tx0\ Tx1\ Tx2 + Tx3 + +Tx4+$.
An ant must visit N leaves and internal vertices N-1 times in between. The vertices are not labeled, indicated as $+$. The constraints on suffix representation uses the idea of Stack Count .
We assign leaves with the Stack Value of +1 and internal vertices with -1. The Stack Count of a tree is the sum of the Stack Value of all its vertices. The Stack Count of any complete tree and subtree is 1. While adding up the vertices of a valid representation in sequential order, Stack Count never subceed 1. The ants are prohibited from constructing incomplete tree by constraints of Stack Count. The standard probability for choosing next city, pheromone update is used . A specific rule is applied for the probability that the ant will bind the branches at the vertex to form a larger branch.

D	FITCH	Neighbor	Ant
-3	85%	78%	82%
3-6	60%	33%	78%
6-	35%	12%	64%

Table 1: 8-leaf simulation comparison

3 Scoring a Tree

The circular tour length, $C(S)$, along the tree edges is calculated from distance matrix $[\delta_{ij}]$ without an explicit knowledge of the correct evolutionary tree. $C(S)$ is used for evaluating tree structure that the ants have found. The details of conventional methods for determining tree structure is found in .

4 Simulated Experiment

Based on a specific evolutionary tree, random sequences of nucleotide are generated by Seq-Gen . The parameters are, maximum distance in the matrix and sequence length. The 8-leaf and 16-leaf trees were simulated.The inputs were fed to FITCH and Neighbor algorithm of the PHYLIP programs for comparison.

Results of the simulation are summarized in Table1. In the simulated experiment, the algorithms showed comparable results against existing software.

4.1 Multiple Alignment Examples

The method was tested on 15 species of Cytochrome P450 CYP050A (http://cpd.ibmh.msk.su/). The comparison is made to FITCH and Neighbor programs. The Ant algorithms gave the best score and created a very feasible structure.