MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **63** (2010) 483-492

ISSN 0340 - 6253

A Method for Constructing Phylogenetic Tree Based on a Dissimilarity Matrix

Wen Zhu, Bo Liao *, Renfa Li

School of Computer and Communication
Hunan University, Changsha Hunan, 410082,China

(Received July 15, 2009)

Abstract. A method for constructing phylogenetic tree based on a dissimilarity matrix is proposed. In the dissimilarity matrix, the smaller element is, the more similar are the species. We translate the dissimilarity matrix to a similarity matrix based on the proposed rules, which is reflective and symmetric. The transitive closure of the obtained similarity matrix is used to construct a phylogenetic tree.

1 Introduction

A phylogenetic tree, is a model of the evolutionary history for a set of species. With more and more DNA and protein sequences have been obtained[1-3] , the problem of inferring the evolutionary history and constructing the phylogenetic tree has become one of the major problems in computational biology. This is because the evolutionary relationship of species provides a great deal of information about their biochemical machinery.

Phylogenetic analysis using biological sequences can be divided into two groups. The algorithms in the first group calculate a matrix representing the distance between each pair of sequences and then transform this matrix into a tree. In the second type of approaches, instead of building a tree, the tree that can best explain the observed

*Corresponding author E-mail: dragonbw@163.com; Fax:+86-731-8821715

sequences under the evolutionary assumption is found by evaluating the fitness of different topologies. For example, Jukes and Cantor[4], Kimura[5], Barry and Hartigan[6], Kishino and Hasegawa[7], Lake[8] proposed various distance measures. Camin and Sokal[9], Eck and Dayhoff[10], Cavalli-Sforza and Edwards[11], and Fitch gave parsimony methods[12]. Felsenstein et al[13-15] proposed maximum likelihood methods.

But, all of these methods require a multiple alignment of the sequences and assume some sort of an evolutionary model. In addition to problems in multiple alignment (computational complexity and inherent ambiguity of the alignment cost criteria), these methods become insufficient for phylogenies using complete genomes. Multiple alignment become misleading due to gene rearrangement, inversion, transposition and translocation at the substring level, unequal length of sequences, etc, and statistical evolutionary models are yet to be suggested for complete genomes. On the other hand, whole genome-based phylogenic analysis are appearing because single gene sequences generally do not possess enough information to construct an evolutionary history of organisms. Factors such as different rates of evolution and horizontal gene transfer make phylogenetic analysis of species using single gene sequences difficult. Recently, Liao proposed a graphical method to construct a phylogenetic tree, which avoids multiple alignment[16,18,19].

Here, a new method for constructing phylogenetic tree based on a dissimilarity matrix is proposed. In the dissimilarity matrix, the smaller element is, the more similar are the species. In our method, we will translate the dissimilarity matrix to a similarity matrix based on the proposed rules, which is reflective and symmetric. The transitive closure of the obtained similarity matrix is used to construct a phylogenetic tree.

2 Construction of phylogenetic tree

Recently, some graphical representations are used to make similarity analysis of gene sequence. Many similarity or dissimilarity matrices are obtained[16-26]. For example, Liao obtained a dissimilarity matrix based on a 3D representation of DNA sequences, which is listed on table 1 [17].

Table 1: The dissimilarity matrix(1.0e+004) for the coding sequences

Species	Chi	Gor	Hyl	L. cat	M. fas	M. fus	M. mul	M. syl	Pon	S. sci	T. syr
Chi	0	0.0107	0.0725	0.2649	0.0827	0.1254	0.1155	0.1811	0.0537	0.2802	0.3299
Gor		0	0.0618	0.2542	0.0720	0.1147	0.1048	0.1704	0.0430	0.2695	0.3192
Hyl			0	0.1924	0.0102	0.0529	0.0430	0.1086	0.0188	0.2077	0.2574
L. cat				0	0.1822	0.1395	0.1494	0.0838	0.2112	0.0153	0.0650
M. fas					0	0.0427	0.0328	0.0984	0.0290	0.1975	0.2472
M. fus						0	0.0099	0.0557	0.0717	0.1548	0.2045
M. mul							0	0.0656	0.0618	0.1647	0.2144
M. syl								0	0.1274	0.0991	0.1488
Pon									0	0.2265	0.2762
S. sci										0	0.0497
T. syr											0

The elements of the similarity matrix indicate the similarity relation of the corresponding species. So the elements can be directly used to construct. While, the elements of the dissimilarity matrix indicate the difference of the corresponding species. So the dissimilarity matrix should be translated to a similarity matrix.

Algorithm.(Algorithm for constructing phylogenetic trees based on the dissimilarity matrix)

Input: the dissimilarity matrix $D = (d_{ij})$ for the coding sequences of some species $\{x_i\}$, for $i = 1, 2, \cdots, n$.

Output: a phylogenetic tree related to these species.

1. Get the similarity matrix R from the input using the following two steps:

Step one: Translating the dissimilarity matrix $D = (d_{ij})$ to a matrix $D' = (d'_{ij})$,where $d'_{ij} = 1$ if $i = j$, whereas for $i \neq j$, $d'_{ij} = 1-(1.0e-004)\times \sum\limits_{k=1,k\neq i,k\neq j}^{n} [\frac{2(d_{ik}\wedge d_{jk})}{(d_{ik}+d_{jk})} log(\frac{2(d_{ik}\vee d_{jk})}{(d_{ik}+d_{jk})}) + \frac{2(d_{ik}\vee d_{jk})}{(d_{ik}+d_{jk})} log(\frac{2(d_{ik}\wedge d_{jk})}{(d_{ik}+d_{jk})})]$

Step two: Translating the matrix $D' = (d'_{ij})$ to similarity matrix $R = (r_{ij})$,where

$$r_{ij} = \frac{\sum_{t=1}^{n} min\{d'_{it}, d'_{jt}\}}{\frac{1}{2}\sum_{t=1}^{n}[d'_{it} + d'_{jt}]}$$

2. Obviously, R is reflective and symmetric, we can get the transitive closure $\bar{R} = R^k$ of it using square method. And $\bar{R}$ is a similarity relation. That is to say, we will obtain the transitive closure $\bar{R}$ by computing $R^2, R^4, R^8, ..., R^{2^k} = R^{2^{k+1}}$.

3. Choose an appropriate sequence $\{\alpha_k\}$, and get the nested sequence of partitions

$$\pi_{\alpha_1}, \pi_{\alpha_2}, \cdots, \pi_{\alpha_k},$$

from which, we construct the phylogenetic tree.

From Table 1, we can get the corresponding matrix D',relation matrix R,transitive closure $\bar{R}$.

$$D' = \begin{pmatrix}
1 & 1.0894 & 7.864 & 14.544 & 8.2142 & 9.518 & 9.5515 & 8.4122 & 4.9737 & 15.272 & 13.553 \\
1.0894 & 1 & 7.3057 & 15.261 & 7.6448 & 9.2815 & 9.2232 & 8.6767 & 9.7197 & 16.038 & 14.5 \\
7.864 & 7.3057 & 1 & 17.508 & 1.2712 & 5.9428 & 5.353 & 8.7072 & 2.1292 & 18.561 & 17.93 \\
14.544 & 15.261 & 17.508 & 1 & 16.852 & 13.037 & 14.249 & 5.6014 & 15.899 & 1.1079 & 2.6605 \\
8.2142 & 7.6448 & 1.2712 & 16.852 & 1 & 4.9125 & 4.5613 & 7.9108 & 2.1872 & 17.925 & 17.396 \\
9.518 & 9.2815 & 5.9428 & 13.037 & 4.9125 & 1 & 1.1322 & 4.9274 & 5.8729 & 14.048 & 13.678 \\
9.5515 & 9.2232 & 5.353 & 14.249 & 4.5613 & 1.1322 & 1 & 5.6221 & 5.3746 & 15.306 & 14.954 \\
8.4122 & 8.6767 & 8.7072 & 5.6014 & 7.9108 & 4.9274 & 5.6221 & 1 & 7.6244 & 6.1433 & 5.5553 \\
4.9737 & 4.7197 & 2.1291 & 15.899 & 2.1872 & 5.8729 & 5.3746 & 7.6244 & 1 & 16.89 & 16.028 \\
15.272 & 16.038 & 18.561 & 1.1079 & 17.925 & 14.048 & 15.306 & 6.1433 & 16.89 & 1 & 2.7535 \\
13.553 & 14.5 & 17.93 & 2.6605 & 17.396 & 13.678 & 14.954 & 5.5553 & 16.028 & 2.7535 & 1
\end{pmatrix}$$

$$R = \begin{pmatrix}
1 & 0.97451 & 0.74186 & 0.49667 & 0.73213 & 0.73995 & 0.75202 & 0.62963 & 0.78841 & 0.48588 & 0.50633 \\
0.97451 & 1 & 0.76746 & 0.47662 & 0.75543 & 0.73544 & 0.75699 & 0.61374 & 0.81241 & 0.46654 & 0.48646 \\
0.74186 & 0.76746 & 1 & 0.3903 & 0.96874 & 0.76184 & 0.7986 & 0.58808 & 0.91486 & 0.38305 & 0.40111 \\
0.49667 & 0.47662 & 0.3903 & 1 & 0.3869 & 0.47101 & 0.45639 & 0.61358 & 0.36548 & 0.96898 & 0.96124 \\
0.73213 & 0.75543 & 0.96874 & 0.3869 & 1 & 0.78673 & 0.82384 & 0.60093 & 0.91416 & 0.37964 & 0.39794 \\
0.73995 & 0.73544 & 0.76184 & 0.47101 & 0.78673 & 1 & 0.96325 & 0.70668 & 0.7568 & 0.45537 & 0.48192 \\
0.75202 & 0.75699 & 0.7986 & 0.45639 & 0.82384 & 0.96325 & 1 & 0.67485 & 0.79601 & 0.44168 & 0.46686 \\
0.62963 & 0.61374 & 0.58808 & 0.61358 & 0.60093 & 0.70668 & 0.67485 & 1 & 0.57054 & 0.59154 & 0.62354 \\
0.78841 & 0.81241 & 0.91486 & 0.36548 & 0.91416 & 0.7568 & 0.79601 & 0.57054 & 1 & 0.35873 & 0.37716 \\
0.48588 & 0.46654 & 0.38305 & 0.96898 & 0.37964 & 0.45537 & 0.44168 & 0.59154 & 0.35873 & 1 & 0.95246 \\
0.50633 & 0.48646 & 0.40111 & 0.96124 & 0.39794 & 0.48192 & 0.46686 & 0.62354 & 0.37716 & 0.95246 & 1
\end{pmatrix}$$

$$R^2 = \begin{pmatrix}
1 & 0.97451 & 0.78841 & 0.61358 & 0.78841 & 0.7568 & 0.78841 & 0.70668 & 0.81241 & 0.59154 & 0.62354 \\
0.97451 & 1 & 0.81241 & 0.61358 & 0.81241 & 0.76184 & 0.79601 & 0.70668 & 0.81241 & 0.59154 & 0.61374 \\
0.78841 & 0.81241 & 1 & 0.58808 & 0.96874 & 0.7986 & 0.82384 & 0.70668 & 0.91486 & 0.58808 & 0.58808 \\
0.61358 & 0.61358 & 0.58808 & 1 & 0.60093 & 0.61358 & 0.61358 & 0.62354 & 0.57054 & 0.96898 & 0.96124 \\
0.78841 & 0.81241 & 0.96874 & 0.60093 & 1 & 0.82384 & 0.82384 & 0.70668 & 0.91486 & 0.59154 & 0.60093 \\
0.7568 & 0.76184 & 0.7986 & 0.61358 & 0.82384 & 1 & 0.96325 & 0.70668 & 0.79601 & 0.59154 & 0.62354 \\
0.78841 & 0.79601 & 0.82384 & 0.61358 & 0.82384 & 0.96325 & 1 & 0.70668 & 0.82384 & 0.59154 & 0.62354 \\
0.70668 & 0.70668 & 0.70668 & 0.62354 & 0.70668 & 0.70668 & 0.70668 & 1 & 0.70668 & 0.62354 & 0.62354 \\
0.81241 & 0.81241 & 0.91486 & 0.57054 & 0.91486 & 0.79601 & 0.82384 & 0.70668 & 1 & 0.57054 & 0.57054 \\
0.59154 & 0.59154 & 0.58808 & 0.96898 & 0.59154 & 0.59154 & 0.59154 & 0.62354 & 0.57054 & 1 & 0.96124 \\
0.62354 & 0.61374 & 0.58808 & 0.96124 & 0.60093 & 0.62354 & 0.62354 & 0.62354 & 0.57054 & 0.96124 & 1
\end{pmatrix}$$

$$R^4 = \begin{pmatrix}
1 & 0.97451 & 0.81241 & 0.62354 & 0.81241 & 0.79601 & 0.81241 & 0.70668 & 0.81241 & 0.62354 & 0.62354 \\
0.97451 & 1 & 0.81241 & 0.62354 & 0.81241 & 0.81241 & 0.81241 & 0.70668 & 0.81241 & 0.62354 & 0.62354 \\
0.81241 & 0.81241 & 1 & 0.62354 & 0.96874 & 0.82384 & 0.82384 & 0.70668 & 0.91486 & 0.62354 & 0.62354 \\
0.62354 & 0.62354 & 0.62354 & 1 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.96898 & 0.96124 \\
0.81241 & 0.81241 & 0.96874 & 0.62354 & 1 & 0.82384 & 0.82384 & 0.70668 & 0.91486 & 0.62354 & 0.62354 \\
0.79601 & 0.81241 & 0.82384 & 0.62354 & 0.82384 & 1 & 0.96325 & 0.70668 & 0.82384 & 0.62354 & 0.62354 \\
0.81241 & 0.81241 & 0.82384 & 0.62354 & 0.82384 & 0.96325 & 1 & 0.70668 & 0.82384 & 0.62354 & 0.62354 \\
0.70668 & 0.70668 & 0.70668 & 0.62354 & 0.70668 & 0.70668 & 0.70668 & 1 & 0.70668 & 0.62354 & 0.62354 \\
0.81241 & 0.81241 & 0.91486 & 0.62354 & 0.91486 & 0.82384 & 0.82384 & 0.70668 & 1 & 0.62354 & 0.62354 \\
0.62354 & 0.62354 & 0.62354 & 0.96898 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 1 & 0.96124 \\
0.62354 & 0.62354 & 0.62354 & 0.96124 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.96124 & 1
\end{pmatrix}$$

$$R^8 = \begin{pmatrix}
1 & 0.97451 & 0.81241 & 0.62354 & 0.81241 & 0.81241 & 0.81241 & 0.70668 & 0.81241 & 0.62354 & 0.62354 \\
0.97451 & 1 & 0.81241 & 0.62354 & 0.81241 & 0.81241 & 0.81241 & 0.70668 & 0.81241 & 0.62354 & 0.62354 \\
0.81241 & 0.81241 & 1 & 0.62354 & 0.96874 & 0.82384 & 0.82384 & 0.70668 & 0.91486 & 0.62354 & 0.62354 \\
0.62354 & 0.62354 & 0.62354 & 1 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.96898 & 0.96124 \\
0.81241 & 0.81241 & 0.96874 & 0.62354 & 1 & 0.82384 & 0.82384 & 0.70668 & 0.91486 & 0.62354 & 0.62354 \\
0.81241 & 0.81241 & 0.82384 & 0.62354 & 0.82384 & 1 & 0.96325 & 0.70668 & 0.82384 & 0.62354 & 0.62354 \\
0.81241 & 0.81241 & 0.82384 & 0.62354 & 0.82384 & 0.96325 & 1 & 0.70668 & 0.82384 & 0.62354 & 0.62354 \\
0.70668 & 0.70668 & 0.70668 & 0.62354 & 0.70668 & 0.70668 & 0.70668 & 1 & 0.70668 & 0.62354 & 0.62354 \\
0.81241 & 0.81241 & 0.91486 & 0.62354 & 0.91486 & 0.82384 & 0.82384 & 0.70668 & 1 & 0.62354 & 0.62354 \\
0.62354 & 0.62354 & 0.62354 & 0.96898 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 1 & 0.96124 \\
0.62354 & 0.62354 & 0.62354 & 0.96124 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.96124 & 1
\end{pmatrix}$$

$$R^{16}=\begin{pmatrix}
1 & 0.97451 & 0.81241 & 0.62354 & 0.81241 & 0.81241 & 0.81241 & 0.70668 & 0.81241 & 0.62354 & 0.62354\\
0.97451 & 1 & 0.81241 & 0.62354 & 0.81241 & 0.81241 & 0.81241 & 0.70668 & 0.81241 & 0.62354 & 0.62354\\
0.81241 & 0.81241 & 1 & 0.62354 & 0.96874 & 0.82384 & 0.82384 & 0.70668 & 0.91486 & 0.62354 & 0.62354\\
0.62354 & 0.62354 & 0.62354 & 1 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.96898 & 0.96124\\
0.81241 & 0.81241 & 0.96874 & 0.62354 & 1 & 0.82384 & 0.82384 & 0.70668 & 0.91486 & 0.62354 & 0.62354\\
0.81241 & 0.81241 & 0.82384 & 0.62354 & 0.82384 & 1 & 0.96325 & 0.70668 & 0.82384 & 0.62354 & 0.62354\\
0.81241 & 0.81241 & 0.82384 & 0.62354 & 0.82384 & 0.96325 & 1 & 0.70668 & 0.82384 & 0.62354 & 0.62354\\
0.70668 & 0.70668 & 0.70668 & 0.62354 & 0.70668 & 0.70668 & 0.70668 & 1 & 0.70668 & 0.62354 & 0.62354\\
0.81241 & 0.81241 & 0.91486 & 0.62354 & 0.91486 & 0.82384 & 0.82384 & 0.70668 & 1 & 0.62354 & 0.62354\\
0.62354 & 0.62354 & 0.62354 & 0.96898 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 1 & 0.96124\\
0.62354 & 0.62354 & 0.62354 & 0.96124 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.96124 & 1
\end{pmatrix}$$

Because $R^{16}=R^{8}$, so we can obtain the transitive closure $\overline{R}$

$$\bar{R}=\begin{pmatrix}
1 & 0.97451 & 0.81241 & 0.62354 & 0.81241 & 0.81241 & 0.81241 & 0.70668 & 0.81241 & 0.62354 & 0.62354\\
0.97451 & 1 & 0.81241 & 0.62354 & 0.81241 & 0.81241 & 0.81241 & 0.70668 & 0.81241 & 0.62354 & 0.62354\\
0.81241 & 0.81241 & 1 & 0.62354 & 0.96874 & 0.82384 & 0.82384 & 0.70668 & 0.91486 & 0.62354 & 0.62354\\
0.62354 & 0.62354 & 0.62354 & 1 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.96898 & 0.96124\\
0.81241 & 0.81241 & 0.96874 & 0.62354 & 1 & 0.82384 & 0.82384 & 0.70668 & 0.91486 & 0.62354 & 0.62354\\
0.81241 & 0.81241 & 0.82384 & 0.62354 & 0.82384 & 1 & 0.96325 & 0.70668 & 0.82384 & 0.62354 & 0.62354\\
0.81241 & 0.81241 & 0.82384 & 0.62354 & 0.82384 & 0.96325 & 1 & 0.70668 & 0.82384 & 0.62354 & 0.62354\\
0.70668 & 0.70668 & 0.70668 & 0.62354 & 0.70668 & 0.70668 & 0.70668 & 1 & 0.70668 & 0.62354 & 0.62354\\
0.81241 & 0.81241 & 0.91486 & 0.62354 & 0.91486 & 0.82384 & 0.82384 & 0.70668 & 1 & 0.62354 & 0.62354\\
0.62354 & 0.62354 & 0.62354 & 0.96898 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 1 & 0.96124\\
0.62354 & 0.62354 & 0.62354 & 0.96124 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.62354 & 0.96124 & 1
\end{pmatrix}$$

Then, by choosing the sequence

$$\{0.97451, 0.96898, 0.96874, 0.96325, 0.96124, 0.91486, 0.82384, 0.81241, 0.70668, 0.62354\}$$

we can get the nested sequence of partitions

$$\{1,2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\},\{11\}$$

$$\{1,2\},\{4,10\},\{3\},\{5\},\{6\},\{7\},\{8\},\{9\},\{11\}$$

$$\{1,2\},\{4,10\},\{3,5\},\{6\},\{7\},\{8\},\{9\},\{11\}$$

$$\{1,2\},\{4,10\},\{3,5\},\{6,7\},\{8\},\{9\},\{11\}$$

$$\{1,2\},\{4,10,11\},\{3,5\},\{6,7\},\{8\},\{9\}$$

$$\{1,2\},\{4,10,11\},\{3,5,9\},\{6,7\},\{8\}$$

$$\{1,2\},\{4,10,11\},\{3,5,9,6,7\},\{8\}$$

$$\{1,2,3,5,9,6,7\},\{4,10,11\},\{8\}$$

$$\{1,2,3,5,9,6,7,8\},\{4,10,11\}$$

$$\{1,2,3,5,9,6,7,8,4,10,11\}$$

which implies the following relations among these species:

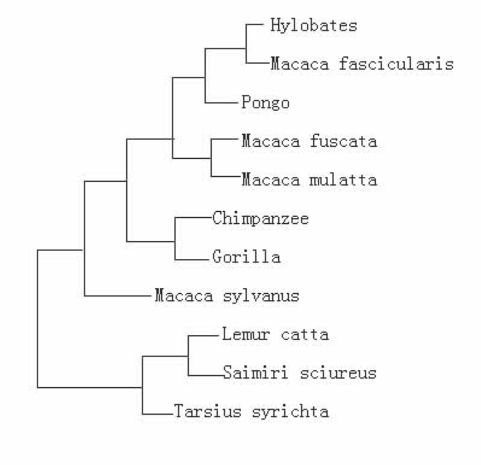

Figure 1: phylogenic tree based on the proposed algorithm.

Using the DRAWGRAM program in the PHYLIP package(http://evolution.genetics. washington.edu/ phylip.html), we can obtain the similar phylogenic tree.

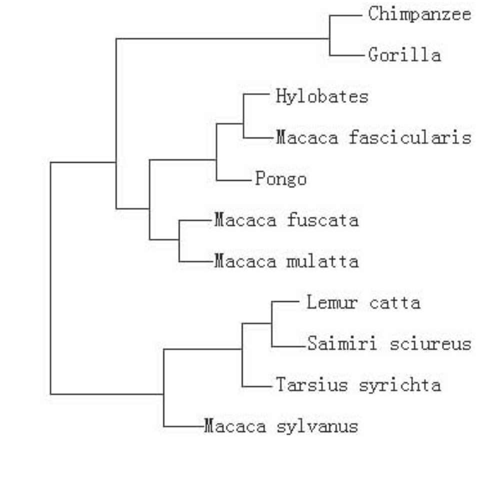

Figure 2: phylogenic tree using DRAWGRAM program.

3 Conclusion

Our algorithm provides a simple method to construct phylogenetic tree by computing the transitive closure based on a dissimilarity matrix. Translating the dissimilarity matrix to a similarity matrix based on the proposed rules is needed. Using the graphical representation of sequence, one can obtain the dissimilarity matrix without needing multiple alignment, so unlike most existing phylogeny construction methods, the proposed method does not require multiple alignment. Also, both computational scientists and molecular biologists can use it to analysis DNA sequences efficiently.

4 Acknowledgment

This work is supported in part by the National Nature Science Foundation of China (Grant 60973082) ,the National Nature Science Foundation of Hunan province (Grant 07JJ5080) and and the Planned Science and Technology Project of Hunan Province (Grant 2009FJ3195).

References

[1] T. Hodge, M. J. T. V. Cope, A myosin family tree, *J. Cell Sci.* **113** (2000) 3353–3354.

[2] N. Saitou, M. Nei, The neighbor–joining method: A new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.* **4** (1987) 406–425.

[3] T. H. Reijmers, F. D. Daeyaert, P. J. Lewi, L. M. Buydens, R. Wehrens, Using genetic algorithms for the construction of phylogenetic trees: application to G-protein coupled receptor sequences, *Biosystems* **49** (1999) 31–43.

[4] T. H. Jukes, C. R. Cantor, *Mammalian Protein Metabolism*, Academic Press, New York, 1969.

[5] M. Kimura, A simple model for estimating evolutionary rates of base substitiutions through comparative studies of nucleotide sequences, *J. Mol. Evol.* **16** (1980) 111–120.

[6] D. Barry, J. A. Hartigan, Statistical analysis of hominoid molecular evolution, *Stat. Sci.* **2** (1987) 191–210.

[7] H. Kishino, M. Hasegawa, Evolution of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoida, *J. Mol. Evol.* **29** (1989) 170–179.

[8] J. A. Lake, Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances, *Proc. Natl. Acad. Sci.* **91** (1994) 1455–1459.

[9] J. Camin, R. Sokal, A method for deducing branching sequences in phylogeny, *Evolution* **19** (1965) 311–326.

[10] R. V. Eck, M. O. Dayhoff, *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Silver Spring, MD, 1966, pp. 161–202.

[11] L. L. Cavalli-Sforza, A. W. F. Edwards, Phylogenetic analysisis: models and estimation procedures, *Evolution* **21** (1967) 550–570.

[12] W. M. Fitch, Toward defining the course of evolution: minimum change for a specific tree topology, *Syst. Zool.* **35** (1971) 406–416.

[13] J. Felsenstein, Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters, *Syst. Zool.* **22** (1973) 240–249.

[14] J. Felsenstein, Evolutionary trees from DNA sequences: a maximum likelihood approach, *J. Mol. Evol.* **17** (1981) 368–376.

[15] J. Felsenstein, G. A. Churchill, A hidden Markov model approach to variation among sites in rate of evolution, *Mol. Bio. Evol.* **13** (1996) 93–104.

[16] W. P. Wang, B. Liao, T. M. Wang, W. Zhu, A graphical method to construct a phylogenetic tree, *Int. J. Quantum Chem.* **106** (2006) 1998–2005.

[17] B. Liao, W. Zhu, Y. Liu, 3D graphical representation of DNA sequence without degeneracy and its application in constructing phylogenetic tree, *MATCH Commun. Math. Comput. Chem.* **56** (2006) 209–216.

[18] B. Liao, X. Y. Xiang, W. Zhu, Coronavirus phylogeny based on 2D graphical representation of DNA sequence, *J. Comput. Chem.* **27** (2006) 1196–1202.

[19] B. Liao, X. Z. Shan, W. Zhu, R. F. Li, Phylogenetic tree construction based on 2D graphical representation, *Chem. Phys. Lett.* **422** (2006) 282–288.

[20] Z. Cao, B. Liao, R. F. Li, A group of 3D graphical representation of DNA sequences based on dual nucleotides, *Int. J. Quantum Chem.* **108** (2008) 1485–1490.

[21] G. H. Huang, B. Liao, Y. F. Li, Z. B. Liu, H-L curve: A novel 2D graphical representation for DNA sequences, *Chem. Phys. Lett.* **462** (2008) 129–132.

[22] W. Y. Chen, B. Liao, Y. S. Liu, W. Zhu, Z. Z. Su, A numerical representation of DNA sequence and its applications, *MATCH Commun. Math. Comput. Chem.* **60** (2008) 291–300.

[23] W. Y. Chen, B. Liao, X. Y. Xiang, W. Zhu, An improved binary representation of DNA sequences and its applications, *MATCH Commun. Math. Comput. Chem.* **61** (2009) 767–780.

[24] Z. B. Liu, B. Liao, W. Zhu, G. H. Huang, A 2-D graphical representation of DNA sequence based on dual nucleotides and its application, *Int. J. Quantum Chem.* **109** (2009) 948–958.

[25] G. H. Huang, B. Liao, R. F. Li, Similarity studies of DNA sequences based on a new 2D graphical representation, *Biophy. Chem.* **143** (2009) 55–59.

[26] W. Chen, Y. S. Zhang, Three distances for rapid similarity analysis of DNA sequences, *MATCH Commun. Math. Comput. Chem.* **61** (2009) 781–788.