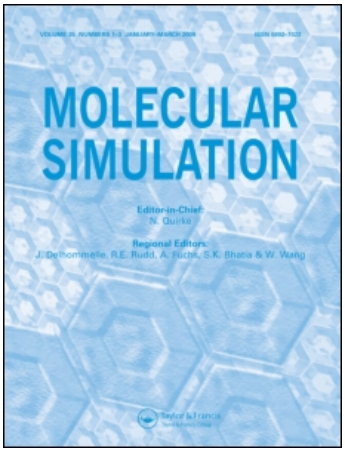

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

On 3DD-curves of DNA sequences

Y. Zhang[ab]; B. Liao[bc]; K. Ding[b]
[a] School of Information and Engineering, Shandong University at Weihai, Weihai, People's Republic of China [b] Department of Mathematics, Graduate School of the Chinese Academy of Sciences, Beijing, People's Republic of China [c] School of Computer and Communication, Hunan University, Changsha, People's Republic of China

To cite this Article Zhang, Y. , Liao, B. and Ding, K.(2006) 'On 3DD-curves of DNA sequences', Molecular Simulation, 32: 1, 29 — 34
To link to this Article: DOI: 10.1080/08927020500517223
URL: http://dx.doi.org/10.1080/08927020500517223

Molecular Simulation, Vol. 32, No. 1, 15 January 2006, 29–34

Taylor & Francis
Taylor & Francis Group

On 3DD-curves of DNA sequences

Y. ZHANG†¶*, B. LIAO‡¶ and K. DING¶

†School of Information and Engineering, Shandong University at Weihai, Weihai 264209, People's Republic of China
‡School of Computer and Communication, Hunan University, Changsha 410082, People's Republic of China
¶Department of Mathematics, Graduate School of the Chinese Academy of Sciences, Beijing 100049, People's Republic of China

(Received September 2005; in final form December 2005)

The 3DD-curves, a new 3D graphical representation of DNA sequences, can resolves degeneracy completely and is mathematically proven to eliminate circuit formation. As an application, we make a comparison for the mitochondrial sequences belonging to 11 different species based on the new 3D graphic representation.

Keywords: DNA; 3DD-curve; Graphical representation; Phylogenetic trees

1. Introduction

Graphical techniques have emerged as a very powerful tool for the visualization and analysis of long DNA sequences. These techniques provide useful insights into local and global characteristics and the occurrences, variations and repetition of the nucleotides along a sequence which are not as easily obtainable by other methods. Several authors outlined different 2D graphical representation of DNA sequences [1–6]. In 2000, Randic *et al.* [7] generalized the 2D to 3D graphical representation. They place the origin of the cartesian (x, y, z) coordinate system in the center of a cube so that the four corners of the cube have the coordinates $(+1, -1, -1)$, $(-1, +1, -1)$, $(-1, -1, +1)$, $(+1, +1, +1)$, which corresponding to four nucleic bases. The H-curve [8] is another 3D graphical representation of DNA sequences. The basic rule for constructing the H-curve is to move one unit along one of four directions (NW, NE, SE and SW) representing four bases in xy-plane, and one for each unit in the z-direction. We can see that the H-curve needs more space since its height always equals the length of the DNA sequence considered. Other 3D graphical representation can be found in [9,10,12,13,15].

One reason that given the representation is that we can get the appreciate representation which satisfy biologist's need by choosing the parameters under conditions and need not consider the non-degeneracy of it.

In this paper, we introduce a novel 3D graphical representation of DNA primary sequences which has one to one correspondence between DNA sequences and DNA graphs. The 3D graphic representation resolves degeneracy completely and is mathematically proven to eliminate circuit formation.

2. 3DD-curve

In a 3D-space, a point or a vector has three components. We will assign the following vectors to the four nucleic bases:

$$\begin{aligned}(v\sqrt{n}, 0, u) &\to A\\ (0, v, w\sqrt{n}) &\to G\\ (w, u\sqrt{n}, 0) &\to C\\ (u, w, v) &\to T\end{aligned}$$

where u, v and w are different integer numbers and n is not perfect square number. So that we can reduce a DNA sequence into a series of nodes P_0, P_1, P_2,...,P_N, whose coordinates x_i, y_i, z_i ($i = 0, 1, 2,\ldots, N$, where N is the length of the DNA sequence being studied) satisfy

$$\begin{cases} x_i = v\sqrt{n}A_i + wC_i + uT_i \\ y_i = u\sqrt{n}C_i + vG_i + wT_i \\ z_i = w\sqrt{n}G_i + uA_i + vT_i \end{cases} \quad (1)$$

where A_i, C_i, G_i, and T_i are the cumulative occurrence numbers of A, C, G and T, respectively, in the subsequence

*Corresponding author. Email: zhangys@sdu.edu.cn

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020500517223

from the first base to the i-th base in the sequence, hence, $A_i + C_i + T_i + G_i = i$. We define $A_0 = C_0 = G_0 = T_0 = 0$.

We call the corresponding plot set be characteristic plot set. The curve connecting all plots of the characteristic plot set in turn is called 3DD-curve (3D curve of DNA).

The method for constructing the 3DD-curve is different from that one for constructing of H-curve. For any DNA sequence, the H-curve always extends along the z-axis until its height reaches the length of the sequences while it expands in the xy-plane. However, in our 3DD-curve, the move in each of the three directions is equiprobable. It is not to move one unit along one of four directions representing the bases (A, C, G, T) but to move different unit for different bases. That is why our 3D graphical representation of DNA sequences is non-degeneracy. That means the 3DD-curve is a 3D space curve constituting the unique representation of a given DNA sequence in the sense that each can be uniquely reconstructed given the other. Based on the 3DD-curve, any DNA sequence can be uniquely described by three independent distributions, i.e. x_i, y_i and z_i. Therefore, the 3DD-curve contains all the information that the corresponding DNA sequence carries. A DNA sequence can be analyzed by studying the corresponding 3DD-curve. One of the advantages of the 3DD-curve is its intuitiveness; the entire 3DD-curve of a genome can be viewed on a computer screen, regardless of genome length, thus allowing both global and local compositional features of genomes to be easily grasped. By combining use of the 3DD-curve with statistical analysis, better results may be obtained.

Now we give three special cases.

Firstly, letting $(\sqrt{n}, 0, 0) \to A$, $(0, 1, -\sqrt{n}) \to G$, $(-1, 0, 0) \to C$, $(0, -1, 1) \to T$, i.e. $v = 1$, $w = -1$ and $u = 0$ in equation (1), then we get the non-degeneracy 3DD-curve corresponding to Leong and Mogenthaler's 2D graphic representation:

$$\begin{cases} x_i = \sqrt{n}A_i - C_i \\ y_i = G_i - T_i \\ z_i = T_i - \sqrt{n}G_i \end{cases} \tag{2}$$

where n is positive real number, but not perfect square number. The projection on the coordinate y-plane is exactly the non-degeneracy form corresponding to Leong and Mogenthaler's 2D graphic representation:

$$\begin{cases} x_i = \sqrt{n}A_i - C_i \\ z_i = T_i - \sqrt{n}G_i \end{cases} \tag{3}$$

We can obtain the Leong and Mogenthaler's 2D graphic representation by letting $n = 1$ in above curve.

Secondly, letting $(0, 0, -1) \to A$, $(0, 0, \sqrt{n}) \to G$, $(1, \sqrt{n}, 0) \to C$, $(-1, 1, 0) \to T$, i.e. $w = 1$, $u = -1$ and $v = 0$ in equation (1), then we get the non-degeneracy 3DD-curve corresponding to Nandy's 2D graphic representation:

$$\begin{cases} x_i = C_i - T_i \\ y_i = T_i - \sqrt{n}C_i \\ z_i = \sqrt{n}G_i - A_i \end{cases} \tag{4}$$

where n is positive real number, but not perfect square number. When $n = 1$ the projection on the coordinate y-plane is exactly the Nandy's 2D graphic representation:

$$\begin{cases} x_i = \mathrm{C}_i - T_i \\ z_i = G_i - A_i \end{cases} \tag{5}$$

finally, letting $(-\sqrt{n}, 0, 1) \to A$, $(0, -1, 0) \to G$, $(0, \sqrt{n}, 0) \to C$, $(1, 0, -1) \to T$, i.e. $u = 1$, $v = -1$ and $w = 0$ in equation (1), then we get the non-degeneracy 3DD-curve corresponding to Gates' 2D graphic representation:

$$\begin{cases} x_i = -\sqrt{n}A_i + T_i \\ y_i = \sqrt{n}C_i - G_i \\ z_i = A_i - T_i \end{cases} \tag{6}$$

where n is positive real number, but not perfect square number. The projection on the coordinate z-plane is exactly the non-degeneracy form corresponding to Gates' 2D graphic representation:

$$\begin{cases} x_i = -\sqrt{n}A_i + T_i \\ y_i = \sqrt{n}C_i - G_i \end{cases} \tag{7}$$

We can obtain the Gates' 2D graphic representation by letting $n = 1$ in above curve.

From the construction of the 3DD-curve, we can see that it can provide more information than existing 2D graphic representation by choosing the appropriate parameters and we can choose the system most appropriate to the problem at hand. Another advantage is that it is non-degeneracy comparing with the most existing 3D graphic representation. That means the 3DD-curve is determined uniquely for any different DNA sequence which will be proved in next section.

In figure 1, we present the 3DD-curve corresponding to the coding sequence of mitochondrial sequences of Chimpanzee V00672 with $u = 1$, $v = 2$, $w = -3$, $n = 3$ and $u = 5$, $v = 1 = 2$, $w = -3$, $n = 3$.

It is easy to see that different parameters can result in different visual clues to DNA sequence. So we should choose the 3DD-curve most appropriate to the problem under consideration.

3. Properties

Property 1. For a given DNA sequence there is a unique 3DD-curve corresponding to it.

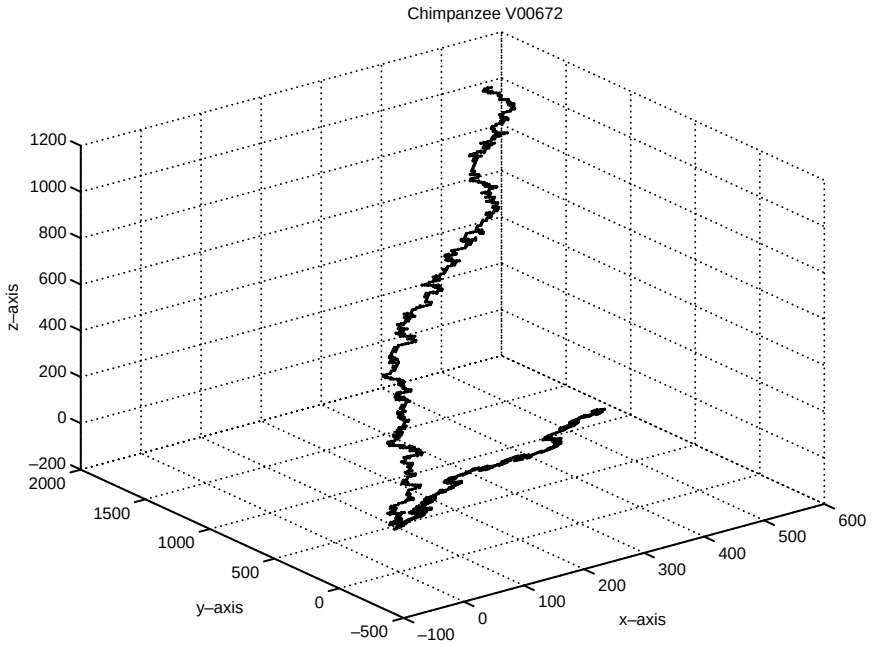

Figure 1. 3DD-curve of chimpanzee V00672.

Proof. Let (x_i, y_i, z_i) be the coordinates of the 3DD-curve corresponding to the i-th base of DNA sequence, then we have

$$\begin{cases} x_i = v\sqrt{n}A_i + wC_i + uT_i \\ y_i = u\sqrt{n}C_i + vG_i + wT_i \\ z_i = w\sqrt{n}G_i + uA_i + vT_i \end{cases}$$

If (x_i, y_i, z_i) can also be expressed as

$$\begin{cases} x_i = v\sqrt{n}A'_i + wC'_i + uT'_i \\ y_i = u\sqrt{n}C'_i + vG'_i + wT'_i \\ z_i = w\sqrt{n}G'_i + uA'_i + vT'_i \end{cases}$$

then, we have

$$\begin{cases} v\sqrt{n}(A_i - A'_i) + w(C_i - C'_i) + u(T_i - T'_i) = 0 \\ u\sqrt{n}(C_i - C'_i) + v(G_i - G'_i) + w(T_i - T'_i) = 0 \\ w\sqrt{n}(G_i - G'_i) + u(A_i - A'_i) + v(T_i - T'_i) = 0 \end{cases}$$

Considering that n is not perfect square number, it is easy to see that

$$A_i = A'_i,\ \ C_i = C'_i,\ \ T_i = T'_i,\ \ G_i = G'_i$$

So, for given x-, y- and z-projection of any point $P = (x, y, z)$ on 3DD-curve, after uniquely determining the number a_p, g_p, c_p and t_p of A, G, C and T from the beginning of the sequence to the point P. By successive x-, y- and z-projection of points on the sequence, we can recover the original DNA sequence uniquely from the DNA graph. By similar method, we can also prove the following property.

Property 2. There is no circuit or degeneracy in 3DD-curve.

The vector pointing to the point P_i from the origin O is denoted by r_i. The component of r_i, i.e. x_i and y_i are calculated by equation (1). Let $\Delta r_i = r_i - r_{i-1}$, then we have Property 3.

Property 3. For any $i = 1, 2, \ldots, N$, where N is the length of the studied DNA sequence, the vector Δr_i has at most four possible direction and the possible directions are summarized as follows (table 1).

4. Similarities/dissimilarities

In this section, we will make a comparison for 11 mitochondrial DNA sequences based on our 3DD-curve [16]. In table 2, the mitochondrial DNA sequences for 11 different species are listed.

Now we give a numerical characterization of the new representation that will facilitate quantitative comparisons of DNA sequences. One of the possibilities to achieve this aim is to characterize the curves by invariants. In order to find some of the invariants sensitive to the form of the curve, one can transform the graphical representation of the curve into another mathematical

Table 1. Four possible directions.

	Δx_i	Δy_i	Δz_i	$\lvert\Delta r_i\rvert$
A	$v\sqrt{n}$	0	u	$\sqrt{nv^2 + u^2}$
C	w	$u\sqrt{n}$	0	$\sqrt{nu^2 + w^2}$
G	0	v	$w\sqrt{n}$	$\sqrt{nw^2 + v^2}$
T	u	v	w	$\sqrt{v^2 + u^2 + w^2}$

Table 2. Database source.

Species	*ID/ACCESSION*	*Length (bp)*	*Database*
Saimiri sciureus	M22655	893	NCBI
Hylobates	V00659	896	NCBI
Lemur catta	M22657	895	NCBI
Macaca fascicular	M22653	896	NCBI
Gorilla	V00658	896	NCBI
Macaca fuscata	M22651	896	NCBI
Macaca mulatta	M22650	896	NCBI
Macaca sylvanus	M22654	896	NCBI
Chimpanzee	V00672	896	NCBI
Orangutan	V00675	895	NCBI
Tarsius syrichta	M22656	895	NCBI

object, a matrix. Once a matrix representation of a DNA sequence is given, some of matrix invariants, e.g. the leading eigenvalues, can be used as descriptors of the sequence [4]. Here, we consider the quotient matrix *E*/*P* and *E*/*G* [4]. The (*i*, *j*) element 5 of matrix *E*/*P* is defined to be the quotient of the Euclidean-distance between vertices *i* and *j* of the 3DD-curve and the sum of the distances between the same pair of vertices. In other words, $[E/P]_{ij} = [ED]_{ij}/\sum_{k=i}^{j-1}[ED]_{k,k+1}$, where $[ED]_{ij}$ is the Euclidean distance between a pair of vertices and the (*i*, *j*) element $[E/G]_{ij}$ of matrix *E*/*G* is defined to be $[ED]_{ij}/|i-j|$.

We choose the leading eigenvalues of quotient matrices *E*/*P* and *E*/*G* as descriptors of DNA sequences. Since the 3DD-curve does not represent the genuine molecular geometry we are not interested in the interpretation of the leading eigenvalues of these matrices, but are interested in them as numerical parameters that may facilitate comparisons of DNA sequences.

A direct comparison of these DNA sequences using computer codes is somewhat less straightforward due to the fact that the DNA sequence have different lengths. We will construct a 4-component vector consisting of the normalized leading eigenvalue of the quotient matrix *E*/*G* or *E*/*P* of the 3DD-curves with different parameters. For our application, we will use the following four 3DD-curves:

Case a. Letting $u = 1$, $v = 2$, $w = 3$ and $n = 3$ in equation (1), then we get 3DD-curve:

$$\begin{cases} x_i = 2\sqrt{3}A_i + 3C_i + T_i \\ y_i = \sqrt{3}\mathrm{C}_i + 2G_i + 3T_i \\ z_i = 3\sqrt{3}G_i + A_i + 2T_i \end{cases}$$

Case b. Letting $u = 3$, $v = 2$, $w = 1$ and $n = 3$ in equation (1), then we get 3DD-curve:

$$\begin{cases} x_i = 2\sqrt{3}A_i + \mathrm{C}_i + 3T_i \\ y_i = 3\sqrt{3}\mathrm{C}_i + 2G_i + T_i \\ z_i = \sqrt{3}G_i + 3A_i + 2T_i \end{cases}$$

Case c. Letting $u = 2$, $v = 3$, $w = 1$ and $n = 3$ in equation (1), then we get 3DD-curve;

$$\begin{cases} x_i = 3\sqrt{3}A_i + C_i + 2T_i \\ y_i = 2\sqrt{3}C_i + 3G_i + T_i \\ z_i = \sqrt{3}G_i + 2A_i + 3T_i \end{cases}$$

Case d. Letting $u = 3$, $v = 1$, $w = 2$ and $n = 3$ in equation (1), then we get 3DD-curve:

$$\begin{cases} x_i = \sqrt{3}A_i + 2C_i + uT_i \\ y_i = 3\sqrt{3}C_i + G_i + 2T_i \\ z_i = 2\sqrt{3}G_i + 3A_i + T_i \end{cases}$$

The analysis of similarity/dissimilarity among these DNA sequences represented by the 4-component vectors is based on the assumption that two DNA sequences are similar if the corresponding 4-component vectors point to a similar direction in the 4D-space and have similar magnitudes. The similarity between these two vectors can be measured by calculating the Euclidean distance between their end points. Clearly, the smaller is the Euclidean distance the more similar are the two DNA sequences.

In table 3, we give the similarities and dissimilarities for the mitochondrial sequences based on the Euclidean distances between the end points of the 4-component vectors of the normalized leading eigenvalues of the *E*/*G*

Table 3. The similarity/dissimilarity matrix for the mitochondrial sequences based on the Euclidean distances between the end points of the 4-component vectors of the normalized leading eigenvalues of the *E*/*G* matrices.

Species	*Chi*	*Gorilla*	*Hyl*	*Lemur*	*M. Fas*	*M. Fus*	*M. Syl*	*Ora*	*S. Sci*	*T. Syr*	*M. Mul*
Chi	0	0.0084	0.0323	0.1388	0.0523	0.0516	0.0955	0.0182	0.1495	0.1858	0.0588
Gorilla		0	0.0338	0.1443	0.0561	0.0568	0.1016	0.0181	0.1529	0.1912	0.0635
Hyl			0	0.1142	0.0246	0.0293	0.0745	0.0499	0.1194	0.1608	0.0337
Lemur				0	0.0897	0.0875	0.0449	0.1548	0.0435	0.0471	0.0811
M. Fas					0	0.0107	0.0508	0.0698	0.0972	0.1364	0.0100
M. Fus						0	0.0457	0.0684	0.0998	0.1345	0.0084
M. Syl							0	0.1108	0.0716	0.0913	0.0410
Ora								0	0.1669	0.2016	0.0758
S. Sci									0	0.0613	0.0918
T. Syr										0	0.1279
M. Mul											0

Table 4. The similarity/dissimilarity matrix for mitochondrial sequence based on Euclidean distances between the end points of the 4-component vectors of the normalized leading eigenvalues of the *E*/*P* matrices.

Species	*Chi*	*Gorilla*	*Hyl*	*Lemur*	*M. Fas*	*M. Fus*	*M. Syl*	*Ora*	*S. Sci*	*T. Syr*	*M. Mul*
Chi	0	0.0033	0.0058	0.0152	0.0049	0.0035	0.0096	0.0062	0.0168	0.0230	0.0048
Gorilla		0	0.0043	0.0176	0.0068	0.0047	0.0122	0.0039	0.0184	0.0254	0.0066
Hyl			0	0.0160	0.0059	0.0047	0.0115	0.0075	0.0155	0.0234	0.0058
Lemur				0	0.0108	0.0135	0.0062	0.0207	0.0064	0.0080	0.0111
M. Fas					0	0.0031	0.0058	0.0102	0.0121	0.0186	0.0010
M. Fus						0	0.0079	0.0074	0.0148	0.0213	0.0024
M. Syl							0	0.0150	0.0102	0.0141	0.0059
Ora								0	0.0219	0.0286	0.0097
S. Sci									0	0.0094	0.0126
T. Syr										0	0.0190
M. Mul											0

matrices. We believe that it is not accidental that the smallest entries in table 2 are associated with the pairs (*Gorilla*, *Chimpanzee*), (*Macaca fascicular*, *Macaca fuscata*), (*Macaca fascicular*, *Macaca mulatta*) and (*Macaca fuscata*, *Macaca mulatta*). On the other hand, *Lemur catta* is very dissimilar to others among the 11 species because its corresponding column has lager entries. We can obtain that *Tarsius syrichta* is dissimilar to others too.

In table 4, we give the similarities and dissimilarities for 11 coding sequences that based on Euclidean distances between the end points of the 4-component vectors of the normalized 7 leading eigenvalues of the *E*/*P* matrices. The results of similarities and dissimilarities for 11 coding sequences based on table 4 are somewhat dissimilar with that based on table 3. Similar phenomenon have presented in [14] and other papers. That means the results based on *E*/*P* matrices are somewhat different from that based on *E*/*G* matrices.

In order to avoid tolerance between the two methods and draw the dendrogram of the 11 different species, we establish the table 5 whose entry is the sum of the corresponding entry of tables $4 + 3$. Comparing tables 3–5, we can find at table 5 based on *E*/*G* matrices and *E*/*P* matrices is the best tolerance for the similarities. Randic use 12-component vector whose components are made up of the normalized leading eigenvalue [11]. In paper [14], we construct similarity/dissimilarity matrix based on different descriptors (Euclidean distances and angle). As one can see the existing methods to analyze the similarity/dissimilarity of DNA sequences are not more convenient than our approach since the alternative descriptor is a 4-component vector while the other approach used higher dimensional vector.

Hayasaka *et al.* [16] calculated the number of nucleotide substitutions for a given pair of species by the six-parameter method. Using the calculated numbers, they constructed a phylogenetic tree by the NJ method, the distance Wagner method and unweighed pair grouping method, respectively. The algorithms for constructing phylogenetic trees are different from each other. These three different methods give phylogenetic trees with the same topology. The topology of the tree, except for the position of the tarsier, is generally in agreement with the widely accepted classification of primates that is based on fossil records and other molecular analysis.

In figure 2, we constructed the phylogenetic trees for the 11 different species by the NJ method. The phylogeny obtained is generally consistent with phylogenetic trees constructed in previous studies. The phylogenetic relationships among primate groups shown by our analyses are generally consistent with results of previous studies. But, all the previous methods require a multiple alignment of the nine sequences and assume some sort of an evolutionary model. In addition to problems in multiple alignment (computational complexity and inherent

Table 5. Sum of the entry of table 4 plus corresponding the entry of table 3.

Species	*Chi*	*Gorilla*	*Hyl*	*Lemur*	*M. Fas*	*M. Fus*	*M. Syl*	*Ora*	*S. Sci*	*T. Syr*	*M. Mul*
Chi	0	0.0117	0.0381	0.1540	0.0572	0.0551	0.1051	0.0244	0.1663	0.2088	0.0636
Gorilla		0	0.0381	0.1619	0.0629	0.0615	0.1138	0.0220	0.1713	0.2166	0.0701
Hyl			0	0.1302	0.0305	0.0340	0.0860	0.0574	0.1349	0.1842	0.3950
Lemur				0	0.1005	0.1010	0.0511	0.1755	0.0499	0.0551	0.0922
M. Fas					0	0.0138	0.0566	0.0800	0.1093	0.1550	0.0110
M. Fus						0	0.0536	0.0758	0.1146	0.1558	0.0108
M. Syl							0	0.1258	0.0818	0.1054	0.0469
Ora								0	0.1888	0.2302	0.0855
S. Sci									0	0.0707	0.1044
T. Syr										0	0.1469
M. Mul											0

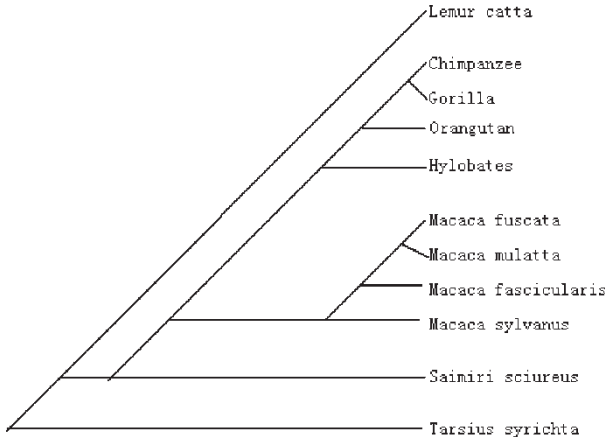

Figure 2. Phylogenetic trees of 11 mitochondrial DNA sequences based on our 3DD-curve.

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 analyses 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. Unlike most existing phylogeny construction methods, the proposed method does not require multiple alignment.

5. Conclusions

This letter introduces a non-degeneracy 3D representation of DNA sequence. Many properties of visual importance in a DNA sequence are preserved in the 3DD-curve. It is useful for visualizing the local and global features of long or short DNA sequences and can facilitate the visual discovery of interesting features in a DNA sequence. The 3DD-curve can also serve as a good annotation tool for biologist.

References

[1] M.A. Gates. A simple way to look at DNA. *J. Theor. Biol.*, **119**, 319 (1986).
[2] A. Nandy. Graphical representation of long DNA sequences. *Curr. Sci.*, **66**, 821 (1994).
[3] P.M. Leong, S. Morgenthaler. Random walk and gap plots of DNA sequences. *Comput. Appl. Biosci.*, **12**, 503 (1995).
[4] M. Randić, M. Vračko, N. Lerš, D. Plavšić. Novel 2-D graphical representation of DNA sequences and their numberical characterization. *Chem. Phys. Lett.*, **368**, 1 (2003).
[5] B. Liao. A 2D graphical representation of DNA sequence. *Chem. Phys. Lett.*, **401**, 196 (2005).
[6] Y. Wu, A.W. Liew, H. Yan, M. Yang. DB-Curve: A novel 2D method of DNA sequence visualization and reprsentation. *Chem. Phys. Lett.*, **367**, 170 (2003).
[7] M. Randić, M. Vracko, A. Nandy, S.C. Basak. On 3-D graphical representation of DNA primary sequence and their numerical characterization. *J. Chem. Inf. Comput. Sci.*, **40**, 1235 (2000).
[8] E. Hamori, J. Ruskin. H-curves, a novel method of representation of nucleotides series especially suited for long DNA sequences. *J. Biol. Chem.*, **258**, 1318 (1983).
[9] C. Yuan, B. Liao, T. Wang. New 3-D graphical representation of DNA sequences and their numerical characterization. *Chem. Phys. Lett.*, **379**, 412 (2003).
[10] C. Li, J. Wang. On a 3-D representation of DNA primary sequences. *Comb. Chem. High Throughput Screen.*, **7**, 23 (2004).
[11] M. Randić, M. Vracko, N. Lers, D. Plavšić. Analysis of similarity/dissimilarity of DNA sequences based on novel 2-D graphical representation. *Chem. Phys. Lett.*, **371**, 202 (2003).
[12] M. Randić. Condensed representation of DNA primary sequences. *J. Chem. Inf. Comput. Sci.*, **40**, 50 (2000).
[13] M. Randić, X.F. Guo, S.C. Basak. On the characterization of DNA primary sequence by triplet of nucleic acid bases. *J. Chem. Inf. Comput. Sci.*, **41**, 619 (2001).
[14] B. Liao, Y. Zhang, K. Ding, T. Wang. Analysis of similarity/dissimilarity of DNA sequences based on a condensed curve representation. *J. Mol. Struct. Theochem.*, **717**, 199 (2005).
[15] B. Liao, T. Wang. 3-D graphical representation of DNA sequences and their numerical characterization. *J. Mol. Struct. Theochem.*, **681**, 209 (2004).
[16] K. Hayasaka, T. Gojobori, S. Horai. Molecular phylogeny and evolution of primate mitochondrial DNA. *Mol. Biol. Evol.*, **5**, 626 (1988).