

On the similarity of DNA primary sequences based on 5-D representation

Bo Liao,* Renfa Li, and Wen Zhu

*Laboratory of Embedded Computing and System, School of Computer and Communication
Hunan University, Changsha Hunan 410082, China*

E-mail: dragonbw@163.com

Xuyu Xiang

Department of Computer Science, Hunan City University, Yiyang 413000, China

Received 9 December 2005; revised 8 February 2006

We propose a 5-D representation of DNA sequences based on the classifications of the four basic acid bases *A*, *T*, *G*, and *C*. The use of the 5-D representation is illustrated by constructing sequence invariants based on the entries in derived sequence matrices restricted to a selected width of a band along the main diagonal. As an application, we make a comparison for the first exon of β -globin gene sequences belonging to eleven different species based on the 5-D representation.

KEY WORDS: DNA, similarity, graphical representation, 5-D representation

1. Introduction

One important task in the study of genome sequences is to determine densities of specific nucleotides and to understand the implications for exons, or coding regions. Several methods for addressing this problem graphically have been advanced [1]. In recent years, graphical representations of DNA were introduced to facilitate comparison of DNA sequences and observing differences in their structure [2–16]. The advantage of graphical representation of DNA sequences is that they allow visual inspection of data, helping in recognizing major differences among similar DNA sequences. Several authors outlined different graphical representation of DNA sequences based on 2-D, 3-D or 4-D [2,3,5–8,10–11,17–19]. But both 2-D and 3-D graphical representation are accompanied with some loss of information due to overlapping and crossing of the curve representing DNA with itself. Randic and Balaban [2] presented a 4-D representation and outlined an approach to analysis the similarity among the coding sequences of the first exon of β -globin gene of eleven different species.

*Corresponding author.

In this paper, we propose a 5-D representation and outline an approach to compare the similarities/dissimilarities of DNA primary sequences. Comparing with Randic and Balaban's 4D representation, one can find that the computation of D/D matrix and the judgement of the position of mutations are easy. We make a comparison for the first exon of β -globin genes sequences belonging to eleven different species based on 5-D representation. We construct a n -component vector (with $n = 5, 10$ or 15) consisting of the average bandwidth of the D/D matrices. The similarities are computed by calculating the Euclidean distance between the end point of the vectors or calculating the correlation angle between the two vectors.

2. 5-D representation of DNA sequences

We construct a map between the bases of DNA sequences and plots in 5-D space, then we will obtain a 5-D representation of the corresponding DNA sequences. In 5-D space, points, vectors and directions have five components, and the unit vectors representing four nucleotides $A, G, C,$ and T are as follows: $A (1, 0, 0, 0, 0)$; $C (0, 0, 1, 0, 0)$; $T (0, 0, 0, 1, 0)$; $G (0, 1, 0, 0, 0)$.

Let $G = g_1g_2 \dots$ be an arbitrary DNA primary sequence, then we have a map ϕ , which maps G into a plot set. Explicitly, $\phi(G) = \phi(g_1)\phi(g_2), \dots$, where

$$\phi(g_i) = (\phi_1(g_i), \phi_2(g_i), \phi_3(g_i), i) = \begin{cases} (1, 0, 0, 0, i), & \text{if } g_i = A, \\ (0, 1, 0, 0, i), & \text{if } g_i = G, \\ (0, 0, 1, 0, i), & \text{if } g_i = C, \\ (0, 0, 0, 1, i), & \text{if } g_i = T. \end{cases}$$

We call the corresponding plot set characteristic plot set. Obviously, there is only one 5-D representation. Because the five directions of 5-D space fully equivalent, the above selection is equivalent to any other permutation of labels and directions, hence this selection should not be viewed as introducing any arbitrary decision to influence numerical analysis that follows. In table 1 we have introduced 5-D coordinates for the first 15 bases of Human and Goat in order to illustrate differences between the initial stages of the two sequences.

Since we have no graphical representation to associate with a random walk in 5-D space, we construct the distance matrix for each such random walk in which any (i, j) entry is the Euclidean distance between the corresponding points in 5-D space given by

$$D_{i,j} = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2 + \Delta s^2 + \Delta v^2}.$$

In fact,

$$D_{i,j} = \begin{cases} |i - j|, & \text{if } a_i = a_j, \\ \sqrt{2 + (i - j)^2}, & \text{if } a_i \neq a_j, \end{cases}$$

Table 1
5-D coordinates for the first 15 bases of human and goat.

No.	Human					Goat						
	Base	x	y	z	s	v	Base	x	y	z	s	v
1	A	1	0	0	0	1	A	1	0	0	0	1
2	T	0	0	0	1	2	T	0	0	0	1	2
3	G	0	1	0	0	3	G	0	1	0	0	3
4	G	0	1	0	0	4	C	0	0	1	0	4
5	T	0	0	0	1	5	T	0	0	0	1	5
6	G	0	1	0	0	6	G	0	1	0	0	6
7	C	0	0	1	0	7	A	1	0	0	0	7
8	A	1	0	0	0	8	C	0	0	1	0	8
9	C	0	0	1	0	9	T	0	0	0	1	9
10	C	0	0	1	0	10	G	0	1	0	0	10
11	T	0	0	0	1	11	G	0	1	0	0	11
12	G	0	1	0	0	12	A	1	0	0	0	12
13	A	1	0	0	0	13	G	0	1	0	0	13
14	C	0	0	1	0	14	G	0	1	0	0	14
15	T	0	0	0	1	15	A	1	0	0	0	15

where a_i, a_j denote the i th and j th base, respectively. Immediately, we can observe that the distance matrix is unique to a DNA sequence.

In table 2 we show a fragment of 92×92 distance matrix corresponding to the 5-D representation of Human. This matrix, which summarizes all the distances between the nucleic bases in the five-dimensional representation, forms our basic information of DNA sequences.

We will do mathematical analysis and make a comparison of distance matrices belonging to different DNA sequences. Constructing suitable sequence and matrix invariants facilitates such comparison. In this contribution we will point out invariants that can be extracted from the distance matrix, such as that shown in table 3.

3. Sequence invariants

The sequence invariant that we will introduce will be explained on the 15×15 fragment of the distance matrix of table 2. One can observe that in each row of the table the entries increase from left to right. The matrix can easily be rearranged by first placing the smallest entry (either 1 or $\sqrt{3}$) next to the main diagonal, then the next smallest entry (2 or $\sqrt{6}$) next to the first, and so on till all entries are arranged in increasing order as we move from the diagonal zero to the right.

Table 2

A 15×15 fragment of the 92×92 distance matrix belonging to the 5-D representation of human.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	$\sqrt{3}$	$\sqrt{6}$	$\sqrt{11}$	$\sqrt{18}$	$\sqrt{27}$	$\sqrt{38}$	7	$\sqrt{66}$	$\sqrt{83}$	$\sqrt{102}$	$\sqrt{123}$	12	$\sqrt{171}$	$\sqrt{198}$
2		0	$\sqrt{3}$	$\sqrt{6}$	3	$\sqrt{18}$	$\sqrt{27}$	$\sqrt{38}$	$\sqrt{51}$	$\sqrt{66}$	9	$\sqrt{102}$	$\sqrt{123}$	$\sqrt{146}$	13
3			0	1	$\sqrt{6}$	3	$\sqrt{18}$	$\sqrt{27}$	$\sqrt{38}$	$\sqrt{51}$	$\sqrt{66}$	9	$\sqrt{102}$	$\sqrt{123}$	$\sqrt{146}$
4				0	$\sqrt{3}$	2	$\sqrt{11}$	$\sqrt{18}$	$\sqrt{27}$	$\sqrt{38}$	$\sqrt{51}$	8	$\sqrt{83}$	$\sqrt{102}$	$\sqrt{123}$
5					0	$\sqrt{3}$	$\sqrt{6}$	$\sqrt{11}$	$\sqrt{18}$	$\sqrt{27}$	6	$\sqrt{51}$	$\sqrt{66}$	$\sqrt{83}$	10
6						0	$\sqrt{3}$	$\sqrt{6}$	$\sqrt{11}$	$\sqrt{18}$	$\sqrt{27}$	6	$\sqrt{51}$	$\sqrt{66}$	$\sqrt{83}$
7							0	$\sqrt{3}$	2	3	$\sqrt{18}$	$\sqrt{27}$	$\sqrt{38}$	7	$\sqrt{66}$
8								0	$\sqrt{3}$	$\sqrt{6}$	$\sqrt{11}$	$\sqrt{18}$	5	$\sqrt{38}$	$\sqrt{51}$
9									0	1	$\sqrt{6}$	$\sqrt{11}$	$\sqrt{18}$	5	$\sqrt{38}$
10										0	$\sqrt{3}$	$\sqrt{6}$	$\sqrt{11}$	4	$\sqrt{27}$
11											0	$\sqrt{3}$	$\sqrt{6}$	$\sqrt{11}$	4
12												0	$\sqrt{3}$	$\sqrt{6}$	$\sqrt{11}$
13													0	$\sqrt{3}$	$\sqrt{6}$
14														0	$\sqrt{3}$
15															0

Table 3

Expression for band average widths 1–14 for the 15×15 fragment of the distance matrices of human and goat.

Band	Human	Goat
1	$(2 + 12\sqrt{3})/14$	$(14\sqrt{3})/14$
2	$(4 + 11\sqrt{6})/13$	$(6 + 10\sqrt{6})/13$
3	$(9 + 9\sqrt{11})/12$	$(15 + 7\sqrt{11})/12$
4	$(8 + 9\sqrt{18})/11$	$(16 + 7\sqrt{18})/11$
5	$(10 + 8\sqrt{27})/10$	$(10 + 8\sqrt{27})/10$
6	$(12 + 7\sqrt{38})/9$	$(6 + 8\sqrt{38})/9$
7	$(16 + 6\sqrt{51})/8$	$(42 + 2\sqrt{51})/8$
8	$(8 + 6\sqrt{66})/7$	$(7\sqrt{66})/7$
9	$(18 + 4\sqrt{83})/6$	$(9 + 5\sqrt{83})/6$
10	$(10 + 4\sqrt{102})/5$	$(20 + 3\sqrt{102})/5$
11	$(4\sqrt{123})/4$	$(11 + 3\sqrt{123})/4$
12	$(12 + 2\sqrt{146})/3$	$(12 + 2\sqrt{146})/3$
13	$(13 + \sqrt{171})/2$	$(13 + \sqrt{171})/2$
14	$\sqrt{198}$	$\sqrt{198}$

Table 4
Numerical value for the band average widths 1–14 for the 15×15 fragment of the distance matrices for the first exon of β -globin gene of eleven species of table 1.

Band	Human	Goat	Gallus	Opossum	Lemur	Mouse	Rabbit	Rat	Bovine	Gorilla	Chimpanzee
0	0	0	0	0	0	0	0	0	0	0	0
1	1.6275	1.7321	1.6275	1.6275	1.6275	1.5752	1.6275	1.5752	1.7321	1.6275	1.6275
2	2.3803	2.4149	2.3803	2.3803	2.3803	2.4149	2.3458	2.3803	2.4149	2.3803	2.3803
3	3.2375	3.1847	3.2638	3.2638	3.2375	3.2375	3.2375	3.2375	3.1847	3.2375	3.2375
4	4.1985	4.1764	4.2426	4.2205	4.1764	4.1764	4.1764	4.1764	4.1764	4.1985	4.1985
5	5.1570	5.1766	5.0981	5.1177	5.1177	5.1570	5.1766	5.1570	5.1766	5.1570	5.1570
6	6.1279	6.1461	6.1461	6.1279	6.1279	6.1279	6.1279	6.1461	6.1461	6.1279	6.1279
7	7.1060	7.0354	7.0884	7.1060	7.1237	7.1060	7.0884	7.1060	7.0354	7.1060	7.1060
8	8.1063	8.1240	8.0709	8.0886	8.1063	8.1063	8.0709	8.1240	8.1240	8.1063	8.1063
9	9.0736	9.0920	9.0920	9.0736	9.0368	9.0920	9.0736	9.0920	9.0920	9.0736	9.0736
10	10.0796	10.0597	10.0796	10.0796	10.0995	10.0597	10.0995	10.0796	10.0597	10.0796	10.0796
11	11.0905	11.0905	11.0905	11.0905	11.0679	11.0905	11.0679	11.0679	11.0905	11.0905	11.0905
12	12.0553	12.0553	12.0553	12.0553	12.0553	12.0830	12.0830	12.0553	12.0553	12.0553	12.0553
13	13.0384	13.0384	13.0384	13.0384	13.0384	13.0384	13.0767	13.0384	13.0384	13.0384	13.0384
14	14.0712	14.0712	14.0712	14.0712	14.0712	14.0712	14.0712	14.0712	14.0712	14.0712	14.0712

In the following we will assume that all distance matrices considered have already been so ordered. We can now consider the sums of elements in diagonal entries parallel to the main diagonal, which consists of zeroes. In the case of table 2, we obtain for the first few neighboring diagonal the following sums: $(2 + 12\sqrt{3})$, $(4 + 11\sqrt{6})$, $(9 + 9\sqrt{11})$, and so on. We will refer to these sums as band invariants of different width, specifically b_1, b_2, b_3 for $(2 + 12\sqrt{3})$, $(4 + 11\sqrt{6})$, $(9 + 9\sqrt{11})$. One can observe that these quantities are not matrix invariants but can always be extracted from any matrix in whatever form it is presented by considering adjacency between sequence elements. If the distance matrix is already in the canonical form based on assigning labels to nucleic acid bases sequentially, band invariants can readily be obtained by summing elements along each of the lines parallel to the main diagonal.

In table 3 we show the band invariants based on 15×15 fragments of the sequences of Human and Goat. As one immediately sees, there is considerable variation in the form and numerical values for bandwidth invariants belonging to different sequences. In table 4 are listed the numerical values for the average bandwidths.

4. Similarities/dissimilarities based on bandwidth averages

We will illustrate the use of the 5-D quantitative characterization of DNA primary sequence with the examination of similarities/dissimilarities among the eleven coding sequences [6–9]. We construct a n -component vectors (with $n = 5$,

Table 5

The similarity/dissimilarity matrix for the eleven species of table 1 based on the angle between the end points of the 5-component vectors of the bandwidth 5.

Species	Human	Goat	Gallus	Opossum	Lemur	Mouse	Rabbit	Rat	Bovine	Gorilla	Chimpanzee
Human	0	0.020573	0.004130	0.002906	0.002637	0.010728	0.004822	0.008080	0.020573	0	0
Goat		0	0.023762	0.023048	0.020122	0.027356	0.022099	0.027685	0	0.020573	0.020573
Gallus			0	0.002610	0.006296	0.012110	0.004754	0.007963	0.023762	0.004130	0.004130
Opossum				0	0.004155	0.011165	0.003404	0.007267	0.023048	0.002906	0.002906
Lemur					0	0.010417	0.005276	0.010417	0.020122	0.002637	0.002637
Mouse						0	0.014238	0.005277	0.027356	0.010728	0.010728
Rabbit							0	0.010432	0.022099	0.004822	0.004822
Rat								0	0.027685	0.008080	0.008080
Bovine									0	0.020573	0.020573
Gorilla										0	0
Chimpanzee											0

10, or 15) consisting of the average bandwidth. The underlying assumption is that if two vectors point to a similar direction in the n -dimensional space, and then the two DNA primary sequences represented by the n -component vectors are similar.

The similarities among such vectors can be computed in two ways: (1) we calculate the Euclidean distance between the end points of the vectors; (2) we calculate the correlation angle of two vectors. The smaller Euclidean distance between the end points of two vectors, the more similar the DNA sequence. And, the smaller correlation angle between two vectors, the more similar the DNA sequence.

From tables 5 to 7, we have listed the similarity/dissimilarity table between the first exons of β -globin gene among 11 species based on cumulative bandwidths of order 5, 10, and 15, that is based on vectors having 5, 10, and 15 components. The three different bandwidths that, we consider corresponding to invariants that consider local DNA fragments of length 5, 10, and 15, all confined to the first 15 DNA nucleic acid bases. Observing tables 5–7, we can find that there is the smallest entry 0, it indicates that the subsequences of Human with length 5, 10, or 15 are the same as the corresponding subsequences of Gorilla and Chimpanzee, and the subsequences of Goat with length 5, 10, or 15 are the same as the corresponding subsequences of Bovine.

A comparison of the three separate similarities may vary. Such as, for the shortest segments, two pairs {Opossum, Gallus} and {Lemur, Rabbit} show the largest similarities (except for Human–Gorilla, Human–Chimpanzee, and Goat–Bovine), but as we increase the length of the segments, the similarity of the pair {Lemur–Rabbit} decreases considerably. On the other hand, the similarity between the pairs Human–Gorilla, Human–Chimpanzee, Goat–Bovine, Human–Mouse, and Rabbit–Human only slightly affected by the increase of the length of the fragment considered and corresponds to the smaller numerical entry for the cumulative bandwidths based on all 15 components of the vectors in table 5.

Table 6

The similarity/dissimilarity matrix for the eleven species of table 1 based on the angle between the end points of the 10-component vectors of the bandwidth 10.

Species	Human	Goat	Gallus	Opossum	Lemur	Mouse	Rabbit	Rat	Bovine	Gorilla	Chimpanzee
Human	0	0.008530	0.005253	0.003136	0.003093	0.003981	0.003085	0.003657	0.008530	0	0
Goat		0	0.010735	0.010117	0.009849	0.010567	0.008955	0.010626	0	0.008530	0.008530
Gallus			0	0.002692	0.006058	0.006887	0.006536	0.006804	0.010735	0.005253	0.005253
Opossum				0	0.003770	0.005354	0.005040	0.005257	0.010117	0.003136	0.003136
Lemur					0	0.005088	0.005343	0.004898	0.009849	0.003093	0.003093
Mouse						0	0.005414	0.002412	0.010567	0.003981	0.003981
Rabbit							0	0.004600	0.008955	0.003085	0.003085
Rat								0	0.010626	0.003657	0.003657
Bovine									0	0.008530	0.008530
Gorilla										0	0
Chimpanzee											0

Table 7

The similarity/dissimilarity matrix for the eleven species of table 1 based on the angle between the end points of the 15-component vectors of the bandwidth 15.

Species	Human	Goat	Gallus	Opossum	Lemur	Mouse	Rabbit	Rat	Bovine	Gorilla	Chimpanzee
Human	0	0.004620	0.002831	0.001703	0.002040	0.002398	0.002538	0.002133	0.004620	0	0
Goat		0	0.005817	0.005478	0.005543	0.005729	0.005280	0.005797	0	0.004622	0.004622
Gallus			0	0.001445	0.003417	0.003854	0.003931	0.003790	0.005807	0.002831	0.002831
Opossum				0	0.002279	0.003073	0.003234	0.002999	0.005478	0.001703	0.001703
Lemur					0	0.003279	0.003113	0.003027	0.005543	0.002040	0.002040
Mouse						0	0.003587	0.001830	0.005729	0.002398	0.002398
Rabbit							0	0.003204	0.005280	0.002538	0.002538
Rat								0	0.005797	0.002133	0.002133
Bovine									0	0.004620	0.004620
Gorilla										0	0
Chimpanzee											0

Table 8

Initial band average width for the distance matrices of the first exon of β -globin gene of eleven species of table 1.

Band	Human	Goat	Gallus	Opossum	Lemur	Mouse	Rabbit	Rat	Bovine	Gorilla	Chimpanzee
1	1.5229	1.5340	1.5310	1.5712	1.5632	1.5353	1.5347	1.5310	1.5082	1.5252	1.5209
2	2.3546	2.3746	2.3596	2.3696	2.3147	2.3518	2.3269	2.3496	2.3585	2.3556	2.3622
3	3.2206	3.1984	3.2241	3.2099	3.2206	3.2227	3.2220	3.2241	3.1945	3.2146	3.2173
4	4.1957	4.1953	4.2095	4.2068	4.1819	4.1860	4.1890	4.2095	4.1893	4.1935	4.1898
5	5.1443	5.1574	5.1398	5.1353	5.1398	5.1543	5.1477	5.1534	5.1526	5.1471	5.1491

Table 9

Similarity/Dissimilarity between the eleven species of table 1 based on the Euclidean distances between the end points of the 5-component vectors of the average bandwidth 5 using the full DNA sequence.

Species	Human	Goat	Gallus	Opossum	Lemur	Mouse	Rabbit	Rat	Bovine	Gorilla	Chimpanzee
Human	0	0.034465	0.017707	0.053634	0.058539	0.018976	0.031067	0.019394	0.031974	0.007414	0.011424
Goat		0	0.037495	0.046495	0.073640	0.034758	0.054466	0.038886	0.031609	0.028464	0.027959
Gallus			0	0.044105	0.061862	0.029048	0.039624	0.016881	0.044374	0.021194	0.025092
Opossum				0	0.061909	0.050627	0.061421	0.050523	0.070249	0.051480	0.055848
Lemur					0	0.048850	0.032800	0.056691	0.076441	0.057798	0.064849
Mouse						0	0.025950	0.024049	0.039854	0.017034	0.019651
Rabbit							0	0.031403	0.049811	0.031453	0.038226
Rat								0	0.043404	0.021344	0.026713
Bovine									0	0.027374	0.026595
Gorilla										0	0.009329
Chimpanzee											0

Table 10

The similarity/dissimilarity matrix for the eleven species of table 1 based on the angle between the end points of the 5-component vectors of the average bandwidth 5 using the full DNA sequence.

Species	Human	Goat	Gallus	Opossum	Lemur	Mouse	Rabbit	Rat	Bovine	Gorilla	Chimpanzee
Human	0	0.004261	0.001938	0.006680	0.007202	0.002355	0.003844	0.001590	0.003839	0.000930	0.001446
Goat		0	0.004742	0.005878	0.008921	0.004376	0.006663	0.004831	0.003318	0.003440	0.003431
Gallus			0	0.005580	0.007273	0.003632	0.004614	0.002002	0.005071	0.002370	0.002995
Opossum				0	0.007241	0.006366	0.007500	0.006358	0.008536	0.006377	0.006976
Lemur					0	0.005758	0.004051	0.006174	0.009687	0.007138	0.008008
Mouse						0	0.003007	0.002708	0.004727	0.002069	0.002454
Rabbit							0	0.002901	0.006301	0.003922	0.004757
Rat								0	0.004491	0.001813	0.002861
Bovine									0	0.003268	0.003081
Gorilla										0	0.001165
Chimpanzee											0

In tables 9 and 10, we show the similarity/dissimilarity table for the eleven DNA sequences based on cumulative bandwidths of order 5 but using the full sequence (table 8). Observing tables 9 and 10, we find that the more similar species pairs are Human–Gorilla, Gorilla–Chimpanzee, and Gorilla–Chimpanzee. Similar results have been obtained by Randić and Balanba [2], Randić et al. [4], Liao and Wang [6–9]. In figure 1 we make a comparison of data in tables 9 and 10, we find the data of table 9 and the data of table 10 correlate well, so we can conclude that tables 9 and 10 almost have the same information.

The similarities based on the full DNA primary sequence are somewhat altered when the full DNA primary sequence is compared with the initial part of

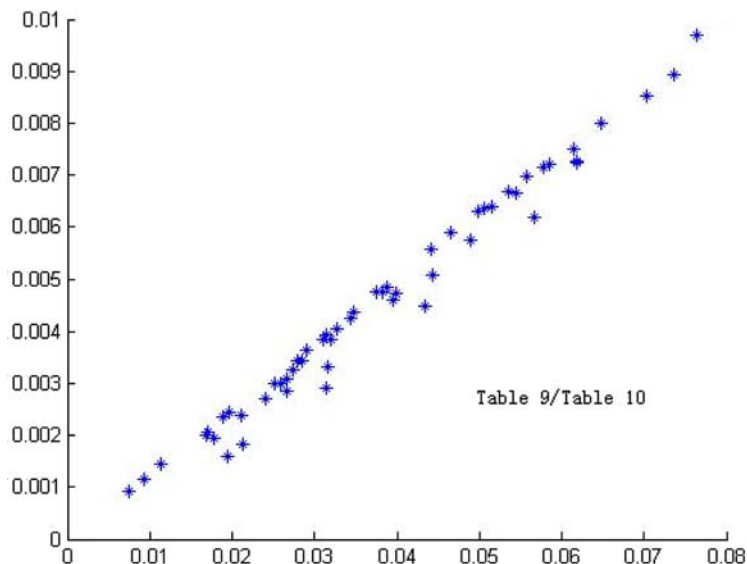


Figure 1. Comparison of tables 9 and 10.

the sequence, as shown in figure 2. Entries that remain close to the line $y = x$ indicate pairs of DNA sequences, which have not varied much after the initial changes. Such are for instance the pairs {Opossum, Gallus}, {Human, Opossum}, and {Lemur, Rabbit}. Entries that are at greater distance from the line $y = x$ indicate DNA sequences that show considerable variation not only at the initial fragments, but also throughout their length. In figure 2, we show a plot of bandwidths 1–5 for the initial fragment of length 15 of Human against the bandwidths 1–5 for the full length of Human. Each species is characterized by a similar correlation that can be approximated either by a straight line or a parabola having different constant terms.

5. Conclusion

We have proposed a 5-D representation of DNA primary sequence and outlined an approach to compare the similarities/dissimilarities of DNA sequences. The advantage of our approach is that it allows visual inspection of data, helping in recognizing major similarities among different DNA sequences. It is well-known that the alignments of DNA sequences are computer intensive that is direct comparison for DNA sequences. The sequence invariant easily computed and compared is applied to compare DNA primary sequences, rather than sequences themselves. In future work, we will use the numerical characterization of this representation to judge the mutations of nucleotides A , T , G , and C .

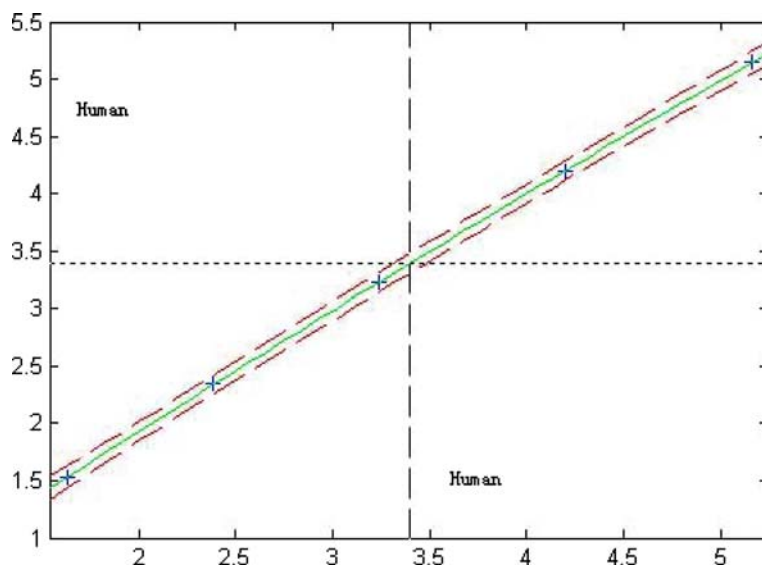


Figure 2. Plot of initial average bandwidths 1–5 as computed from the complete DNA sequence of human versus the corresponding bandwidths of the subsequence with length 15.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (Grant 10571019), the China Postdoctoral Science Foundation and the National Natural Science Foundation of Hunan University.

References

- [1] A. Arneodo, Y.D. Aubenton Carafa, B. Audit, E. Bacry, J.F. Muzy and C. Thermes, *Physica A* 249 (1998) 439–448.
- [2] M. Randic and A.T. Balanba, *J. Chem. Inf. Comput. Sci.* 40 (2000) 50–56.
- [3] M. Randic, M. Vracko, N. Lers and D. Plavsic, *Chem. Phys. Lett.* 368 (2003) 1–6.
- [4] M. Randic, M. Vracko, N. Lers and D. Plavsic, *Chem. Phys. Lett.* 371 (2003) 202–207.
- [5] Q.X. Yuan, B. Liao and T.M. Wang, *Chem. Phys. Lett.* 379 (2003) 412–417.
- [6] B. Liao and T.M. Wang, *J. comput. Chem.* 25(11) (2004) 1364–1368.
- [7] B. Liao and T.M. Wang, *J. Mol. Struct. THEOCHEM* 681 (2004) 209–212.
- [8] B. Liao and T.M. Wang, *Chem. Phys. Lett.* 388 (2004) 195–200.
- [9] B. Liao and T.M. Wang, *J. Chem. Inf. Comput. Sci.* 44 (2004) 1666–1670.
- [10] M. Randic, M. Vracko, A. Nandy and S.C. Basak, *J. Chem. Inf. Comput. Sci.* 40 (2000) 1235–1244.
- [11] A. Nandy, *Curr. Sci.* 66 (1994) 309–314.
- [12] R. Zhang and C.T. Zhang, *J. Biomol. Str. Dyn.* 11(4) (1994) 767–782.
- [13] M.L. Lan and M.S.T. Carpendale, In *Proceedings of IEEE Visualization* (1998) 443–446.
- [14] E. Hamori and J. Ruskin, *J. Biol. Chem.* 258 (1983) 1318–1327.
- [15] B. Liao and K.Q. Ding, *J. Comput. Chem.* 14(26) (2005) 1519–1523.

- [16] B. Liao, M.S. Tan and K.Q. Ding, Chem. Phys. Lett. 414 (2005) 296–300.
- [17] B. Liao, Chem. Phys. Lett. 401 (2005) 196–199.
- [18] B. Liao, M.S. Tan and K.Q. Ding, Chem. Phys. Lett. 402 (2005) 380–383.
- [19] B. Liao, Y.S. Zhang, K.Q. Ding and T.M. Wang, J. Mol. Struct. THEOCHEM 717 (2005) 199–203.