

## Weighted Alignment Free Dissimilarity Metric for Promoter Sequence Comparison

Kouser<sup>1</sup>, Lalitha Rangarajan<sup>2</sup>.  
<sup>1,2</sup> University of Mysore, Mysore-570006, India

### Article Info

#### Article history:

Received Feb 10<sup>th</sup>, 2015

Revised Apr 7<sup>th</sup>, 2015

Accepted Apr 10<sup>th</sup>, 2015

#### Keyword:

Weighted  
Alignment free  
Non coding regions  
Promoter comparison  
Dissimilarity measure

### ABSTRACT

Comparative sequence analysis has been a powerful tool in bioinformatics which interprets knowledge about the functionality of a sequence, making use of its structural information. Among the non coding regions of DNA, the comparison of promoter sequences has received a great deal of attention in medical science as promoter regions play a crucial role in gene regulation. In this work we propose an alignment free sequence comparison metric for comparison of promoter sequences. We use the binary and decimal position specific motif matrices (PSMM) of the promoters which were created for our experiments using the TFSEARCH tool. Simple weighted algorithm is used to compute the dissimilarity between the PSMMs of promoter sequences, thereby analyzing its underlying homology and functionality. The NCBI database was used to obtain the promoter sequences of 500 nucleotides upstream the transcription start site (TSS) of enzyme pyruvate kinase (PKLR) from the glycolysis pathway of different organisms for one experiment and all the enzymes from the glycolysis pathway of organism human for the other. The proposed dissimilarity metric is successful in bringing out differences on both the datasets and the results regarding similarities and differences in promoter sequences could be essential to have a clear knowledge of transcription regulation process in different organisms. The results reveal some useful findings which can be extended for a broader investigation.

Copyright © 2015 International Journal for Computational Biology,  
<http://www.ijcb.in>, All rights reserved.

### Corresponding Author:

Kouser,  
Department of Studies in Computer  
Science, University of Mysore,  
Mysore, India.  
Email: kauserd@gmail.com



### How to Cite:

First Author name *et. al.* Title of the Manuscript.  
IJCB. 2015; Volume 4 (Issue 2): Page 13-20.

## 1. INTRODUCTION

Understanding the accumulated knowledge related to huge data has gained importance in the recent years. Extracting knowledge from such huge data is more challenging than the analysis of simple conventional data. One of the most analyzed massive data is the biological sequences. Promoter sequences in the non-coding regions of the genes are one such data which are less explored when compared to the coding regions of the genes. Promoters can prove to be one of the keys to understand the underlying hypotheses of gene expression and regulation which is mainly controlled by transcription factors (TFs), that is, proteins that bind to promoter regions at specific sites (TFBSs) and regulate the process of transcription initiation.

The knowledge that governs the RNA synthesis using RNA polymerase is positioned in the promoter region that lies between 200-2000 nucleotides upstream of the Transcription Start Site (TSS) of a gene. The TFs interact with the sequence specific elements or motifs, which are 5-12 nucleotides in length. The motifs appear to be arranged in a specific configuration that confers on each gene an individualized spatial or temporal transcription program [1]. The lack of nucleotide sequence conservation between functionally related promoter regions makes it challenging to come out with an efficient computational model for promoter sequence analysis. Thus, the promoter regions of genes with similar expression pattern may not show sequence similarity even though they may be regulated by similar configuration of TFs [2].

A considerable amount of work has been carried out in aligning coding regions of DNA sequences for finding homology between different species. A variety of computational methods are available for sequence alignment namely, Bayesian methods [3], Dot matrix [4] [5], Scoring matrices [4], Dynamic programming [6], Hidden Markov Models [7] [8], Neural networks [9], Genetic algorithms [10], Word-based techniques etc. Plenty of alignment programs based on these methods exist, notably Clustal W, Clustal X, GRAIL, BLASTX, FASTA and MEME. Most of these methods use nucleotide comparison which is not suited for promoter regions. Few tools that align promoter sequences are CONREAL, Monkey and AVID. Despite the recent progress in phylogenetic foot printing techniques, lack of nucleotide conservation partially explains the limited success of the available computational models for promoter comparison [11].

In promoters, functionality is based on motif conservation and not nucleotide conservation. Hence, this type of comparison is not suited for promoters. There have been many successful attempts on similarity analysis and comparison of promoter sequences which perform alignment and few are alignment free [6][12][13][14]. However, we propose an alignment free sequence comparison metric for promoter sequence comparison.

In this work, we have extracted promoter sequences consisting of one prokaryote and nine eukaryotes. We use the NCBI database to extract the entire possible promoter of enzyme pyruvate kinase (dataset 1) from the glycolysis pathway and also the promoters of 10 enzymes from glycolysis pathway of organism human (dataset 2). Only few reports regarding the comparison of promoter sequences of enzymes of a metabolic pathway exist. These obtained promoters are varying in length (500bps – 35,000bps). For analysis we have considered 500bps from each of the promoter sequence and then used TFSEARCH tool to obtain the motif details and later created the normalized position specific motif matrices (PSMM) from the same. This results in matrices of size 73x10, 79x10 for each promoter of dataset 1 and dataset 2 respectively. The total number of motifs extracted by the tool is 73 in dataset 1 and 79 motifs were extracted from dataset 2. The matrix is a record of the presence (entry in the cell is 1 or more) or absence (entry in the cell is 0) of a motif in various positions 1-50, 51-100, ... 451-500 of the 500 bps considered. Then we apply the proposed alignment free sequence comparison method for the analysis of these promoter sequences. The steps are described in detail in the next section.

## 2. MATERIALS AND METHODS

The overall schema of the method adopted is as described in the Fig 1. The PSSM creation and the proposed dissimilarity measures are discussed in detail in the next section.

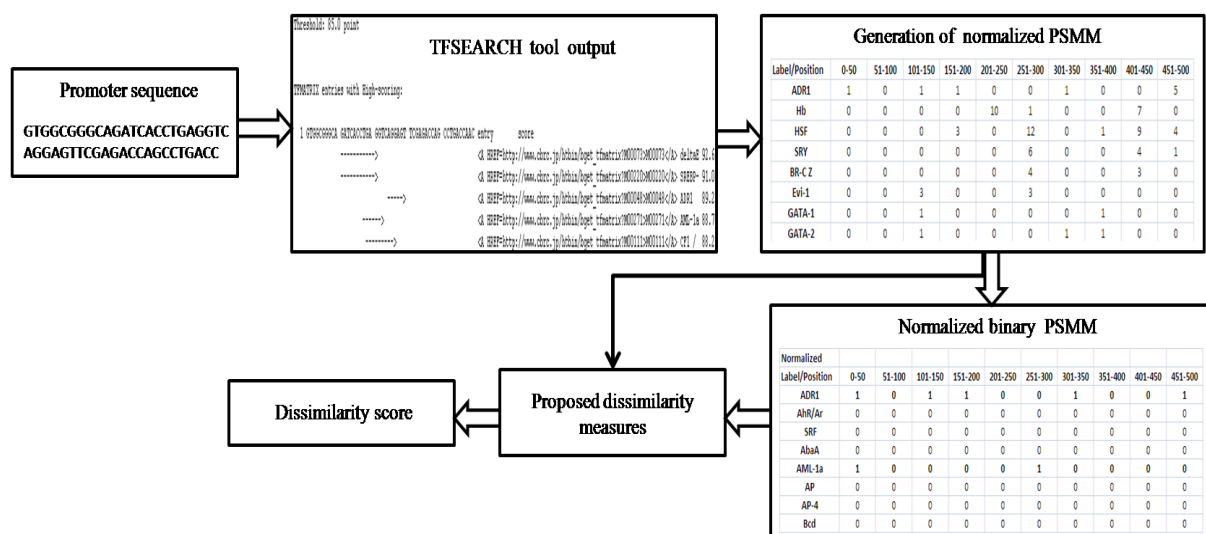


Figure 1. Architecture of the proposed model

## 2.1. The PSMM Creation

The PSMM of promoter sequences of both the dataset is obtained by the following steps. At first, we obtain the nucleotides in the region behind the start of PKLR gene till the end/start of the gene behind it for forward strands/backward strands, from the NCBI. This constitutes the entire promoter of that particular gene. Then, we spot the first 'TATA box' behind initiator codon (ATG) of the gene and then take 500bps behind that TATA for our experimentation. Later, these sequences are submitted to the 'TFSEARCH' [15] tool with a default threshold of 85% (specified in tool) to obtain the details of all the motifs with their position and score. This process is carried out for all the organisms listed below in Table 1 and the enzymes listed in Table 2.

**Table 1 - Organisms list (dataset 1)**

1. *Homo sapien* (human)
2. *Gorilla gorilla* (gorilla)
3. *Macaca mulatta* (rhesus monkey)
4. *Bos taurus* (cattle)
5. *Felis catus* (cat)
6. *Pan troglodytes* (chimpanzee)
7. *Canis lupus* (dog)
8. *Rattus norvegicus* (rat)
9. *Drosophila melanogaster* (fruit fly)
10. *Pseudomonas aeruginosa* (bacterium)

**Table 2 - Enzymes list (dataset 2)**

1. HK1 hexokinase 1
2. GPI glucose-6-phosphate isomerase
3. PFKM phosphofructokinase
4. ALDOA aldolase A, fructose-bisphosphate
5. TPI1 triosephosphate isomerase 1
6. GAPDH glyceraldehyde-3-phosphate dehydrogenase
7. PGK1 phosphoglycerate kinase 1
8. PGAM2 phosphoglycerate mutase 2 (muscle)
9. ENO1 enolase 1, (alpha)
10. PKM pyruvate kinase (muscle)

The PSMM matrix is generated by considering all the unique motifs identified by the TFSEARCH tool and are greater than 5bp in length. The matrix has 10 columns representing positions 1-50, 51-100, up to .....451-500 of the considered 500 bps. The number of rows depends on the number of unique motifs present in that particular promoter. We build an initial matrix where the multiple occurrences of a motif are counted. We have considered its binary matrix indicating just presence or absence of a motif in a position for our 1st experiment and matrix with actual number of motif occurrences for 2nd experiment. In case the motif is spread over two different ranges then we put '1' in the cell/cells where at least 40% of the motif is present. Every promoter has different number of motifs resulting in varying number of rows. Hence before we submit the matrix to our algorithm we do initial pre processing by normalizing the number of rows by simply taking union of all the motifs from different promoters. The input matrix of the organism *Homo sapien* is as shown in Table 3.

**Table 3 – Partial *Homo sapien* PSMM**

Motif/ Positions	0- 50	51- 100	-	-	-	-	-	-	-	451- 500
<b>ADR1</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>5</b>
<b>AhR/Ar</b>	0	0	0	0	0	0	0	0	0	0
<b>SRF</b>	0	0	0	0	0	0	0	0	0	0
<b>AbaA</b>	0	0	0	0	0	0	0	0	0	0
<b>AML-1a</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>AP</b>	0	0	0	0	0	0	0	0	0	0
<b>AP-4</b>	0	0	0	0	0	0	0	0	0	0
<b>Bcd</b>	0	0	0	0	0	0	0	0	0	0
<b>BR-C Z</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>3</b>	<b>0</b>
<b>C/EBPa</b>	0	0	0	0	0	0	0	0	0	0
<b>C/EBPb</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>
<b>C/EBP</b>	0	0	0	0	0	0	0	0	0	0
<b>cap</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>CDP CR</b>	0	0	0	0	0	0	0	0	0	0
<b>CdxA</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>CdxA-1</b>	0	0	0	0	0	0	0	0	0	0
<b>c-Ets</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>CF1 /</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

## 2.2. Proposed Dissimilarity Metric

A dissimilarity measure  $D$  is considered a metric if it satisfies the following relations  $\forall$  sequences  $A$ ,  $B$  and  $C$  [16] [17] [18].

1. *Nonnegativity* :  $D(A, B) \geq 0$
2. *Reflexivity* :  $D(A, B) = 0$  iff  $A = B$
3. *Symmetry* :  $D(A, B) = D(B, A)$
4. *Triangle Inequality* :  $D(A, B) + D(B, C) \geq D(A, C)$

### Metric: Weighted column sum

Let  $A$  and  $B$  be the PSSMs of the two promoters under comparison. We perform logical 'bit XOR' operation as in equation 1 on these matrices. This operation results in a matrix 'X' containing 1 in the bit position if corresponding bits of matrices  $A$ ,  $B$  are dissimilar and 0 if the bits are similar.

$$X = A \oplus B \quad (1)$$

The column sum 'CS' of each column 'j' in the matrix 'X' is computed using the equation 2 and then the weighted column sum 'WCS' is calculated by multiplying each CS with numerical weights in the increasing order starting with 1 as described in the equation 3. Later, the average weighted column sum 'AWCS' is calculated by dividing ACS by the size of matrix 'X'. The AWCS value constitutes the dissimilarity score between the promoters  $A$  and  $B$ .

$$CS(j) = \sum_{i=1}^r X(i, j), \quad j = 1, 2, 3, \dots, c \quad (2)$$

$$WCS = \sum_{j=1}^c CS(j) * k, \quad k = 1, 2, \dots, c \quad (3)$$

$$AWCS = WCS / (r * c) \quad (4)$$

Where 'r' is the number of rows and 'c' is the number of columns in matrix 'X'.  $CS(j)$  measures the dissimilarity present between the promoters due to various motifs in the position 'j'. Position 'j' is assigned weight inversely proportional to its distance from TATA box. WCS denotes the total weighted dissimilarity of the two promoters and AWCS gives the normalized dissimilarity.

## 3. RESULTS AND ANALYSIS

### 3.1. Experiment 1 (binary matrices ignoring multiple occurrences of a motif)

The results show lower dissimilarity for organisms belonging to the same family. This can be seen in Fig 3.b for organisms 1, 2 and 6 which belong to family 'Hominidae'. There is considerably higher dissimilarity for organisms of different families. The dissimilarity is considerably high between a prokaryote and a eukaryote in dataset 1. The results on dataset 2 show that the dissimilarity is quite uniform suggesting existence of some similarity between the enzymes of the glycolysis pathway as plotted in Fig 4.b. Table 4 and Table 5 are the dissimilarity matrices of dataset 1 and dataset 2 respectively.

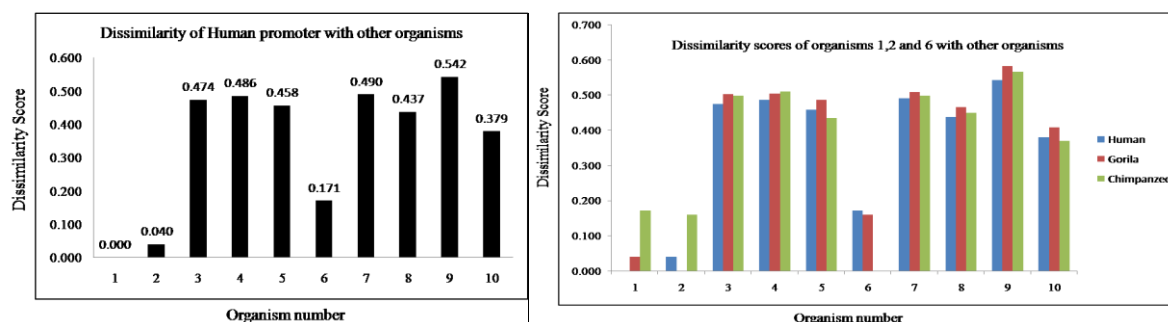


Figure 3. a. Dissimilarity of human promoter with the rest of 9 organisms (dataset1). Figure 3. b. Dissimilarity of organisms 1,2,6 with the rest of 7 organisms.

Table 4 – The dissimilarity matrix of experiment 1 on dataset 1

	1	2	3	4	5	6	7	8	9	10
1	0.000	0.040	0.474	0.486	0.458	0.171	0.490	0.437	0.542	0.379
2	0.040	0.000	0.503	0.504	0.486	0.159	0.508	0.466	0.582	0.408
3	0.474	0.503	0.000	0.423	0.433	0.497	0.436	0.379	0.548	0.368
4	0.486	0.504	0.423	0.000	0.404	0.510	0.484	0.452	0.593	0.389
5	0.458	0.486	0.433	0.404	0.000	0.434	0.392	0.355	0.468	0.303
6	0.171	0.159	0.497	0.510	0.434	0.000	0.497	0.449	0.566	0.370
7	0.490	0.508	0.436	0.484	0.392	0.497	0.000	0.429	0.548	0.374
8	0.437	0.466	0.379	0.452	0.355	0.449	0.429	0.000	0.489	0.315
9	0.542	0.582	0.548	0.593	0.468	0.566	0.548	0.489	0.000	0.432
10	0.379	0.408	0.368	0.389	0.303	0.370	0.374	0.315	0.432	0.000

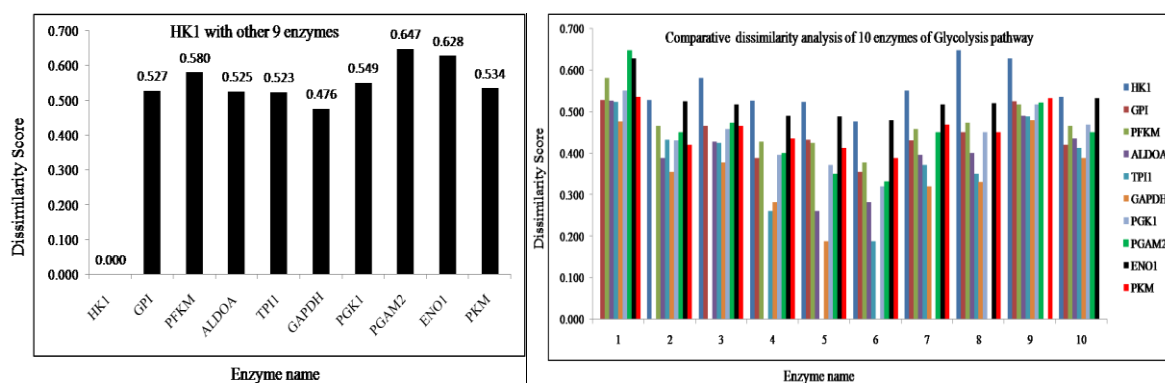


Figure 4. a. Dissimilarity score plotted of HK1 promoter with the promoters of rest of the 9 enzymes of glycolysis (dataset2). Figure 4. b. Dissimilarity of HK1 with rest 9 enzymes.

Table 5 – The dissimilarity matrix of experiment 1 on dataset 2

	HK1	GPI	PFKM	ALDOA	TPI1	GAPDH	PGK1	PGAM2	ENO1	PKM
HK1	0.000	0.527	0.580	0.525	0.523	0.476	0.549	0.647	0.628	0.534
GPI	0.527	0.000	0.466	0.389	0.432	0.354	0.430	0.449	0.524	0.420
PFKM	0.580	0.466	0.000	0.427	0.424	0.377	0.458	0.472	0.516	0.466
ALDOA	0.525	0.389	0.427	0.000	0.261	0.282	0.396	0.400	0.490	0.434
TPI1	0.523	0.432	0.424	0.261	0.000	0.189	0.371	0.349	0.487	0.411
GAPDH	0.476	0.354	0.377	0.282	0.189	0.000	0.319	0.330	0.478	0.387
PGK1	0.549	0.430	0.458	0.396	0.371	0.319	0.000	0.449	0.516	0.468
PGAM2	0.647	0.449	0.472	0.400	0.349	0.330	0.449	0.000	0.520	0.449
ENO1	0.628	0.524	0.516	0.490	0.487	0.478	0.516	0.520	0.000	0.532
PKM	0.534	0.420	0.466	0.434	0.411	0.387	0.468	0.449	0.532	0.000

**3.2. Experiment 2 (matrices with actual number of motif occurrences)**

The pattern of dissimilarity results obtained on dataset 1 and dataset 2 when the decimal PSSM was fed as input to the algorithm is similar to the results obtained in experiment 1 as plotted in Fig 5.a and 5.b. In this case the dissimilarity values were scaled up as the multiple occurrences of the motif were taken into account. In dataset 2, we observe significant similarity between the promoter sequences of all enzymes involved in the glycolysis pathway as plotted in Fig 6. All the enzymes had HSF, CdXA, ADR1 and cap in common in their promoters. This observation also suggests that the expression of the enzymes in a single metabolic pathway is likely to be regulated by similar pattern of TFBs in a promoter sequence. However, this similarity analysis should be carried out and compared for different metabolic pathways which would give an insight into the specific or ubiquitous nature of the regulatory motifs in promoter sequences.

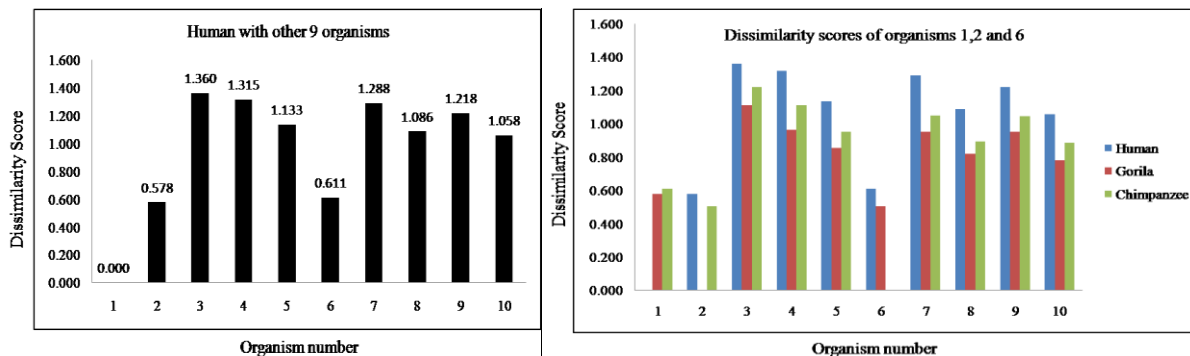


Figure 5.a Dissimilarity of human promoter with the rest of 9 organisms (dataset1). Figure 5.b. Dissimilarity of organisms 1, 2, 6 with the rest of 7 organisms.

Table 6 – The dissimilarity matrix of experiment 2 on dataset 1.

	1	2	3	4	5	6	7	8	9	10
1	0.000	0.578	1.360	1.315	1.133	0.611	1.288	1.086	1.218	1.058
2	0.578	0.000	1.111	0.964	0.856	0.504	0.951	0.818	0.952	0.781
3	1.360	1.111	0.000	0.927	0.693	1.218	0.905	0.753	0.841	0.684
4	1.315	0.964	0.927	0.000	0.758	1.112	0.910	0.738	0.889	0.649
5	1.133	0.856	0.693	0.758	0.000	0.952	0.612	0.488	0.611	0.429
6	0.611	0.504	1.218	1.112	0.952	0.000	1.049	0.895	1.045	0.885
7	1.288	0.951	0.905	0.910	0.612	1.049	0.000	0.634	0.727	0.608
8	1.086	0.818	0.753	0.738	0.488	0.895	0.634	0.000	0.592	0.393
9	1.218	0.952	0.841	0.889	0.611	1.045	0.727	0.592	0.000	0.497
10	1.058	0.781	0.684	0.649	0.429	0.885	0.608	0.393	0.497	0.000

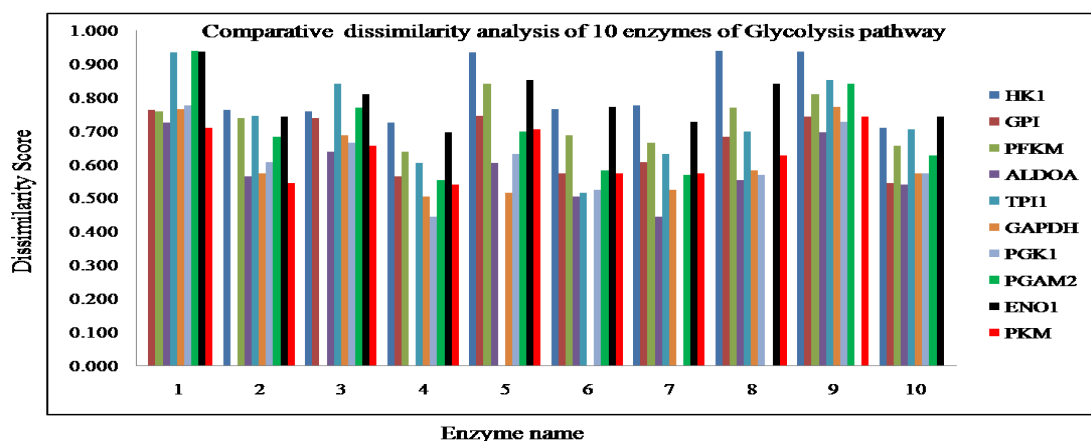


Figure 6. Dissimilarity of HK1 with rest 9 enzymes.

Table 7 – The dissimilarity matrix of experiment 2 on dataset 2.

	HK1	GPI	PFKM	ALDOA	TPII	GAPDH	PGK1	PGAM2	ENO1	PKM
HK1	0.000	0.761	0.758	0.724	0.933	0.763	0.776	0.937	0.935	0.708
GPI	0.761	0.000	0.737	0.563	0.744	0.572	0.605	0.682	0.742	0.544
PFKM	0.758	0.737	0.000	0.637	0.841	0.686	0.663	0.768	0.808	0.656
ALDOA	0.724	0.563	0.637	0.000	0.604	0.503	0.444	0.552	0.695	0.538
TPII	0.933	0.744	0.841	0.604	0.000	0.514	0.630	0.697	0.851	0.704
GAPDH	0.763	0.572	0.686	0.503	0.514	0.000	0.524	0.581	0.770	0.572
PGK1	0.776	0.605	0.663	0.444	0.630	0.524	0.000	0.568	0.727	0.572
PGAM2	0.937	0.682	0.768	0.552	0.697	0.581	0.568	0.000	0.839	0.627
ENO1	0.935	0.742	0.808	0.695	0.851	0.770	0.727	0.839	0.000	0.742
PKM	0.708	0.544	0.656	0.538	0.704	0.572	0.572	0.627	0.742	0.000

#### 4. CONCLUSION

This paper successfully demonstrates the use of proposed weighted column sum dissimilarity metric for promoter sequence similarity analysis. Demarcation between prokaryotes and eukaryotes has been observed. These differences in prokaryotic and eukaryotic promoter sequence motifs suggest the distinct patterns of transcription regulation and this could be helpful in designing pathogen specific drugs. Moreover, the information regarding the similarities and differences in promoter sequences could be essential to have a clear knowledge of transcription regulation process in different organisms. These results might also lead to some inferences on the evolutionary process of transcription from prokaryotes to eukaryotes.

We also notice that the promoter regions of genes with dissimilar motif pattern and arrangement may show sequence similarity even though they may not be regulated by similar motif configuration. All these results are from analysis of normalized promoter PSMM. We can also extend the experiments onto varied size PSMM and compare the obtained results. Perhaps a combination of non coding region (promoter) and coding region similarity score can give us a better understanding of the underlying homology existing between these genes.

#### ACKNOWLEDGEMENTS

It's our immense pleasure to acknowledge Arun ChandraShekar (President, R&D, Bhat BioTech, Bangalore, India) and Sunitha K (CSIR –SRF, Department of Biochemistry, University of Mysore, Mysore, India) for their valuable guidance and suggestions.

#### REFERENCES

- [1] Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol.* 2003 Sep;20(9):1377-419. Cited in PubMed; PMID 12777501.
- [2] Reiter LT, Potocki L, Chein S, Gribskov M, Bier E. A Systematic Analysis of Human Disease – Associated Gene Sequences in *Drosophila Melanogaster*. *Genome Res.* 2001;71:1114-125.
- [3] Liang KC, Wang X, Anastassiou D. Bayesian Basecalling for DNA Sequence Analysis Using Hidden Markov Models. *IEEE/ACM Trans Comput Biol Bioinform.* 2007 Jul-Sep;4(3):430-40. Cited in PubMed; PMID 17666762.
- [4] Dong X, Sung SY, Sung WK, Tan CL. Constraint based method for finding motifs in DNA sequences. *IEEE Trans Symposium on Bioinform and Bioeng.* 2004;483-90.
- [5] Leung HNM, Chin FYL. Discovering DNA Motifs with Nucleotide Dependency. *IEEE Trans Symposium on Bioinform and Bioeng.* 2006;70-80.
- [6] Meera A. Computational Models for DNA Sequence Alignment- Some New Approaches. Doctoral Thesis. University of Mysore. 2011.
- [7] Eddy SR. Hidden Markov Models. *Curr Opin Struct Biol.* 1996 Jun;6(3):361-5. Cited in PubMed; PMID 8804822.
- [8] Ji X, Ling JL, Sun Z. Mining gene expression data using a novel approach based on hidden Markov models. *FEBS Lett.* 2003 May 8;542(1-3):125-31. Cited in PubMed; PMID 12729911.
- [9] Chandra V, Girijadevi R, Nair AS, Pillai SS, Pillai RM. MTar : a computational microRNA target prediction architecture for human transcriptome. *BMC Bioinformatics.* 2010 Jan 18;11 Suppl 1:S2. Cited in PubMed; PMID 20122191.



- [10] Chan TM, Leung KS, Lee KH, Lio P. Generic Spaced DNA Motif Discovery Using Genetic Algorithm. Congress on Evolutionary Computation (CEC) IEEE Trans. 2010:1-8.
- [11] Meera A, Rangarajan L, Shilpa N. New Distance Measure for Sequence Comparison using Cumulative Frequency Distribution. I.J.C.A. 2011;19:13-18. DOI - 10.5120/2335-3043.
- [12] Deyneko IV, Kel AE, Blocker H, Kauer G. Signal-theoretical DNA similarity measure revealing unexpected similarities of E. coli promoters. In Silico Biol. 2005;5(5-6):547-55. Cited in PubMed; PMID 16268796.
- [13] Hu J, Liang X, Zhao H, Chen D. The Analysis of Similarity for Promoter Sequence Structures in yeast Genes. BMEI IEEE Trans. 2012;919-22.
- [14] Vinga S, Almeida J. Alignment-free sequence comparison- a review. Bioinformatics. 2003;513-23.
- [15] Web URL: <http://www.cbrc.jp/research/db/TFSEARCH.html>
- [16] Goshtasby AA. Image Registration - Similarity and Dissimilarity Measures - Chapter 2. Springer Advances in Computer Vision and Pattern Recognition. 2012; 07-66. DOI 10.1007/978-1-4471-2458-0.
- [17] Duda RO, Hart PE, Stork DG. Pattern Classification. 2<sup>nd</sup> edn. Wiley-Interscience. New York. 2001;187.
- [18] Theodoridis S, Koutroumbas K. Pattern Classification. 4<sup>th</sup> edn. Academic Press. New York. 2009;602-06.