ejpmr, 2018, 5(11), 380-383

SJIF Impact Factor 4.897

EUROPEAN JOURNAL OF PHARMACEUTICAL AND MEDICAL RESEARCH

www.ejpmr.com

Research Article

ISSN 2394-3211

EJPMR

IMPLEMENTING BIOINFORMATICS FOR THE STUDY OF DIABETES MELLITUS TYPE II, IN SIMILARITY SEARCH, AND CONSTRUCTION OF PHYLOGENETIC TREE.

[1]Priyanka Dongre, [1]Ashwini Jarange, [1*]Shubham Mete, [1]Sudarshan Padale, [1]S. N. Harke and [2]Yusuf Talib

[1]Institute of Biosciences and Technology, MGM campus, Aurangabad. (M.S) 431001.
[1]MGM's G. Y. Pathrikar College of C.S. and I.T., MGM Campus, Aurangabad. (M.S) 431001.
[2]Department of Biotechnology & Bioinformatics, Dr. Rafiq Zakaria Campus, Maulana Azad College of PG and Research Center, Rauza Bagh, Aurangabad.(M.S) 431001.

***Corresponding Author: Shubham Mete**

Institute of Biosciences and Technology, MGM campus, Aurangabad. (M.S) 431001.

Article Received on 03/09/2018 **Article Revised on 24/09/2018** **Article Accepted on 14/10/2018**

ABSTRACT

Diabetes mellitus, describes a group of metabolic diseases in which the person has high amount of blood glucose (blood sugar), either because insulin production is inadequate, or body's cells do not respond properly to insulin, or both. Patients with high blood sugar will usually experience polyuria (frequent urination), they will become increasingly thirsty (polydipsia) and hungry (polyphagia). Diabetes is a condition in which blood sugar level increases. In 2017 it was estimated that over 425 million people (aged between 20 and 79 years) over the world had diabetes. Diabetes is mainly classified in two types: Type I and Type II. In Type I Diabetes the body is not able produce insulin. Approximately 10% of all diabetes cases are type I, In Type II Diabetes the body does not produce enough insulin for proper function. Approximately 90% cases of diabetes worldwide are of type II. Bioinformatics one of the field in life science plays an important role to study diabetes and its effective impact on the life of human kind. In these field various types of methods, algorithm, programs are available to search for similarity (alignment), to provide the 3D coordinates (structure prediction), and to provide the Pedigree analysis (phylogenetic Tree Construction). To implement above said methods for sequence, structure, Tree construction as sample study we took Diabetes as a marker. Three samples were studied based on their BMI (Body mass Index), BSL (Blood Sugar Level), LP (lipid Profile), and Glycelated Haemoglobin (HbA1C). In the given study approach is to correlate three 3 conditions Healthy, Prediabetic and Diabetic.

KEYWORDS: Bioinformatics, Diabetes, Alignment, Structure Prediction, Phylogenetic tree.

INTRODUCTION

Diabetes: Diabetes is devastating disease that is characterized by high glucose levels in the blood and has been recorded in the medical literature since as early as 1500 BC.[2] Diabetics either do not produce enough insulin to process their intake of glucose or the body does not use the insulin efficiently enough to control glucose levels[2]. Untreated, diabetes can cause a number of health problems including, blindness, loss of circulation resulting in limb amputation, high blood pressure, heart disease, and kidney failure.[2]

Alignment

Best method to analyse the sequence is similarity searching which gives the values to correlate with other fragments, organisms, species, phylum and kingdom, here The BLAST program (Stephen Altschul) was utilized for analysis, these program works on Needleman Wunsch Algorithm, where sequences are breaked into fragments, then Scoring matrices (BLOSUM 62) were generated, then log odd ratio, and lastly KA equation (E-Value) is given.

The target is to find high-scoring ungapped segments among related sequences. The existence of such segments above a given threshold indicates pairwise similarity beyond random chance, which helps to discriminate related sequences from unrelated sequences in a database.[1]

Structure Prediction

Protein Structure Prediction is the process of prediction of the three dimensional structure of a protein from its amino acid sequence. It is the inference of the three-dimensional structure of a protein from its amino acid sequence—that is, the prediction of its folding and its secondary and tertiary structure from its primary structure. Structure prediction is fundamentally different from the inverse problem of protein design. Protein structure prediction is one of the most important goals

pursued by bioinformatics and theoretical chemistry; it is highly important in medicine (for example, in drug design) and biotechnology (for example, in the design of novel enzymes). The QMEAN Z-score in the provides an estimate of the "degree of nativeness" of the structural features observed in the model on a global scale. QMEAN Z-scores around zero (0) indicate good agreement between the model structure and experimental structures of similar size. Scores of -4.0 or below are an indication of models with low quality [Graph: 1].

Phylogenetic tree

Phylogeny is classified as relationship between two species. The resulting relationship is represented as binary tree. Two main types of trees are: (i) Rooted trees—where all nodes are derived from single node. And (ii) Unrooted trees—where it is not clear that where the nodes originated from.

Our goal in this paper is to look at how bioinformatics techniques have been applied specifically to diabetes research and make comparisons between them to establish a proper understanding of how bioinformatics has impacted this important field of study.

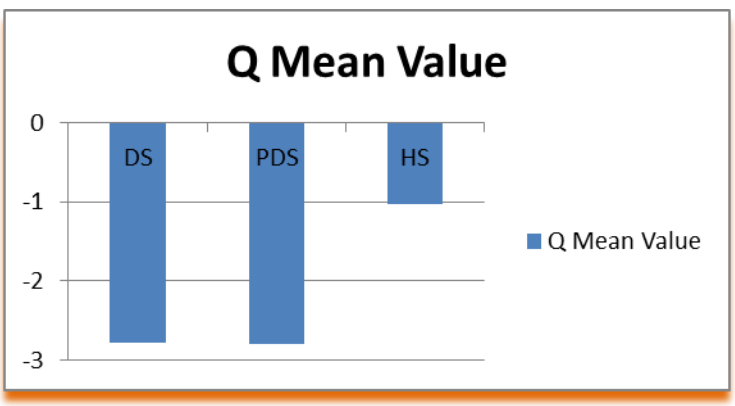
2. PROGRAMS

Different softwares and methods are available by which information of common ancestor can be drawn for that firstly target sequence is obtained in specific fasta file form biological database (NCBI), obtained sequences are processed by Different softwares. these sequence has processed for sequence alignment, similar sequence has been shortlisted on the basis of Expectation Value (E-Value) (Table 1), then those shortlisted sequences are performed by using FASTA33 ,SWISS-PROT (ExPASy) for to predict the structure of protein sequence for homology modeling using SWISS-MODEL server.

For phylogenetic tree construction those shortlisted sequence has been performed by using CLUSTALX drawn in the form of fasta using ClustalX (Provides output in .dnd, .phy) and Phylip.

Table 1.

Sr. No.	Description	E-Value	QMEAN	Protein Structure	Uniprot Accession
1.	Diabetic Patient	1.9E-30	-2.78		A0A024RBF6 (A0A024RBF6_HUMAN)
2.	Pre-Diabetic Patient	2.5E-29	-2.79		A0A024RBF6 (A0A024RBF6_HUMAN)
3.	Healthy Patient	8.2E-30	-1.02		A0A024RBF6 (A0A024RBF6_HUMAN)



Graph: 1

3. PHYLOGENETIC TREE CONSTRUCTION

Once the sequence alignment is completed the samples were processed for the Phylogenetic tree construction, for which here we used Clustal X and PHYLIP program to get the pedigree analysis.

After completion of again offline sequence similarity search using Clustal X, three output files are generated named .dnd,.aln,.phy, that will help to construct phylogenetic tree using Maximum Likelihood [In fig. 2] and Maximum Parsimony method [In fig. 3], in rooted and unrooted form[In fig. 1]. The sampled output shows the correlation based on similarity and dissimilarity between three samples [HS, DS, PDS]. Those outputs show the different branch length from obtained sequence gives the evolutionary divergence.

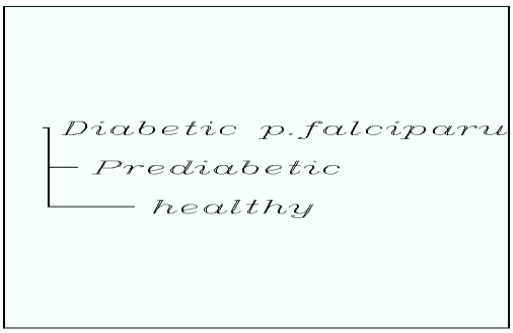


Fig 1: Rooted Phylogenetic tree

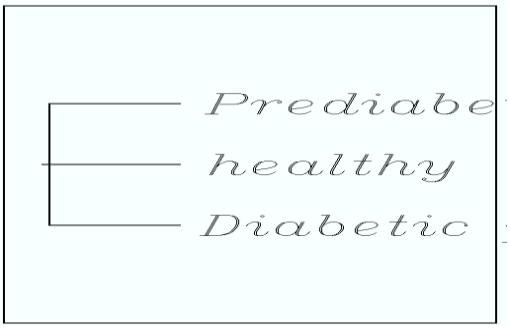


Fig 2: Rooted Phylogenetic tree by using Maximum Likelihood

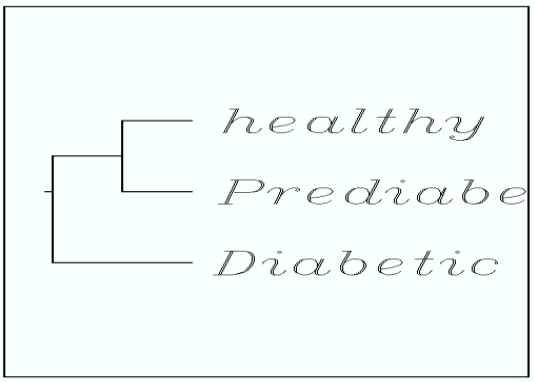


Fig 3: Rooted Phylogenetic tree by using Maximum Parsimony

4. STRUCTURE PREDICTION

Predicting the secondary structure is one of the important aspect in Biological world, that will help to understand the availability of atoms in the molecule, energy level, stability of the molecule, and mobility. In this work, Based on similarity search and expected E-values, some sequences have been shortlisted, and ExPasy was utilized to predict the secondary structure, by which different chains of helix, sheets and thread can be seen. Q-Mean provided from the Swiss-Prot shows the accuracy and similarity between the said different sequences of samples [Table-1].

RESULTS AND CONCLUSION

Using different methods, algorithm, and programs in bioinformatics led to analyse the sequence, structure and predicting them. In these study, implementing Bioinformatics and their methods, various tools led to conclude the similarity, structure robustness and there activity in causing disease of Diabetes Mellitus Type II [Table 1]. Q-Mean shows the quality score of the said structures from DS to HS, ranging from -1..02 to -2.79, which concludes the quality of the structure between experimental structure and target database structure, as nearer to Zero(0) (Graph : 1) will have optimum and good quality score for the perspective of functionality, mobility and energy level of three samples. After performing the swiss model score generated led to conclude the average quality of the structure as -4.0 or below led to conclude low quality of the structure, which is of no use for further analysis.

Phylogenetic tree constructed in Pedigree analysis [fig. 1, fig. 2, and fig. 3] shows the implemented algorithm UPGMA, MP,ML prove the correlation between DS, HS and PDS.

Which will led to practitioners to make watch on Pre diabetic patients to control to convert into healthy and not moving towards diabetic one. One more aspect can be implemented in future about to correlated the method for functional finding in the sequences. Also to target the SNPs and Pseudogene involved in it.

ETHICAL MATTERS

Samples used in this work is approved by ethical committee, and informed consent are noted from the patients.

REFERENCES

1. Robert Kelley1 and Eric C. Rouchka1 TR-ULBL-(2007) Bioinformatics Techniques Used in Diabetes Research.
2. H.-Y. LV, W.-G. Zhou, and C.-G. Zhou, "A Discrete Particle Swarm Optimization Algorithm for Phylogenetic Tree Reconstruction," 2004; 2650-2654.
3. J. Pevsner, "Molecular Phylogeny and Evolution,", 1 ed John Wiley and Sons, Inc, 2003; 368-388.

4. Sneath PHA and Sokal RR, "Numerical Taxonomy," in *Numerical Taxonomy* San Francisco, CA: Freeman, 1973; 230-234.
5. N. Saitou and M. Nei, "The neighbor-joining method: a new method for reconstructing phylogenetic trees," *Mol. Biol. Evol*, July1987; 4(4): 406-425.
6. W. M. Fitch and E. Margoliash, "Construction of phylogenetic trees," *Science*, Jan.1967; 155(760): 279-284.
7. W. Fitch, "Toward defining the course of evolution: minimum change for a specified tree topology.," *Syst. Zool*, Jan.1971; 20: 406-416. J. Felsenstein, "Evolutionary trees from DNA sequences: a maximum likelihood approach," *J. Mol. Evol*, 1981; 17(6): 368-376.
8. Uma Makheswari M and Sudarsanam D. A review on Bioinformatics for diabetes mellitus, 2012.
9. Suparna Deepak, Rosalind Marita (2005) Genomics Of Type 2 Diabetes.
10. H.-Y. LV, W.-G. Zhou, and C.-G. Zhou, "A Discrete Particle Swarm Optimization Algorithm for Phylogenetic Tree Reconstruction," 2004; 2650-2654.
11. J. Pevsner, "Molecular Phylogeny and Evolution,", 1 ed John Wiley and Sons, Inc., 2003, pp. 368-388. Sneath PHA and Sokal RR, "Numerical Taxonomy," in *Numerical Taxonomy* San Francisco, CA: Freeman, 1973; 230-234.
12. N. Saitou and M. Nei, "The neighbor-joining method: a new method for reconstructing phylogenetic trees," *Mol. Biol. Evol*, July1987; 4(4): 406-425. W. M. Fitch and E. Margoliash, "Construction of phylogenetic trees," *Science*, Jan.1967; 155(760): 279-284.
13. W. Fitch, "Toward defining the course of evolution: minimum change for a specified tree topology.," *Syst. Zool.*, vol. 20, pp. 406-416, Jan.1971. J. Felsenstein, "Evolutionary trees from DNA sequences: a maximum likelihood approach," *J. Mol. Evol*, 1981; 17(6): 368-376.
14. Durbin R., S. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Reprint ed. Cambridge: Cambridge University Press, 1999.
15. B. G. Hall, *Phylogenetic Trees Made Easy: A How-to Manual*, 3 ed Sinauer Associates, Inc., 2007.
16. A. N. Egan and K. A. Crandall, "Theory of Phylogenetic Estimation," in *Evolutionary Genetics: Concepts and Case Studies*, 1 ed Oxford University Press, USA, 2006; 426-436.
17. M. Dorigo and G. Di Caro, "The Ant Colony Optimization Meta-Heuristic," in *New Ideas in Optimization*. Corne D., M. Dorigo, and F. Glover, Eds. London: McGraw-Hill, 1999; 11-32.
18. L. Nakhleh, G. Jin, F. Zhao, and J. Mellor-Crummey, "Reconstructing phylogenetic networks using maximum parsimony," *Proc. IEEE Comput. Syst. Bioinform. Conf*, 2005; 93-102.
19. M. Perretto and H. S. Lopes, "Reconstruction of phylogenetic trees using the ant colony optimization algorithm," *Genet. Mol. Res*, 2005; 4(3): 581-589.
20. C. Kumnorkaew, K. Ku, and P. Ruenglertpanyakul, "Application of ant colony optimization to evolutionary tree construction," 2004.
21. S. Ando and H. Iba, "Ant algorithm for construction of evolutionary tree," Honolulu, HI, USA, 2007; 1552-1557.
22. M. Dorigo and T. Stutzle, *Ant Colony Optimization* Mit Press, 2004. J. Guo, L. Chen, L. Qin, and C. Wang, "A self-adaptive ant colony algorithm for phylogenetic tree construction,", 2006; 01 ed: 78.
23. J. Kennedy and R. C. Eberhart, "Particle swarm optimization,"
24. Piscataway, NJ, USA: IEEE Service Center, 1995; 1942-1948.