RESEARCH PAPER | Bioinformatics | Volume : 4 | Issue : 7 | July 2014 | ISSN - 2249-555X

Database Development on Inborn Errors of Nucleotide Metabolism in *Homo Sapiens* and Tool Designing for Nucleotide Sequence Alignment

KEYWORDS: Homo sapiens, IEMs, inborn errors, nucleotides

Satyamvada Maurya
Deptt. of, Bioinformatics Banasthali University, Banasthali Vidyapith - 304 022, India

Mansaf Alam
Department of Computer Science Jamia Millia Islamia, New Delhi-110025, India

ABSTRACT *Inborn errors of metabolism (IEMs) are individually rare, but collectively numerous. Recently; IEMs were considered a specialty of pediatricians. Indeed, the term "inborn" has meant for a long time, a disease which starts in the newborn period or at least in childhood. Although most IEMs can have mild forms starting in adolescence or late in adulthood, this concept of "adult onset IEMs" has not reached the medical community until recently. They refer to single gene disorders where in loss of function of a single enzyme results in abnormalities in synthesis or catabolism of proteins, carbohydrates or fats, which results in a disruption in a metabolic pathway. In have developed a computational tool, for in born errors of nucleotide sequence alignment. This tool is based on the Dynamic Programming Algorithm. These data indicate the impartment of similarity, dissimilarity and evolutionary relationship to each other.*

INTRODUCTION

Bioinformatics deals with organizing and presenting information in effective ways. With the globalisation of the Internet and the data deluge from the Genome sequencing projects bioinformatics is going through a period of explosive growth and development. The WWW facilitates the sharing of this treasure and has changed the nature of learning by providing increased access to resources in a variety of media.

Genetic analysis can be used generally to describe methods both used in and resulting from the sciences of genetics and molecular biology, or to applications resulting from this research. Genetic analysis may be done to identify genetic/inherited disorders and also to make a differential diagnosis in certain somatic diseases such as cancer. Genetic analyses of cancer include detection of mutations, fusion genes, and DNA copy number changes. Genetic analyses include but are not limited to molecular technologies such as PCR, RT-PCR, DNA sequencing, and DNA microarrays, and cytogenetic methods such as karyotyping and fluorescence in situ hybridization.

Inborn errors of metabolism (IEMs) are individually rare, but collectively numerous. Until recently; IEMs were considered a specialty of paediatricians. Indeed, the term "inborn" has meant for a long time, a disease which starts in the newborn period or at least in childhood. Although most IEMs can have mild forms starting in adolescence or late in adulthood, this concept of "adult onset IEMs" has not reached the medical community until recently. They refer to single gene disorders wherein loss of function of a single enzyme results in abnormalities in synthesis or catabolism of proteins, carbohydrates or fats, which results in a disruption in a metabolic pathway. This results in toxic accumulations of substrates before the disruption, intermediates from alternative pathways, and/or defects in energy production and utilization. Nearly every metabolic disease has several forms that vary in age of onset, clinical severity and mode of inheritance. The mode of inheritance determines the male to female ratio of affected and many IEMs have multiple forms that differ in their mode of inheritance. The metabolic requirements for the nucleotides and their cognate bases can be met by either dietary intake or synthesis de novo from low molecular weight precursors. Indeed, the ability to salvage nucleotides from sources within the body alleviates any nutritional requirement for nucleotides, thus the purine and pyrimidine bases are not required in the diet. The salvage pathways are a major source of nucleotides for synthesis of DNA, RNA and enzyme co-factors.

Technology is driving the field of human genetics research with advances in techniques to generate high-throughput data that interrogate various levels of biological regulation (Holzinger, 2013). The striking increase in the number of people with metabolic syndrome (MetS) as a result of the global epidemic of obesity and diabetes. Increasing evidence suggests that uric acid may play a role in MetS (Coutant etal. 2012). To assess the prevalence of MetS in a large cohort from Israel and its association with hyperuricemia using the latest three definitions of MetS. (Cohen et al. 2012).

In inborn errors of metabolism (IEMs) are hereditary metabolic defects, which are encountered in almost all major metabolic pathways occurring in man. Many IEMs are screened for in neonates through metabolomic analysis of dried blood spot samples. To enable the mapping of these metabolomic data onto the published human metabolic reconstruction, they added missing reactions and pathways involved in acylcarnitine (AC) and fatty acid oxidation (FAO) metabolism. (Sahoo etal. 2012). The identification of a new mutation responsible for ausing human severe combined immune deficiency syndrome (SCID). In a large consanguineous Israeli Arab family, this served as a diagnostic tool and enabled us to carry out preimplantation genetic diagnosis (PGD). They also demonstrated that PGD for homozygosity alleles is feasible (Tomashov-Matar etal. 2012). VAAST (the Variant Annotation, Analysis & Search Tool) is a probabilistic search tool for identifying damaged genes and their disease-causing variants in personal genome sequences (Yandell etal. 2011). Murine air pouch is a bursa-like space that resembles the human synovial membrane. Injection of monosodium urate (MSU) crystals into the pouch elicits an acute inflammatory response similar to human gout. They conducted the present study to identify mRNAs that were highly regulated by MSU crystals in the pouch membrane. (Pessler etal. 2008). Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE) is an autosomal recessive disease caused by loss-of-function mutations in the gene encoding thymidine phosphorylase (TP) (Martí etal. 2003). Compares erythrocyte nucleotide levels in patients with eight different inherited purine or pyrimidine enzyme defects identified amongst a variety of patients referred predominantly for investigation of severe neurological abnormalities, or immunodeficiency syndromes (Simmonds et al., 1988). Collecting millions of genetic variations is feasible with the advanced genotyping technology. With a huge amount of genetic variations data in hand; developing efficient algorithms to carry out the gene-gene interaction analysis in a timely manner has become one of the key problems in

Volume : 4 | Issue : 7 | July 2014 | ISSN - 2249-555X

genome-wide association studies (GWAS) (Yung, et al.2011).

MATERIALS AND METHODS

The programming language of the present system is coded in Microsoft Visual Studio 2010 used as front end and Microsoft SQL Server 2012 used as a back end. The application can run on Pc with at least core i3 processor or higher version 4 GB RAM and 500GB Hard Disk or higher version windows 7. The information gather from the field was entered in Microsoft SQL Server 2012 Data base management system, which also served as back end of the software. The Gene sequence from are retrieved from the Gen bank database of NCBI.

Web pages designed by the help of .NET which is uses the C# language. Firstly I designed the Login page then designed the Home, About, Tool, Drugs, Pathways, Search and Logout pages. Write the text all the pages which the pages designed already. Login is the initial module of this tool. This module is for verification of users. It requires input username and password which are stored in login table. Then it matches the user identification to the valid stored data in table. If this matching is successful then that user can obtained welcome user, otherwise invalid user with an error message.

This tool is based on dynamic programming algorithm. Sequence alignment is the procedure of comparing two (pairwise alignment) or more (multiple alignment) sequences by searching for a series of characters that are in the same order in all sequences. Two sequences can be aligned by writing them across a page in two rows. Identical or similar characters are placed in the same column, and non identical ones can either be placed in the same column as a mismatch or against a gap (-) in the other sequence. Sequences that are aligned in this manner are said to be similar. Sequence alignment is useful for discovering functional, structural, and evolutionary information in biological sequences. Take two sequences and determine the best alignment between them. The total score of the alignment depends on each column of the alignment. If the column has two identical characters, it will receive value +1 (a match). Different characters will give the column value -1 (a mismatch). Finally a gap in a column drops down its value to -2 (Gap Penalty). The best alignment will be one with the maximum total score. M =size of Seq1 and N= size of Seq2 ,the computation is arranged into an (N+1) × (M+1) array where entry (j,i) contains similarity between Seq2[1.....j] and Seq1[1.....i]. The algorithm computes the value for entry (j,i) by looking at just three previous entries. These are divided into three steps-

The first class contains three methods that describe the steps of dynamic programming algorithm. The first method is named Intialization_Step, this method prepares the matrix a [i,j] that holds the similarity between arbitrary prefixes of the two sequences. The algorithm starts with shorter prefixes and uses previously computed results to solve the problem for larger prefixes.

The second method named Get_Max computes the value of the cell (j,i) by this equation. where p(j,i)= +1 if Seq2[j]=Seq1[i] (match Score) and p(j,i)= -1 if Seq2[j]!=Seq1[i].

$$\begin{Bmatrix} a[j,i-1]+Gap \\ a[j-1,i-1]+p(j,i) \\ a[j-1,i]+Gap \end{Bmatrix}$$

The third method is named Trace back Step. This method will produce the alignment by traversing the cell matrix (N-1, M-1) back towards the initial entry of the cell matrix (1, 1).

Connected the pages with the database for extract all the information which I want.

RESULTS AND DISCUSSION

I compared gene sequences using this tool and obtained following results. Computational approaches to sequence alignment generally fall into two categories: global alignments and local alignments. Calculating a global alignment is a form of global optimization that "forces" the alignment to span the entire length of all query sequences. By contrast, local alignments identify regions of similarity within long sequences that are often widely divergent overall. Local alignments are often preferable, but can be more difficult to calculate because of the additional challenge of identifying the regions of similarity. A variety of computational algorithms have been applied to the sequence alignment problem. These include slow but formally correct methods like dynamic programming. These also include efficient, heuristic algorithms or probabilistic methods designed for large-scale database search.

This tool is design for "Gene Analysis" .Firstly I register in the "Nucleotide Metabolic Genetic Database" page which is given below.

a. Registration:
I have been created account in the "Nucleotide Metabolic Genetic Database" in the help of registration page (Figure 1).

b. Login:
After registration I open the Login Page and do the login in this page.The result is given below. (Figure 2).

c. Home Page:
This is Home Page of Tool for Gene Analysis. (Figure 3).

d. Tool:
This page is DNA Sequence Alignment. Gene analysis tool calculated several measures matches, mismatches, insertion and deletion. Those properties are explained below. (Figure 4).

d. 1.Matches:
I give the score of match value is 1.Match value is when I compared two sequences where these sequence are match to each other then count it.

d. 2. Mismatches:
I give the score of mismatch is -1.Mismatch value is when I compared two sequences and found the mismatch .Total mismatch value count it.

d. 3. Insertion:
I give the score for insertion is -2.Total insertions count it.

d. 4. Deletion:
I give the score for deletion is -2.Total deletion counts it.

Insert the Sequence 1:
>gi|224589803:66218240-66360071 Homo sapiens chromosome 12, GRCh37.p10 Primary Assembly

CTTGAATCTTGGGGCAGGAACTCAGAAAACTTC-
CAGCCCGGGCAGCGCGCGCTTGGTGCAAGACTCAG-
GAGCTAGCAGCCCGTCCCCCTCCGACTCTCCGGTGC-
CGCCGCTGCCTGCTCCCGCCACCCTAGGAGGCGCGGT-
GCCACCCACTACTCTGTCCTCTGCCTGTGCTCCGTGC-
CCGACCCTATCCCGGCGGAGTCTCCCCATCCTCCTTT-
GCTTTCCGACTGCCCAAGGCACTTTCAATCTCAATCTCTT
CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
CTCTCTCTCTCTCGCAGGGTGGGGGGAAGAGGAGGAG-
GAATTCTTTCCCCGCC

Insert the Sequence 2:
CCTTTGCTTTCCGACTGCCCAAGGCACTTTCAATCT-
CAATCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TCTCTCTCTCTCTCTCTCTCTCTCTCTCGCAGGGT
GGGGGGAAGAGGAGGAGGAATTCTTTCCCCGC-

Volume : 4 | Issue : 7 | July 2014 | ISSN - 2249-555X

CTAACATTTCAAGGGACACAATTCACTCCAAGTCTCTTC-
CCTTTCCAAGCCGCTTCCGAAGTGCTCCCGGTGCC-
CGCAACTCCTGATCCCAACCCGCGAGAGGAGCCTCT-
GCGACCTCAAAGCCTCTCTTCCTTCTCCCTCGCTTC-
CCTCCTCCTCTTGCTACCTCCACCTCCACCGCCAC-
CTCCACCTCCGGCACCCACCCACCGCCGCCGCCGC-
CACCGGCAGCGCCTCCTCCTCTCCTCCTCCTCCC-
CTCTTCTCTTTTTGGCAGCCGCTGGACG

In this page I put the sequences and give the match score is 1, mismatch score is -1and gap penalty is -2.

e. Output:The out put is given by tool is matches are 212 and gap penalty are 0. (Figure5).

CONCLUSION

In bioinformatics, a sequence alignment is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. Aligned sequences of nucleotide or amino acid residues are typically represented as rows within a matrix. Gaps are inserted between the residues so that identical or similar characters are aligned in successive columns. If two sequences in an alignment share a common ancestor, mismatches can be interpreted as point mutations and gaps as indels (that is, insertion or deletion mutations) introduced in one or both lineages in the time since they diverged from one another. In sequence alignments of proteins, the degree of similarity between amino acids occupying a particular position in the sequence can be interpreted as a rough measure of how conserved a particular region or sequence motif is among lineages. The absence of substitutions, or the presence of only very conservative substitutions (that is, the substitution of amino acids whose side chains have similar biochemical properties) in a particular region of the sequence, suggest that this region has structural or functional importance. Although DNA and RNA nucleotide bases are more similar to each other than are amino acids, the conservation of base pairs can indicate a similar functional or structural role. When I compared two sequences then I concluded that this sequence is these sequences are very similar to each other. They show the functional and evolutionary relationship to each other. The sequences are showing the highest similarity to one another.

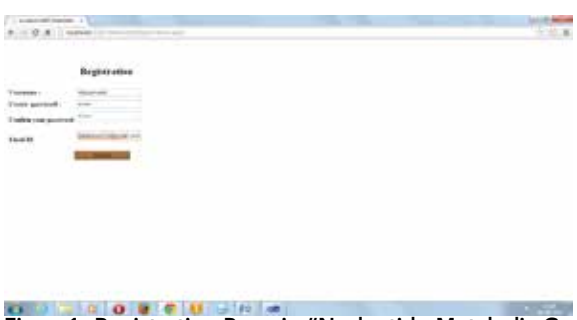

Figure1: Registration Page in "Nucleotide Metabolic Genetic Database".

Figure2: Login Page in "Nucleotide Metabolic Genetic Database".

Figure3: Home Page of "Gene Analysis Tool".

Figure4: DNA Sequence Alignment.

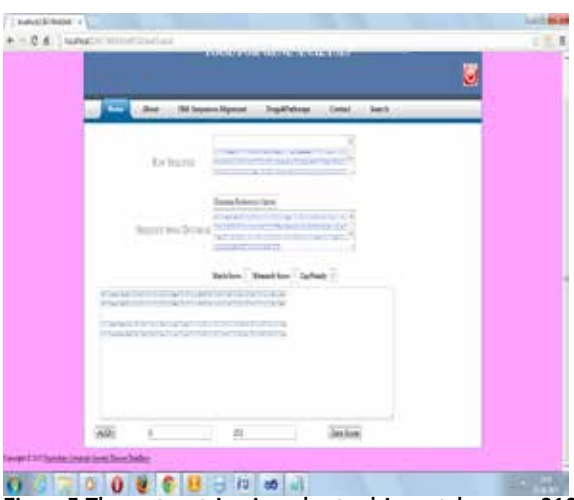

Figure5:The out put is given by tool is matches are 212 and gap penalty are 0

Volume : 4 | Issue : 7 | July 2014 | ISSN - 2249-555X

REFERENCE

1. Cohen E, Krause I, Fraser A, Goldberg E, Garty M. (2012), "Hyperuricemia and metabolic syndrome: lessons from a large cohort from Israel. Isr Med Assoc J. 14 (11), 676-80. | 2. Coutant S, Cabot C, Lefebvre A, Léonard M, Prieur-Gaston E, Campion D, Lecroq T, Dauchel H. EVA (2012), "Exome Variation Analyzer, an efficient and versatile tool for filtering strategies in medical genomics". BMC Bioinformatics. 13 Suppl, 14:S9. doi: 10.1186/1471-2105-13-S14-S9. | 3. Simmonds H.A., Fairbanksa L.D., Morrisa G.S., Websterb, E.H. (1988), "Harleyc Altered erythrocyte nucleotide patterns are characteristic of inherited disorders of purine or pyrimidine metabolism". Clinica Chimica Acta on sciVerse Science Direct Volume 171, (2–3), 197–210. | 4. Holzinger ER, Dudek SM, Frase AT, Krauss RM, Medina MW, Ritchie MD. (2013), "Athena: a tool for meta-dimensional analysis applied to genotypes and gene expression data to predict hdl cholesterol levels". Pacific Symposium on Biocomputing, 385-396. | 5. Martí R, Nishigaki Y, Vilá MR, Hirano M. (2003), "Alteration of Nucleotide Metabolism: A New Mechanism for Mitochondrial Disorders". Clin Chem Lab Med. 41(7), 845-51. | 6. Milosevic D, Rinat C, Batinic D, Frishberg Y. (2002), "Genetic analysis--a diagnostic tool for primary hyperoxaluria type I". Pediatr Nephrol. 17(11), 896-898. | 7. Pessler F, Mayer CT, Jung SM, Behrens EM, Dai L, Menetski JP, Schumacher HR. 2008), "Identification of novel monosodium urate crystal regulated mRNAs by transcript profiling of dissected murine air pouch membranes". Arthritis Res Ther. 10(3), R64. doi: 10.1186/ar2435. | 8. Sahoo S, Franzson L, Jonsson JJ, (2012), "Thiele I. A compendium of inborn errors of metabolism mapped onto the human metabolic network". Mol Biosyst. 8(10), 2545-2558. | 9. Tomashov-Matar R, Biran G, Lagovsky I, Kotler N, Stein A, Fisch B, Sapir O, Shohat M. (2012), Severe combined immunodeficiency (SCID): from the detection of a new mutation to preimplantation genetic diagnosis. J Assist Reprod. Genet. 29(7), 687-92. | 10. Yandell M, Huff C, Hu H, Singleton M, Moore B, Xing J, Jorde LB, Reese MG. (2011), "Figure A probabilistic disease-gene finder for personal genomes". Genome Res. 21(9), 1529-42. | 11. Yung LS, Yang C, Wan X, Yu W. GBOOST (2011), "A GPU-based tool for detecting gene-gene interactions in genome-wide case control studies". Bioinformatics 27(9), 1309-131 |