Abhishek Upadhyay¹, Ilu Kumari², Dr. Suryakant Chaturvedi³ and Dr. Virendra Upadhyay⁴

¹Department of Biotechnology, Madhav Institute of Technology and Science, Race Course Road, Gola Ka Mandir, Gwalior (M.P.) India.

² Indian Academy of Biotechnology, Patna (Bihar), India.

³ Department of Biological Sciences, Mahatma Gandhi Chitrakoot Gramodaya University, Chitrakoot, Satna

(M.P.), India.

⁴ Department of Physical Sciences, Mahatma Gandhi Chitrakoot Gramodaya University, Chitrakoot, Satna(M.P.), India.

Abstract: Increasing the activity of defective cystic fibrosis transmembranes conductance regulator (CFTR) protein is a potential treatment for cystic fibrosis. The life expectancy of people with cystic fibrosis (CF), a lethal inherited disease, has been greatly extended by advances in therapy. Currently, there are a number of potential drugs for treatment of CF lung disease in clinical trials. These therapies are targeted at all points in the pathogenesis of lung disease, from gene transfer to drugs that treat mucus, infection and inflammation in the airways. An exciting development is that of modulation of the abnormal protein that causes CF, the cystic fibrosis transmembrane conductance regulator (CFTR), where drugs are targeted at specific defects in CFTR transcription, processing or functioning. A number of antibacterial agents formulated for inhalation are at various stages of study or newly approved, which should improve options for chronic management of airway infection Cystic fibrosis (CF) is a pleiotropic disease, originating from mutations in the CF transmembrane conductance regulator (CFTR). Both higher adherence to a Mediterranean-type diet and more physical activity have been independently associated with lower Alzheimer disease (AD) risk. Therefore comparative genomics analysis of these diseases is done. In this context Exons and CNS were observed in all three sequences. This paper provides information meant to increase an understanding of the public-health impact of cystic fibrosis, sickle cell and Alzheimer's diseases including incidence and prevalence, mortality, lifetime risks, costs, and impact on family caregivers.

Keywords: Fibrosis Transmembrane Conductance Regulator (CFTR) protein, Pleiotropic disease, Alzheimer disease (AD) and Comparative Genomics.

Introduction:

Genomics is a discipline in genetics concerned with the study of the genomes of organisms. The field includes efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping. The field also includes studies of intra genomic phenomena such as heterosis, epistasis, pleiotropy and other interactions between loci and alleles within the genome. There are two types of genomics-Structural and Functional genomics. Cystic fibrosis is an inherited disease of your secretory glands, including the glands that make mucus and sweat. "Inherited" means that the disease is passed through the genes from parents to children. People who have cystic fibrosis inherit two faulty cystic fibrosis genes-one from each parent. The parents likely don't have the disease themselves. ^[1] Cystic fibrosis mostly affects the lungs, pancreas, liver, intestines, sinuses, and sex organs. The symptoms of cystic fibrosis vary from person to person and over time. Sometimes you will have few symptoms. CFTR conducts thiocyanate (SCN) ions and it is important because, in several ways, they can limit potentially harmful accumulations of hydrogen peroxide (H_2O_2) and hypochlorite (OCl⁻). Oxidation of SCN⁻ to hypothiocyanite (OSCN⁻), consumes H₂O₂. Second, SCN⁻ even at low concentrations competes effectively with Cl⁻ for myeloperoxidase (MPO) (which is released by white blood cells), thus limiting OCl⁻ production by the enzyme. Third, SCN⁻ can rapidly reduce OCl⁻ without catalysis. It shows that SCN⁻ protect a lung cell line from injuries caused by H₂O₂; and that SCN⁻ protects from OCl⁻ made by MPO⁻ BCL11A serves as a barrier to HbF reactivation by known HbF inducing agents. Inactivation of BCL11A in SCD transgenic mice corrects the hematologic and pathologic defects associated with SCD through high-level pan cellular HbF induction. Thus, interference with HbF silencing by manipulation of a single target protein is sufficient to reverse SCD^{. [2]} Other times, your symptoms may become more severe. Sickle cell disease is an inherited blood disorder that affects nearly 100,000 people in the United States. Red blood cells contain hemoglobin, a protein

that carries oxygen in the blood. Normal red blood cells are round and flexible, which enables them to travel through small blood vessels to deliver oxygen to all parts of the body. ^[3] Alzheimer's disease is an irreversible, progressive brain disease. It is characterized by the development of amyloid plaques and neurofibrillary tangles, the loss of connections between nerve cells, or neurons, in the brain, and the death of these nerve cells. There are two types of Alzheimer's—early-onset and late-onset. Both types have a genetic component. ^[4]

Methods:

- 1. Sequence Retrieval from NCBI: Sequence details of the target gene for cystic fibrosis are retrieved from NCBI.
- 2. Similarity search is performed using BLAST.
- 3. The sequence of gene is saved in FASTA format for further use in comparison.
- 4. Genes from other diseases are selected so that it can be compared with gene causing cystic fibrosis using comparative tools. Alzheimer's disease and sickle cell disease are chosen and their sequence details of target gene are retrieved from NCBI.
- 5. Alzheimer's disease: gene sequence is retrieved from NCBI.
- 6. Gene sequence is saved in FASTA format for further use.
- 7. Sickle cell Disease: gene sequence is retrieved from NCBI.
- 8. Gene sequence is saved in FASTA format.
- 9. Now, with these sequences in FASTA format, comparison is done using tools.

Results:

T-COFFEE: Result of T-Coffee tool which is a multiple sequence alignment program which allows us to combine results obtained with several alignment methods, is shown below. First sequence is entered or pasted in any supported format, then it is submitted to get results of comparison.



EMBL-EBI		Ente	r Text Here	Find	Help Feedback			
Databases Tools	Research Trai	ning Indu	stry About Us	Help	Site Index 🔝 🎒			
= Help	EBI > Tools > Multiple Seq	uence Alignment >	T-Coffee					
= Jalview	Alignments Result Summary Guide Tree Submission Details Submit Another Job							
 Related Applications 	Alignment							
Multiple Sequence	Download Alignment	File Hide Colo	rs					
Phylogeny	CLUSTAL W (1.83) multiple sequence alignment							
	gi 166795291 ref N gi 28302128 ref NM gi 90421312 ref NM	G_007376.1 _000518.4 _000492.3	AAAAACTGCCCA ACA AATTGGAAGCAAATG-ACA * **	NTCACTGGGG NTCACAGCAGGTCAGAGA	AAAAGAATTTT TTT AAAAGGGTTGAGCG **			
	gi 166795291 ref N gi 28302128 ref NM gi 90421312 ref NM	G_007376.1 _000518.4 _000492.3	<mark>A</mark> -TTATGCAAAGCT GCAGGCACCCAGAGTAGTA	TCAACGCTATTTACA 	GTTT <mark>AGA</mark> C GCTT GCTTG <mark>A</mark> GCCC <mark>AGA</mark> C * *			
	gi 166795291 ref N gi 28302128 ref NM gi 90421312 ref NM	G_007376.1 _000518.4 _000492.3	TTTTGTAGCTATTGA GGCCCTAGCAGGGACCCC	AG-G-CTGACATT CTGA AGCGCCCGAGAGACCATG * **	G <mark>AGA</mark> T CAGAGGTCGCCTCT			
	gi 166795291 ref N gi 28302128 ref NM gi 90421312 ref NM	G_007376.1 _000518.4 _000492.3	<mark>A-AAGAAGTTAATCATGTO</mark> 	CCTT-CTGTC TCC <mark>AAA</mark> CTTTTTTTC <mark>A</mark> G	TTGG <mark>A</mark> GG <mark>A</mark> GGT CTGGACCAGACCAA			
	gi 166795291 ref N gi 28302128 ref NM gi 90421312 ref NM	G_007376.1 _000518.4 _000492.3	<mark>AGAAAGAGAT</mark> C C A TTTTGAGG <mark>AAAGGATACA</mark> C	GAGAATGAATACAATT CAA GACAGCGCCTGGAATTGT **	CAGG-ATCTA CAGACATATACCAA			

Result is obtained as follows, where red color of bases shows the sequence similarity

T-COFFEE, Version_8.93(Thu Aug 5 18:09:23 CEST 2010) Cedric Notredame CPU TIME:0 sec. SCORE=33 BAD AVG GOOD qi|166795291|re <mark>gi|28302128|ref</mark> gi|90421312|ref 35 19 35 cons 33 gi |166795291|re gi |28302128|ref gi |90421312|ref --<mark>ACTG</mark>C<mark>CCATCACTG</mark>--------<mark>AA</mark>-AA-** cons * GGGAAAAGAATT<mark>TT</mark>----<mark>A-</mark>TTATGCAAAGCT gi|166795291|re gi |28302128|ref gi |90421312|ref -----TTT-----G<mark>C</mark>ACCCAGGTA ** cons gi | 166795291 | re gi | 28302128 | ref -----<mark>TCAACGCTATTTACAGTT</mark>-----<mark>TAGA</mark> -----<mark>GCTT</mark>-----GTAGG<mark>TCTTTGGCATTAGGAGCCTT</mark>GAGCC<mark>CAGA</mark> gi 90421312 ref * * cons CT---TTTGTAGCTATTGAAG-G-CTGA gi|166795291|re gi|28302128|ref gi|90421312|ref CTGA-----CGGCCCCAGCAGGGACCCCAGCCCGAGAGAGAC

T-Coffee result in colour format is shown below.

PAIRWISE SEQUENCE ALIGNMENT-EMBOSS NEEDLE: It is used to identify regions of similarity that may indicate functional, structural or evolutionary relationships between two biological sequences. Result of EMBOSS Needle are as follows. Sequence is entered and submitted.

EMBL-EBI		Enter Text Here	Find	
Databases Tools	Research Training	Industry About Us	s Help	Site Index 🔂 🚄
= Hein	EBI > Tools > Pairwise Sequence	e Alignment > EMBOSS Needle		
= EMBOSS website	EMBOSS Needle - Pa	airwise Sequence Alignm	ient	
= Programmatic Access = Download	EMBOSS Needle reads two alignment to file.	nucleotide or protein input sequ	uences and writes their optim	nal global sequence
	Use this tool			
 Similar Applications 	STEP 1 - Enter your nucleo	tide sequences		
EMBOSS Needle	Enter or paste your first nuc	leotide sequence in any suppo	orted format:	
related literature Search for EMBOSS Needle related literature in Medline more	G CTGTCAAGCCGTGTTCTA A AATTTGATGAAGGACTTG G GCTAATCTGGGAGTTGTT T CAGGCTGGGCTAGGGAGA	GATAAAATAAGTATTGGACAA CATTGGCACATTTCGTGTGGA ACAGGCGTCTGCCTTCTGTGG ATGATGATGAAGTACAGAGAT	CTTGTTAGTCTCCTTTCCAA .TCGCTCCTTTGCAAGTGGCJ .ACTTGGTTTCCTGATAGTCA .CAGAGAGCTGGGAAGATCA(ACAACCTGAAC ACTCCTCATGG
	Or, upload a file:	Browse		
	AND Enter or posts your cocond	nucleatide coquence in any o	upported format	
	CCTGAGAACTTCAGGCTC A CCCCACCAGTGCAGGCTG	CTGGGCAACGTGCTGGTCTGT CCTATCAGAAAGTGGTGGCTG	GTGCTGGCCCATCACTTTGC	3CAAAGAATTC 📩
	A CTAAGCTCGCTTTCTTGC T	TGTCCAATTTCTATTAAAGGT	TCCTTTGTTCCCTAAGTCCJ	астастааас
	GGGGGGATATTATGAAGGG	CCTTGAGCATCTGGATTCTGC	CTAATAAAAAACATTTATT1	FTCATTGC 💌
	Or, upload a file:	Browse		

Result of EMBOSS Needle of sequence of Alzheimer's disease and Cystic fibrosis.

# # #	Score: 911.0			
N	G_007376.1	1	AAAAACTGCCCATCACTGGGGAAAAGAATTITATTATGCAAAGCTTCAAC	50
N	M_000492.3	1		0
N	G_007376.1	51	GCTATTTACAGTTTAGACTTTTGTAGCTATTGAAGGCTGACATTGAGATA	100
N	M_000492.3	1	AATTG	5
N	G_007376.1	101	AAGAAGTTAATCATGTCCTTCTGTCTTGGAGGAGGTAGAAAGA	143
N	M_000492.3	6	GAAGCAAATGACATCACAGCAGGTCAGAGAAAAAGG	41
N	G_007376.1	144	GATGAGAATGAATACA-ATTCAGGATCTACTTCTGGTCTTTGATG	187
N	M_000492.3	42	GTTGAGCGGCAGGCACCCAGAGT-AGTAGGTCTTTGGCATT	81
N	G_007376.1	188	AGGAGTTAGCACACGGTTCTGGGAGGAAAGAC	219
N	M_000492.3	82	AGGAGCTTGAGCCCAGACGGCCCTAGCAGGGACCCCAGCGCCCGAGAGAC	131
N	G_007376.1	220	AGGTTAAGAGGCATGTGAAACTCT	243
N	M_000492.3	132	CATGCAGAGGTCGCCTCTGGAAAAGGCCAGCGTTGTCTCCAAACTTTTTT	181
N	G_007376.1	244	ATAC-GTCACTGCGTCTGC	264
N	M_000492.3	182	TCAGCTGGACCAGACCAATTTTGAGGAAAGGATACAGACAGCGCCTGG	229
N	G_007376.1	265	CAACGTACATGATACCCAGCAAGCTCAC-	292
N	M_000492.3	230	AATTGTCAGACAT-ATACCAAATCCCTTCTGTTGATTCTGCTGACA	274
N	G_007376.1	293	ATCTTCATGGAAAGCATGG	311
N	M_000492.3	275	ATCTATCTGAAAAATTGGAAAGAGAATGGGATAGAGAGCTGGCTTCAAAG	324

Here, ----- represents gaps and dot will represent the mismatch of the sequences whereas lines between two sequences shows the similarity. In the shown figure number of similar bases are 145. Mentioned figure shows that there are 41 mismatches. Result of EMBOSS Needle for the sequence of Sickle cell disease and Alzheimer's disease.

# Score: 571.5 # #			
#			
NM_000518.4	1		0
NG_007376.1	1	AAAAACTGCCCATCACTGGGGAAAAGAATTTTATTATGCAAAGCTTCAAC	50
NM_000518.4	1		0
NG_007376.1	51	GCTATTTACAGTTTAGACTTTTGTAGCTATTGAAGGCTGACATTGAGATA	100
NM_000518.4	1		0
NG_007376.1	101	AAGAAGTTAATCATGTCCTTCTGTCTTGGAGGAGGTAGAAAGAGATGAGA	150
NM_000518.4	1		0
NG_007376.1	151	ATGAATACAATTCAGGATCTACTTCTGGTCTTTGATGAGGAGTTAGCACA	200
NM_000518.4	1		0
NG_007376.1	201	CGGTTCTGGGAGGAAAGACAGGTTAAGAGGCATGTGAAACTCTCAAATAC	250
NM_000518.4	1	ACATTTGCTTCTGACAACTGTGTTCACTAGCAACCTCAA	41
NG_007376.1	251	TI TITTI TITI TITTI TITTI TITTI TITTI TITTI TITI	292
NM_000518.4	42	ACAGACACCATGGTGCATCTGACTCCTGAGGAGAAGTCT	80
NG_007376.1	293	ATCTTCATGGAAAGCATGGTAATTCCCAACACTACCGGAAGTCT	336
NM_000518.4	81	GCCGTTACTGCCCTGTGGGGGCAAGGTGAACGTGGATGA	118
NG_007376.1	337 GGAGTGGCTAAGTAATCCATATATTCAACCAGGAAGC	373
NM_000518.4	119	AGTTGGTGGTGAGGCCCTGGGCAGGCTGCTGGT	151
NG 007376.1	374	II II.I.I.IIIIIIIIIIIIIIIII AGCTAAAGAAATATTCTAATTACCTAGGAAGGTTTCTGATITCAAA	419
_			

Here, ---- represents gaps and dots represents the mismatch whereas line between the two sequences represents the similarity between the sequences. In the shown figure there are 89 similar bases. Mentioned figure shows that there are 31 mismatches.

CLUSTALW2: It is a general purpose multiple sequence alignment program for DNA or proteins. To compare the sequences, sequences are first submitted in supportable form as shown below.

EMBL-EBI			Enter Text H	lere		Find		
Databases Tools	Research	Training	Industry	About Us	Help		Site Index 🚦	🐋 <i></i>
= Help = FAQ = Clustal website = Jalview = Programmatic Access = Download	EBI > Tools > Mult ClustalW2 - ClustalW2 is a g New version! Cl Use this tool	iple Sequence Al Multiple Se Jeneral purpos Justal Omega is	ignment > Clustel quence Alig e multiple sequ now available t	ment ence alignment (for protein seque	program for D inces - give it	NA or protein a tryl	16.	
Related Applications Pairwise Sequence Alignment Multiple Sequence Alignment Phylogeny Clustal related Reference	STEP 1 - Enter Enter or paste A CCCCACCAG A CTAAGCTCGG T GGGGGGATAT	your input see a set of DNA recaescree rateaagecce rateaagecce	Sequences Sequence TATCAGAAAG TTCGAATTTCT.	es in any suppor TGGTGGCTGGT ATTAAAGGTTC GGATTCTGCCT.	ted format: GTGGCTAATC CTTTGTTCCC AATAAAAAA	SCCCTGGCC STAAGTCCA CATTTATTT	CACAAGTATC ACTACTAAAC TCATTGC	
Search for Clustel related itterature in Medine more	Or, upload a fill STEP 2 - Set y Alignment Typ <i>The default set</i> (More options) STEP 3 - Set y	e: our Pairwise A a: Slow tings will fulfill t (Click here our Multiple Se	lignment Option Fast he needs of mo , if you want to v quence Alignm	Browse IS of users and, for view or change th ent Options	that reason, a he default sett.	ire not visible ings)	Local Ar	rea Connect
	The default set	tinga will fulfill t	he needs of mo	at users and, for	that reason, a	re not visible	Speed: 100.0	Mbps

After submission of sequences, one link is obtained through which we can access the comparison results.



© Copyright European Bioinformatics Institute 2002-2011. All Rights and Trademarks Reserved.

EMBL-EBI		Enter Text Here		Find	Help Feedback	
Databases Tools	Research Training	Industry	About Us Help		Site Index <u>ର</u> 🎒	
= Help	EBL > Tools > Multiple Sequence Alig	;nment > ClustalW2				
= FAQ	ClustalW2 Results					
= Jalview	Alignments Result Summar	y Guide Tree	Submission Details	Submit Another Job	0	
	Alignment					
 Related Applications 	Download Alignment File Hi	de Colors				
Multiple Sequence Alignment Phylogeny	CLUSTAL 2.1 multiple sequ	ence alignment				
	gi 90421312 ref NM_000492	.3] AATT	GGAAGCAAATGACATCA	CAGCAGGTCAGAGAA.	AAAGGGTTG <mark>A</mark> GCGG 50	
	gi 28302128 ref NM_000518 gi 166795291 ref NG_00737	6.1	AAAAACTGCCCATCA	CTGGGG <mark>AAAA</mark> GAATT	FTATT-ATGC <mark>AAA</mark> G 43	
	gi 90421312 ref NM_000492	.3 CAGG	CACCCAGAGTAGTAGGT	CTTTGGCATTAGGAG	CTTG <mark>A</mark> GCCC <mark>AGA</mark> CG 100	
	gi 166795291 ref NG_00737	6.1) CTTC.	AACGCTATTTACAG	TTTAGACTTTTGTAG	CTATTGAAGGCT 88	
	gi 90421312 ref NM_000492 gi 28302128 ref NM_000518	.3 GCCC	T <mark>a</mark> gc <mark>a</mark> gggaccccagcg	CCCGAGAGACCATGC.	AGAGGTCGCCTCTG 150	
	gi 166795291 ref NG_00737	6.1) GACA	FTGAGATAAAGAAGTTA	ATCATGTCCTTCTGT	CTTGG <mark>A</mark> GG <mark>A</mark> GGT <mark>A</mark> G 138	
	gi 90421312 ref NM_000492 gi 28302128 ref NM_000518	.3 GAAA	AGGCCAGCGTTGTCTCC	AAACTTTTTTTCAGC	TGG <mark>ACCAGACCAA</mark> T 200	
	gi 166795291 ref NG_00737	6.1 AAAG	AGATGAGAATGA	ATACAATTCAGGATC'	F <mark>A</mark> CTTCTGGTCTTT 183	
	gi 90421312 ref NM_000492 gi 28302128 ref NM_000492	.3 TTTG	AGGAAAGGATACAGACA	GCGCCTGG <mark>AA</mark> TTGTC.	AGACATATACCAAA 250	
	gi 166795291 ref NG_00737	6.1 GATG	AGGAGTTAGCACACG	GTTCTGGG <mark>A</mark> GG <mark>AA</mark>	AGACAGGTTAAGAG 229	

Here, red color of bases shows the similarity of the sequences. There are 30 similar bases in the shown figure. DOTLET: It is a program for comparing sequences by the diagonal plot method. The sequences are entered and then computed to get the results as follows.

Sequence of Alzheimer's disease is compared with Cystic fibrosis



Each pixel represents a score. High score means good match. Pixel's colour depends on how similar the two sequences are. Darker the pixel, lower the score. Histogram window represents the frequency of each score, on linear (blue) and logarithmic (purple). Lowest possible score on the left and the highest on the right. Larger peak represents low score. Here in horizontal column sequence1 is used and second sequence is used in vertical column. 1:1 zoom is used for observing the results. 41 percent grey scale is used out of 100 to view the results. The grey scale can be adjusted accordingly.

Comparison of sequence of Sickle cell disease with Alzheimer's disease



Below the comparison window, sequence similarity can be seen covering few bases at a time. GENOME VISTA: It uses a computational strategy where query sequence contigs are anchored on the base genome by local alignment matches and then globally aligned to candidate regions by AVID program. Here, human march 2006 genome is used as a base genome for comparison of sequences.

Result of genome VISTA for Cystic fibrosis.

When sequences are submitted then one link is obtained through which result can be accessed as follows.

VIST	Ά		Tools for	Comparati∨	e Genomi	cs	🔏 About Us	💭 <u>Cite Us</u>	🙆 <u>Contact Us</u>
VISTA Home	Custo	m Alignment	Browser	Enhancer DB	Downloads	Publications	Training	Help	
mVISTA r ⁱ	VISTA	GenomeVIST/	wgVISTA	Phylo-VISTA					
GenomeVI Submit About Genom 	GenomeVISTA GenomeVISTA Submit: » Submit Your request has been submitted. The results will be available at: About GenomeVISTA About GenomeVISTA http://pipeline.lbl.gov/cgi-bin/textBrowser2?act=summary&run=u380-NXtnYKqv&base=hg18								
Cite Your Unique Run Identifier is u380-NXtnYKqv Once they are ready, you can retrieve the results by following the link above or by entering the Run Identifier on the GenomeVista page									
JGIS <u>Doeja</u>	oint Genor	ne Institute	C	1997-2011 The Reg	ents of the Univer	sity of California		U.S. DEPARTME	GY Office of Science

When link is opened the result is obtained in vista point format.

Human Mar. 2006 (chr7:117,015,024-117,039,	D99)	View with: 📊 Vie	sta Browser 🔀 Vista Point	🔩 Synteny 💉 Dot Plot 📪 Help
clade: genome: Vertebrate Vertebrate	release: alignment:	[381] 🕑 s	add alignment: Submit 7 Vertebrates [80]	Add R
<				
chr7 🕑 chr7:117,015,024-117,039,099	Go -10x -3x -1.5x -	+1.5x +3x +10x 44 4	▶ ▶ length: 24,076 bp	
mb 20 40	1 60 1	1 1 80 100 1 1		140
CFTR	•			
sequence1				

		Show alignments longer than	50 bp 💌 🗐 Download RefSeq genes 🏾 🛎
sequence1 [381]			
	🔁 <u>alignment in PDF</u> get CNSs	s: Human Mar. 2006 - sequence1 🛛 🚺 🍕	Page 1 of 1 🕨 🕅 😂 Displaying 1 - 2 of 2
Location on Human Mar. 2006 Mar 2006	Tools	Location on sequence1 Unknown	
chr7:117,015,024-117,019,948 (+) Sequence (softmasked) length: 4,925 bp	rankVISTA pairwise alignments: Alignment:Human Mar. 2006 - sequence1 MFA: Human Mar. 2006 - sequence1 CNS: Human Mar. 2006 - sequence1 Vista: Human Mar. 2006 - sequence1 PDF: Human Mar. 2006 - sequence1	gi]90421312 ref NM_000492.3 :1,712-2,623 (+) <u>Sequence</u> length: 912 bp <u>Vista Point</u> <u>Vista Browser</u>	

Result in VISTA browser format is obtained as follows.



Result in VISTA Track format is shown below.



Result for Alzheimer's disease.

Result as VISTA point is shown below.

chr21 💌	chr21:26,468,	674-26,470,003	Go -10x	-3x -1.5x +1.5	x +3x +10x	44 4 >	▶ length: 1,330) bp	
mb	5	10 1	15 1	20 1	25	30 1	35 1	40 1	45
	I	11			1 1		I	11 1	1 1 1
sequence1									<u>14</u>
									T

		Show alignments longer than	50 bp 💌 🚯 Download RefSeq genes 🔻
sequence1 [428]			
	🔀 <u>alignment in PDF</u> get CNSs	s: Human Mar. 2006 - sequence1 🛛 🚺 🍕	Page 1 of 1 🕨 🕅 😂 Displaying 1 - 1 of 1
Location on Human Mar. 2006 Mar 2006	Tools	Location on sequence1 Unknown	
chr21:26,468,674-26,470,003 (+) Sequence (softmasked) length: 1,330 bp	rankVISTA pairwise alignments: Alignment <u>Human Mar. 2006 - sequence1</u> MFA: <u>Human Mar. 2006 - sequence1</u> CNS: <u>Human Mar. 2006 - sequence1</u> rVista: <u>Human Mar. 2006 - sequence1</u> PDF: <u>Human Mar. 2006 - sequence1</u>	gi 166795291 ref NG_007376.1 :1-1,330 (-) <u>Sequence</u> length: 1,330 bp <u>Vista Point</u> <u>Vista Browser</u>	

Result in VISTA Browser format is as follows





Result for Sickel cell anemia. Result in VISTA point format.

chr11 🔽 chr11:5,203,272-5,204,876	Go -10x -3x -1.5x +1.5x	+3x +10x 44 + >> length:	1,605 bp
mb > < 20	40 60 1 1	80 100 1 1	120
Нвв			
sequence1	<u> </u>		······································
		Show alignments longer than	50 bp 💌 🔂 Download RefSeq genes
sequence1 [441]			
	🔁 alignment in PDF get CNSs	🛚 Human Mar. 2006 - sequence1 🛛 🚺 🍕	Page 1 of 1 🕨 🕅 😂 Displaying 1 - 2 of 2
Location on Human Mar. 2006 Mar 2006	Tools	Location on sequence1 Unknown	
chr11:5,203,272-5,203,533 (+) <u>Sequence (softmasked)</u> length: 262 bp	rankVISTA pairwise alignments:	gi[28302128 ref]NM_000518.4 :365-626 (-) <u>Sequence</u> length: 262 bp	• • • • • • • • • • • • • • • • • • •
	Alignment <u>Human Mar. 2006 - sequence1</u> MFA: <u>Human Mar. 2006 - sequence1</u> CNS: <u>Human Mar. 2006 - sequence1</u>	<mark>I Vista Point</mark> <mark>III Vista Browser</mark>	
	PDF: Human Mar. 2006 - sequence1	i	J 🔍

Result in VISTA Browser format.



Comparing sequences of the three diseases by using VISTA TOOL, it is clear that untranslated sequence is observed only in Sickle cell disease at position Chromosome11 between 5,203,272 -5,204,876. Exons and CNS is observed in all three sequences. So, by using these tools we analyze gene of our interest and compare it with many other sequences to get the similarity results which is biologically very useful in different aspects of life.

Discussion:

When sequencing is done using different tools, then it is noticeable that two or more sequence varies with respect to each other by very few bases. Knowing these variations and their positions, possible actions can be taken to exploit these sequences or if these causes disease then drugs can be made accordingly. Different tools give results in different format. Like in T coffee and in Clustal-W color variation shows the sequence variation. In pair wise alignment tool ----represents gaps and ... represents mismatch. In Dotlet each pixel represents a score and high score means good match. In genome vista similarity is shown graph wise. Gene similarity results are very useful and can be used for different research work.

Conclusion:

Sequence comparison based on T-Coffee Version 8.9 was having consensus similarity of 33 score in which first sequence was having 35 score, second sequence was having 19 score and third sequence of score 35. When sequence is compared with T-Coffee then multiple sequences can be taken at a time while with the Pair wise Alignment Tool, two sequences is taken at a time. In T-Coffee similarity is Maximum for Adenine but it is not in the case of Pair wise Alignment Tool. In T-Coffee result, score is obtained for the similarity but score is not obtained in Pair wise Alignment TOOL. During the genomic sequence comparison of Cystic Fibrosis with Alzheimer's disease 19 amino acids were found to be similar in one window. While in case of AD with SCD there were 20 amino acids similar. So the maximum genomic similarity is for AD with SCD. When sequence of three diseases is compared by VISTA TOOL, then UTR is observed only in SCD at position of Chromosome 11 between 5,203,272 -5,204,876. Exons and CNS are observed in all three sequences.

Acknowledgement:

Authors are thankful to Madhav Institute of Technology and Science (MITS), Gwalior (M.P.) for providing computing facilities. Abhishek Upadhyay is also thankful to his parents for endowing him with their exceptional genes.

References:

- A CFTR Potentiator in Patients with Cystic Fibrosis and the G551D Mutation, Bonnie W. Ramsey, M.D., N Engl J Med 2011; 365:1663-3,16November 2011.
- [2]. Emerging Therapies in Cystic Fibrosis, Paula Anderson, MD, Posted: 09/10/2010; Ther Adv Resp Dis. 2010; 4(3):177-185. © 2010 Sage Publications, Inc.
- [3]. the antioxidant role of thiocyanate in the pathogenesis of cystic fibrosis and other inflammation-related diseases, Communicated by Clay M. Armstrong, University of Pennsylvania Medical Center, Philadelphia, PA, October 3, 2009 (received for review June 3, 2009).
- [4]. Allergenic Hematopoietic Stem-Cell Transplantation for Sickle Cell Disease, Matthew M. Hsieh, M.D., Elizabeth M. Kang, M.D., Courtney D. Fitzhugh, N Engl J Med 2009; 361:2309-2317December 10, 2009.
- [5]. Alzheimer's & Dementia: The Journal of the Alzheimer's Association, Volume 6, Issue 1, Pages 1-10.e1, January 2010.
- [6]. Alzheimer's & Dementia: The Journal of the Alzheimer's Association, Volume 5, Issue 3, Pages 234-270, May 2009.