IOSR Journal of Pharmacy and Biological Sciences (IOSRJPBS)
ISSN : 2278-3008 Volume 1, Issue (July-August 2012), PP 26-38
www.iosrjournals.org

Genomic Analysis and comparative genomics in reference to cystic fibrosis, sickle cell and Alzheimer's diseases.

Abhishek Upadhyay[1], Ilu Kumari[2], Dr. Suryakant Chaturvedi[3] and Dr. Virendra Upadhyay[4]

[1]*Department 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 H_2O_2. 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 H_2O_2; 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

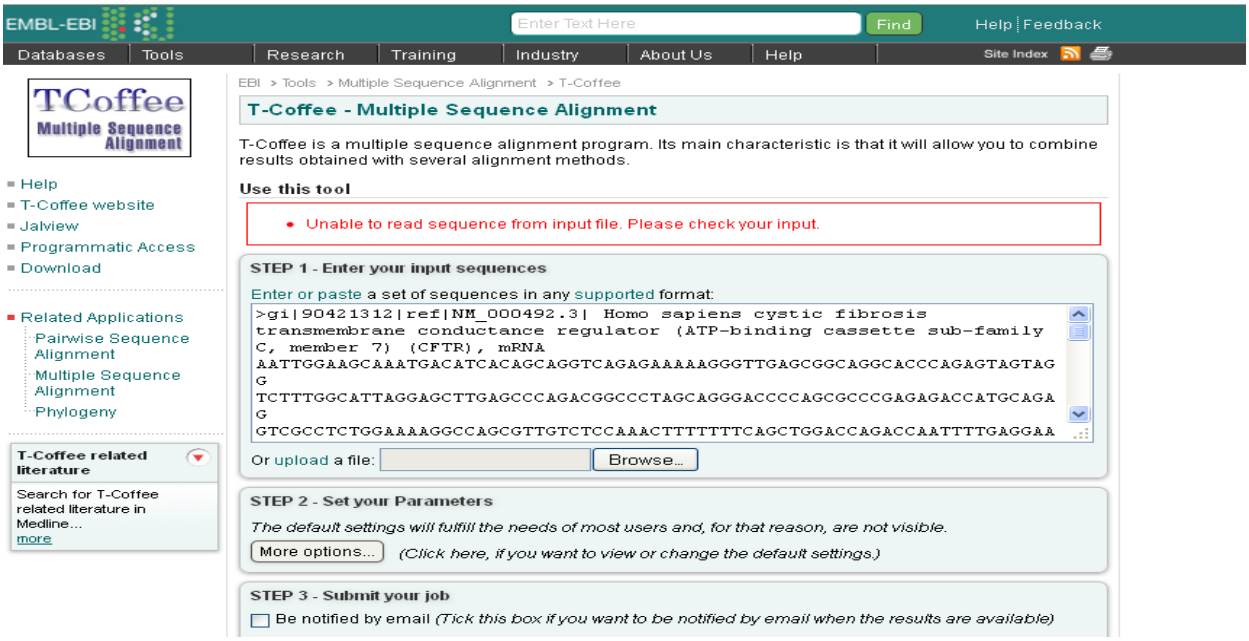
that carries oxygen in the blood. Normal red blood cells are round and flexible, which enables them to travel through small blood vessels to deliver oxygen to all parts of the body. [3] Alzheimer's disease is an irreversible, progressive brain disease. It is characterized by the development of amyloid 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]

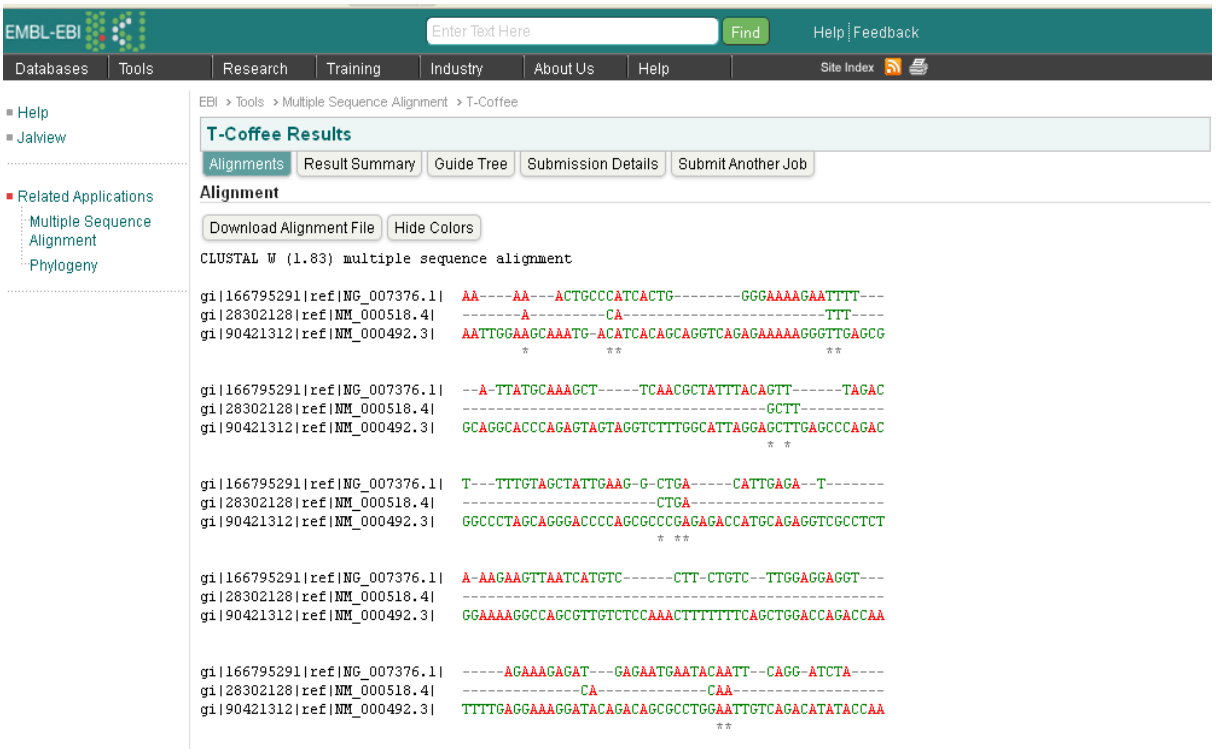
Methods:

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.

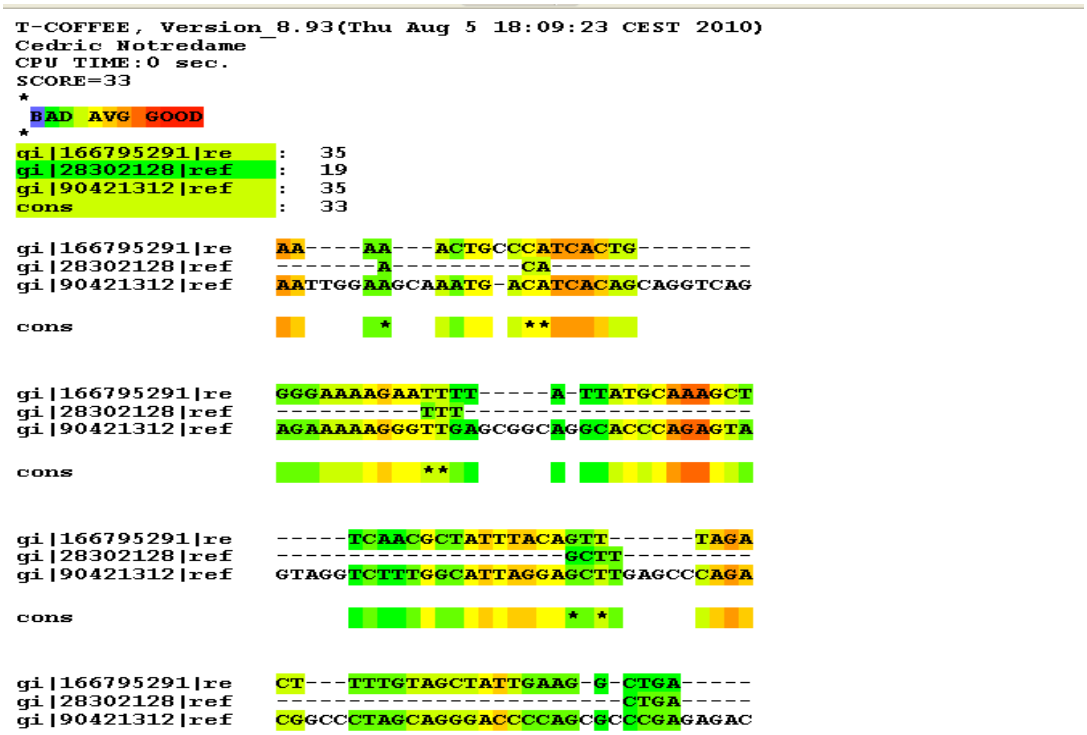
Results:

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.



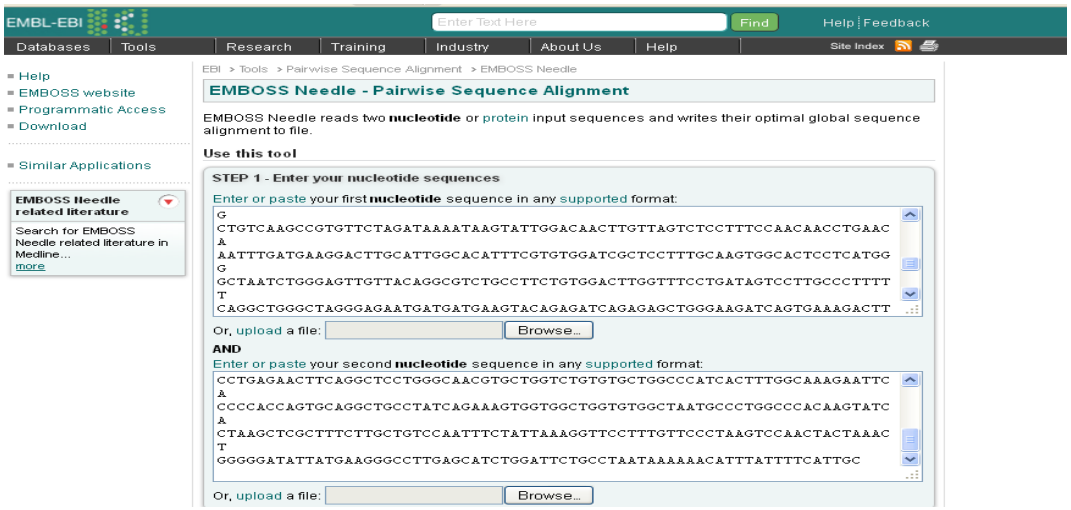


Result is obtained as follows, where red color of bases shows the sequence similarity



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.



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

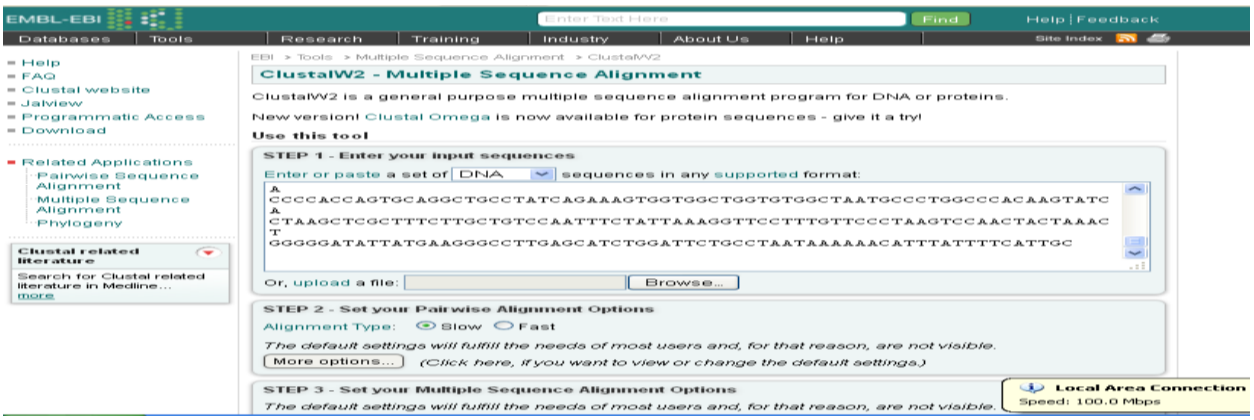
```
# Score: 911.0
#
#
#=======================================

NG_007376.1         1 AAAAACTGCCCATCACTGGGGAAAAGAATTTTATTATGCAAAGCTTCAAC     50

NM_000492.3         1 ------------------------------------------------------      0

NG_007376.1        51 GCTATTTACAGTTTAGACTTTTGTAGCTATTGAAGGCTGACATTGAGATA    100
                                                        .||||
NM_000492.3         1 ------------------------------------------AATTG-----      5

NG_007376.1       101 AAGAAGTTAAT--CATGTCCTTCTGTCTTGGAGGAGGT-----AGAAAGA    143
                        ||||..|||  |||..|            ||.||||     |.|||||
NM_000492.3         6 --GAAGCAAATGACATCAC-----------AGCAGGTCAGAGAAAAAGG     41

NG_007376.1       144 GATGA---GAATGAATACA-ATTCAGGATCTACTTCTGGTCTTTG--ATG    187
                      |.|||   |.|.|.|..|| |.| ||.|         ||||||||  ||.
NM_000492.3        42 GTTGAGCGGCAGGCACCCAGAGT-AGTA---------GGTCTTTGGCATT     81

NG_007376.1       188 AGGAGTTAG----CACACGGTTCTGGGAG--------------GAAAGAC    219
                      |||||.|.|    ||.||||..||.|.||               ||.||||
NM_000492.3        82 AGGAGCTTGAGCCCAGACGGCCCTAGCAGGGACCCCAGCGCCCGAGAGAC    131

NG_007376.1       220 -------AGGT--------TAAGAGGCATGTG--------AAACTCT---    243
                              ||||        .||.||||..|.|        ||||||.|
NM_000492.3       132 CATGCAGAGGTCGCCTCTGGAAAAGGCCAGCGTTGTCTCCAAACTTTTTT    181

NG_007376.1       244 ---------------CAA-------------ATAC-GTCACTGCGTCTGC    264
                                      |||             |||| |  ||.|||.|||.
NM_000492.3       182 TCAGCTGGACCAGACCAATTTTGAGGAAAGGATACAG--ACAGCGCCTGG    229

NG_007376.1       265 CAACGT---ACATGATACCCAGCAA------------------GCTCAC-    292
                      .|..||   |||| ||||    |||                   |||.||
NM_000492.3       230 AATTGTCAGACAT-ATAC----CAAATCCCTTCTGTTGATTCTGCTGACA    274

NG_007376.1       293 ----ATCTTCA----TGGAA--AGCATGG---------------------    311
                          ||||..|    |||||  ||.||||
NM_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.

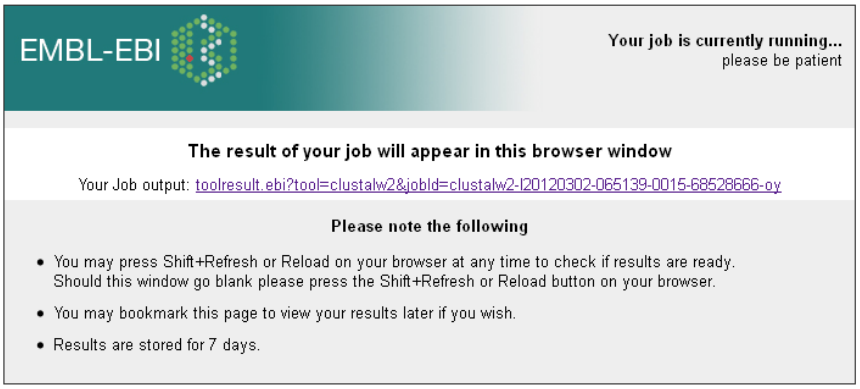
```
# Score: 571.5
#
#
#=======================================

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 ---ACATTTGCTTCTGACACAACTGTGTTCA------CTAGCAACCTCAA     41
                         ||   |||.|||| .||||   ||.||      |.||||.||||.
NG_007376.1      251 GTCAC---TGCGTCTG--CCAAC---GTACATGATACCCAGCAAGCTCAC    292

NM_000518.4       42 A----CA--GACACCATGGT-----GCATCTGACTCCTGAGGAGAAGTCT     80
                     |    ||  ||.|.||||||     .||.|  |||.|.|    |||||||
NG_007376.1      293 ATCTTCATGGAAAGCATGGTAATTCCCAAC--ACTACCG----GAAGTCT    336

NM_000518.4       81 GCCGTTACTGCCCTGTGGGGCAAGG------------TGAACGTGGATGA    118
                     |             |.|.|||.|.|            |.||||..|||.|.
NG_007376.1      337 G-------------GAGTGGCTAAGTAATCCATATATTCAACCAGGAAGC    373

NM_000518.4      119 AG-----------TTGGTGGTGAGGCCCTGGGCAGGCTGCTGGT------    151
                       ||           ||    .|.|...|||.||.|||.|.|||.|
NG_007376.1      374 AGCTAAAGAAATATT----CTAATTACCTAGGAAGGTTTCTGATTTCAAA    419
```

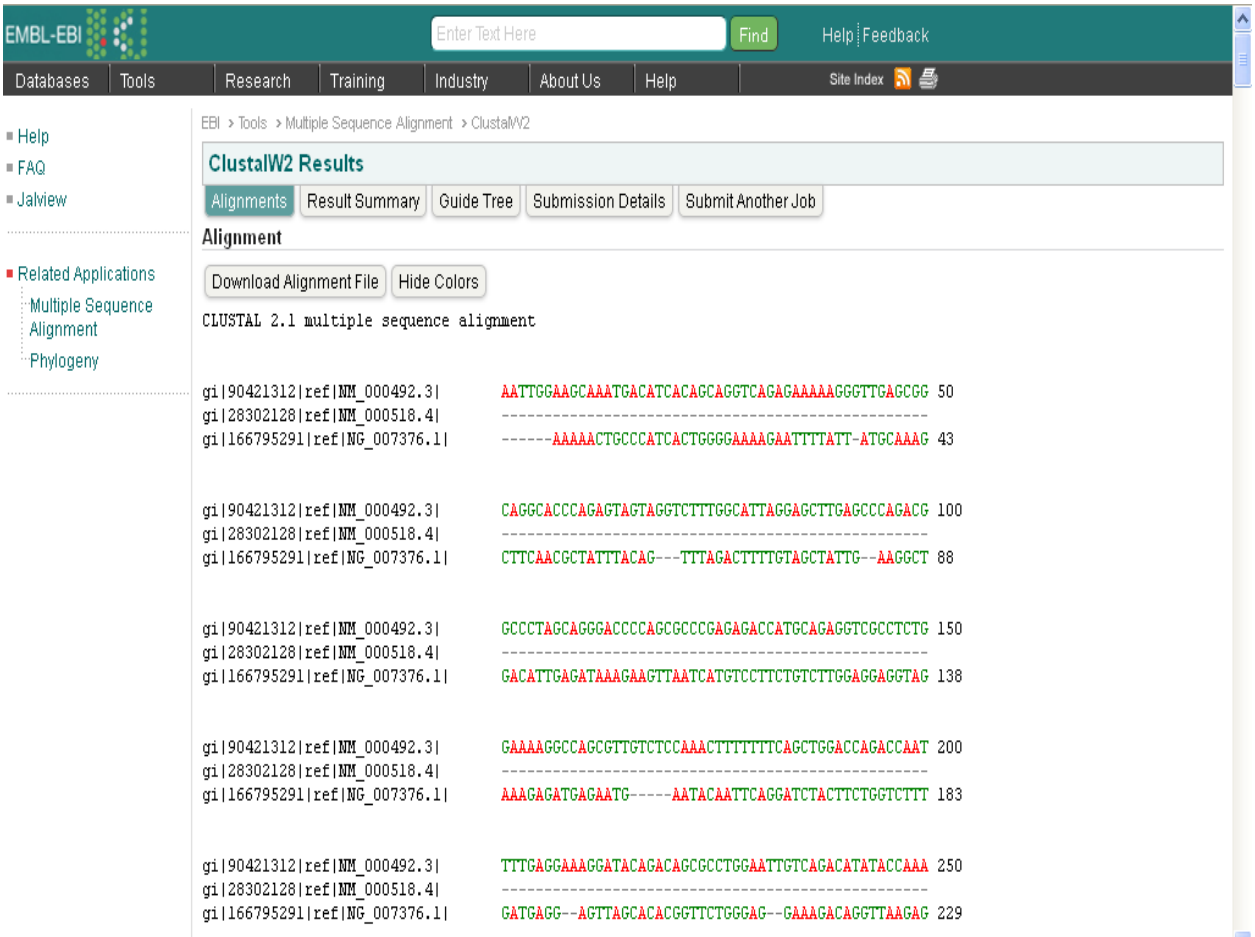
Here, ---- represents gaps and dots represents the mismatch whereas line between the two sequences represents the similarity between the sequences. In the shown figure there are 89 similar bases. Mentioned figure shows that there are 31 mismatches.

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.

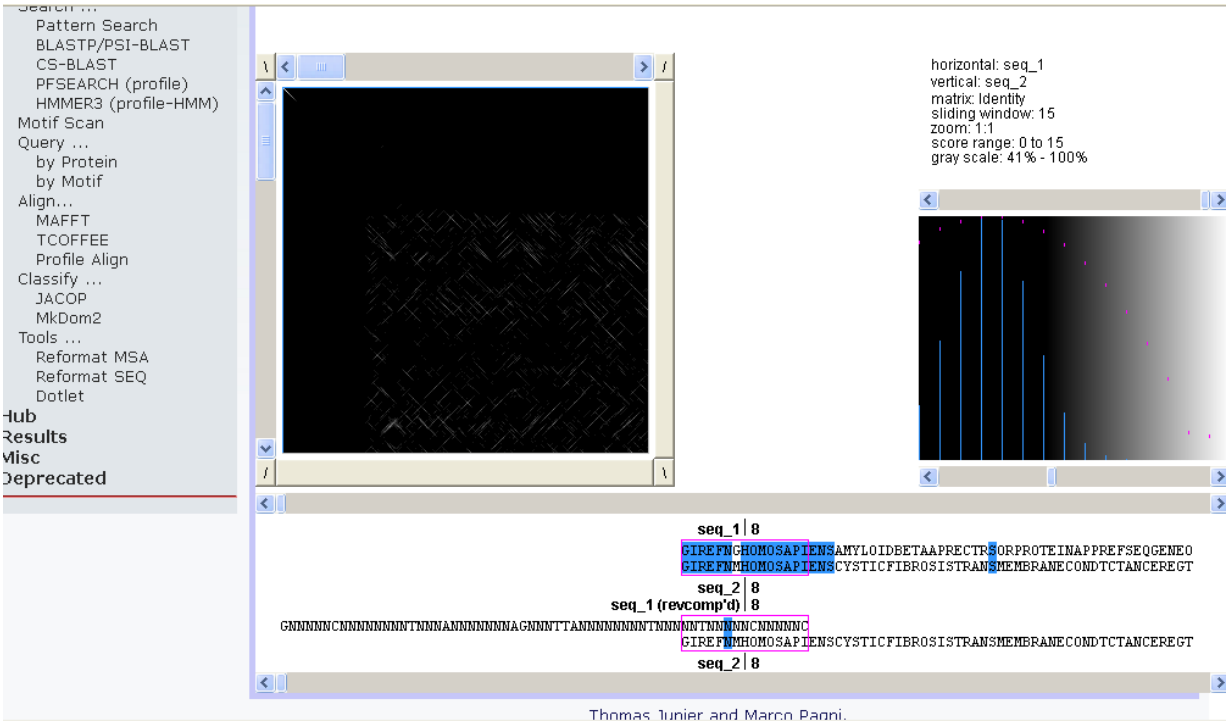


After submission of sequences, one link is obtained through which we can access the comparison results.



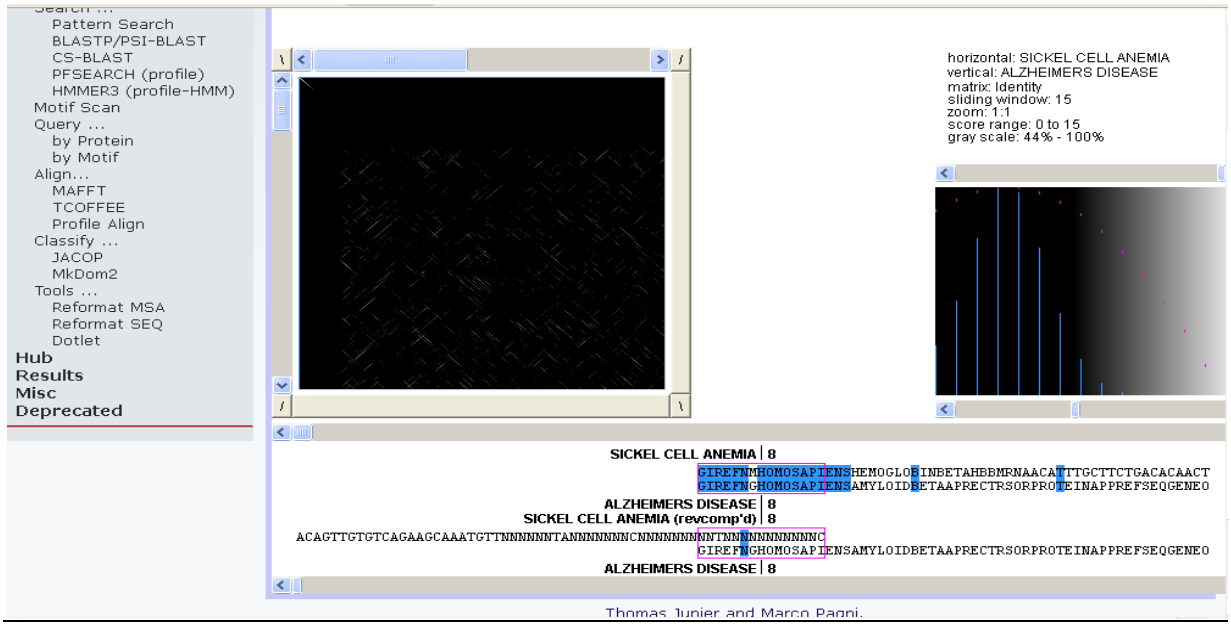


Here, red color of bases shows the similarity of the sequences. There are 30 similar bases in the shown figure.
DOTLET: It is a program for comparing sequences by the diagonal plot method. The sequences are entered and then computed to get the results as follows.
Sequence of Alzheimer's disease is compared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

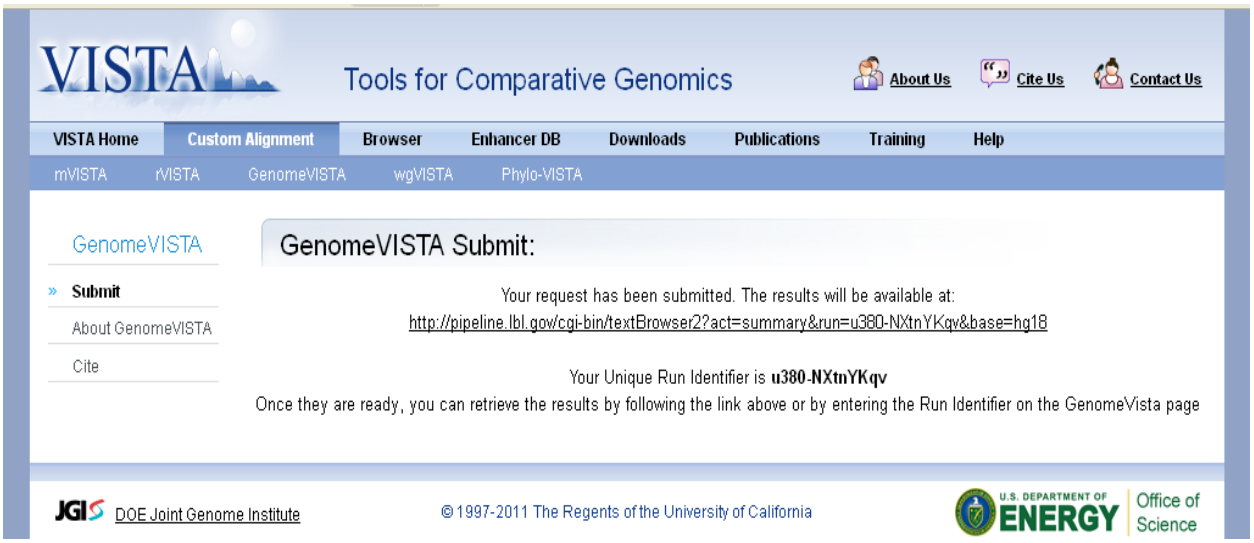


Below the comparison window, sequence similarity can be seen covering few bases at a time.

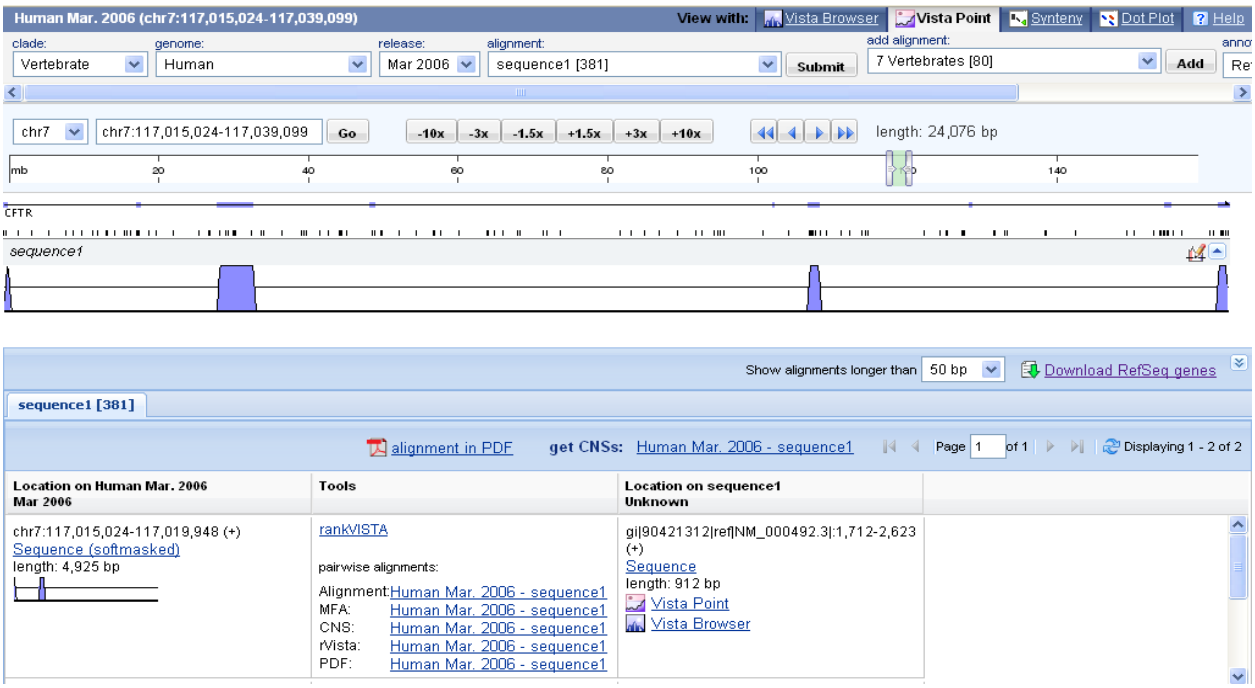
GENOME VISTA: It uses a computational strategy where query sequence contigs are anchored on the base genome by local alignment matches and then globally aligned to candidate regions by AVID program. Here, human march 2006 genome is used as a base genome for comparison of sequences.

Result of genome VISTA for Cystic fibrosis.

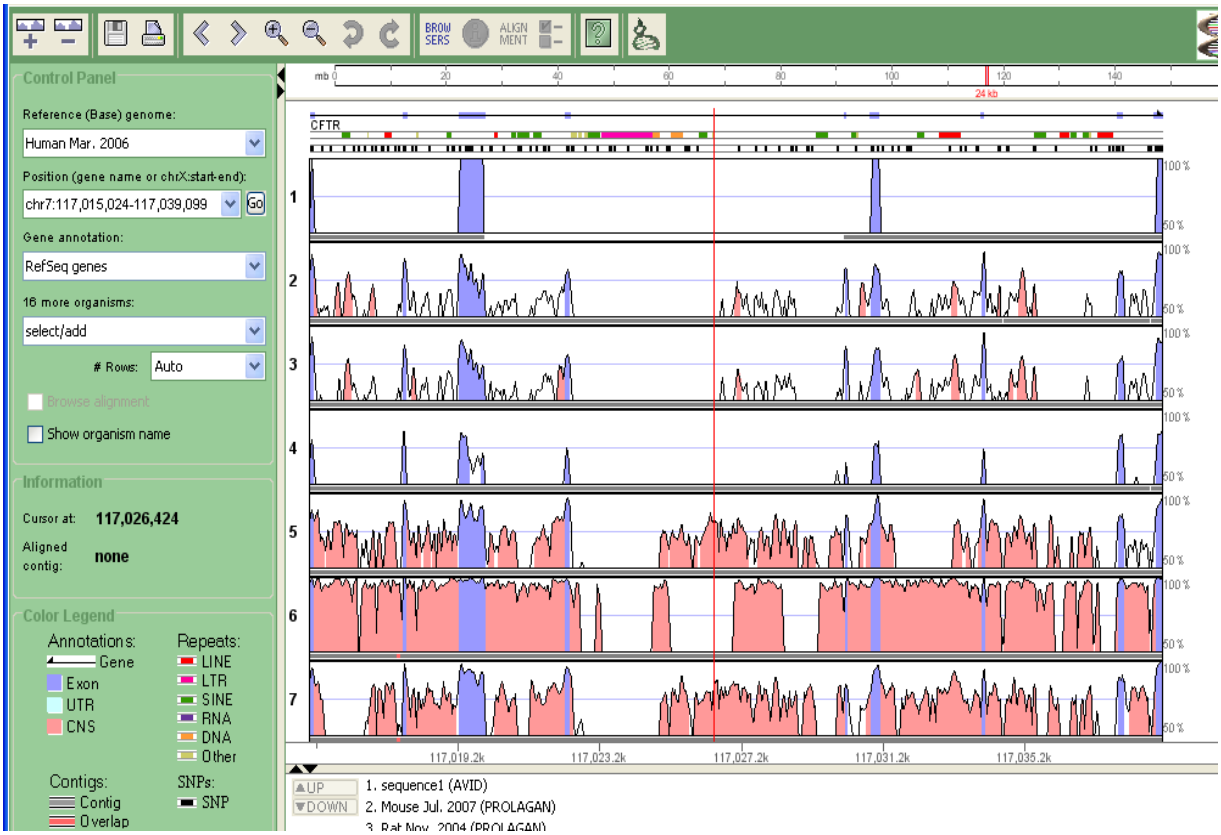
When sequences are submitted then one link is obtained through which result can be accessed as follows.



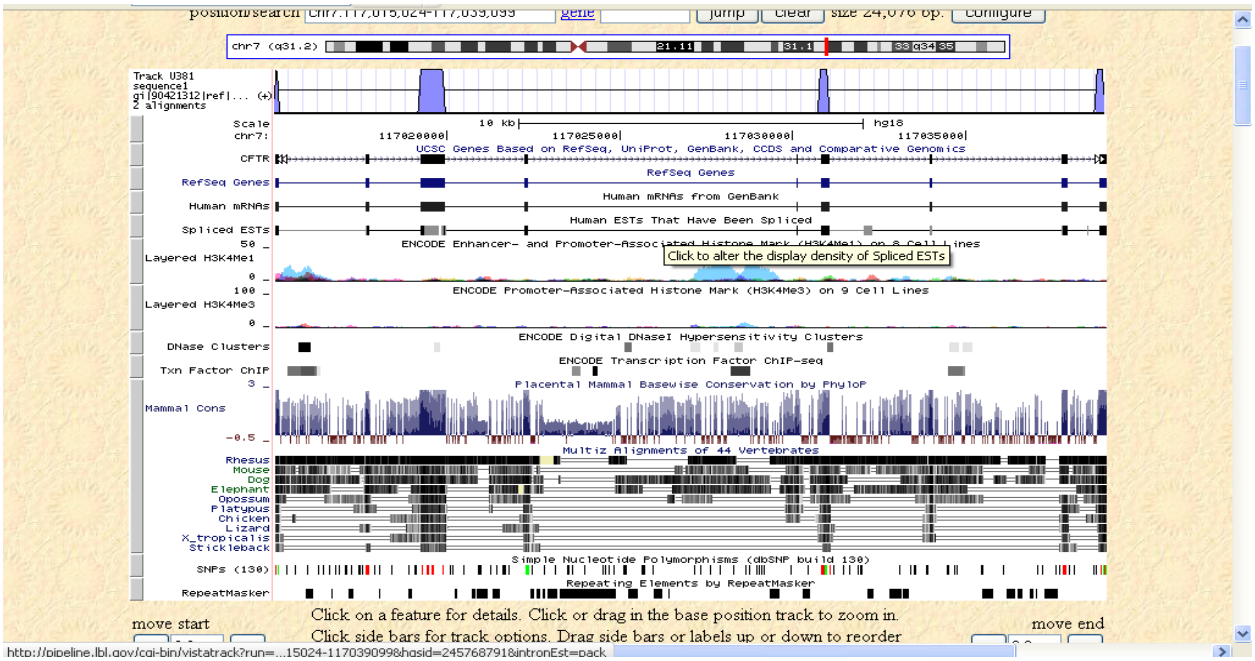
When link is opened the result is obtained in vista point format.



Result in VISTA browser format is obtained as follows.

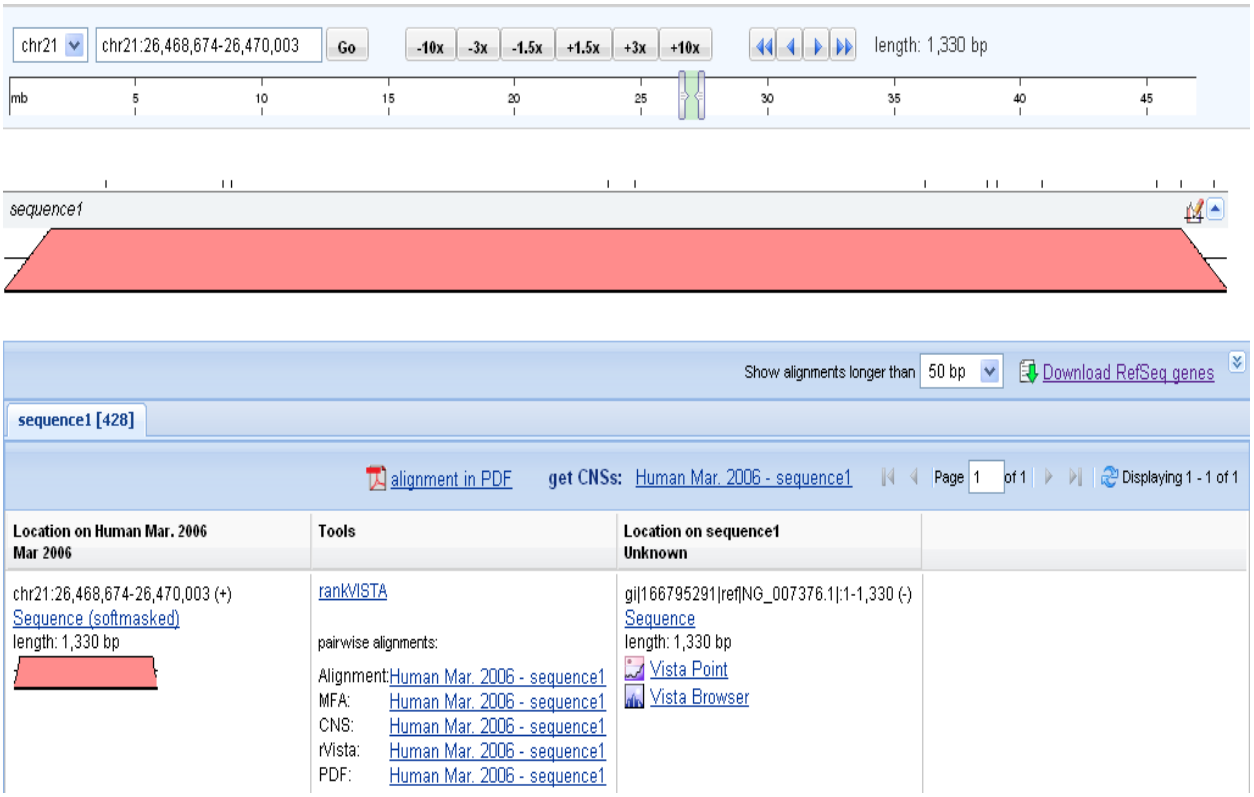


Result in VISTA Track format is shown below.

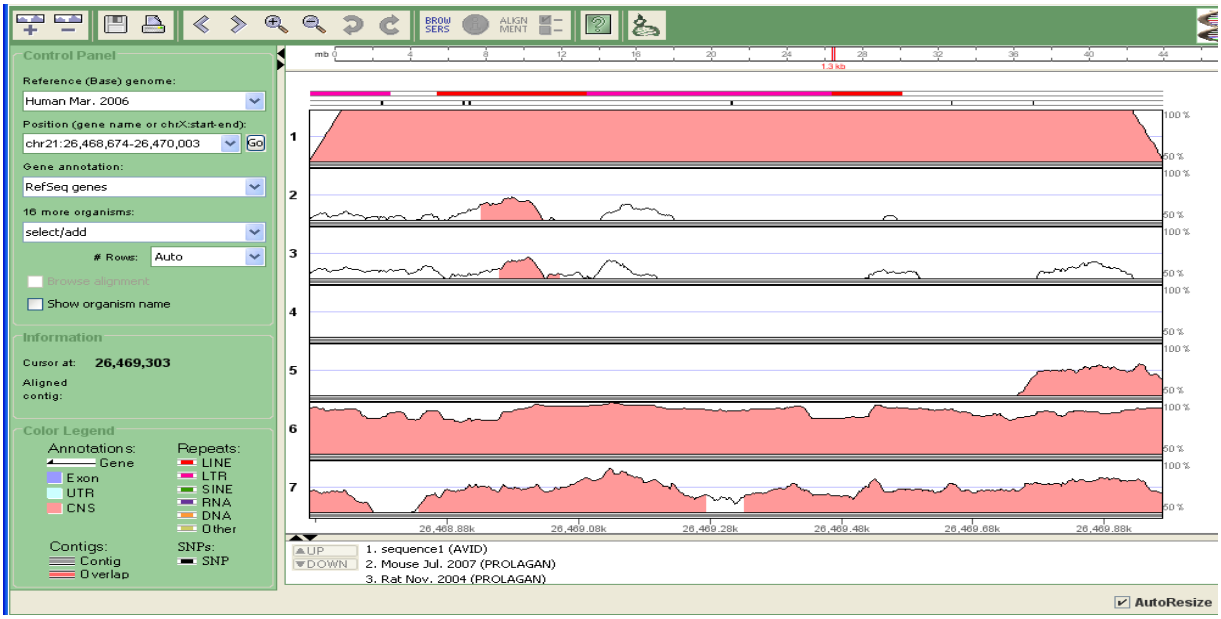


Result for Alzheimer's disease.

Result as VISTA point is shown below.

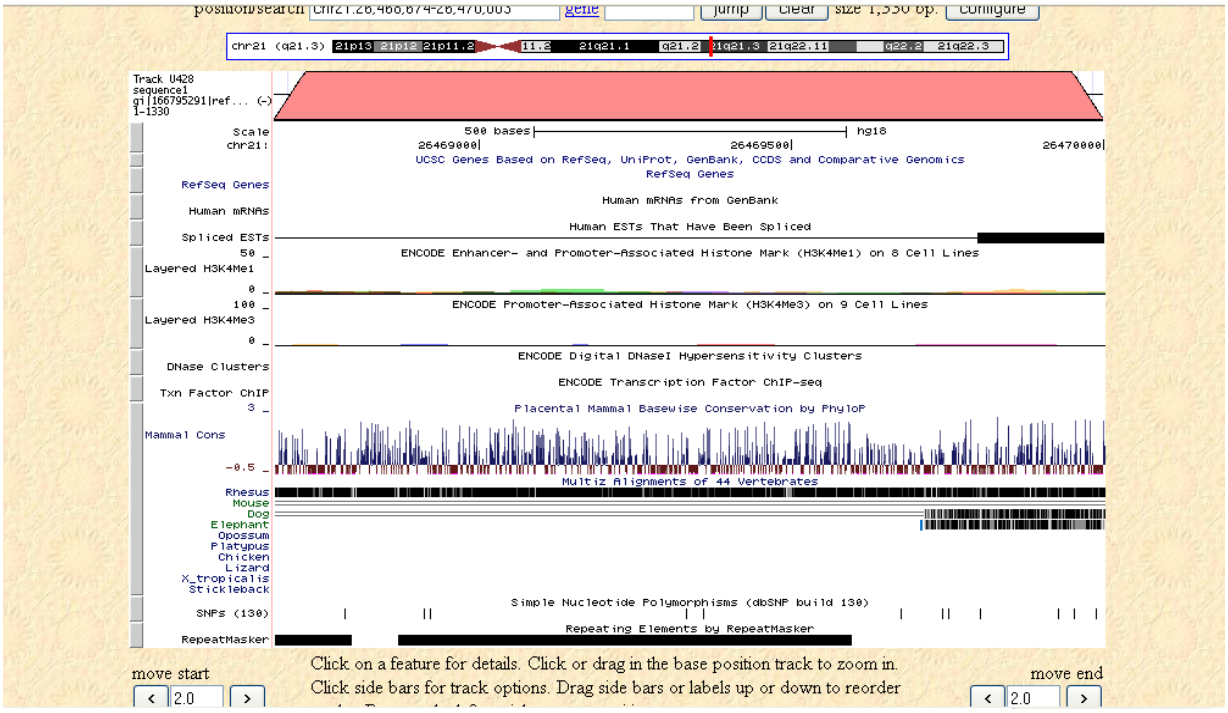


Result in VISTA Browser format is as follows

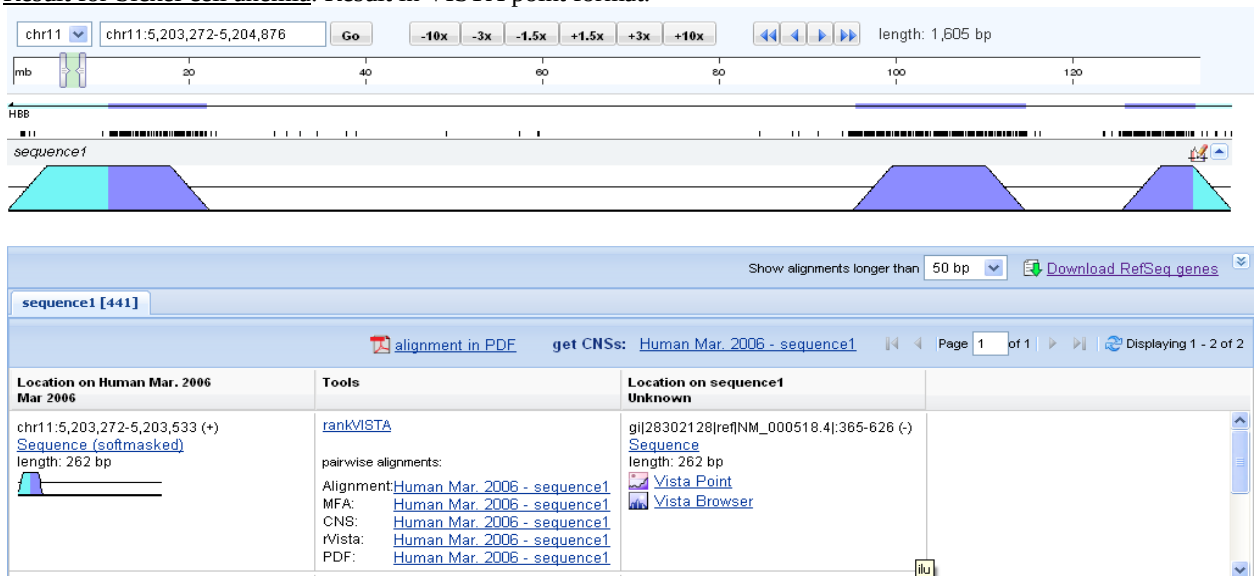


.

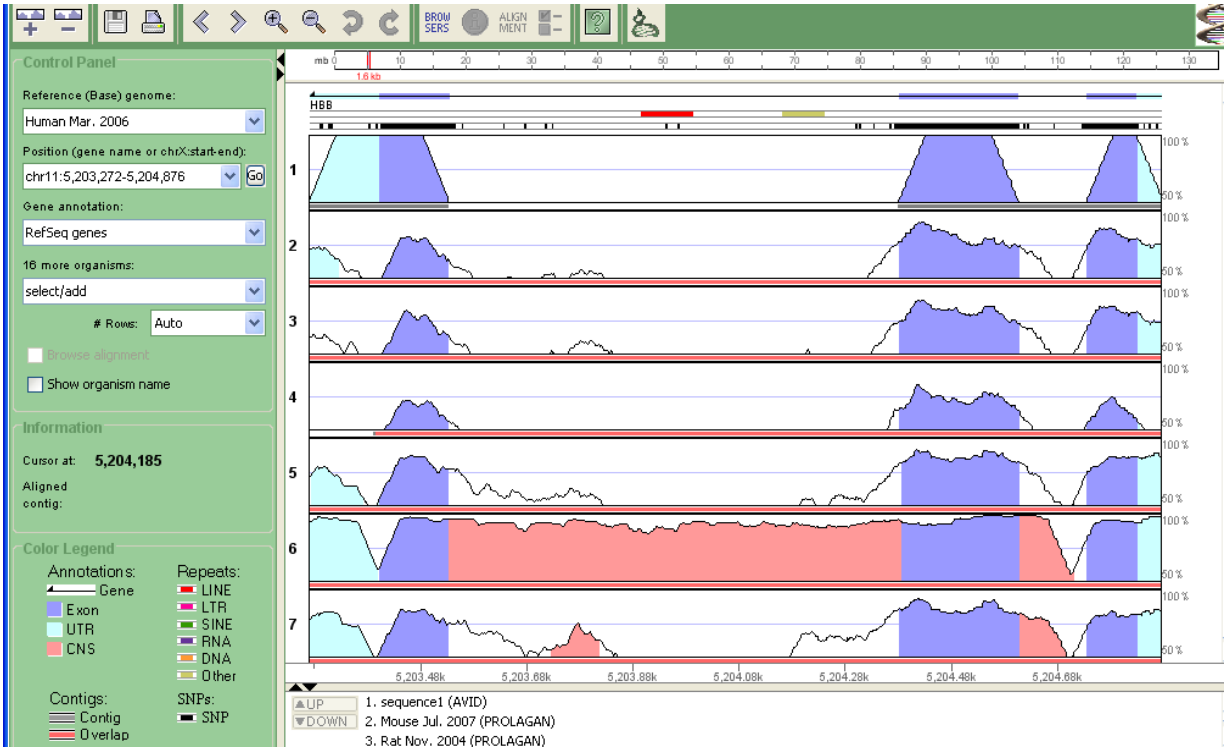
Result in VISTA track format



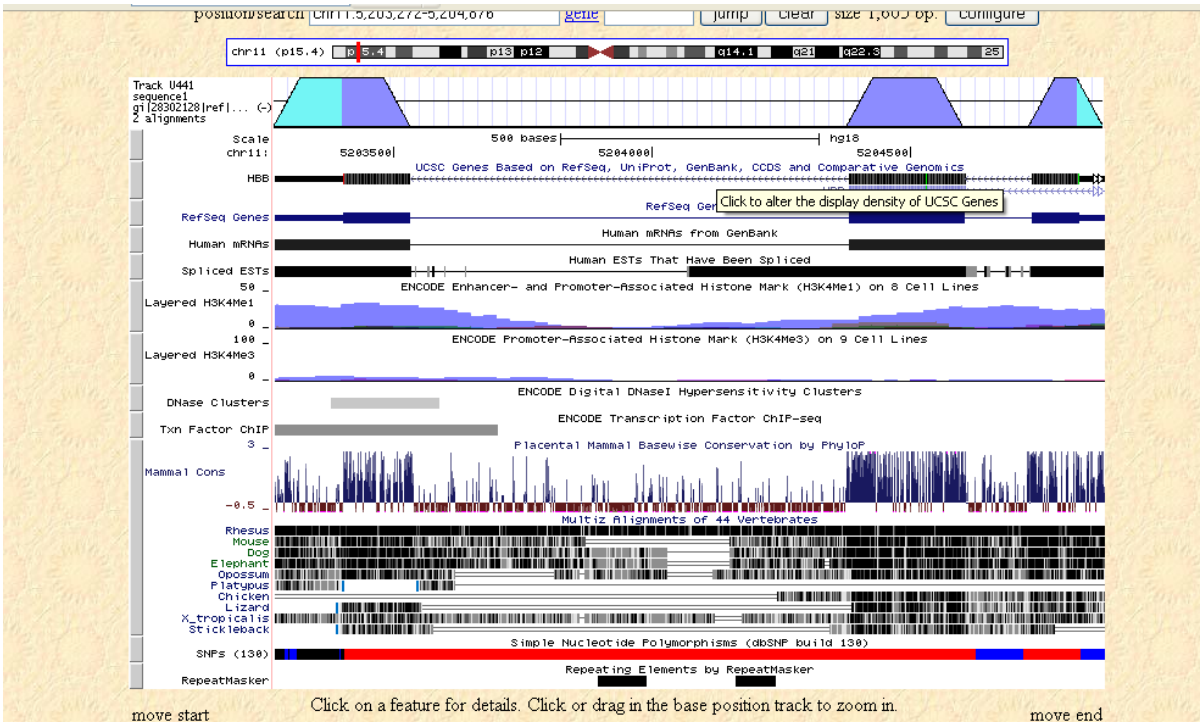
Result for Sickel cell anemia. Result in VISTA point format.



Result in VISTA Browser format.



Result in VISTA track 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:

[1]. 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.