

Original Research Article

<https://doi.org/10.20546/ijcmas.2017.608.152>

## Inquest of the SNP in Cystic Fibrosis – A Bioinformatic Approach

Balakrishnaraja Rengaraju<sup>1\*</sup>, Keerthana<sup>1</sup>, Akila<sup>1</sup>, K. Pavithra<sup>1</sup>, D. Vinotha<sup>1</sup>,  
Sri Harsha Challapalli<sup>2</sup> and Arunava Das<sup>1</sup>

<sup>1</sup>Bannari Amman Institute of Technology, Sathyamangalam-638401, India

<sup>2</sup>ETH Technical University, Zurich-8093, Switzerland

\*Corresponding author

### A B S T R A C T

#### Keywords

SNP, CFTR gene,  
Cystic fibrosis,  
BLAST,  
SWISSPROT.

#### Article Info

Accepted:  
17 June 2017  
Available Online:  
10 August 2017

Single nucleotide polymorphisms or SNPs are DNA sequence variations occurring due to an alteration of a single nucleotide of the genome sequence. These variations have a major impact on human response to diseases and therapeutic drugs. Although SNPs may or may not cause a disease, they can help us to determine the probability of occurrence of the disease. One such well-known lethal genetic disorder is Cystic Fibrosis. It is an autosomal recessive disease caused by an alteration in the CFTR gene (Cystic Fibrosis Trans membrane conductance Regulator gene), which codes for a chloride channel protein, found in the membranes lining the respiratory and digestive tracts. Bearing some common symptoms is the disorder chronic rhino sinusitis, which can also be traced to the alteration in the same CFTR gene. A comparative study of the two diseases indicates that the former requires two copies of the altered gene whereas a single copy may influence the occurrence of the latter. Our SNP analysis of this gene reveals that a point mutation occurs at locus 5618 of the nucleotide sequence leading to the conversion of the amino acid histidine to proline thus affecting the three-dimensional structure of the protein.

### Introduction

Along the length of the helical strand are strewn, the basic entities of inheritance genes that are constituted of nucleotide sequences. These sequences comprise the different combinations of the alphabets of life -A, T, G and C. The arrangement of these nitrogenous bases remains the same in almost all the genomes of a species except for certain loci that may exhibit variations. Such alterations are due to change in a single nucleotide and hence these are referred to as Single Nucleotide Polymorphism or SNPs. For example, the change of *GGAATC* to *GGAATT* can be considered as an SNP. At least 1% of the population must have these variations for

them to be considered as an SNP. In the three billion base human genome, there is one SNP for every 100-300 bases. They can be found either in the coding or non-coding regions of the genome. In the former case they can affect the biological properties of a protein occasionally leading to diseased conditions. Even in the latter case, when they are not responsible for the disease, they may serve as genetic markers due to their proximity to the mutated regions. These markers can aid in designing therapeutic drugs for various genetic disorders. SNPs are indicative of how a person responds to diseases, microbes, toxins, drugs and therapies. Hence they are of

great value in pharmacogenomics, diagnostic and biomedical research. SNPs can also help unravel the evolutionary pathway because these are the consequences of past mutations, which prove unique when two individuals are compared given by Riordan *et al.*, (1989)

Cystic fibrosis is an inherited disease believed to be caused by an SNP. It is characterized by abnormal fluid secretion. Cystic fibrosis (CF) is a recessively inherited condition caused by mutation of the CFTR gene given by Southern *et al.*, (2007). Cystic Fibrosis Trans membrane conductance Regulator (CFTR) gene plays a key role in chloride ion transport across the membrane. The alteration in the same gene can also result in a different disorder of the mucous membrane known as chronic rhino sinusitis. The gene sequence of CFTR can be analyzed using Bioinformatics tools and the exact position of the SNP can be determined. Being a killer disease, cystic fibrosis has always been an area of extensive research. Studies have been carried out to identify its root cause and the possible remedies. Cystic fibrosis can interfere with the normal functioning of the liver, lungs, gastrointestinal tract, pancreas, the sweat glands and the male reproductive system. Pathogens are associated with chronic lung infections and increased risk of death in patients with cystic fibrosis (CF) given by Isabel Sá-Correia *et al.*, (2017). The chemical properties of the mucus are altered resulting in the secretion of thicker-than-normal mucus that obstructs the airways and the ducts of digestive organs. The major symptom is the presence of an excessive amount of salt in sweat

As this is caused by a recessive gene, for a person to be afflicted, causes this disease the gene must be inherited from both parents. If a person has just a single copy, he is a carrier of the disease. When two such carriers have a baby, there is 25% chance of the child not to have the disease at all (homozygous

dominant), 25% probability that the child will show the symptoms of cystic fibrosis (homozygous recessive) and the other 50% that he or she is a carrier (heterozygous) as illustrated in table 1. The clinical manifestations of CF involved several systems. The most common symptom was recurrent pulmonary infections given by Xu *et al.*, (2017). Although the 3D structure of the entire CFTR has not been experimentally determined, a model depicting the structure of CFTR has been proposed. CFTR is made up of five domains: two membrane-spanning domains (MSD1 and MSD2) that form the chloride ion channel, two nucleotide-binding domains (NBD1 and NBD2) that bind and hydrolyze ATP, and a regulatory (R) domain given by Xiaorong Zhang *et al.*, (2011) (Fig. 1). It is a chronic inflammatory disease Apart from the CFTR gene, cystic fibrosis is also affected by environmental factors and genetic factors. Mutations in the gene responsible for cystic fibrosis may also be associated with the development of another genetic disorder namely, chronic rhino sinusitis in general population. It is a long-lasting inflammatory disease of the nasal and associated mucosa, having a variety of signs and symptoms. Purulent discharges anteriorly and posteriorly are the significant findings in its diagnosis. Large number of patients has been associated with rhinitis, either of sensitized or transferable source. Physiologic testing in them showed normal indices of nasal epithelial sodium transport with a slight reduction in CFTR mediated chloride conductance.

Though the same gene is involved in both cystic fibrosis and chronic rhino sinusitis, the alterations in the gene or may not be the same. Those distressed with prolonged rhino sinusitis can transmit a single copy of the CF associated CFTR gene or another difference in the similar gene known as or may not be the same. Those distressed

from prolonged rhino sinusitis could transmit a single copy of the CF related CFTR gene or another difference in the similar gene known as M470V.

### **Insight of CFTR gene**

The CFTR gene is located at q31.2 on the long arm (q) of human chromosome 7. It is 250000 base pairs long and contains 27 exons. Lap Chee Tsui and John Riordan first isolated the gene in 1989. The coding sequence lie within 4443 base pairs. The CFTR protein is 1480 amino acids long and has a molecular weight of 168173 Da. It is classified as an ABC (ATP-Binding Cassette) transports molecules of sugars, peptides, inorganic phosphates and metal cations.

It is especially involved in the transport of chloride ions across the cellular membrane. Model studies on the protein show that it consists of five domains, two membrane spanning domains that form the chloride ion channel, two nucleotide binding domains that bind and hydrolyze ATP and the regulatory domain containing several serine residues that can be phosphorylated by the protein kinase called PKA that is activated by the secondary messenger cAMP. The presence of the regulatory domain is a unique feature of the CFTR gene and the modification by addition or removal of phosphate groups can regulate the movement of chloride ions across the membrane. With normal CFTR once the protein is synthesized it is transported to endoplasmic reticulum and the Golgi bodies for post translational modification before being integrated into the cell membrane. Modification in the gene effects in acute loss of chloride ion conveyance that interrupts the balance between sodium and chloride ions that is needed to sustain a thin mucus layer. This inequity creates a thick mucus layer thus flooring the way for prolonged infections.

### **SNP analysis of the CFTR gene**

Exploiting the tools of bioinformatics one can store and search data, analyze and determine the relationship between macromolecular sequences, structures and biochemical pathways. Generally, we encounter three major forms of DNA sequences viz. genomic DNA; cDNA and recombinant DNA. cDNA is comprised only of exons reverse transcribed from mRNA and rDNA is a hybrid of a cloning vector and a foreign DNA, synthesized *in vitro*. Our analysis is mainly of the genomic DNA sequences that contain both the coding (exons) and non-coding (introns) sequences. Standard format for the expression of the DNA and protein sequences come in various forms of which the most widely used are: NBRF/PIR (National Biomedical Research Foundation / Protein Information Resource) format, FASTA format and the GDE format. The available databases can be classified as under primary databases that contain raw sequence data, secondary sequences containing information on sequence patterns and organism-specific databases pertaining to a particular species. The DNA sequence of interest can be obtained from any of the three primary sequence databases, which are repositories of enormous amount of information Gen Bank (NCBI), the Nucleotide Sequence Database (EMBL) and the DNA Databank of Japan (DDBJ). The sequences thus obtained can be crosschecked using SWISS-PROT that is a collection of confirmed protein sequences that are being curated and are up-dated regularly. BLAST and FASTA are used to compare the sequence of interest with the database sequences. These are preferred to dynamic programming algorithms due to their greater speed. While BLAST gives the local alignment and can identify many high-scoring regions of similarity, FASTA provides the global alignment and only identifies the highest-scoring region. The high-scoring local alignments are known as

high scoring segment pairs (HSPs). BLAST relates 3 protein sequences or 11 nucleotide sequence simultaneously whereas FASTA performs the same for 2 protein sequences or 6 nucleotide sequences. This sequence comparison can provide a knowledge almost the protein function. Once the position of the SNP has been known, it is essential to find if it appears in the coding region of the gene. Protein encoding genes are featured by the presence of an Open Reading Frame (ORF), consisting of a series of sense codons that have a start and a stop codon.

Using the ORF Finder in the NCBI, we can locate the exact position of the exons in the given sequence. Further, we can find the mutated nucleotide and the corresponding amino acid. It has been proposed that genes interacting with CF Tran membrane conductance regulator (CFTR) and epithelial sodium channel (ENaC) are potential modifiers. Therefore, the impact of single-nucleotide polymorphisms (SNPs) of several of these interactors on CF disease outcome was assessed by Gallati *et al.*, (2012). Materials involved: DNA sequence of the CFTR gene; GenBank (NCBI) database; SWISS PROT; SNP database; ORF Finder of NCBI.

## Materials and Methods

The GenBank ID of the CFTR gene (1080) was obtained from the search engine of the NCBI homepage. The nucleotide sequence was found using the access code NM000492. This sequence was blasted with an SNP database<sup>2</sup>. The BLAST result showed variations at 3 loci namely 5618, 5732 and 5826. To check if these SNPs lie within the coding regions, the sequences were fed to the ORF Finder<sup>1</sup>. The exon of interest was obtained and the codes were read. To confirm the protein, the nucleotide sequence was translated and blasted with SWISS PROT (ID:

P13569).

## Results and Discussion

### Sequence extraction

The nucleotide sequence obtained from the NCBI database and the bioinformatics tools were used to find out the mutated protein responsible for the cystic fibrosis gene.

NUCLEOTIDE SEQUENCE NM 000492.  
Homo sapiens cyst...[gi:6995995]  
>gi|90421312|ref|NM\_000492.3| Homo sapiens cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) (CFTR), mRNA  
AATTGGAAGCAAATGACATCACAGCA GGTAGAGAAAAAGGGTTGAGCGGCA GGCACCCAGAGTAGTAGGTCTTGGCA TTAGGAGCTTGAGCCCAGACGGCCCTA GCAGGGACCCCCAGCGCCCGAGAGACC ATGCAGAGGTCGCCTCTGGAAAAGGCC AGCGTTGTCTCCAAACTTTTTTCAGCT GGACCAGACCAATTGAGGAAAGGA TACAGACAGCGCCTGGAATTGTCAGAC ATATACCAAATCCCTCTGTTGATTCTG CTGACAATCTATCTGAAAAATTGGAAA GAGAATGGGATAGAGAGCTGGCTTCA AAGAAAAATCCTAAACTCATTAATGCC CTCGGCGATGTTTTCTGGAGATTAA TGTTCTATGGAATCTTTTATATTAGG GGAAGTCACCAAAGCAGTACAGCCTCT CTTACTGGGAAGAACATAGCTTCTA TGACCCGGATAACAAGGAGGAACGCT CTATCGCGATTATCTAGGCATAGGCT TATGCCTCTCTTATTGTGAGGACACT GCTCCTACACCCAGCCATTGGCCTT CATCACATTGGAATGCAGATGAGAATA GCTATGTTAGTTGATTATAAGAAG ACTTTAAAGCTGTCAAGCCGTGTTCTA GATAAAATAAGTATTGGACAACCTGTT AGTCTCCTTCCAACAACCTGAACAAA TTTGATGAAGGACTTGCATTGGCACAT TTCGTGTGGATCGCTCCTTGCAAGTG GCACTCCTCATGGGGCTAATCTGGGAG

TTGTTACAGCGTCTGCCCTGTGGA  
CTTGGTTCTGATAGTCCTGCCCTT  
TTCAGGCTGGCTAGGGAGAATGATGA  
TGAAGTACAGAGATCAGAGAGCTGGG  
AAGATCAGTGAAGACTGTGATTACC  
TCAGAAATGATTGAAAATATCCAATCT  
GTTAAGGCATACTGCTGGAGAAGC  
AATGGAAAAATGATTGAAAACCTAA  
GACAAACAGAACTGAAACTGACTCGG  
AAGGCAGCCTATGTGAGATACTCAAT  
AGCTCAGCCTCTTCTCAGGGTTCT  
TTGTGGTGTCTTATCTGTGCTCCCTA  
TGCACTAATCAAAGGAATCATCCTCCG  
GAAAATATTACCAACCCTCATTCTG  
CATTGTTCTGCATGGCGGTCACTCG  
GCAATTCCCTGGCTGTACAAACATG  
GTATGACTCTGGAGCAATAAACAA  
AATACAGGATTCTTACAAAAGCAAGA  
ATATAAGACATTGGAATATAACTAAC  
GAACAGAAGTAGTGATGGAGAATG  
TAACAGCCTCTGGGAGGGAGGATTG  
GGGAATTATTGAGAAAGCAAAACAA  
AACAAATAACAATAGAAAAACTCTAAT  
GGTGATGACAGCCTCTCAGTAAT  
TTCTCACTTCTGGTACTCCTGTCTGA  
AAGATATTAAATTCAAGATAGAAAGAG  
GACAGTTGGCGGTTGCTGGATCCA  
CTGGAGCAGGCAAGACTTCACTCTAA  
TGGTATTATGGGAGAACTGGAGCCTT  
CAGAGGGTAAATTAAAGCACAGTGG  
AGAATTTCATTCTGTTCTCAGTTTCT  
GGATTATGCCTGGCACCATTAAGAAA  
ATATCATCTTGGTCTTCTATGATGA  
ATATAGATAAGCGTCATCAAAGC  
ATGCCAACTAGAAGAGGACATCTCAA  
GTTTGCAGAGAAAGACAATATAGTCT  
TGGAGAAGGTGGAATCACACTGAGTG  
GAGGTCAACGAGCAAGAATTCTTAG  
CAAGAGCAGTACAAAGATGCTGATT  
TGTATTATTAGACTCTCCTTGGATA  
CCTAGATGTTAACAGAAAAAGAAAT  
ATTGAAAGCTGTCTGAAACTGAT  
GGCTAACAAACTAGGATTTGGTCAC  
TTCTAAATGGAACATTAAAGAAAGC  
TGACAAAATATTAATTGATGAAGG  
TAGCAGCTATTGATGGGACATTTC

GAACCTCCAAAATCTACAGCCAGACTTT  
AGCTCAAAACTCATGGGATGTGATTCT  
TTCGACCAATTAGTGCAGAAAGAAGA  
AATTCAATCCTAACTGAGACCTTACAC  
CGTTCTCATTAGAAGGAGATGCTCCT  
GTCTCCTGGACAGAAACAAAAAAACA  
ATCTTTAAACAGACTGGAGAGTTGG  
GGAAAAAAAGGAAGAATTCTATTCTCAA  
TCCAATCAACTCTATACGAAAATTTC  
CATTGTGCAAAAGACTCCCTACAAAT  
GAATGGCATCGAAGAGGATTCTGATGA  
GCCTTAGAGAGAAGGCTGTCTTAGT  
ACCAAGATTCTGAGCAGGGAGAGCGA  
TACTGCCTCGCATCAGCGTGATCAGCA  
CTGGCCCCACGCTTCAGGCACGAAGGA  
GGCAGTCTGCTGAACCTGATGACAC  
ACTCAGTTAACCAAGGTCAAGAACATT  
ACCGAAAGACAACAGCATCCACACGA  
AAAGTGTCACTGGCCCCTCAGGCAAAC  
TTGACTGAACGGATATATTCAAGA  
AGGTTATCTCAAGAAACTGGCTTGAA  
ATAAGTGAAGAAATTACGAAGAAGA  
CTTAAAGGAGTGCTTTTGATGATAT  
GGAGAGCATACCAGCAGTGACTACAT  
GGAACACATACCTCGATATATTACTG  
TCCACAAGAGCTTAATTGGCTGCTAA  
TTTGGTGTCTAGTAATTGGCTGCTGAGA  
GGTGGCTGCTTCTTGGTTGTGCTGTGG  
CTCCTGGAAACACTCCTCTCAAGAC  
AAAGGGAAATGACTCATAGTAGAAA  
TAACAGCTATGCAGTGATTATCACCAG  
CACCAGTTGTATTATGTGTTTACATT  
TACGTGGAGTAGCCGACACTTGCTT  
GCTATGGGATTCTCAGAGGTCTACCA  
CTGGTGCATACTCTAACACAGTGTG  
AAAATTACACCACAAATGTTACAT  
TCTGTTCTCAAGCACCTATGTCAACCC  
TCAACACGTTGAAAGCAGGTGGGATT  
TTAATAGATTCTCAAAGATATAGCAA  
TTTGGATGACCTCTGCCTCTTACCAT  
ATTGACTTCATCCAGTTGTTATTAATT  
GTGATTGGAGCTAGCAGTTGTCGCA  
GTTTACAACCCTACATCTTGGCTGCAA  
CAGTGCCAGTGTAGTGGCTTTATTA  
TGTGAGAGCATATTCCCTCAAACCT  
CACAGCAACTCAAACAACTGGAATCTG

AAGGCAGGAGTCCAATTTCACTCATC  
TTGTTACAAGCTAAAAGGACTATGGA  
CACTCGTGCCTCGGACGGCAGCCTT  
ACTTGAAACTCTGTTCCACAAAGCTC  
TGAATTACATACTGCCAAGTGGTTCTT  
GTACCTGTCAACACTGCGCTGGTCCA  
AATGAGAATAGAAATGATTTGTCA  
CTTCTCATGCTGTTACCTCATTCC  
ATTTAACAAACAGGAGAAGGAGAAGG  
AAGAGTTGGTATTATCCTGACTTAGC  
CATGAATATCATGAGTACATTGAGTG  
GGCTGTAAACTCCAGCATAGATGTGGA  
TAGCTTGATGCGATCTGTGAGCCGAGT  
CTTAAAGTTATTGACATGCCAACAGA  
AGGTAAACCTACCAAGTCAACCAAACC  
ATACAAGAATGCCAACCTCTGAAAGT  
TATGATTATTGAGAATTCACACGTGAA  
GAAAGATGACATCTGGCCCTCAGGGG  
GCCAAATGACTGTCAAAGATCTCACAG  
CAAAATACACAGAAGGTGGAAATGCC  
ATATTAGAGAACATTCTCTCAATA  
AGTCCTGGCCAGAGGGTGGGCTCTG  
GGAAGAACTGGATCAGGAAAGAGTAC  
TTTGTATCAGCTTTTGAGACTACTG  
AACACTGAAGGAGAACATCCAGATCGA  
TGGTGTGTCTGGGATTCAATAACTT  
GCAACAGTGGAGGAAAGCCTTGGAG  
TGATACCACAGAAAGTATTATTTTTC  
TGGAACATTAGAAAAAAACTTGGATCC  
CTATGAACAGTGGAGTGTCAAGAACAT  
ATGGAAAGTTGCAGATGAGGTTGGGCT  
CAGATCTGTGATAGAACAGTTCTG  
GAAGCTTGACTTGTCCCTGTGGATGG  
GGGCTGTGCTTAAGCCATGGCACAA  
GCAGTTGATGTGCTGGCTAGATCTGT  
TCTCAGTAAGGCGAAGATCTGCTGCT  
TGATGAACCCAGTGCTCATTGGATCC  
AGTAACATACCAAATAATTAGAAGAA  
CTCTAAAACAAGCATTGCTGATTGCA  
CAGTAATTCTCTGTGAACACAGGATAG  
AAGCAATGCTGGAATGCCAACAAATT  
TGGTCATAGAAGAGAACAAAGTGC  
CAGTACGATTCCATCCAGAAACTGCTG  
AACGAGAGGAGGCCCTTCCGGCAAGCC  
ATCAGCCCCCTCCGACAGGGTGAAGCTC  
TTTCCCCACCGGAACCAAGCAAGTGC

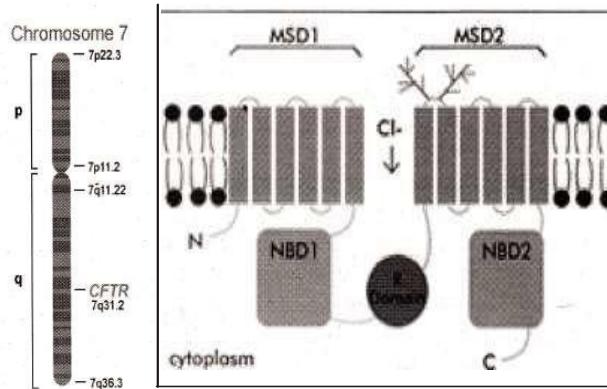
AAGTCTAACGCCAGATTGCTGCTCTG  
AAAGAGGAGACAGAAGAAGAGGTGCA  
AGATACAAGGCTTAGAGAGCAGCAT  
AAATGTTGACATGGGACATTGCTCAT  
GGAATTGGAGCTCGTGGACAGTCACC  
TCATGGAATTGGAGCTGTTGAAACAGT  
TACCTCTGCCTCAGAAAACAAGGATGA  
ATTAAGTTTTTTAAAAAAGAAACA  
TTTGTAAAGGGATTGAGGACACTGA  
TATGGGTCTTGATAAATGGCTCCTGG  
CAATAGTCAAATTGTGTGAAAGGTACT  
TCAAATCCTGAAAGATTACCACTTGT  
GTTTGCAAGCCAGATTTCTGAAAAA  
CCCTGCCATGTGCTAGTAATTGGAAA  
GGCAGCTCTAAATGTCAATCAGCCTAG  
TTGATCAGCTATTGCTAGTGAACACT  
CGTTAATTGTAGTGGAGAAGAAC  
TGAAATCATACTCTTAGGGTTATGAT  
TAAGTAATGATAACTGGAAACTTCAGC  
GGTTATATAAGCTGTATTCTTTTC  
TCTCCTCTCCCCATGATGTTAGAAC  
ACAACATATTGTTGCTAAGCATTCC  
AACTATCTCATTCCAAGCAAGTATTA  
GAATACCACAGGAACCACAAGACTGC  
ACATCAAATATGCCCAATTCAACATC  
TAGTGAGCAGTCAGGAAAGAGAACTT  
CCAGATCCTGGAAATCAGGGTTAGTAT  
TGTCCAGGTCTACCAAAATCTCAATA  
TTTCAGATAATCACAATACATCCCTTA  
CCTGGAAAGGGCTGTTATAATCTTTC  
ACAGGGGACAGGATGGTCCCTGATG  
AAGAAGTTGATATGCCTTTCCCAACT  
CCAGAAAGTGACAAGCTCACAGACCTT  
TGAACATAGAGTTAGCTGGAAAAGTAT  
GTTAGTGCACAGGACAGCC  
CTTCTTCCACAGAACGCTCCAGGTAGA  
GGGTGTGTAAGTAGATAGGCCATGGGC  
ACTGTGGGTAGACACACATGAAGTCCA  
AGCATTAGATGTATAGGTTGATGGTG  
GTATGTTTCAGGCTAGATGTATGTAC  
TTCATGCTGTCTACACTAACAGAGAGAAT  
GAGAGACACACTGAAGAACCAAT  
CATGAATTAGTTATATGCTTCTGTT  
TATAATTGTGAAGCAAAATTCTTCT  
CTAGGAAATATTATTTAATAATGTTT  
CAAACATATACAATGCTGTATT

AAAAGAATGATTATGAATTACATTGT  
ATAAAATAATTTTATATTGAAATATT  
GACTTTT

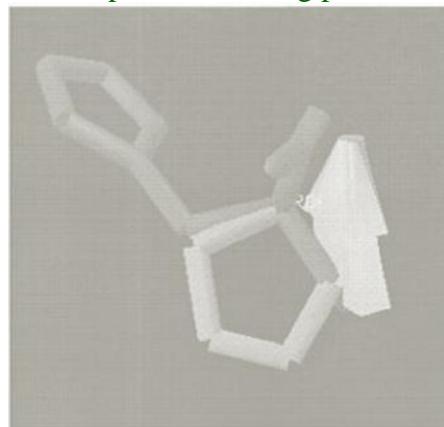
ATGGCACTAGTATTCTATGAAATATT  
ATGTTAAAACGGGACAGGGGAGAAC  
CTAGGGTGATATTAACCAGGGGCCATG  
AATCACCTTTGGTCTGGAGGGAGCC

TTGGGGCTGATGCAGTTGTCGCCACA  
GCTGTATGATTCCCAGCCAGCACAGCC  
TCTTAGATGCAGTCTGAAGAAGATGG  
TACCACCACTGACTGTTCCATCAA  
GGGTACACTGCCTCTCAACTCCAAAC  
TGACTCTTAAGAAGACTGCATTATATT  
TATTACTGTAAGAAAATATCAGTTGTC  
AATAAAATCCATACATTGTGTGAAA

**Fig.1** Diagrammatic representation of the cystic fibrosis condition through gene regulation



**Fig.2** CFTR protein showing point mutation



**Table.1** Explanation of the cystic fibrosis based on genetic pattern

		MOTHER	
FATHER	N	N	n
		Non- carrier NN	Carrier Nn
	n	Carrier Nn	Cystic Fibrosis nn

This result shows the comparison of the CFTR gene with the most similar SNP database sequence. As is evident from the identities value, out of the 430 nucleotides, there are three positions at which the bases differ. It is also necessary to note that two positive strands of DNA have been compared. The E value (Expect value) of O.O is indicative of the highest similarity.

This shows that the locus 5618 appears in the coding region. The triplet 'cac' codes for the amino acid Histidine. 5539 atgtatagttgtatggt ggtatgtttcaggctagatgt atgtac MYRLMVV CFQARCMY 5584ttcatgctgtcacactaagagaga tgagagacacactgaagaa FMLSTLRENERHTEE 5629 gcacccaatc atgaatttag APIMN\*

## Translated protein sequence

'>spiP135691CFTR \_HUMAN Cystic fibrosis  
Trans membrane conductance regulator (CFTR)  
(CAMP dependent chloride channel) -Homo  
sapiens (Human).

MQRSPLEKASVVSKLFFSWTEPILRKGYR  
QRLELDIYQIPSVDSADNLSEKLEREWDR  
ELASKKNPKLINALRRCCFWRFMFYGINFLY  
LGEVTKAQPLLLGRIASYDPDNKEERSI  
AIYLGIGLCLLFIVRTLLLHPAIFGLHHIGM  
QMRIAMFSLIYKKTLLKSSRVLDKISIGQL  
VSLLSNNLNKFDEGLALAHFVWIAPLQVA  
LLMGLIWELLQASAFCGLGFLIVLALFQA  
GLGRMMMKYRDQRAGKISERLVITSEMIE  
NIQSVKAYCWEEMEKMENLRQTELKLT  
RKAAYVRYFNSSAFFSGFFVVFLSVPY  
LIKGIILRKIFTTISFCIVLRMAVTRQFPWA  
VQTWYDSDLGAINKIQDFLQKQEYKTLEYN  
LTTEVVMENVTAFWEEGFELFEKAKQ  
NNNRKTSNGDDSLFFSNFSLLGTPVJKDI  
NFKIERGQLAVAGSTGAGKTLLMMIMG  
ELEPSEGKIKHSGRISFCSQFSWIMPGTIKE  
NIIFGVSYDEYRYRSVIACQLEEDISKFAE  
KDNLVLGEGGITLSGGQRARISLARAVYK  
DADLYLLDSPFGYLDVLTEKEIFESCVCKL  
MANKTRILVTSKMEHLKKADKILILHEGSS  
YFYGTFSELQNLQPDFSSKLMGCDSFDQFS  
AERRNSILTETLHRFSLEGDAPVSWTETKK  
QSFKQTGEFGEKRKNSILNPINSIRKFSIVQ

KTPLQMNGIEEDSDEPLERRSLVPDSEQG  
 EAILPRISVISTGPTLQARRRQSVLNLMTHS  
 VNQGQNIHRKTTASTRKVSLAPQANLTEL  
 DIYSRRLSQETGLEISEEINEEDLKECFFDD  
 MESIPAVTTWNTYLRYITVHKSLIFVLIWC  
 LVIFLAEVAASLVVLWLLGNTPLQDKGNS  
 THSRNNSYAVIITSTSSYYVFYIVGVADT  
 LLAMGFFRGLPLVHTLITVSKILHHKMLHS  
 VLQAPMSTLNTLKAGGILNRFSKDIAILDD  
 LLPLTIFDFIQLLLIVIGAIAVVAVLQPYIFV  
 ATPVIVAFIMLRAYFLQTSQL, KQJESEG  
 RSPIFTHLVTSLKGLWTLRAFGRQPYFETL  
 FHKALNLHTANWFLYLSTLRFQMRIMIEMI  
 FVIFFIAVTFISILTTGEGEGRVGIILTLAMNI  
 MSTLQWAVNSSIDVDSLMSRSRVFKFID  
 MPTEGKPTKSTKPYKNGQLSKVMIIENSH  
 VKKDDIWPSGGQMTVKDLTAKYTEGGNA  
 ILENISFSISPGQRVGLLGRTGSGKSTLLSA  
 FLRLLNTEGEIQIDGVSWDSITLQQWRKAF  
 GVIPQKVIFSGTFRKNLDPYEQWSDQEIW  
 KVADEVGLRSVIEQFPGKLDLFDLVDGGCV  
 LSHGHKQLMCLARSVLSKAKILLDEPSA  
 HLDPVTYQIIRRTLQAFADCTVILCEHRIE  
 AMLECQQFLVIEENKVRQYDSIQKLLNER  
 SLFRQAISPSDRVKLFPHRNSSKCKSKPQIA  
 ALKEETEEEVQDTRL

This is the protein sequence obtained from SWISS-PROT.

Figure 2 shows the E-chain of the protein molecule. The normal protein contains two histidine residues at positions 608 and 620. The histidine at 620 is converted into proline due to a point mutation of 'cac' to 'ccc'. The amino acid histidine is depicted in red and the proline residue is sketched in green.

In conclusion, the SNP analysis clearly proves that the disorder, cystic fibrosis is the result of a change in a single nucleotide of the CFTR gene. This study underlines the significance of

single nucleotide polymorphisms.

Though, environmental and lifestyle factors play a major role in causing a disease, the genetic factors are more important to be considered as they can be predetermined and accordingly treated. This dysfunction can also be attributed to the disease chronic rhino sinusitis.

This type of analysis using bioinformatics is a novel approach to the identification and characterization of the mutated genes. It is accurate and faster than the wet lab techniques. This strategy can be followed for a variety of other genetic disorders and can serve useful in drug designing.

## References

- Four case reports of Chinese cystic fibrosis patients and literature review (2017)
- A survey of newborn screening for cystic fibrosis in Europe (2007)
- Exploring cystic fibrosis using bioinformatics tools: A module designed for the freshman biology course (2011)
- Identification of SNPs in the cystic fibrosis interactome influencing pulmonary progression in cystic fibrosis (2012)
- Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, et al., *Science* 1989; 245:1066-73. (AN: 89368940)
- Structure of O-Antigen and Hybrid Biosynthetic Locus in *Burkholderia cenocepacia* Clonal Variants Recovered from a Cystic Fibrosis Patient (2017).

## How to cite this article:

Balakrishnaraja Rengaraju, Keerthana, Akila, K. Pavithra, D. Vinotha, Sri Harsha Challapalli and Arunava Das. 2017. Inquest of the SNP in Cystic Fibrosis – A Bioinformatic Approach. *Int.J.Curr.Microbiol.App.Sci*. 6(8): 1255-1263. doi: <https://doi.org/10.20546/ijcmas.2017.608.152>