GENOMIC AND EVOLUTIONARY INVESTIGATIONS ON MT-ND GENE FAMILY FOR ITS REGULATORY ROLE IN ALZHEIMER`S DISEASE

Dissertation submitted in partial fulfilment of the requirement for the degree of

MASTER OF SCIENCE

IN

BIOTECHNOLOGY

By

Gargi Prasher

217811

Under the guidance of

Dr. Tiratha Raj Singh

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

WAKNAGHAT, SOLAN, 173234, H.P., INDIA

CERTIFICATE OF ORIGINALITY

This is to certify that the thesis titled **"Genomic and Evolutionary investigations on MT-ND gene family for its regulatory role in Alzheimer's Disease"** is an original work of the student and is being submitted in partial fulfilment for the award of the Degree of **Master of Science (Biotechnology)**. This dissertation thesis has not been submitted earlier either to this University or to any other University/ Institution for the fulfilment of the requirement of any course of study.

(Dr. Tiratha Raj Singh) **(Gargi Prasher)**

Professor
Dept. of Biotechnology and Bioinformatics,
JUIT, Waknaghat
HP-173234, India

UNDERTAKING FROM THE CANDIDATE

I, Gargi Prasher, present the project entitled titled **"Genomic and Evolutionary investigations on MT-ND gene family for its regulatory role in Alzheimer's Disease."** Though care has been taken while writing this report, there may be still some errors (typographical or otherwise) which are inadvertent on my part.

Place: **Jaypee University of Information Technology, Waknaghat, Distt. Solan, H.P.**

(Gargi Prasher)

CERTIFICATE OF UNIVERSITY FACULTY GUIDE

Enrollment No. - 217811

This is to certify that Ms. Gargi Prasher of M.Sc. (Biotechnology) has completed this Research/ Dissertation Project under my supervision in partial fulfilment for the award of Master of Science Degree in Biotechnology from Jaypee University of Information Technology, Waknaghat, Distt. Solan, Himachal Pradesh.

Course Name: DISSERTATION

Course Code: 20MS9BT491

(Dr. Tiratha Raj Singh) **(Gargi Prasher)**

Professor
Dept. of Biotechnology and Bioinformatics,
JUIT, Waknaghat
HP-173234, India

PLACE: Jaypee University of Information Technology, Waknaghat, Solan, H.P.

DATE:

ACKNOWLEDGEMENT

I, Gargi Prasher would like to express my profound gratitude to my guide **Dr. Tiratha Raj Singh** for his guidance, support, and constant encouragement throughout the course of this project work. He has been more than just my project guide; at times a mentor to rescue me out of my doubts. He has always helped me to work hard and taught me how to implement different ideas to deal with the problem. Moreover, he taught me to not give up and many other valuable lessons.

Furthermore, I would like to acknowledge Vice-Chancellor **Prof. (Dr.) Rajendra Kumar Sharma, Prof. (Dr.) Ashok Kumar Gupta** , Dean of academics & research for providing me with an opportunity to be a part of the institute and to complete my Master's Degree.

I also want to mention the HOD of Biotechnology and Bioinformatics **Prof. (Dr.) Sudhir Syal** has been a source of immense motivation and inspiration both for my academic and personal life. He was never, and I know will never be, more than just a phone call away. He has helped me in almost every aspect I have asked him for.

In addition, I would like to thank all the faculty members of the BT/BI Department of JUIT, who have helped me whenever I needed and would like to thank all the lab engineers and specially Ms. Somlata Sharma ma'am for providing me with a workplace and for always motivating me.

I would also like to appreciate my classmates those have played a great role in shaping this project work. They have been my constant support and cheered me up at hard times. They helped me whenever I had any doubts. Thanks a lot!

A big thanks to all open-source tool providers because of which I was able to complete my project.

I would like to thank the almighty God for his grace throughout my life. Last but not the least I would like to thank my mother and Father who have always supported me through thick and thin and have been a constant source of encouragement and support.

LIST OF TABLES

Table Number **Title**

LIST OF FIGURES

LIST OF ABBREVATIONS

AD – Alzheimer’s disease

Aβ- Amyloid Beta

nsSNPs – Non synonyms single nucleotide polymorphism

CUB- Codon usage bias

MT-ND - Mitochondrially encoded NADH dehydrogenase

OXPHOS- Oxidative phosphorylation

mtDNA – Mitochondrial Deoxyribonucleic acid

ATP – Adenosine triphosphate

SAD – Sporadic Alzheimer’s disease

FAD – Familial Alzheimer`s disease

TAMM41 - TAM41 Mitochondrial Translocator Assembly and Maintenance Homolog

PTGS1 - Prostaglandin-Endoperoxide Synthase 1

ROS – Reactive oxygen species

TABLE OF CONTENT

ABSTRACT

In Alzheimer's disease (AD), amyloid-beta(Aβ)-induced neuronal damage is characterised by mitochondrial dysfunction. Researchers are now considering the potential which are related to mitochondria and intramitochondrial. Given the new emphasis on the intracellular biology of Amyloid beta and its precursor protein (APP), Amyloid beta could trigger which cause neurotoxicity. In order to promote new perspectives on the development of AD, we study the crucial part that mitochondrial dysfunction plays in AD with Codon usage analysis and with SNPs detection (CUB) is the term describing the phenomena of inconsistent use of similar codons, where a particular codon is preferred over another. Two key evolutionary factors, namely mutation pressure and selection, are known to influence CUB. Since there has not been any research published yet, we employed a range of methods to comprehend the style in codon utilisation in MT-ND genes, PTGS1, and TAMM41 (nuclear-related MT-genes), which are involved in complex I of the respiratory chain (RC). Protein structure and function are altered by single nucleotide polymorphisms (SNPs) associated with illness. The SNPs of human mitochondrial ND GENES linked with illness were thoroughly analysed in the current study. Four extremely harmful nsSNPs (L285P, L71P, D393H, and Y59C) were found by SNP screening using a range of both sequence and structure-based methods. Both a conservation study and malignant evaluation shows most of the disease condition nsSNPs are found at exceptionally stable residues.

CHAPTER-1

INTRODUCTION

1.1 Aloysius "Alöis" Alzheimer carefully outlined the signs and symptoms of his 51-year-old patient Auguste Deter at the public facility in the German city of Frankfurt, which included troubles with memory and other cognitive disabilities [1]. Auguste Deter's brain was histologically analysed using silver's properties staining, and AD detected the development of NFTs, neuritic plaques, and amyloid beta after her demise. Such outcomes constituted the distinguishing feature of the illness, later named Alzheimer`s Disease. AD is the most prevalent form of dementia globally. One person is diagnosis with dementia every three seconds. In 2018, there were approximately 50 million dementia sufferers worldwide, of whom two-thirds had AD. The necessity for a more thorough investigation to enhance patient care and potential treatments is highlighted by the rising cases and high mortality of AD. It has been seen that mitochondrial dysfunction is an apparent feature that needs to be studied. Through oxidative phosphorylation, the mitochondria produce the majority of cellular ATP (OXPHOS). Most of the nuclear DNA-encoded genes are required to form each of the five OXPHOS complexes (nDNA). On the other hand, only a small subset of OXPHOS-related genes is encoded by mtDNA. Numerous neurodegenerative diseases and typical aging have both been linked to damaged mtDNA. Furthermore, neurological abnormalities are common in people with legitimate mitochondrial illnesses.[1] Numerous pathogenic insults can harm mtDNA, but oxidative damage caused by mitochondrial reactive oxygen species has received considerable attention. The physiologic roles of (ROS) in cellular viability and overall well-being are still unresolved even though ROS may trigger cell destruction and that there is a connection between high ROS and damage from oxidation in neurodegenerative diseases.

Although some studies observed no differences in the overall load of brain mtDNA point mutations between AD and control, other studies reported contradictory findings. More sensitive studies have revealed that the lack of cell-specific mtDNA differentiation may also be a factor in the variability.[2]

Higher oxidative damage in the AD brain was thought to be the cause of increased mtDNA mutations. In fact, mtDNA suffered an average of three times more oxidative damage in the brains of AD patients compared to age-matched controls, and mtDNA contained almost ten times more oxidised bases than nuclear DNA. As a matter of fact, individuals with preclinical Alzheimer's disease (PCAD) and MCI had much greater quantities of oxidised nucleic acids in

their mtDNA, showing that this was a symptom of the disease at an early stage. It is yet unknown how these discoveries relate to the aetiology of human AD. [1]

1.2 Overall, these investigations indicate that mtDNA variabilities, mutations, and alteration are likely crucial in the pathophysiology of AD. The mitochondrial cascade theory states that inherited mtDNA variants define one's vulnerability, and that the phenotypic manifestations are determined by the accumulation of brain somatic mtDNA modifications and mutations through time that reflect the impact of the environment.

1.3 There are conflicting results about the relationship between Alzheimer's disease (AD) and mitochondrial DNA (mtDNA) polymorphisms.[2] Therefore, there is a need to understand the role of mitochondria in the regulatory processed associated with the AD's initiation and progression.

OBJECTIVES

This work has been majorly categorized into three objectives as:

- Analyzing the mitochondrial- NADH dehydrogenase (MT-ND) gene family through various bioinformatic approaches.
- Genomic Investigations for its plausible Regulatory role in AD.
- Functional and Evolutionary Analysis of involved genes, proteins, and its applications.

CHAPTER-2

LITERATURE SURVEY

- **Genetic Variables**

AD can be classed as either sporadic or familial. Since the cause of onset is unclear in about 99% of cases with AD. Aging continues to be the key factor. Symptoms typically start to appear between the ages of 60 and 65. late-onset AD or Sporadic AD are undoubtedly influenced by both genetics and environment. The three apolipoprotein E (APOE) gene variations e2, e3, and e4 are the primary genetic causes of SAD.[1] Cholesterol and other lipids are packaged and transported into the bloodstream by the APOE protein. Each person inherits one of the three APOE gene types, often the e3 version, from each of their parents. Compared to inheriting e3, having e4 greatly raises the likelihood of developing AD, whilst having e2 may lower the risk. The likelihood of developing AD is made higher by having a single copy of the e4 gene. by three times compared to two copies of the e3 gene, whereas two copies of the e4 gene increase the risk by 8–12 times. The e4 version of the APOE gene does not, however, ensure that AD will develop. In the past ten years, studies using whole genome sequencing and genome-wide association data have discovered a few novel genetic factors that are linked to a high risk of LOAD, including CD33, ABCA7, complement C3b/C4b receptor 1, CR1, TREM2.

Most of them are particularly or is predominantly found in brain, with TREM2. A vital component of microglial phagocytosis, chemotaxis, survival, and proliferation is the cell surface receptor known as TREM2. The risk of AD is increased two to four times by R47H, a mutation causing TREM2 to lose function, which is comparable to the risk brought on by inheriting a single copy of the APOE e4 variation. All three APOE isoforms as well as other lipoprotein particles are bound by TREM2.[3] When TREM2 is mutated due to a disease, less TREM2-APOE binding occurs, which is thought to have an impact on AD pathogenesis. The loss of microglia's ability to maintain homeostasis and subsequent neurodegeneration are caused by APOE-TREM2 crosstalk, which causes a transcriptomic modification in AD microglia. The signs of familial AD (FAD) appear before the age of 60, or perhaps before 55. In comparison to people with no family history of AD, those in the following sibling or even the next generation are more prone to have the condition if one or more family members have been diagnosed with AD (Alzheimer's Association, 2018). Presenilin 1 (PSEN1), presenilin 2 (PSEN2), and the amyloid precursor protein (APP) all contain mutations that are related to FAD. Each of the above genetic elements linked to SAD and FAD and the mismatch between amyloid-(Aβ) plaque formation and elimination are important. [1]

- **Pathology**

Amyloid Beta plaque aggregation and build-up, along with NFTs are the primary cause of AD, according to compelling, though not conclusive, evidence from research conducted in the last 20 years, The "amyloid hypothesis" or "amyloid cascade hypothesis" refers to this idea. NFTs develop inside neurons, whereas A plaques amass outside of cells. A plaque and NFT build-up cause neurodegeneration, which results in the death of synaptic and neurotransmitters.[1] A-beta is formed through the progressive degradation of amyloid precursor protein with the BACE1, a secretase, a protein complex including presellin1, presellin2 as its enzymatic component 37–43 amino acid A–peptides are produced as a result of processing APP, with A–42 being the most hazardous. Due to the A peptides' resistance to proteolytic breakdown and an imbalance between their synthesis and elimination, A builds up and causes A -beta plaques to form. A-beta is capable of spontaneous self-assembly and coexistence with many physical forms. One of the variants has intermediate assemblies made up of oligomers of 2–6 peptides. A can also form fibrils, which come together to form sheets with pleats and become insoluble fibres that form advance amyloid plaques. The most damaging forms of A are thought to be intermediate amyloids and soluble oligomers.[2] In the brain, neurolysin and insulin degrading enzyme (IDE) break down A peptides. Ageing and AD both result in decreased expression of IDE and neurolysin, which upsets the homeostasis of A-beta. NFT production and accumulation, which are primarily composed of tau protein that has been hyperphosphorylated, is another defining feature of AD. The soluble protein tau, which is found in neurons, is crucial for preserving the stability of microtubules. In addition to facilitating healthy axonal transport and neuronal growth, microtubules give neurons structural support Tau becomes insoluble, loses its affinity for microtubules, and forms paired helical filamentous formations when it is hyperphosphorylated.[3] Tau's turnover is further decreased in comparison to normal tau due to hyperphosphorylation, which also makes tau resistant to being degraded by calcium triggered neutral proteases. Like A-beta plaques, hyperphosphorylated tau intermediate aggregation is hazardous and causes cognitive impairment Changes in tau protein are typically thought to be downstream of A in the pathology of AD, however in certain cases, since memory loss caused by tau gene mutations lacks amyloid plaques, it's possible that tau and A-beta plaques work independently or perhaps even together to increase each other's toxicity. Recent research investigations indicate both A-beta and tau plaques have an immediate impact in the pathophysiology of AD. research has highlighted the critical role that inflammation plays in the pathophysiology of AD. Microglia, known as the brain's native immune system cells, are important in AD-related activation. Microglia gather in close proximity to amyloid and phago cytoses it to get rid of them. If the plaques become complexed with apoE, clustering, and lipo-

proteins with a low density, the process succeeds better.[4] Via CD 36 and the toll-type receptor 4 and 6 heterodimer, A-beta engages with microglia. Inflammasome NLRP3 appears to be activated by this interaction, leading to the release of IL-1. Long-term production of these cytokines promotes disease pathophysiology and is cytotoxic.

- **Function and structure of mitochondrial DNA**

The super-coiled, elliptical dual-stranded human mtDNA, it measures 16,569 base pairs and encodes 37-genes required for oxidative phosphorylation and MT-protein synthesis, was first discovered in 1963. The inner mitochondrial membrane is home to five multi-subunit enzyme complexes that make up the OXPHOS system. The mtDNA codes for 13protein subunits, In additional to these 13 parts, the mtDNA also encodes for 22 tRNAs and 2 rRNAs, one or more of the necessary subunits for the Complex I, Complex III, and cytochrome operator. The heavier strand and lighter strand of mtDNA are differentiated as such because the first one is guanine rich and the one that follows is cytosine rich. 9-genes are encoded on the L strand after the initial 28 genes, which are all on the H-strand.[5]

- **Multiple copies in mitochondria are found in cells**

The bulk of cells contain 1,000 mitochondrial genomes. The nuclear genome, in contrast, is only present in two copies per cell. The requirement for energy by a cell often affects the numbers of mtDNA molecules. The process of replicating mtDNA is unrelated to Just a handful of recognised enzymes play a part as well as the cell phase. Two of these actors are the mtDNA helicase Twinkle and the mitochondrial polymerase POLG; Both mtDNA deficiencies and infections of the mitochondria have been attributed to polymorphisms in either of these genes. A single cell frequently contains both mutant and the normal form strands of mtDNA as in result of this high copy number (mtDNA heteroplasmy). Because of the independence of mtDNA replication from the cell cycle and the possibility of mtDNA segregation Heteroplasmy levels change dynamically during replication and can fluctuate over the course of an organism's both post-mitotic and mitotic cells and tissues throughout their existence. The percentage of heteroplasmy is a significant determinant in determining the clinical severity of mitochondrial disorders, along with the type of mutation.[5] For reduced mitochondrial function and phenotypic development, The fraction of mutated mtDNA must pass a biological threshold. Although the mutation affects this threshold, the cell heteroplasmy threshold values can range between - and tissue-type. A discernible phenotype must be present between 70% and 90%. The central nervous system is impacted by mitochondrial DNA disorders. Studies conducted over the past few decades have shown that Increased ROS and mtDNA damage don't seem to

be related in any way. There are several mitochondrial disorders that also include CNS impairment, however heteroplasmy levels can differ.

Additionally, several neurodegenerative illnesses, many of which have altered or damaged mtDNA and some of which have generalised oxidative damage, are defined by mitochondrial malfunction. The pathogenesis of many neurodegenerative diseases and disorders that include mtDNA will be covered in the sections that follow, along with how mtDNA mutations may impact ROS levels. In the past 30 years, more than 330 harmful point mutations in human mtDNA have been discovered. The mtDNA molecule contains these mutations all over it. Mitochondrial encephalopathies are caused by several of these point mutations. Investigations into the causes and effects of these mutations are ongoing.[5]

- **Age-related neurodegeneration, mitochondrial DNA abnormalities, reactive oxygen species**

oxidative damage, and mitochondrial malfunction are associated with age-related neurodegenerative disorders. Correlations between disease progression and higher ROS in various age-related neurodegenerative disorders suggest mitochondrial malfunction. There is, however, no solid proof that mtDNA damage or mutation, particularly the widespread deletion, contributes to illness development.[2]

The most common late-onset progressive neurodegenerative condition AD, bears a connection to defects in cytochrome c oxidase. The cortex and hippocampus are the two areas where amyloid beta fragments most frequently form cytotoxic plaques. A-beta fragments have a deleterious impact on mitochondrial function, indicating that A poisoning is the cause of mitochondrial malfunction. Cytotoxic plaques made of amyloid beta fragments are more frequently found in the cortex and hippocampus. Amyloid beta fragments impair mitochondrial function, indicating that the dysfunction of the mitochondria is a result of the toxicity of amyloid beta. According to the "mitochondrial cascade hypothesis for AD", a person’s genes control the intrinsic synthesis of ROS, which in turn regulates how severe oxidative damage takes place. The respiratory chain's protein-making genes are controlled by this. Furthermore, it was hypothesised that when oxidative mtDNA damage accumulated, this would eventually result in decreasing Amyloid beta toxicity, elevated oxidative stress, and ATP levels would all contribute to the neurodegenerative process.[3]

Even though the pathophysiology of the disease has not been linked to a causative mtDNA mutation in AD patients, it has been discovered that AD patients had a higher overall burden

of rare point mutations than juvenile control systems, but not regulates of the same age. While evaluating for the presence of mtDNA damage in victims, the late phases of AD deterioration, where motor neurons with large amounts of mtDNA damage may have essentially gone, are of interest. In the hippocampus of early-stage of the AD patients, there was an increase in the frequency of mtDNA mutations; However, it was found that replication errors, not oxidative damage, were to blame for these mutations. The prevalence of the common mutation in AD patients as well as controls with comparable ages has been the subject to several investigations have produced negative results. Furthermore, there is controversy surrounding research examining different mtDNA haplo-groups as potential AD risk variables. [1]

CHAPTER-3

MATERIALS AND METHODS

3.1 RETRIEVING SEQUENCE DATA

The National Centre for Biotechnology Information (NCBI) Genome repository

(http://www.ncbi.nlm.nih.gov/Genbank/) provided the CDS of all the MT ND, TAMM41, PTGS1 for this investigation and all these sequences were downloaded in FASTA format. Details are available in Table 1.

Table:1. Representation of nucleotide and protein length obtained from NCBI.

MITOCHONDRIAL AND OTHER PROTEIN-CODING GENES USED IN THIS STUDY	NUCLEOTIDE (Length in bp)	PROTEIN (Length in aa)
MT-ND1	955bp	318aa
MT-ND2	1,041bp	347aa
MT-ND3	345bp	115aa
MT-ND4	1,377bp	459aa
MT-ND4L	296bp	98aa
MT-ND5	1,811bp	603aa
MT-ND6	524bp	172aa
TAMM41	1,359bp	452aa
PTGS1	1,614bp	537aa

3.2 STRING database analysis of protein-coding genes

A complex network of functional relationships between biomolecules is essential for cellular life. Due to their flexibility, specificity, and variety, protein-protein interactions among these connections are highly significant. With its STRING database, assemble all previously

identified and projected protein relationships, encompassing both physical and functional exchanges. In order to do this, STRING gathers and grades data. from several sources, such as automatic text mining databases of interaction, scientific literature experiments and pathways/complexes with annotations, computational predictions of interactions based on co-expression and consistent evidence of interactions, and preserved genomic context between different organisms. With more than 14 000 creatures in the upcoming version 11.5 of the library, STRING promises to provide comprehensive coverage.[6]

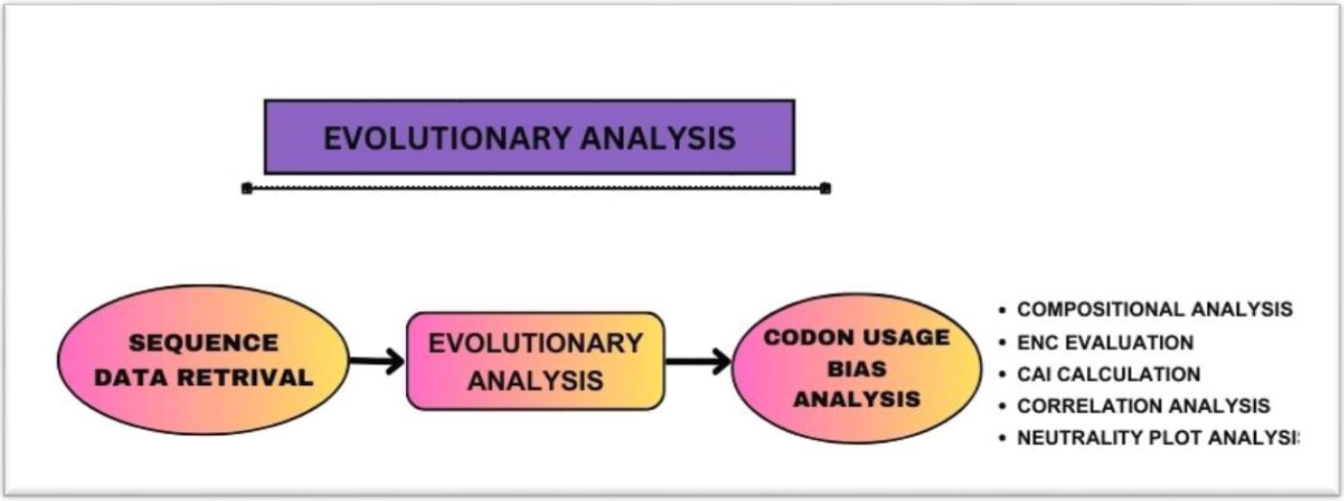


FIGURE:1. Diagrammatic representation of the processes involved in evolutionary analysis.

3.3 CODON USAGE BIAS (CUB) ANALYSIS

3.3.1 Compositional Analysis

Using CAIcal Software, the compositional characteristics of the MT ND gene family, TAMM41, PTGS1 gene were examined. The overall nucleotide sequence composition (A, C, T, G, GC%), the third position's nucleotide makeup (A3, C3, T3, G3, GC3%), the content of GC The top three codon positions, respectively, and the overall GC content (AT, GC) were some of these features. [6]

Table:2. List of compositions of MT-ND family obtained from various tools and calculations.

Gene	A%	T%	G%	C%	%GC	%A(3)	%T(3)	%G(3)	%C(3)	%GC(3)
ND1	28	26	11	35	47.649	36.991	11.285	5.329	46.395	51.724
ND2	31	27	9	33	42.939	36.888	16.427	5.476	41.21	46.686
ND3	29	32	10	29	40.58	42.609	14.783	4.348	38.261	42.609
ND4	30	27	9	34	44.299	38.344	14.161	3.704	43.791	47.495
ND4L	29	29	12	30	42.761	39.394	16.162	6.061	38.384	44.444
ND5	30	26	10	34	44.868	35.43	15.232	3.642	45.695	49.338
ND6	33	24	9	34	42.476	20	41.143	35.429	3.429	38.857
TAMM41	26	28	22	24	46.799	21.192	33.775	17.881	27.152	45.033
PTGS1	22	26	25	27	35.311	31.356	32.203	16.102	20.339	36.441

3.3.2 Effective number of codons (ENC)

It can be applied to evaluate codon usage bias. A greater ENC value indicates less codon use bias, and vice versa. For Table 1, the ENC of a cds can be calculated as Nc = 2 + 7/F2 + 3/F3 + 6/F4 + 1/F6 + 1/F8 wherein Fk is the mean of Fk readings for k-folding deformed amino acids. The COUSIN tool was used to determine the ENC values.[7]

Table:3. Effective number of codon data obtained from COUSIN tool.

GENE	ENC
ND1	34.281
ND2	40.103
ND3	31.587
ND4	38.624
ND4L	32.757
ND5	41.596
ND6	31.988
TAMM41	47.152
PTGS1	35.937

3.3.3 CAI Evaluation

Based on the codon sequence of a gene uses the CAI as a quantitative technique to forecast the degree of gene expression. The CAI value is between 0 and 1. The highest relative adaptiveness values are simply found in the most frequent codons, and greater CAI sequences are favoured over lower CAI sequences. The COUSIN tool was used to calculate the CAI values. [8]

Table:4. Codon adaptive index data obtained from COUSIN/CaLcaI calculator.

GENE	CAI
ND1	0.795
ND2	0.721
ND3	0.791
ND4	0.774
ND4L	0.751
ND5	0.768
ND6	0.598
PTGS1	0.475
TAMM41	0.499

3.3.4 Correlation Analysis

Correlation analysis was carried out between several aspects of the data to determine the connection between identical codon usage behaviours and sequence the material's composition. The Spearman's rank correlation analysis is used to do this analysis. All statistical operations were performed using the Windows version of SPSS 16.0.[9]

3.3.5 Neutrality plot Analysis

A regression equation of GC12 on GC3 is taken to describe a relationship between the GC coding for proteins genes' components. This technique is referred to as neutrality plot analysis. The scatter plot in this instance plots the GC12 values as the ordinate and the GC3 values as the X-axis. Alternatively put, the graph measures the influence of mutation pressure and natural0selection on the pattern of codon usage If it is determined that every gene has a narrow GC3 distribution, the plot may indicate minimal mutational bias or higher GC content conservation.[9],[10]

Table:5. Data for neutrality plot analysis where the average of GC12% (GC contents at the first and second codon positions) taken in account with the respect of GC1, GC2, GC3%

GENE	AVERAGE [GC12]	%GC (1)	%GC (2)	%GC (3)
ND1	45.6115	49.53	41.693	51.724
ND2	41.066	41.21	40.922	46.686
ND3	39.565	46.087	33.043	42.609
ND4	42.7015	45.752	39.651	47.495
ND4L	41.919	51.515	32.323	44.444
ND5	42.6325	45.199	40.066	49.338
ND6	44.286	51.429	37.143	38.857
TAMM41	47.682	48.344	47.02	45.033
PTGS1	34.7455	35.593	33.898	36.441

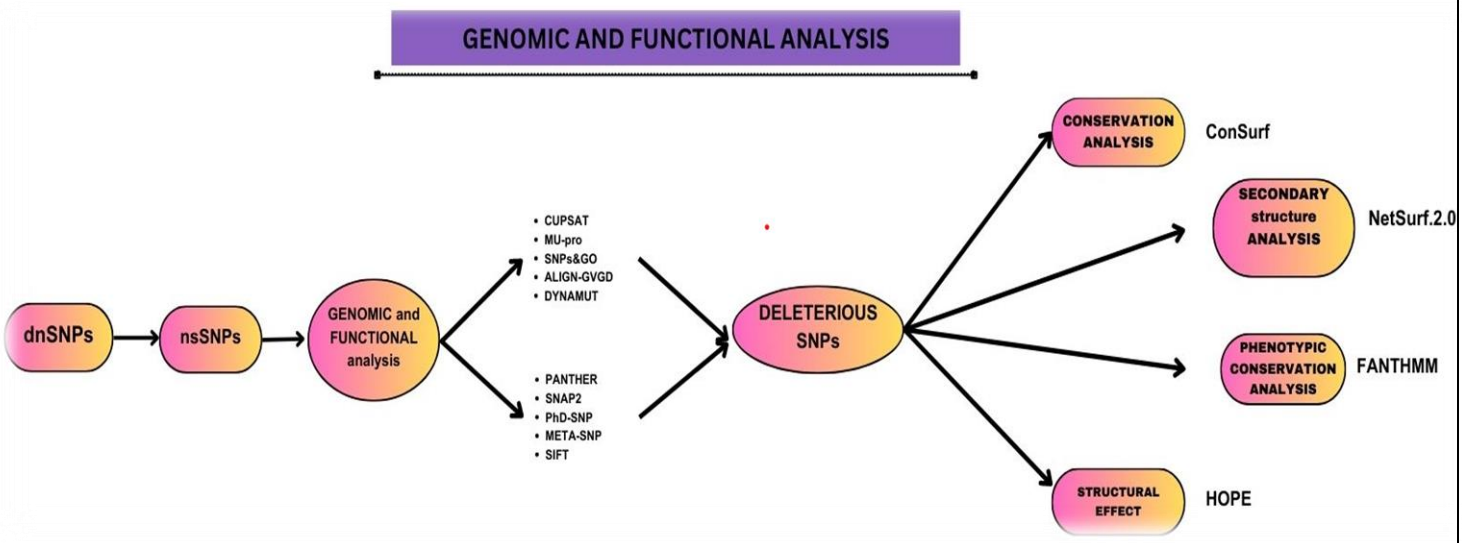


FIGURE:1s. Diagrammatic representation of the processes involved in genomic and functional analysis with the respect of tools used.

3.4 IDENTIFYING DELETERIOUS MISSESNSE nsSNPs

3.4.1 DETECTION OF DELETERIOUS nsSNPs USING SEQUENCE BASED TOOLS

3.4.1.1 PANTHER

PANTHER (protein analysis through evolutionary relationship) categorization system in bioinformatics is a sizable biological database that has been carefully maintained and contains gene and protein categories and its functionally related sub-categories.[10] It is applied in classification and gene products and determine their functions. The GO Reference Genomic Project, which was started to coordinate proteins and their genes for enormous throughput research, includes PANTHER as its member [8] 176 pathways are included in PANTHER utilising the Cell Designer tool. The following file types are available for downloading PANTHER paths.

- Systems Biology Graphical Notation (SBGN-ML)
- Systems Biology Markup Language (SBML)
- Bio PAX

We used this tool from bio. Tools/panther http://www.pantherdb.org/tools/ Where VARIENT rs IDs are the input, selecting homo-sapiens as the needed organism and its functional classification is viewed in the gene list.

3.4.1.2. SNAP-2

In order to classify genetic variation as having an effect (+100, strongly anticipated) or being neutral (-100, strongly predicted), SNAP2, a bioinformatic tool based on neural networks, examines the solvent accessibility of native and mutant proteins.[11] uses neural networks to predict the impact of single amino acid changes (SNPs). demonstrates prediction results as a heatmap.[12]

We used this tool from bio. Tools/snap2 https://github.com/Rostlab/SNAP2 where the protein sequence is in a FASTA format and we go for run prediction.

3.4.1.3. PhD-SNP

The PhD-SNP is web service provides an intuitive user interface for predicting effects of SNVs both in code and without coding domains.[14] PhD-SNP's standalone version is simple to install and use on typical laptop computers. It can predict the consequences of 1,000 SNVs in around two minutes. When the programme is running in web mode, this time rises based on the network speed.[15] Despite having basic input features, PhD-SNP performs on par with cutting-edge techniques that demand more data and resources. This makes PhD-SNP both a baseline benchmark tool for comparing predictors based on less complex input variables and a trustworthy and movable instrument for evaluating the effects of new variations.[16]

We used this tool using https://snps.biofold.org/phd-snp/phd-snp.html where the nucleotide number of proteins in FASTA format is input and the query will be submitted after selecting normal residue, and mutated residue at the specific position of the Amino acid.

3.4.1.4. META-SNP

In human DNA, SNPs One of the most prevalent kind of variants in genes. A haplotype is the set of SNPs present in each of a diploid organism's two copies of a certain chromosome.[17]

Numerous uses for haplotype data include medication development and the diagnosis of genetic diseases.[18] These meta servers aggregate the outcomes predicted by several in silico technologies to create a consensus forecast for a specific SNV. Meta-SNP integrates the results of in silico technologies using a random forest technique. It is determined whether a mutation is "Disease" or "Neutral" [19].

We used this tool using http://snps.biofold.org/meta-snp where amino acid sequence and the list of mutations are put as input and after submit the query the output will be send via email.

3.4.1.5. SIFT

To locate, describe, and match local features in images, David Lowe developed the scale-invariant feature transform (SIFT) in 1999.[20] The SIFT feature descriptor can robustly distinguish objects even among clutter and when partially obscured since it is invariant to uniform scaling, orientation, light variations, and affine distortion.[20][21][22] The original SIFT method is briefly described in this section, along with a few competing approaches for object recognition in the presence of partial occlusion and clutter. In order to choose local scale invariant reference frames for the SIFT descriptor, receptive field measurements of pictures are used.[22]

We used this tool using https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html Where Protein sequence and the substitutions of the interest are used as input file the output file will sort the INTOLERANT mutations from the tolerant ones.

3.4.2. DELETERIOUS nsSNPs INDENTIFICATION USING STRUCTCTURE BASED TOOLS

Five Structure based tools were used to analyse the variation in the stability of the protein-coding genes of mitochondria caused by the missense SNPs that were analysed as deleterious by the sequenced based tools. CUPSAT, MUpro, SNPs&GO3d, Align-GVGD, DynaMut are the five tools that analysed the protein stability and dynamics by missense SNPs.

3.4.2.1. CUPSAT

The online communication CUPSA service analyses and forecasts modifications to protein equilibrium caused by single specific amino acid mutations in points. .[23] This tool makes predictions about the DeltaDeltaG variation in the free energy of unfolded among the original and defective proteins utilising torsion angle the ability and structural environment-specific atom capabilities.[23] The position of the residue must be altered, and the protein structure must be in Protein Data Bank format. [24] It also examines how well the mutant amino-acids can adjust to the measured torsion angles.[23]

We used this tool using http://cupsat.tu-bs.de/ where we predict mutant stability from existing PDB structure. 5XTC is the input for the MT-ND family gene, we go for the one amino acid stability prediction, input is set for prediction stability.

3.4.2.2. MU-pro

Another online bioinformatical tool that forecasts a protein's stability is MUpro. This web server was created using two machine learning techniques that forecast the frequency of single site amino acid mutation on protein stability: Support Vector Machines and Neural Networks.[26][27] Both of them computed a score that ranged from 1 to 1 as the prediction reliability after being trained using a large amount of mutation data with an accuracy of more than 84%.[28]

We used this tool using http://mupro.proteomics.ics.uci.edu/ where protein id 5XTC is used as an input, where original amino means wild type of amino acid and substituent amino acid is the mutated amino acid. Input is ready to predict the result.

3.4.2.3. SNPs & GO

A technique called SNPs&GO uses protein functional annotation to predict harmful Single Amino Acid Polymorphisms (SAPs).[29] The web server implementation of SNPs&GO (WS-SNPs&GO) is presented in this paper. For a specific protein, the server's input consists of its sequence and/or three-dimensional structure (where available), a group of target variants, and its functional Gene Ontology (GO) keywords. The server is based on Support Vector Machines (SVM).[30] The server's output lists the likelihood that each protein variant will be linked to human diseases.[31]

We used this tool using http://snps-and-go.biocomp.unibo.it/snps-and-go/index.html where the input parameter for the output is the UNIPROT the accession Information, the change site, the unmodified type residue, and the replacing residue.

3.4.2.4. Align GVGD

Align-GVGD accepts MSAs, and for each alignment column with a substitution, It determines a score for sustainability, a biochemical distance score (extension of the pairwise Grantham difference, GD), and both.[32][33] Based on the observed values of GD and GV substitutions, substitutions are divided into seven classes: C0,15,25,35,45,55,65from those that are less inclined to impede functionality to those that are most probably to do so. [34]

We used this tool using http://agvgd.hci.utah.edu/ where the input file is of protein sequence with the list of mutations.

3.4.2.5. DynaMut

DynaMut, a web server that implements two independent, well-recognized normal mode techniques, which samples configurations to investigate and display protein dynamics, is used to evaluate the impact of polymorphisms on protein dynamics and stability brought on by changes in vibrational analysis entropy..[35] By attaining a correlation of upto 0.700 in bliend test , we show that our approach surpasses competing approaches at predicting the impact of mutations on stability and flexibilityof protein (P-value 0.001). Additionally, DynaMut offers a complete suite for the investigation and visualisation of protein motion and flexibility.[37]

We used this tool using http://biosig.unimelb.edu.au/dynamut/ where we search the mutation effect analysis. Where the PDB id is used as the input with the detailed list of mutation, by selecting the chains of protein.

3.5 EVOLUTIONARY CONSERVATION ANALYSIS

A tool named ConSurf is used to map areas of protein surfaces with known 3D structures that have undergone evolutionary conservation.[41] It employs the alignment of sequence homologues of proteins whose structures are known to create phylogenetic trees. The assumed amino acid swaps that took place during the protein's evolution are then inferred from the trees.[42] The physicochemical distance between the swapped amino acid is then used to weight each exchange. Sequence data is used in the ConSurf computation, but the findings are more instructive when shown on the macromolecule's 3D structure or a model of it.[43]

We used this tool using https://consurfdb.tau.ac.il/ where the 5XTC is the protein ID (PDB) as input and select chain for different MT-ND family genes. We had performed the conservation analysis on those four mutations which are detected as deleterious in the structure-based analysis.

3.6 ANALYSIS OF SECONDARY STRUCTURE

The term "protein secondary structure" refers to the local spatial shape of the polypeptide backbone absent of side chains.[47] Beta twists and omega loops also occur, but α helices and β sheets are the two structural elements that are most frequently observed.[48]. The SOPMA bioinformatical programme is utilised with the default parameter settings. To determine a protein's secondary structure, this server combines a neural network method (PHD) and a self-optimized prediction method.[40]

We used this tool using https://npsa-prabi.ibcp.fr/ where protein sequence in FASTA format without header is the main input. Submit the query with the default parameters.

3.6.1 STRUCTURAL EFFECTS

NetSurf.2.0 tool was used to know, if there are any structural changes arises after the mutation.[51]

We used this tool using https://services.healthtech.dtu.dk/services/NetSurfP-2.0/ where the protein sequence is the input, after submitting the query the outputs are visible on the screen. After that open this link in the new tab make changes in the protein sequence where you a have got the mutation. As I made changes at position 285 from L to P (L285P). similarly do this for all the mutations we get from structure-based tools.

3.7PHENOTYPIC CONSEQUENCES OF THE SCREENED MUTATIONS ANALYSIS

The investigation of the phenotypic effects of 4 mutations was done using the FATHMM programme. The phenotypic associations chosen were those from the "Human Phenotype Ontology" and the prediction technique utilised was the unweighted algorithm under the "Inherited Disease" section.[54] The PDB ID and replacement make up the input. FATHMM gives each prediction a confidence value (a p-score) in order to facilitate interpretation and focus investigation on a small number of high-confidence forecasts.[55]

3.8 STRUCTURAL EFFECTS OF DELETERIOUS nsSNPs

HOPE server was used to visualize and analyze the mutations. HOPE stands for Have Your Protein Explained server. HOPE elucidates the molecular basis of a trait associated with a disease produced by changes in human proteins.[59]

CHAPTER-4

RESULTS

FUNCTIONAL AND EVOLUTIONARY ANALYSIS

4.1 STRING DATABASE ANALYSIS

4.1.1. INTERACTION ANALYSIS

Figure 2. shows the 13-MT protein-coding genes interaction network retrieved from the STRING database. Then the network was built between the 7 MT-ND family genes, TAMM41 and PTGS1 as these are nuclear related mitochondrial genes present in the mitochondrial inner membrane and shows their impact in Alzheimer's Disease. The input was MT-ND, TAMM41, PTGS1 and the organism was "HOMO SAPIENS". The protein coding genes are represented by nodes whereas their interactions with the edges, different colours of edges have different meaning. Blue colour edge represent that the interaction is from curated databases, purple colour represents the experimentally determined interaction, green, red, dark blue colours represent the predicted interactions. Text mining interaction is represented by the yellow colour, co-expression by brown and protein homology by light blue.

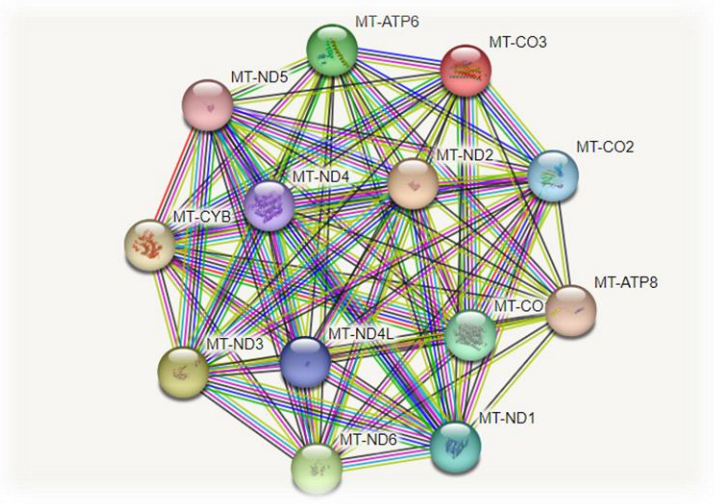


FIGURE:2. shows the interaction between the 13 -protein-coding genes of mitochondria.

Here the interaction between the genes shows their physical and functional relation. Now we perform the interaction analysis between the 7 protein - coding genes of mitochondria, TAMM41 and PTGS1. Where PTGS1 shows its physical and functional relation with other

mitochondrial genes but TAMM41 does not shows any interaction at the PPI enrichment p-value: <1.0e-16 as shown in figure3.

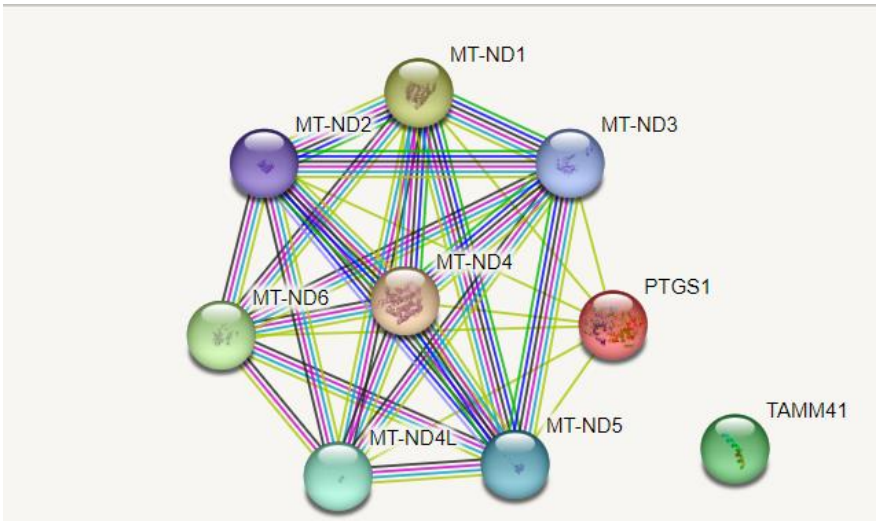


FIGURE:3. shows the interaction of query genes i.e, MT-ND family genes and, PTGS1, TAMM41. Which shows the that there is no interaction of TAMM41 Gene with early interacting partners.

The STRING database's enrichment tools allow us to create a sizable network with 145 nodes and 5291 edges, an average node degree of 71, and an average local clustering coefficient of 0.82. The network's PPI enrichment p-value of 1.0e-16 indicates that it had significantly more interactions than predicted. Which leads to this extensive node network where TAMM41 demonstrates its interaction with additional guilty genes.

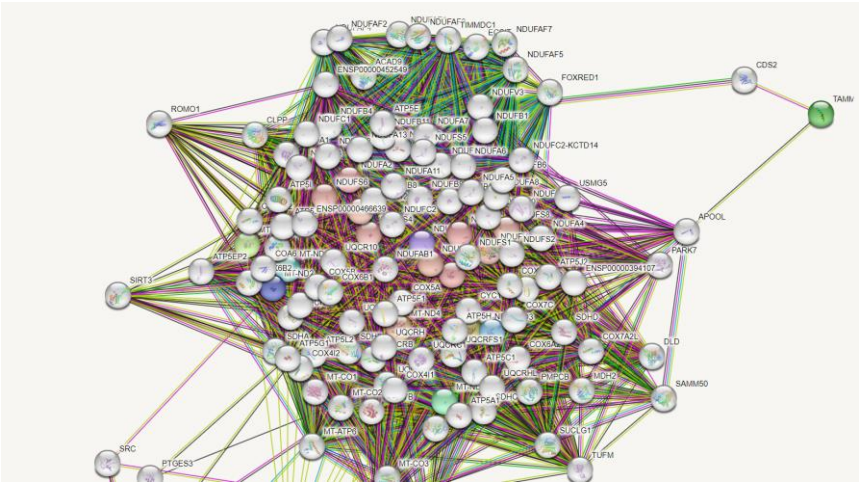


FIGURE:4. Shows the interaction of TAMM41 with the cluster after 145 nodes.

4.2. CODON USAGE BIAS ANALYSIS

4.2.1 COMPOSITIONAL ANALYSIS

compositional features of the 7-protein-coding mitochondrial genes, TAMM41, PTGS1.

One of the most significant variables that could have a major impact on the research of CUB is the base composition of the genes. [2]. So, here we observed the base composition of 7 mitochondrial ND genes, TAMM41, PTGS1 from the data shown in table 2 figure3, it was observed that the average AT level was greater than the average GC material overall. (34.05556 ,31.83333 respectively) similar trend is seen between the overall mean AT percentage at 3rd codon position and GC percentage at 3rd codon position (55.263 and 44.736 respectively). At 3rd codon position A, T, G, C Trend shows C3% is slightly greater than A3% (33.85067 and 33.57822 respectively) and T3% is greater than G3% (21.68567 and 10.88578 respectively). Trend goes like C3%> A3%> T3%> G3%. The similar trend is seen between the bases at the first codon position C%> A%> T3%>G% (31.11111, 28.66667, 27.2222, and 13 respectively).

This points to a disparate dispersion of the codons A, T, G, C in all the supplied genes, indicating a stronger preference for C in the composition. Therefore, base G shows the lowest contribution compared to other bases, according to an analysis of the overall nucleotide composition.

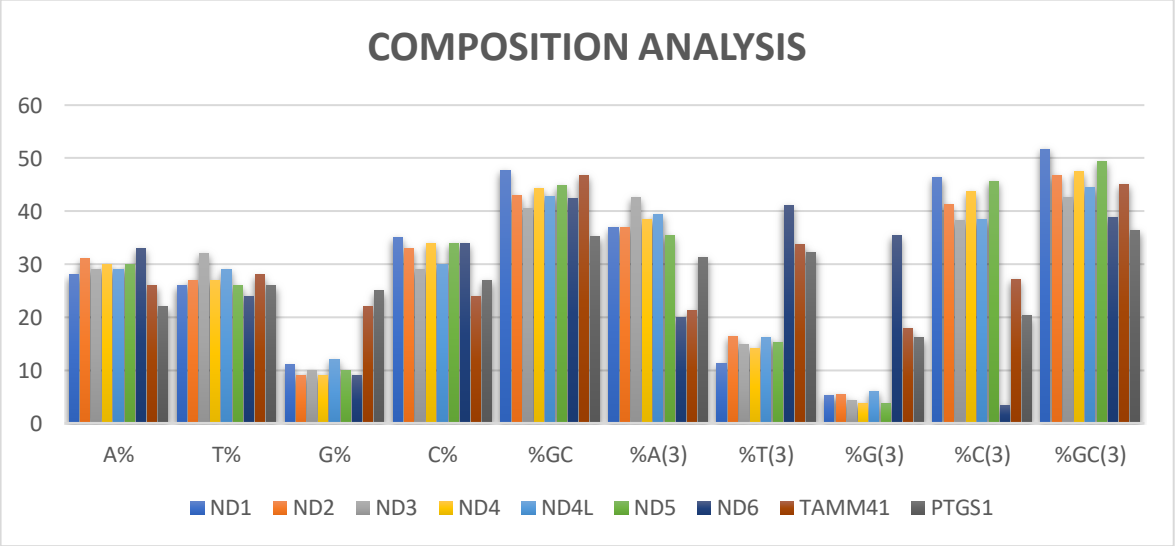


FIGURE:5. Comparative analysis of nucleotide compositions of MT-ND family genes, TAMM41, PTGS1

4.2.2 ENC ANALYSIS

To gauge the degree of codon, use fluctuation [8] among 7 mitochondrial ND genes, TAMM41, PTGS1.The ENC value was calculated for each gene as shown in table 3. The ENC value ranged between 31.988 to 47.152, with the mean value of 37.1139 shown in figure 6, which is less than 40(ENC <40) represents unstable. ENC shows the conserved genomic composition of a genome [8]. In real ENC is inversely related to the expression of the genes [8]. This means if the value of ENC is lower than 40 then it indicates the higher expression of genes and high preference to codon usage bias. our results show the lower ENC Value that indicated the high gene expression and high codon usage which results in slightly biased and would be affect the base composition.

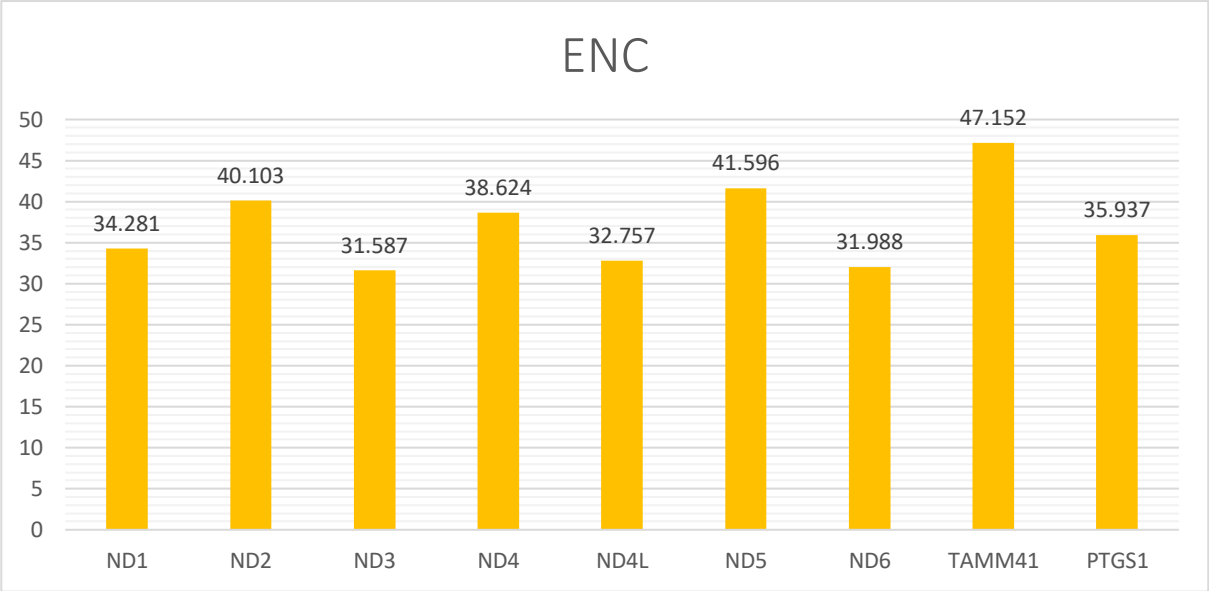


FIGURE:6. Comparative analysis ENC of MT-ND family genes, TAMM41, PTGS1

4.2.3 CODON ADAPTIVE INDEX ANALYSIS

Codon adaptive index (CAI) is frequently used to gauge gene expression levels and evaluate adaptations. [8]. Higher gene expression shoes the strong codon usage bias. As opposed to the CAI, which measures the deviation of a gene's coding regions, we calculated the ENC value, which displays the deviation from the uniform bias. The MT-ND genes, TAMM41, and PTGS1 coding sequences have been used to determine CAI in this case. CAI values ranges from 0.475 to 0.795 (CAI should range between 0 - 1) with the mean of 0.6857. which shows the higher CAI value and results in higher gene expression of the gene as shows in figure7.

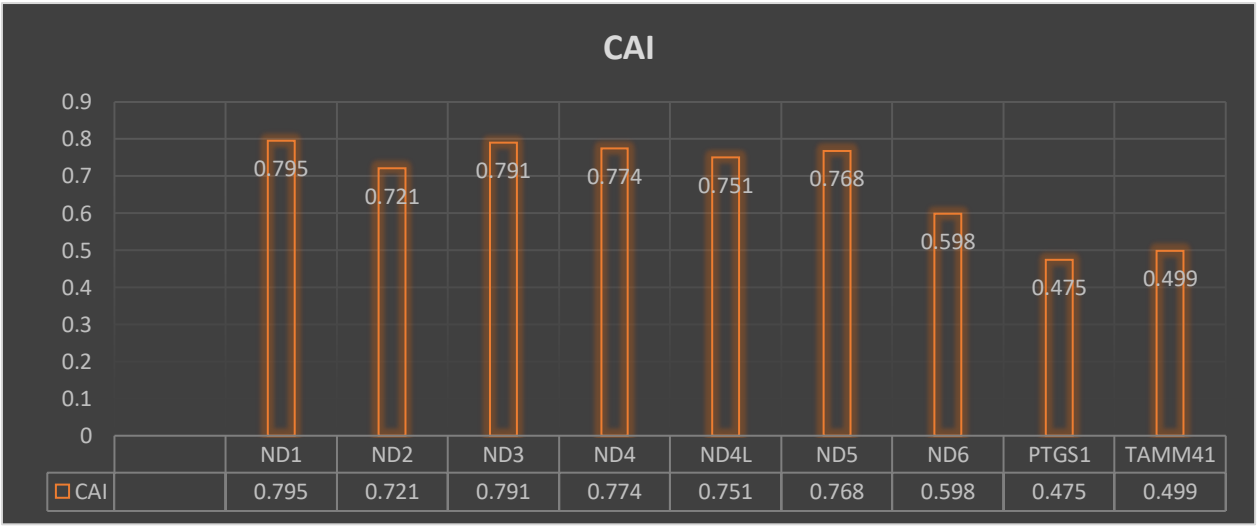


FIGURE:7. Comparative analysis CAI of MT-ND family genes, TAMM41, PTGS1

4.2.2 CORRELATION ANALYSIS

- **Correlation between AT and GC:**

In codon usage bias analysis, correlation plays an important role to know the significant effect. Firstly, the Study of the relationship among the AT with GC values of the protein coding genes of Mitochondria, PTGS1, TAMM41. The +VE GC value demonstrates the depth of the base G over C nucleotide, whereas the -VE correlation means the depth of nucleotide C above G. similarly the trend is followed by the AT bases. After analysing the average AT and GC values across every gene we analyse that AT and GC shows negative correlation within themselves (-151 pearson correlation, with the significance of 0.698) which suggests the preference of T more than A, and C is more preferred than G.

- **Correlation between nucleotides at 1st and 3rd codon positions:**

Secondly it is performed between the all the DNA bases at first codon site and third codon position. From our study we observed that A shows positive correlation with base A3 with 0.041 pearson correlation value and at significant value of 0.917, which is greater than 0.01 and 0.05, which indicates its less significant appearance. T shows negative pearson correlation with T3 (-0.431). G shows its positive pearson correlation value (0.260) with G3. C also commits positive correlation value with C3 (0.289). whereas GC shows its positive correlation with the GC3 (0.829**). These values shows that the bases in some genes are significantly affected while others are not.

- **Correlation between codon usage bias and SCS and CAI:**

Scaled chi-square is one of the another parameter of codon bias, it gives path to CUB [5]. Whereas the comparable index among CUB and its reference set is measured by the codon adaptable value. They both are directional measures and both shows positive correlation with each other. The correlation analysis was performed between the different nucleotide skews, after evaluating the relationship between the gene expression level and codon utilisation bias, we found that CAI significantly -VE correlates with G, T3, G3 (-0.838**, -0.885**, and -0.702* respectively) whereas shoes a significant +VE relation with C, A3, and C3 (0.0675*, 0.814**, and 0.742*respectively). SCS shows a negative relation with T, G, AT, GC, T3, and G3, (-0.435, -0.264, -0.130, -0.018, -0.148, -0. 129 respectively) but the significance level wasn`t related to 0.001, 0.005.

This demonstrates that the level of protein expression is regulated by either reduced or raised nucleotide skews.

- **Correlation between length and bias in utilising codons:**

The Scaled-chi-square values are statistically +ve correlated among length of the genes pretended to be associated in Alzheimer`s disease (r=0.770*, p=0.15).

4.2.4 NEUTRALITY PLOT ANALYSIS

By plotting GC12 versus GC3 a neutrality plot was created. Plot demonstrates a significant association among GC12, GC, and as the regressive line's gradient gets close to one, it implies that the dominant evolutionary force is the influence of mutation pressure. If there is no discernible relationship between them, it seems that the process of selection by nature might have been involved. [2]. TAMM41, PTGS1, and 7 MT-ND genes' three GC codon locations' relationships should be shown. Using GC12 as the Y-axis and GC3 as the X-axis, a plot was created.

The analysis shows a significant relation within GC12 and GC3. Moreover, it shows the regression coefficient value of GC12 on GC3 as 0.3207, which indicates the mutation pressure of 32.07% and the magnitude of natural selection will 67.93%. this suggests the role of natural selection was higher than mutation pressure. As we know we are working with the mitochondrial genes where the mutations are less.

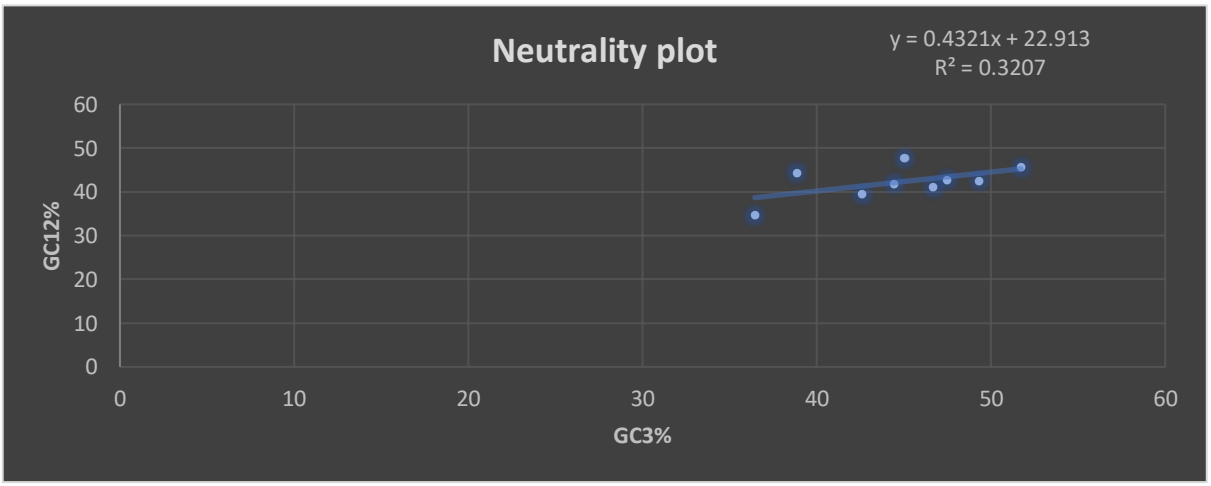


FIGURE 8: neutrality plot analysis of MT-ND family genes, TAMM41, PTGS1

- To know the effect of mutation pressure we will further perform the genomic analysis in where we will try to find the deleterious SNPs.

4.3 Prediction of deleterious missense nsSNPs (sequence-based tool)

- **MT-ND1**

- In order to distinguish harmful nsSNPs from neutral ones, many web servers were set up. First, the missense nsSNPs from the NCBI-SNP database were sent to the PANTHER software programme,[40] which can be used to determine whether a specific change in an amino acid has an effect on a protein's biological functions. By using a carefully curated database of the gene or protein family, this programme can detect and categorise the various gene products' roles. PANTHER identified 23 nsSNPs as potentially harmful out of 27 nsSNPs.[38] tools such as The HIDDEN MARKOV MODEL (HMM), which determines whether the nsSNP is related with a disease, is the basis for SNAP2, a web server.[40] PhD-SNP is a predictor of human harmful SNPs based on support vector machines (SVM). Out of 27, they projected 10, 14, or nsSNPs as harmful.[38] Other methods include Meta-SNP and SIFT, which respectively indicate 13,12 nsSNPs as harmful.

Table:6. Table of nsSNPs deemed to be damaging by 5-sequence-based tool in all the mitochondrial protein-coding genes MT-ND1 GENE.

VARIENT ID	MUTATION	PANTHER	SNAP2	PhD-SNP	Meta-SNP	SIFT
rs199476118	A52T	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs199476119	L285P	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs199476121	Y277C	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs199476122	G131S	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs199476123	E214K	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs199476124	Y215H	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs199476125	E143K	probably damage	EFFECT	DISEASED	DISEASED	DISEASED

rs397515508	A132T	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs397515612	E59K	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs1603218926	R25Q	probably damage	EFFECT	DISEASED	DISEASED	DISEASED

- **MT-ND2**

The five sequence-based techniques were utilised to analyse the harmful nsSNPs in the MT-ND2 gene. Web-based tools PANTHER, SNAP-2, PhD-SNP, Meta-SNP, and SIFT classify 10, 3, 3, 2, and 2 of the total 31 nsSNPs as harmful, respectively.

Table:7. Table of nsSNPs deemed to be damaging by 5-sequence-based tool in all the mitochondrial protein-coding genes of MT-ND2 GENE.

VARIENT ID	MUTATION	PANTHER	SNAP2	PhD-SNP	Meta-SNP	SIFT
rs199476115	G259S	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs267606889	L71P	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs387906426	I57M	probably damage	EFFECT	DISEASED	DISEASED	DISEASED

- **MT-ND3**

The five sequence-based techniques were utilised to analyse the harmful nsSNPs in the MT-ND3 gene. Web-based tools PANTHER, SNAP-2, PhD-SNP, Meta-SNP, and SIFT identify 4, 7, 5, 3, and 2 of the total 38 nsSNPs as harmful, respectively.

Table:8. Table of nsSNPs deemed to be damaging by 5-sequence-based tool in all the mitochondrial protein-coding genes of MT-ND3 GENE.

VARIENT ID	MUTATION	PANTHER	SNAP2	PhD-SNP	Meta-SNP	SIFT
rs587776438	D66N	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs587780529	Q26K	probably damage	EFFECT	DISEASED	DISEASED	DISEASED

- **MT-ND4**

To analyse the harmful nsSNPs in the MT-ND4 gene, five sequence-based techniques were employed. In all, 45 nsSNPs were analysed using the web-based programmes PANTHER, SNAP-2, PhD-SNP, Meta-SNP, and SIFT. Of them, 4, 4, 3, 3, 4 were shown to be harmful.

Table:9. Table of nsSNPs deemed to be damaging by 5-sequence-based tool in all the mitochondrial protein-coding genes of MT-ND4 GENE.

VARIENT ID	MUTATION	PANTHER	SNAP2	PhD-SNP	Meta-SNP	SIFT
rs28384199	R340G	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs28384199	R340S	probably damage	EFFECT	DISEASED	DISEASED	DISEASED

rs199476112	R340H	probably damage	EFFECT	DISEASED	DISEASED	DISEASED

- **MT-ND4L**

The five sequence-based techniques were utilised to analyse the harmful nsSNPs in the MT-ND4l gene. Web-based tools PANTHER, SNAP-2, PhD-SNP, Meta-SNP, and SIFT identify 4, 3, 3, 3, and 3 of the total 41 nsSNPs as harmful, respectively.

Table:10. table of nsSNPs deemed to be damaging by 5-sequence-based tool in all the mitochondrial protein-coding genes of MT-ND4L GENE.

VARIENT ID	MUTATION	PANTHER	SNAP2	PhD-SNP	Meta-SNP	SIFT
rs267606892	C32R	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs387906422	M1T	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs1603221804	G86E	probably damage	EFFECT	DISEASED	DISEASED	DISEASED

- **MT-ND5**

The five sequence-based techniques were utilised to analyse the harmful nsSNPs in the MT-ND5 gene. PANTHER, SNAP-2, PhD-SNP, Meta-SNP, and SIFT are web-based tools that classify 20, 13, 13, and 13 out of a total of 48 nsSNPs as harmful, respectively.

Table:11. table of nsSNPs deemed to be damaging by 5-sequence-based tool in all the mitochondrial protein-coding genes of MT-ND5 GENE.

VARIENT ID	MUTATION	PANTHER	SNAP2	PhD-SNP	Meta-SNP	SIFT
rs199476108	M237L	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs199974018	A236T	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs200145866	A171V	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs200855215	G465E	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs200873900	G239S	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs267606893	F124L	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs267606895	E145G	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs267606897	S250C	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs267606898	D393H	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs267606898	D393N	probably damage	EFFECT	DISEASED	DISEASED	DISEASED

- **MT-ND6**

The five sequence-based techniques were utilised to analyse the harmful nsSNPs in the MT-ND6 gene. Web-based tools PANTHER, SNAP-2, PhD-SNP, Meta-SNP, and SIFT classify 4, 4, 11, 9, and 7 of the total 23 nsSNPs as harmful, respectively.

Table:12. table of nsSNPs deemed to be damaging by 5-sequence-based tool in all the mitochondrial protein-coding genes MT-ND6 GENE.

VARIENT ID	MUTATION	PANTHER	SNAP2	PhD-SNP	Meta-SNP	SIFT
rs199476105	A72V	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs199476109	M63V	probably damage	EFFECT	DISEASED	DISEASED	DISEASED
rs869025186	Y59C	probably damage	EFFECT	DISEASED	DISEASED	DISEASED

Sequence-based search of nsSNPs in the MT-ND family genes found 253 total missense nsSNPs, of which 34 were shown to be deleterious nsSNPs. Now, utilising methods based on structure, these 34 nsSNPs were further examined to know their structural interferences.

4.4 stability change prediction in MT-ND family protein-coding genes

CUPSAT is a structure-based method that analyses the environment of the mutation site's amino acids using amino acid atom potential and torsion angle to anticipate how a point mutation would affect a protein's stability.[38] This distinguishes the environment of the amino acids depending on the suitability of each solvent and the distinctiveness of its secondary structure. The MUpro tool uses several machine learning programmes to forecast the effects of single AA changes on stability of protein.[38] A web-based programme called Align-GVGD uses the physicochemical properties of AA and protein MSAs to determine is there any mis-sense mutations is benign or harmful. Four, five, six, seven, and nine nsSNPs were identified as structurally harmful mutations by CUPSAT, MUpro, SNPs&GO, DynaMut, and Align-GVGD, respectively.[38] **Out of 34 nsSNPs tested using sequence-based methods, four (L285P, L71P, D393H, and Y59C) were demonstrated to be dangerous to the structure's stability. by utilising all five of the above-mentioned tools based on structural analysis.**

Table:13. List of nsSNPs deemed to be damaging by 5-structure based tool in all the mitochondrial protein-coding genes.

GENE NAME	VARIENT ID	MUTATION	CUPSAT	MUpro	DynaMut	SNPs&GO	Align-GVGD
MT- ND1	rs199476119	L285P	destabilizing	Decrease	destabilizing	Diseased	Most likely
MT-ND2	rs267606889	L71P	Destabilizing	Decrease	destabilizing	Diseased	Most likely
MT-ND5	rs267606898	D393H	Destabilizing	Decrease	destabilizing	Diseased	Most likely
MT-ND6	rs869025186	Y59C	destabilizing	decrease	destabilizing	diseased	Most likely

4.5 Conservation Analysis of the mitochondrial MT-ND family protein-coding genes.

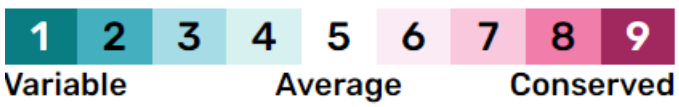
A broad estimation of the damage that detrimental mutations can make to the protein's structure as well as its function characteristics is provided by the conservation level of the residues. It is quite likely that a harmful mutation at a residue that is highly conserved would have negative effects in nature. The protein structure (5XTC with side chains s, i, l, and m sequentially for MT-ND1,2,5,6, respectively) was provided as the input for ConSurf, and the Bayesian approach was utilised for computation. The residues' conservation rankings might range from

1 to 9. A residue with a score of "1" indicates that it is highly changeable, whereas one with a value of "9" indicates that it is highly conserved. The result we obtained from our analysis is shown in table14.

Table:14. this shows the evolutionary conservation of the deleterious nsSNPs of MT-ND genes.

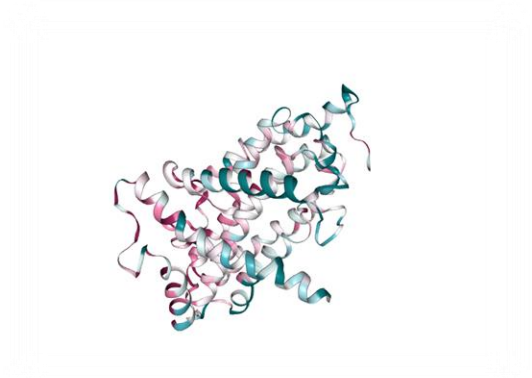
GENE NAME	PDB ID with SIDE CHAIN	MUTATION	SCORE	PREDICTION
MT-ND1	5XTC (chain s)	L285P	7	Conserved
MT-ND2	5XTC (chain i)	L71P	8	Conserved
MT-ND5	5XTC (chain l)	D393H	7	Conserved
MT-ND6	5XTC (chain m)	Y59C	9	Conserved

According to the ConSurf data, all four mutations (L285P, L71P, D393H, and Y59C) are conserved in nature, but the Y59C mutation in the MT-ND6 gene shows the highest conservation score value of 9. ConSurf tool results have been validated using multiple sequence alignment.

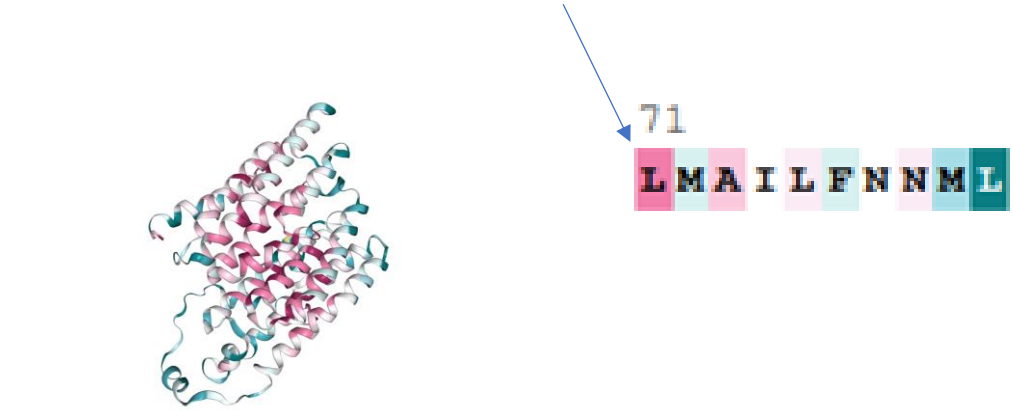


GENE STRUCTURE in ConSurf **MUTATION POSITION IN SEQUENCE**

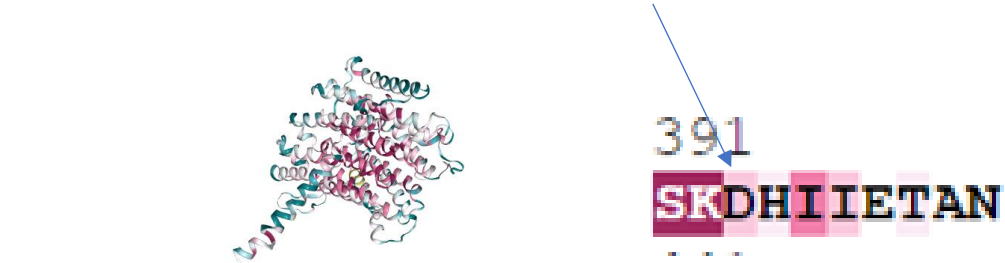
mutation: L285P



mutation: L71P



mutation: D393H



mutation: Y59C

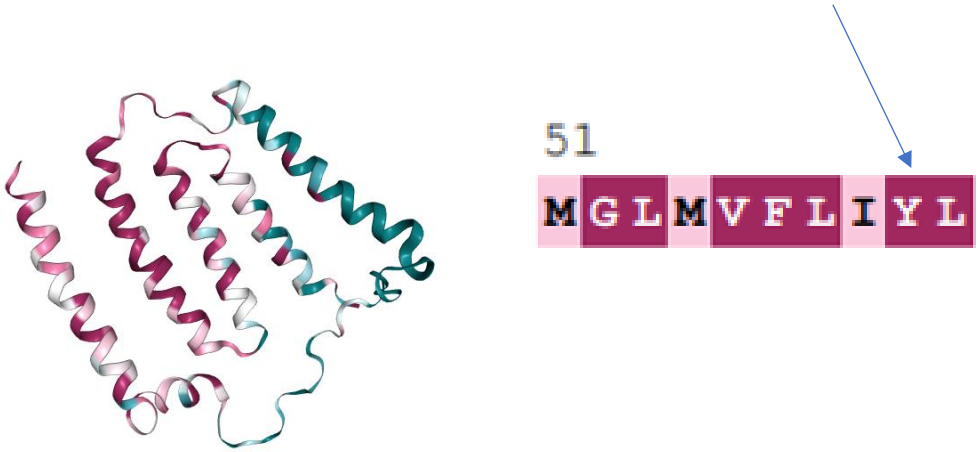


FIGURE:9. Conservation analysis of mutations listed to be deleterious structure-based tool.

4.6 SECONDARY STRUCTURE ANALYSIS

SOPMA tool is used to analyse the secondary structure of the genes (MT-ND1, MT-ND2, MT-ND5, MT-ND6) where the mutations take place (L285P, L71P, D393H, Y59C respectively).

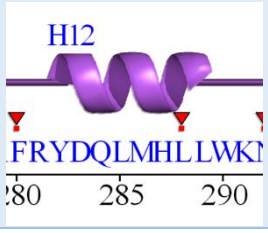
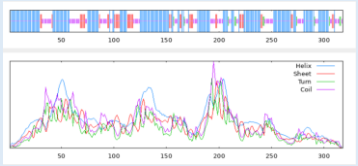
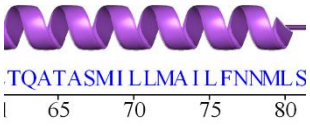
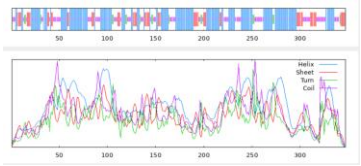
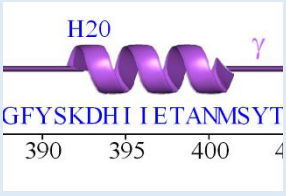
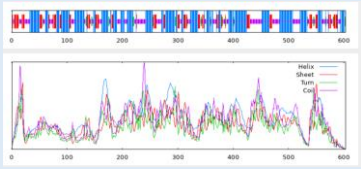
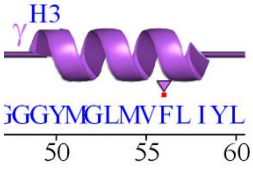
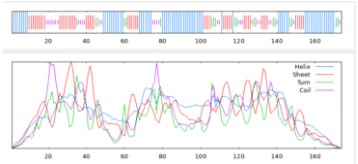
- **MT-ND1** protein-coding gene of length 318 AA (amino acid) is put as an input where 180 AA (56.0%) were found in the Alpha Helix., 42AA (13.21%) in the Extended Strand, 12AA (3.77%) in the Beta Turn and 84AA (26.42) in the Random Coil.

- **MT-ND2** protein-coding gene of length 347 AA (amino acid) is put as an input where 175 AA (50.43%) were found in the Alpha Helix., 59AA (17.00%) in the Extended Strand, 18 AA (5.19 %) in the Beta Turn and 95 AA (27.38%) in the Random Coil.

- **MT-ND5** protein-coding gene of length 603 AA (amino acid) is put as an input where 280 AA (46.43%) were found in the Alpha Helix., 113AA (18.74%) in the Extended Strand, 28 AA (4.64 %) in the Beta Turn and 182 AA (30.18%) in the Random Coil.

- **MT-ND6** protein-coding gene of length 174 AA (amino acid) is put as an input where 74 AA (42.53%) were found in the Alpha Helix., 51AA (29.31%) in the Extended Strand, 19 AA (10.92%) in the Beta Turn and 30 AA (17.24%) in the Random Coil.

Table:15. Secondary structures prediction of MT-ND genes.

GENE NAME Protein length	MUTATION	Secondary Structure	PREDICTED TO BE PRESENT IN	SOPMA ANALYSIS
MT-ND1 (318AA) chain s	L885P		H12-HELICES	
MT-ND2 (347AA) chain i	L71P		H3-HELICES	
MT-ND5 (603AA) chain l	D393H		H20-HELICES	
MT-ND6 (174AA) chain m	Y59C		BETWEEN H3-H4 HELICES (LOOPS)	

4.6.1. STRUCTURAL EFFECTS

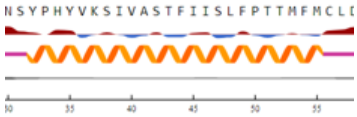
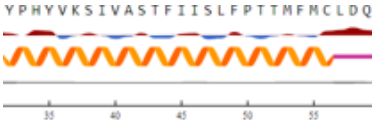
Using NetSurf.2.0 tool we get to know about the structural effects, that occur after mutation in these protein-coding genes. As the result of this analysis, it is seen that 3 (L285P, L71P, D393H) out of 4 shows the changes in their structure. In L285P mutation of MT-ND1gene where there are 10 helixes, the 2nd helix which starts from residue L and ends at residue T in wild type sequence, does not end with same affinity in the mutated one, whereas in the mutated one, this helix ends one amino acid prior to that

MT-ND1 GENE without MUTATION **MT-ND1 GENE with MUTATION**

of original one. i.e., now it starts from same residue L but ends at residue F. it converts the residue at position 37 from helix to coil. This will affect the binding affinity at this binding site.

MT-ND2 GENE without MUTATION **MT-ND2 GENE with MUTATION**

In L71P mutation of MT-ND2 GENE, after mutation it is seen that the 1st helix which starts from residue P and ends at residue L, but after mutation residue at position 23 converted from coil to helixes.

MT-ND5 GENE without MUTATION		MT-ND5 GENE with MUTATION
	**D393H**	

In D393H mutation of MT-ND5 gene, the residue at position 56 converted from coil to helix. **These mutations might be damaging due to the change induced with respect to the helix. As this affects its binding affinity at these positions.**

4.7 Analysis of Phenotypic Consequences of deleterious nsSNPs

Based on the mutation score, FATHMM provides a prediction. There has been little to no change in the underlying amino acid if the score is close to zero.

Table:16. Using the FATHMM tool, phenotypic repercussions of alterations are predicted.

VARIENT ID	GENE NAME	MUTATION	PREDICTION	SCORE	DO INFORMATION	HPO Information
rs1999476119 rs267606889 rs267606898	MT-ND1 MT-ND2 MT-ND5	L285P L71P D393H	DAMAGING DAMAGING DAMAGING	-6.68 -8.21 -4.58	Anatomical entity disease Neurological Disorders Disease of the Brain illness.	Phenotypic abnormality

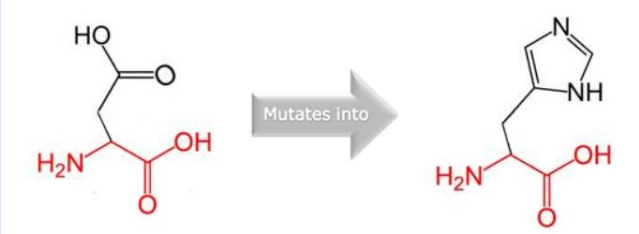
4.8 Structural Effects of deleterious nsSNPs

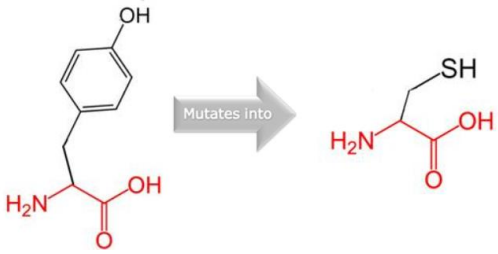
Input for HOPE server consists of mutation and the sequence. HOPE starts to gather data from a variety of data sources. The result highlights the affects of the submitted mutation in the protein's 3D structure with the proper description of amino acid properties and domains as shown in

Table 17: HOPE server data on structural impact on mutational genes

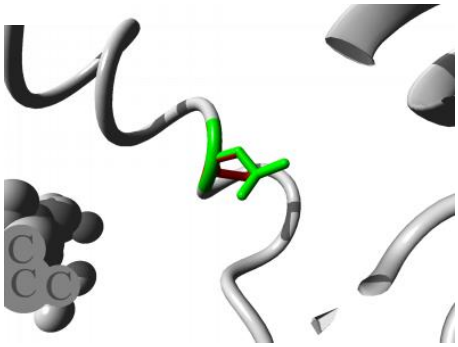
Mutation	*Structure alteration*	*Amino-Acid Characteristics*	*Domain*
L285P		More compact than the wild-type residue is the mutant residue. The size of the mutated one and original AA does not match. More compact than the wild-type residue is the mutant residue.	core of a domain contains the residue. The essential structural integrity of this domain may be disturbed by the variations between the wild-type and mutant residue.

L71P		Greater in size than the wild-type residue is the mutant residue. The size of the mutated one and original AA does not match. Protein's centre concealed the wild-type residue. Because it is larger, the mutant residue probably won't fit.	The core of a domain contains the residue. The fundamental structure of this domain might be disturbed by the variations between the wild-type and mutant residue.	

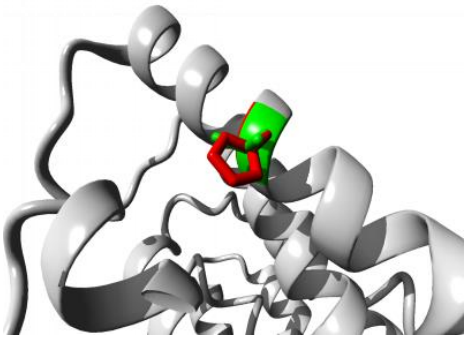
D393H		The mutant residue is Larger than the original residue. The original and mutated AA are different in size. The mutated residue is far larger than the wild-type residue.	The mutated residue is in a domain that is important for the activity of the protein and in contact with residues in another domain. It is possible that this interaction is important for the correct function of the protein.	

Y59C		mutant residue is smaller than the wild-type residue. The original and mutated amino acids differ in size. The mutated residue is smaller than the original residue.	The mutated residue is in a domain that is important for the activity of the protein and in contact with residues in another domain. It is possible that this interaction is important for the correct function of the protein.

L285P



L71P



D393H

Y59

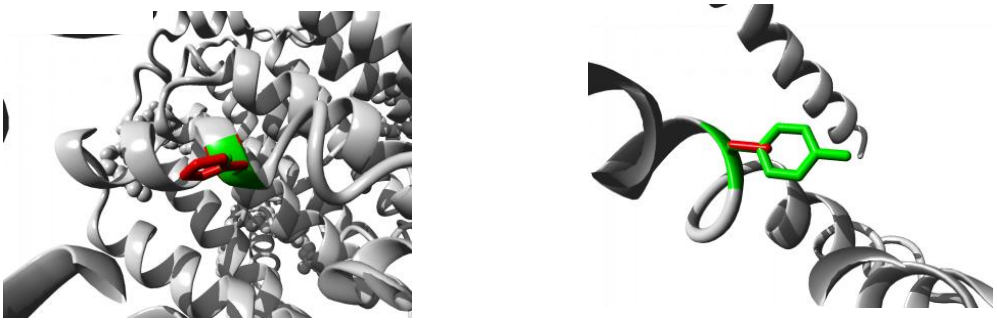


FIGURE:10. Visualization using HOPE server of both mutated amino acid residues (red) and wild type amino acid residue (green)

DISCUSSION

Thanks to mitochondrial genomes' abundantly available code for proteins sequences of data, we were able to examine the patterns of codon usage. Numerous studies have focused on the codon use and evolutionary factors of mitochondrial genomes. [61,62] However, as far as our search is made no research on the mitochondrial ND genes' codon usage pattern has been published. A deeper understanding of the molecular evolutionary viewpoint would result from research on the DNA count and codon utilization pattern of mitochondrial ND genes.

Base composition is a crucial component of a genome and is thought to be the primary factor influencing codon use. The third position is typically preferred by GC-rich organisms [63,64]. As opposed to this, AT-rich organisms favour A or T in the third position [65]. The total CUB was bi-ased towards nucleotides A, T in the all MT-ND genes, PTGS1, TAMM41, where AT was abundant. Our results are satisfactory because the MT-ND1 gene has a strong predilection for AT, as described in a publication [60], and the Alzheimer's brain is AT-rich, as documented in numerous research. It is found that AT% is 55.26% which means GC content is 44.73%. Similar patterns were seen in human mitochondrial DNA across continents, with A and T nucleotides being more preferred [66].

Only if the frequencies of A and T at the third codon location are equivalent to those of G and C, according to Zhang et al. (2013), is it believed that mutational pressure is the only evolutionary force that has shaped the synonymous codon usage bias. Our findings showed an AT bias for the genes, pointing to the possibility that other mechanisms, such as natural selection, may be involved in base composition. The total number of nucleotides and their associated synonymous locations were therefore correlated, and both substantial positive and negative correlations were discovered. While there was a negative association between T and T3, there was a positive correlation between A and A3, G and G3, and Cand C3.

Between GC and GC3, a sizable positive connection was observed. All these findings suggested that compositional restrictions may have been a major factor in determining the pattern in CUB. Contrarily, DNA composition at the third codon site and total nucleotide composition in human mitochondrial protein coding genes also exhibited a positive association, demonstrating that both modes of selections had a role in the synonym’s codon utilization [66].

Calculating ENC values allows one to gauge the extent of codon use bias. Therefore, a strong CUB caused by smaller ENC value was suggested by the analysis of ENC value in all the MT-

ND, TAMM41, and PTGS1 genes. ENC's value should be lower than 40. Comparatively, the CAI value should fall between 0 and 1. The average CAI score in our study was 0.685, indicating greater gene expression. Gene expression is inversely correlated with the ENC, whereas CAI is directly correlated with gene expression. Here, we concur that the gene expression is higher based on our analysis of ENC and CAI.

Whereas, points were distributed with a small range according to neutral plot analysis, and the slant of the regress was additionally quite near to zero, indicating that it is possible that the codon use bias of the MT-ND, TAMM41, PTGS1 gene was caused by both Type of selection. This result was in line with the article's results that the codon use patterns of 13 human protein-coding genes were influenced equally by mutation pressure and selection. [66].

As MT-ND genes exhibit the effects of mutation pressure, our next task was to identify SNPs in these genes at the structural and sequence levels. where we determined 4 nsSNPs to be harmful. The nsSNPs (L285P), (L71P), (D393H), and (Y59C) were identified as harmful, and we further analysed these nsSNPs to determine their conservation, oncogenic nature, and structural effects. where the harmful nsSNPs (L285P, L71P, and D393H) were found in the FANTHMM tool and are linked to brain and central nervous system diseases.

CONCLUSION

In the current investigation, considerable CUB was found when synonymous codon use in the genes MT ND, PTGS1, and TAMM41 was examined. The third codon position was preferred by either A or T in all of the genes' most frequent codons, highlighting the crucial importance of compositional constraint. Additionally, neutrality plot analysis revealed that natural selection and very little mutation pressure both played a significant influence in the gene's evolution. The new investigation has increased our understanding of codon usage strategies and the development of the poorly described genes. The MT-ND gene's nsSNPs have been examined in the current study since it is linked to a number of complex disorders. Four extremely harmful SNPs (L285P, L71P, D393H, and Y59C) have been found among the 34 nsSNPs of ND Genes. Based on a variety of analyses, it was determined that the L285P, L71P, and D393H were the most dangerous. Future research projects that aim to battle the diseases by studying MT-ND1, MT-ND2, and MT-ND5 are anticipated to benefit from the findings of the current study. This study is expected to serve as a model for other nsSNP analyses of a similar nature soon.

REFERENCES

[1] Agrawal I, Jha S. Mitochondrial dysfunction and Alzheimer's disease: Role of microglia. Frontiers in aging neuroscience. 2020:252.

[2] Zhang X, Farrell JJ, Tong T, Hu J, Zhu C, Alzheimer's Disease Sequencing Project, Wang LS, Mayeux R, Haines JL, Pericak-Vance MA, Schellenberg GD. Association of mitochondrial variants and haplogroups identified by whole exome sequencing with Alzheimer's disease. Alzheimer's & dementia. 2022 Feb;18(2):294-306.

[3] Onyango I, Khan S, Miller B, Swerdlow R, Trimmer P, Bennett Jr J. Mitochondrial genomic contribution to mitochondrial dysfunction in Alzheimer's disease. Journal of Alzheimer's Disease. 2006 Jan 1;9(2):183-93.

[4] Nissanka N, Moraes CT. Mitochondrial DNA damage and reactive oxygen species in neurodegenerative disease. FEBS letters. 2018 Mar;592(5):728-42.

[5] Arun S, Liu L, Donmez G. Mitochondrial biology and neurological diseases. Current neuropharmacology. 2016 Feb 1;14(2):143-54.

[6] Shukla R, Kumar A, Kelvin DJ, Singh TR. Disruption of DYRK1A-induced hyperphosphorylation of amyloid-beta and tau protein in Alzheimer's disease: An integrative molecular modeling approach. Frontiers in Molecular Biosciences. 2022;9.

[7] Murthy TK, Shukla R, Durga Prasad N, Swetha P, Shreyas S, Singh TR, Pattabiraman R, Nair SS, Mathew BB, Kumar KM. Comprehensive analysis of non-synonymous missense SNPs of human galactose mutarotase (GALM) gene: an integrated computational approach. Journal of Biomolecular Structure and Dynamics. 2022 Dec 20:1-5.

[8] Saxena S, Murthy TK, Chandramohan V, Yadav AK, Singh TR. Structural and functional analysis of disease-associated mutations in GOT1 gene: An in-silico study. Computers in Biology and Medicine. 2021 Sep 1; 136:104695.

[9] X. Sun, Q. Yang, X. Xia, An improved implementation of effective Number of Codons (N c), Mol. Biol. Evol. 30 (2012) 191–196.

[10] Khandia R, Saeed M, Alharbi AM, Ashraf GM, Greig NH, Kamal MA. Codon Usage Bias Correlates With Gene Length in Neurodegeneration Associated Genes. Frontiers in Neuroscience. 2022;16.

[11] N. Sueoka, Directional mutation pressure and neutral molecular evolution, Proc. Natl. Acad. Sci. 85 (1988) 2653–2657.

[12] N. Sueoka, Two aspects of DNA base composition: g+ C content and translationcoupled deviation from intra-strand rule of A= T and G= C, J. Mol. Evol. 49 (1999) 49–62

[13] Mi H, Ebert D, Muruganujan A, Mills C, Albou LP, Mushayamaha T, Thomas PD. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic acids research. 2021 Jan 8;49(D1):D394-403.

[14] Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic acids research. 2017 Jan 4;45(D1):D183-9.

[15] Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nature protocols. 2013 Aug;8(8):1551-66.

[16] AbdulAzeez S, Borgio JF. In-silico computing of the most deleterious nsSNPs in HBA1 gene. PloS one. 2016 Jan 29;11(1):e0147702.

[16] Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants. BMC genomics. 2015 Dec;16(8):1-2.

[18] Wong DJ, Harris S, Sahni A, Bedford JR, Cortes L, Shawyer R, Wilson AM, Lindsay HA, Campbell D, Popham S, Barneto LM. Developing and validating subjective and objective risk-assessment measures for predicting mortality after major surgery: an international prospective cohort study. PLoS Medicine. 2020 Oct 15;17(10):e1003253.

[19] Capriotti E, Fariselli P. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. Nucleic acids research. 2017 Jul 3;45(W1):W247-52.

[20] Mah JT, Low ES, Lee E. In silico SNP analysis and bioinformatics tools: a review of the state of the art to aid drug discovery. Drug discovery today. 2011 Sep 1;16(17-18):800-9.

[21] Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS computational biology. 2014 Jan 16;10(1):e1003440.

[22] Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. BMC genomics. 2013 May;14:1-9.

[23] Leong IU, Stuckey A, Lai D, Skinner JR, Love DR. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. BMC medical genetics. 2015 Dec;16:1-3.

[24] Cubuk C, Garrett A, Choi S, King L, Loveday C, Torr B, Burghel GJ, Durkie M, Callaway A, Robinson R, Drummond J. Clinical likelihood ratios and balanced accuracy for 44 in silico tools against multiple large-scale functional assays of cancer susceptibility genes. Genetics in Medicine. 2021 Nov;23(11):2096-104.

[25] Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic acids research. 2012 Jul 1;40(W1):W452-7.

[26] Yurcik W. Visualizing NetFlows for Security at Line Speed: The SIFT Tool Suite. InLISA 2005 Dec 4 (pp. 169-176).

[27] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. 2009 Jul;4(7):1073-81.

[28] Parthiban V, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. Nucleic acids research. 2006 Jul 1;34(suppl_2):W239-42.

[29] Azeem D, Choudhary HB, Choudhary KB, Mateen RM, Hussain M. Pathogenicity Prediction of Potential Variants of Iba57 Gene Causing Heredity Spastic Paraplegia Through Mitochondrial Dysfunction by using Computational Tools. THE JOURNAL OF MICROBIOLOGY AND MOLECULAR GENETICS. 2021 Dec 31;2(3):39-54.

[30] Firtina S, Kutlu A, Yozlu M, Cepeci BN, Isikgil B, Ar MC, Sayitoglu M. Integrative Analysis of the TNFRSF13B Variants in Common Variable Immune Deficiency.

[31] Ali A, Rehman MU, Ahmad SM, Mehraj T, Hussain I, Nadeem A, Mir MU, Ganie SA. In Silico Tools for Analysis of Single-Nucleotide Polymorphisms in the Bovine Transferrin Gene. Animals. 2022 Mar 10;12(6):693.

[32] Ghosh P, Ghosh S, Behera B, Behera JK, Bhattacharya M, Kar NB. In-silico analysis of non-synonymous SNPs of human LDLR gene and their impact on familial hypercholesterolemia. Gene Reports. 2021 Jun 1;23:101127.

[33] Saih A, Baba H, Bouqdayr M, Ghazal H, Hamdi S, Kettani A, Wakrim L. In silico analysis of high-risk missense variants in human ACE2 gene and susceptibility to SARS-CoV-2 infection. BioMed research international. 2021 Apr 9;2021.

[34] Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC genomics. 2013 May;14:1-7.

[35] Manfredi M, Savojardo C, Martelli PL, Casadio R. E-SNPs&GO: embedding of protein sequence and function improves the annotation of human pathogenic variants. Bioinformatics. 2022 Dec 1;38(23):5168-74.

[36] Frousios K, Iliopoulos CS, Schlitt T, Simpson MA. Predicting the functional consequences of non-synonymous DNA sequence variants—evaluation of bioinformatics tools and development of a consensus strategy. Genomics. 2013 Oct 1;102(4):223-8.

[37] Fortuno C, James PA, Young EL, Feng B, Olivier M, Pesaran T, Tavtigian SV, Spurdle AB. Improved, ACMG-compliant, in silico prediction of pathogenicity for missense substitutions encoded by TP53 variants. Human mutation. 2018 Aug;39(8):1061-9.

[38] Ernst C, Hahnen E, Engel C, Nothnagel M, Weber J, Schmutzler RK, Hauke J. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. BMC medical genomics. 2018 Dec;11:1-0.

[39] Thompson BA, Greenblatt MS, Vallee MP, Herkert JC, Tessereau C, Young EL, Adzhubey IA, Li B, Bell R, Feng B, Mooney SD. Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. Human mutation. 2013 Jan;34(1):255-65.

[40] Rodrigues CH, Pires DE, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic acids research. 2018 Jul 2;46(W1):W350-5.

[41] Weffenstette D, Walcott KR. DYNAMUT: AMutation TESTING TOOL FOR INDUSTRY-LEVEL EMBEDDED SYSTEM APPLICATIONS. International Journal of Embedded Systems and Applications. 2018;8(1-3):19-34.

[42] Joshi A, Paul S. Phylogenetic analysis of the novel coronavirus reveals important variants in Indian strains. BioRxiv. 2020 Apr 15:2020-04.

[43] Saxena S, Murthy TK, Chandramohan V, Yadav AK, Singh TR. Structural and functional analysis of disease-associated mutations in GOT1 gene: An in silico study. Computers in Biology and Medicine. 2021 Sep 1;136:104695.

[44] Venkata Subbiah H, Ramesh Babu P, Subbiah U. Determination of deleterious single-nucleotide polymorphisms of human LYZ C gene: an in silico study. Journal of Genetic Engineering and Biotechnology. 2022 Jul 1;20(1):92.

[46] Armon A, Graur D, Ben-Tal N. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. Journal of molecular biology. 2001 Mar 16;307(1):447-63.

[47] Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics. 2003 Jan 1;19(1):163-4.

[48] Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic acids research. 2016 Jul 8;44(W1):W344-50.

[49] Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics. 1995 Dec 1;11(6):681-4.

[50] Hundertmark M, Popova AV, Rausch S, Seckler R, Hincha DK. Influence of drying on the secondary structure of intrinsically disordered and globular proteins. Biochemical and biophysical research communications. 2012 Jan 6;417(1):122-8.

[51] Puglisi I, Lo Cicero L, Lo Piero AR. The glutathione S-transferase gene superfamily: an in-silico approach to study the post translational regulation. Biodegradation. 2013 Jul; 24:471-85.

[52] Harris RA, Popov KM, Zhao Y, Kedishvili NY, Shimomura Y, Crabb DW. A new family of protein kinases—the mitochondrial protein kinases. Advances in enzyme regulation. 1995 Jan 1;35:147-62.

[53] McFadden CS, France SC, Sánchez JA, Alderslade P. A molecular phylogenetic analysis of the Octocorallia (Cnidaria: Anthozoa) based on mitochondrial protein-coding sequences. Molecular phylogenetics and evolution. 2006 Dec 1;41(3):513-27.

[54] Yu M. Computational analysis on two putative mitochondrial protein-coding genes from the Emydura subglobosa genome: A functional annotation approach. Plos one. 2022 Aug 18;17(8):e0268031.

[55] Temperley RJ, Wydro M, Lightowlers RN, Chrzanowska-Lightowlers ZM. Human mitochondrial mRNAs—like members of all families, similar but different. Biochimica et Biophysica Acta (BBA)-Bioenergetics. 2010 Jun 1;1797(6-7):1081-5.

[56] Karthikeyan V, Vickram AS, Manian R. Estimation of varicocele associated human ARG2 and NOS1 proteins and computational analysis on the effect of its nsSNPs. International Journal of Biological Macromolecules. 2020 Dec 1;164:735-47.

[57] Alipoor B, Ghaedi H, Omrani MD, Bastami M, Meshkani R, Golmohammadi T. A Bioinformatics Approach to Prioritize Single Nucleotide Polymorphisms in TLRs Signaling Pathway Genes. International journal of molecular and cellular medicine. 2016;5(2):65.

[58] Jalbani AA, Abbasi S, Menghwar GD, Yasmin A. Towards an Approach for Web Surfing in Unison. In2011 Developments in E-systems Engineering 2011 Dec 6 (pp. 267-270). IEEE.

[59] Pouladi N, Abdolahi S, Farajzadeh D, Hosseinpour Feizi MA. Haplotype and linkage disequilibrium of TP53-WRAP53 locus in Iranian-Azeri women with breast cancer. PLoS One. 2019 Aug 6;14(8):e0220727.

[60] Umair M, Palander O, Bilal M, Almuzzaini B, Alam Q, Ahmad F, Younus M, Khan A, Waqas A, Rafeeq MM, Alfadhel M. Biallelic variant in DACH1, encoding Dachshund Homolog 1, defines a novel candidate locus for recessive postaxial polydactyly type A. Genomics. 2021 Jul 1;113(4):2495-502.

[61] Almatrafi AM, Hibshi AM, Basit S. Association of Homozygous PROP1 Mutation in a Saudi Family with Combined Pituitary Hormone Deficiency. Medicina. 2023 Feb 27;59(3):474.

[62] Anandarajah G, Hight E. Spirituality and medical practice: using the HOPE questions as a practical tool for spiritual assessment. American family physician. 2001 Jan 1;63(1):81.

[63] Bressan RA, Iacoponi E, de Assis JC, Shergill SS. Hope is a therapeutic tool. Bmj. 2017 Dec 13;359.

[61] I. Irisarri, D. San Mauro, F. Abascal, A. Ohler, M. Vences, R. Zardoya, The origin of modern frogs (Neobatrachia) was accompanied by acceleration in mitochondrial and nuclear substitution rates, BMC Genomics 13 (2012) 626.

[62] Y. Xia, Y. Zheng, I. Miura, P.B. Wong, R.W. Murphy, X. Zeng, The evolution of mitochondrial genomes in modern frogs (Neobatrachia): nonadaptive evolution of mitochondrial genome reorganization, BMC Genomics 15 (2014) 691.

[63] R. Hershberg, D.A. Petrov, General rules for optimal codon choice, PLoS Genet. 5 (2009) e1000556.

[64] A. Kawabe, N.T. Miyashita, Patterns of codon usage bias in three dicot and four monocot plant species, Genes Genet. Syst. 78 (2003) 343–352.

[65] J.G. Waterkeyn, C. Gauci, A.F. Cowman, M.W. Lightowlers, Codon usage inTaeniaSpecies, Exp. Parasitol. 88 (1998) 76–78.

[66] S. Chakraborty, A. Uddin, T.H. Mazumder, M.N. Choudhury, A.K. Malakar, P. Paul, B. Halder, H. Deka, G.A. Mazumder, R.A. Barbhuiya, Codon usage and expression level of human mitochondrial 13 protein coding genes across six continents, Mitochondrion 42 (2018) 64–76.