Comparative and Functional Genomics Studies for Gut Microbiota and its Association with Alzheimer's Disease.

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

MASTER OF

SCIENCE

IN

MICROBIOLOGY

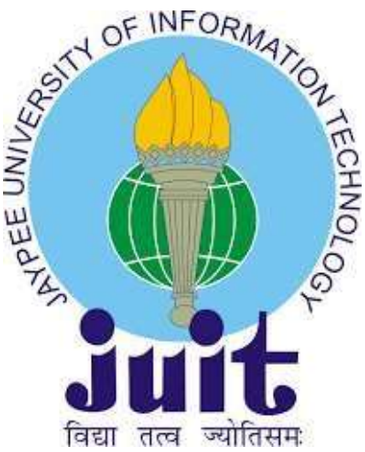
By

Kushum

225112002

Under the guidance of

Prof. Tiratha Raj Singh



Jaypee University of Information Technology

Department of biotechnology and Bioinformatics

Waknaghat, Solan, 173234, H.P., INDIA

TABLE OF CONTENTS

DECLARATION

I hereby declare that work reported in the M.Sc. thesis entitled "**Comparative and functional genomics studies for gut microbiota and its association with Alzheimer's disease"** submitted at Jaypee University of Information Technology, Solan, India, is an a uthentic record of my work carried out under the supervision of Dr. Tiratha Raj Singh. I have not submitted this work elsewhere for any other degree or diploma.I am fully responsible for the contents of my M.Sc. thesis.

kushum Date:

Enrollment Number-225112002

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology Solan, India

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the M.Sc. thesis entitled **"Comparative and functional genomics studies for gut microbiota and its association with Alzheimer's disease"** submitted by Ms. Kushum (225112002) at Jaypee University of Information Technology, Solan, India, is a bonafide record of her original work carried out from July 2023 to May 2024 under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

Supervisor:

Dr.Tiratha Raj Singh Date -

Professor

Department of Biotechnology &Bioinformatics

Jaypee University of Information Technology

Solan, India

ACKNOWLEDGMENT

This project report has been kept on track and been seen through to completion with the support and encouragement of numerous people.

First and foremost, I would like to express my utmost gratitude and appreciation to my project supervisor, **Dr. Tiratha Raj Singh** for his guidance, supervision, and assistance throughout my project work. His expertise and ever-ready guidance contributed a major part in making this project a success.

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 also would like to thank all the lab engineers and specially **Mrs. Somlata Sharma** ma'am for providing me with a workplace and for always motivating me.

My sincere thanks also goes to **Dr. Sudhir Syal**, Head of the Department, Department of Biotechnology and Bioinformatics, JUIT, Solan, for allowing me to work on such a diverse project.

Last but not least I would like to thanks my **Parents** whose unwavering love and support provided a constant source of strength throughout his journey.

Kushum (225112002)

M.Sc. Microbiology II year (4^{th} semester)
Jaypee University of Information
Technology Solan, HP

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBERATIONS

AD - Alzheimer's disease
SCFAs – short-chain-fatty acids
IBD- Inflammatory bowel disease
FMT – fecal microbiota transplantation
Aβ – amyloid beta peptide
LPS – lipopolysaccharides
MSA- multiple sequence alignment
CSF- cerebrospinal fluid
COUSIN - codon usage similarity index
AUCA- Automated codon usage analysis
STRING- search tool for the retrieval of interacting genes/proteins
MRI- magnetic resonance imagining
NMDA- N-methyl cholinesterase inhibitors
CUB- codon usage bias

ABSTRACT

Alzheimer's disease (AD) presents a major challenge to global healthcare systems, as its causes and progression remain not fully understood. Recent research has increasingly implicated the gut microbiota in the pathogenesis of AD, suggesting a potential avenue for therapeutic interventions. This thesis explores the comparative and functional aspects of gut microbiota composition in individuals with Alzheimer's disease (AD), aiming to identify microbial signatures associated with the condition.

Using next-generation sequencing techniques, we characterized the gut microbiota composition in a cohort of individuals with AD, focusing on 20 specific microorganisms identified through National Center for Biotechnology Information (NCBI) databases. By constructing phylogenetic trees based on these microbial sequences, we gained insights into the evolutionary relationships among gut microorganisms in the context of AD.

Furthermore, we employed codon bias analysis to investigate potential functional implications of microbial communities in AD pathology. By examining the preferential usage of synonymous codons within microbial genomes, we elucidated potential mechanisms by which gut microbiota may contribute to AD development and progression.

Our findings reveal distinct microbial signatures associated with AD, providing novel insights into the complex interplay between gut microbiota and neurodegenerative diseases. Understanding the comparative and functional aspects of gut microbiota in AD opens up new avenues for therapeutic interventions and diagnostic strategies aimed at modulating the gut-brain axis to mitigate AD progression.

CHAPTER 1

INTRODUCTION

1.1 Background

The gut microbiota plays a crucial role in maintaining human health by performing a wide array of essential functions, including vitamin production and the regulation of gene expression. This intricate microbial community, primarily residing in the colon, comprises a diverse array of microorganisms, estimated to range from 400 to 1500 species. These microorganisms interact in a symbiotic relationship with the host, contributing to various physiological processes crucial for overall well-being. Central to the functionality of the gut microbiota is its ability to efficiently utilize available substrates for growth and metabolism, thereby ensuring its survival and proliferation within the dynamic environment of the gastrointestinal tract.

Within the complex ecosystem of the gut microbiota, several dominant bacterial phyla, including Bacteroidetes, Firmicutes, Actinobacteria, and Proteobacteria, prevail. These phyla encompass a multitude of genera, each occupying specific ecological niches and exhibiting diverse metabolic capabilities. For instance, Bacteroides and Prevotella, prominent genera within the Bacteroidetes phylum, are adept at metabolizing mucins, the glycoproteins present in mucus, while members of the Firmicutes phylum, such as Faecalibacterium and Roseburia, excel in fermenting dietary fibers into short-chain fatty acids.

The metabolic versatility of gut microbes is underscored by their ability to utilize a wide range of substrates, including host-derived excreta, dietary components, and complex carbohydrates. This metabolic flexibility enables the gut microbiota to thrive in the dynamic milieu of the gastrointestinal tract, where nutrient availability fluctuates rapidly. Moreover, the redundancy in metabolic capacities across different microbial taxa confers resilience to the gut microbiota, allowing it to adapt to varying dietary compositions and environmental perturbations [1].

The concept of enterotypes, proposed in 2011, delineates distinct microbial community compositions based on the prevalence of key microbial taxa. These enterotypes, characterized by variations in microbial diversity and metabolic functions, have been associated with long-term dietary patterns, with Bacteroides-dominated enterotypes correlating with high-protein and animal fat diets, while Prevotella-dominated enterotypes are linked to carbohydrate-rich diets. This underscores the profound influence of diet on shaping gut microbiota composition and functionality.

Dysbiosis, characterized by perturbations in gut microbiota composition and function, has been implicated in the pathogenesis of various diseases, spanning inflammatory bowel diseases (IBD), autoimmune conditions, and metabolic disorders. Understanding the intricate relationship between gut microbiota dysbiosis and disease development offers novel insights into potential therapeutic strategies aimed at restoring microbial homeostasis to improve host health. In the subsequent sections, we delve into the associations between dysbiosis and specific diseases, elucidating the role of the gut microbiota in disease pathogenesis and highlighting avenues for therapeutic intervention and management.[2]

1.2 Pathogenesis

The gut microbiota, a complex ecosystem of microorganisms residing in the gastrointestinal tract, plays a fundamental role in maintaining host health through its diverse interactions with the immune system, metabolism, and overall physiology. Dysbiosis, defined as disruptions in the composition and function of the gut microbiota, has been linked to the pathogenesis of numerous diseases, including gastrointestinal disorders, metabolic conditions, autoimmune diseases, and neurological disorders. This thesis aims to elucidate the pathogenesis of gut microbiota dysbiosis, investigating the intricate mechanisms behind changes in microbial community structure and function and their implications for human health and disease. [3].

1.2.1 Factors Influencing Gut Microbiota Composition

The composition of the gut microbiota is influenced by a myriad of factors, including host genetics, diet, lifestyle, medication use, environmental exposures, and early-life events. Host genetics shape the initial colonization of the gut microbiota, influencing the abundance and diversity of microbial taxa. Diet serves as a key determinant of gut microbiota composition, with dietary components serving as substrates for microbial metabolism and shaping microbial community structure. Lifestyle factors, such as exercise, stress, and sleep patterns, also impact gut microbiota composition, highlighting the dynamic nature of the gut microbial ecosystem [4].

1.2.2 Dynamic Interactions Within the Gut Microbiota

The gut microbiota consists of a diverse array of microorganisms, including bacteria, viruses, fungi, archaea, and other microorganisms, which interact within a complex network of metabolic and signaling pathways. Microbial metabolites, such as short-chain fatty acids (SCFAs), neurotransmitters, and secondary bile acids, mediate host-microbe interactions by modulating immune function, metabolism, and neurological processes. Dysbiosis disrupts these dynamic interactions, resulting in abnormal immune activation, metabolic dysfunction, and neurological disturbances. [5]

1.2.3 Mechanisms of Gut Microbiota Dysbiosis

Multiple mechanisms contribute to gut microbiota dysbiosis, including alterations in microbial community structure, dysregulated host-microbe interactions, and environmental perturbations. Antibiotic use, dietary changes, and stress can disrupt gut microbiota composition, leading to shifts in microbial diversity and abundance. Dysfunctional host immune responses, impaired epithelial barrier function, and alterations in mucosal immune signaling pathways also contribute to gut microbiota dysbiosis, perpetuating a cycle of inflammation and microbial dysregulation[6].

1.2.4 Consequences of Gut Microbiota Dysbiosis

Gut microbiota dysbiosis has far-reaching consequences for host health, contributing to the pathogenesis of various diseases. Inflammatory bowel diseases (IBD), including Crohn's disease and ulcerative colitis, are characterized by dysregulated immune responses to gut microbiota dysbiosis, leading to chronic intestinal inflammation and tissue damage. Metabolic diseases, such as obesity and type 2 diabetes, are associated with alterations in gut microbiota composition and function, influencing

energy metabolism, adiposity, and insulin sensitivity. Furthermore, dysbiosis has been implicated in the pathogenesis of autoimmune conditions, neurological disorders, and even cancer, highlighting the broad impact of gut microbiota dysbiosis on human health.

1.2.5 Therapeutic Strategies for Gut Microbiota Dysbiosis

Understanding the mechanisms underlying gut microbiota dysbiosis offers promising opportunities for therapeutic intervention. Strategies aimed at modulating gut microbiota composition, such as probiotics, prebiotics, dietary interventions, and fecal microbiota transplantation (FMT), hold potential for restoring microbial homeostasis and ameliorating disease symptoms. Targeting host-microbe interactions, mucosal immune signaling pathways, and microbial metabolites represents additional therapeutic approaches for addressing gut microbiota dysbiosis and its associated diseases [7].

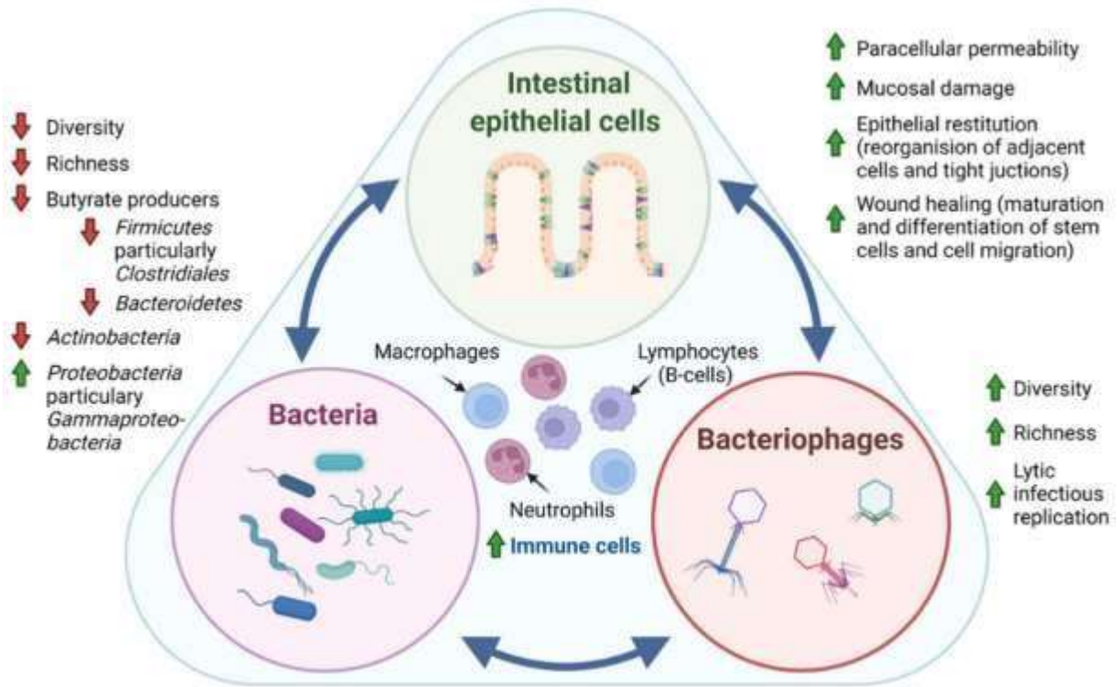


Fig 1.1: Pathogenesis of Gut Microbiota

Site : M. T.-H. Truong, H.-P. Phasitm, T. T.-L. Nguyen, P.-N. Tran, D. M.-H. Nguyen, D. T.-C. Nguyen, T. V.-H. Pham, H. Q. Nguyen, N. H.-T. Nguyen and L. Q.-S. Pham, "Insights into the Antimicrobial Resistance and Virulence of Edwardsiella ictaluri: A Systematic Analysis Using Genomic and Phenotypic Data," Microorganisms, vol. 10, no. 7, Article 1371, pp. 1-17, Jul. 2022. DOI: 10.3390/microorganisms10071371.

CHAPTER 2

ALZHEIMR’s DISEASE

2.1 Background

The most common type of dementia, known as Alzheimer's disease (AD) after the German physician Alois Alzheimer, is typified by a progressive loss of cognitive abilities. Because to the aggregation of amyloid-beta peptide (Aβ), it is characterized by the build-up of neuritic plaques and neurofibrillary tangles, mainly in the brain's medial temporal lobe and neocortical areas. Alois Alzheimer initially noticed the etiology of Alzheimer's disease after examining the brain of a patient who had experienced memory loss and personality changes prior to passing away. This led to the diagnosis of the patient's illness as a severe cerebral cortex issue. The phrase "Alzheimer's disease" was later created by Emil Kraepelin and used in his psychiatric handbook.[8]

The progressive cognitive decline associated with Alzheimer's disease (AD) can be caused by cerebral disorders like AD itself, as well as other conditions like infections, intoxications, abnormalities of the circulatory system that lower blood flow to the brain, deficiencies in nutrition (like low vitamin B12), tumors, and more. Worldwide, there are estimated to be 50 million people with AD, and this figure is expected to increase every five years, perhaps reaching 152 million by 2050. AD affects people individually, as well as their families and the economy as a whole; the estimated yearly worldwide expenses are $1 trillion. There is presently no known cure for Alzheimer's disease, however there are therapies that lessen its symptoms.[9].

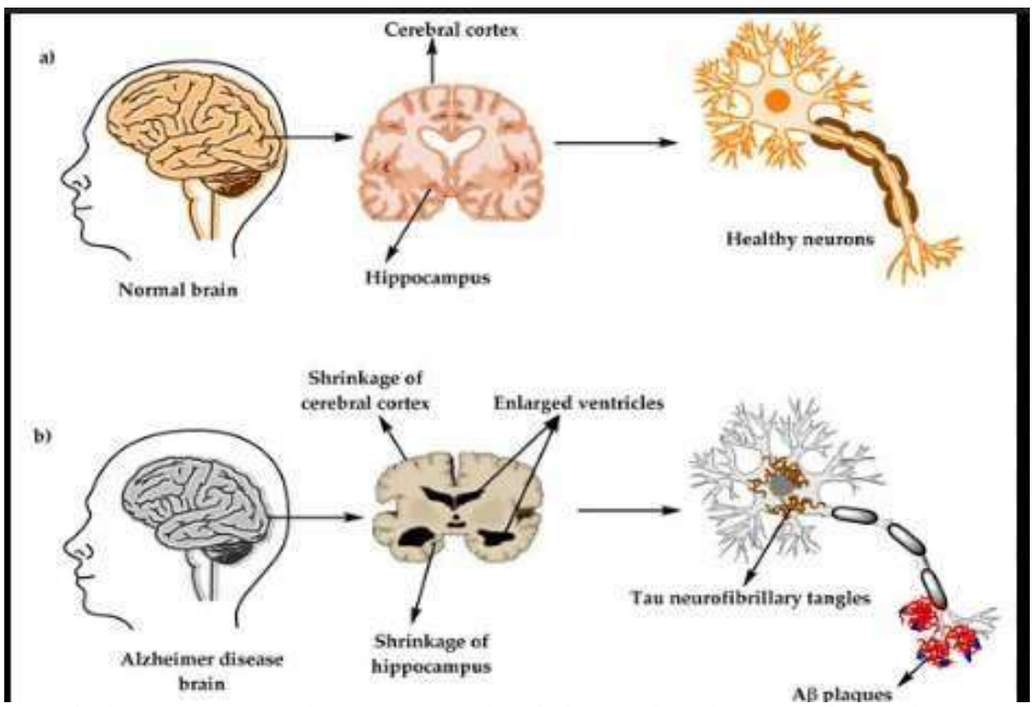


Fig.2.1 : comparison between normal and abnormal Alzheimer’s disease brain

Site: R. Pourdarbani, S. S. Hosseini, E. Hosseinzadeh, M. Mehrara, S. Masoum and S. M. Nabavi, "Applications of Machine Learning in Targeting Protein-Protein Interactions with a Focus on Small Molecules," Molecules, vol. 25, no. 24, Article 5789, pp. 1-21, Dec. 2020. DOI: 10.3390/molecules25245789.

2.2 Association with Alzheimer's Disease

1. Amyloid Beta Deposit: Aβ deposition is a hallmark feature of AD pathology, contributing to synaptic dysfunction and neuronal toxicity. Recent studies have implicated gut dysbiosis in promoting Aβ accumulation in the brain, with alterations in gut microbial composition and function influencing Aβ production and clearance mechanisms. Dysbiotic gut microbiota may exacerbate Aβ pathology through microbial metabolites, inflammatory mediators, and disruption of gut-brain axis communication pathways.
2. Tau Phosphorylation: Hyperphosphorylation of tau protein leads to the formation of neurofibrillary tangles, contributing to neuronal degeneration in AD. Growing evidence suggests that gut microbiota dysbiosis may influence tau pathology through modulating systemic inflammation, oxidative stress, and neuroinflammatory signaling pathways. Moreover, microbial metabolites such as short-chain fatty acids (SCFAs) and lipopolysaccharides (LPS) have been implicated in tau phosphorylation and neuroinflammation.
3. Neuroinflammation: Chronic neuroinflammation is a key pathological feature of AD, characterized by microglial activation, cytokine release, and immune cell infiltration in the brain. Dysbiotic gut microbiota can induce systemic inflammation and disrupt the blood-brain barrier, facilitating the infiltration of peripheral immune cells into the brain parenchyma. Moreover, microbial-derived inflammatory mediators may directly activate microglia and astrocytes, exacerbating neuroinflammatory responses in AD.
4. Metabolic Dysfunction: Metabolic dysfunction, including insulin resistance and dyslipidemia, has been implicated in AD pathogenesis. Gut microbiota dysbiosis can modulate host metabolism through influencing energy harvest, nutrient absorption, and systemic inflammation. Altered gut microbial composition has been associated with metabolic disorders, which may contribute to AD risk and progression through dysregulation of metabolic pathways and insulin signaling in the brain.
5. Oxidative Stress: Oxidative stress plays a critical role in AD pathogenesis, contributing to neuronal damage and synaptic dysfunction. Dysbiotic gut microbiota can promote oxidative stress through the production of reactive oxygen species (ROS) and depletion of antioxidant defenses. Furthermore, gut-derived metabolites such as trimethylamine N-oxide (TMAO) have been implicated in oxidative damage and neuroinflammation in AD [10].

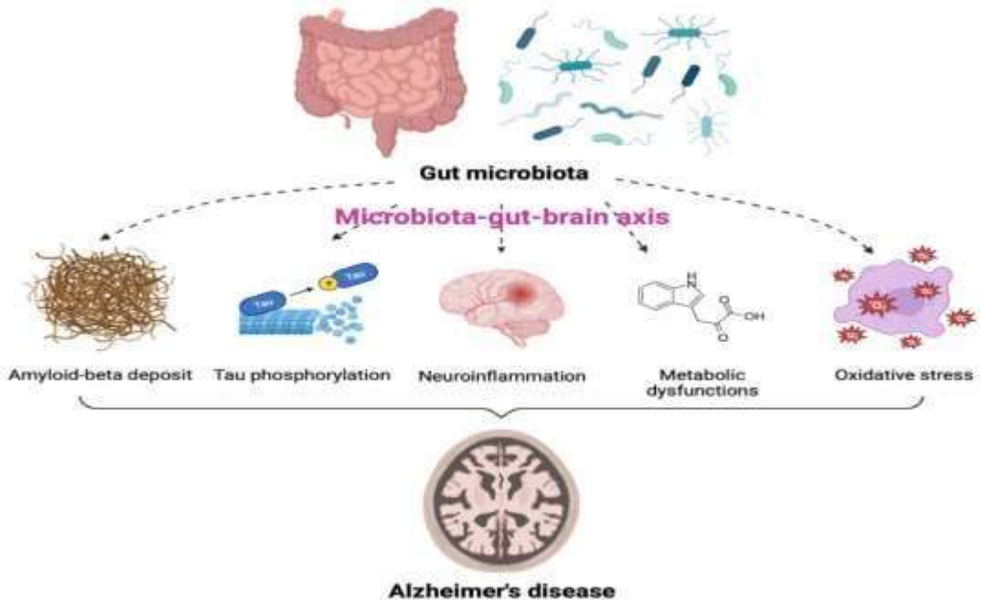


Fig.2.2 : Gut microbiota and its association with Alzheimer's disease

Site : S. Rodrigues, A. Vilela, A. Oliveira and J. M. S. Rocha, "Machine Learning in Yeast Industrial Biotechnology: Applications, Challenges, and Future Perspectives," International Journal of Molecular Sciences, vol. 24, no. 4, Article 4047, pp. 1-28, Feb. 2023. DOI: 10.3390/ijms24044047.

CHAPTER 3

RESULTS

- Comparative and Functional Genomics
- Other Bioinformatics analysis:

✓ Identify genes & proteins from literature and databases.

✓ Gene and protein level analysis (sequence and structure

✓ based studies)

✓ Functional and evolutionary annotations

20 microorganisms present in gut microbiota

Name of microorganisms :	Gene :	Protein :	Location :	Pubmed id :	Characteristics :
Bacteroides fragilis	polysaccharide A (psA)	Bft Bacteroides fragilis toxin	large intestine colon	35404220	Gram -negative bacteria
Escherichia coli	lac Zgene	beta - galactosidase	large intestine colon (Human gastrointestinal tract)	34457996	Gram -negative bacteria
Faecalibacterium praunsitzii	Fprau_RS06095	fprau_RS06095 Butyrate	Human gut microbiota	28045459	Gram - positive bacteria
Bifidobacterium longum	"gro EL"	chaperonin Gro EL	Human gut microbiota	34707577	Gram - positive bacteria
Akkermansia muciniphila	Amuc RS04450	Amuc_ 1100	Human gut microbiota	32937828	Gram -negative bacteria
Lactobacillus acidophilus	lac S	Lactose transport protein	Human gut	30525953	Gram - positive bacteria
Clostridium difficile	"tcd A"	toxin A (tcd A) protein	Gut microbiota	35980280	Gram - positive bacteria
Ruminococcus bromii	GH 3	GH3 enzyme (glycosidase hydrolase	Human gut microbiota	22343308	Gram - positive bacteria
Streptococcus thermophilus	lac Z gene	beta - galactosidase	Human gut microbiota	34707577	Gram - positive bacteria
Enterococcus faecalis	"Efa A" GENE	Efa A protein	Gastrointestinal tracts of human and animals	36157114	Gram - positive bacteria
Prevotella spp.	PUL gene	CaZymes	Human gut microbiota	33461110	Gram -negative bacteria
Roseburia spp.	butyryl CoA: acetate CoA-transferase gene	carbohydrate active enzymes	Human gut microbiota	35324673	Gram - positive bacteria
Fusobacterium spp.	"fim A " gene	Fim A protein	Both oral and gut microbiota	28214091	Gram -negative bacteria
Parabacteriodes spp.	"psa A" gene	Psb A protein	Human gut microbiota and other gut bacteria	29734011	Gram -negative bacteria
veillonella spp.	"vlg" gene	Veillonella lipoprotein G. lipoproteins	Human oral cavity and gastrointestinal tract	26486646	Gram -negative bacteria
Eubacterium spp.	"acs" gene	Acetyl- CoA synthetase enzyme	Human gut microbiota	27007700	Gram - positive bacteria
Blautia spp.	" phn gene cluster"	Phosphonate utilization proteins	Human gut microbiota	35053205	Gram - positive bacteria
Methanobrevibacter smithi	"mcr A" gene	Methyl- co-enzymes M reductase	Human gut microbiota	31031713	Gram - positive bacteria or gram- negative bacteria
Klebsiella spp.	"wca G" gene	Capsular polysaccharides	Human gut microbiota	34442694	Gram -negative bacteria
Lactobacillus rhamnosus	"fol A " gene	dihydrofolate reductase enzyme	Human gut	33985438	Gram- positive bacteria

Firmicutes	"rpo B " gene	RNA polymerase beta subunit	large intestine colon	33246175	Gram- positive bacteria

Table 3.1 : 20 microorganisms present in gut microbiota

The gut microbiota, a diverse community of microorganisms residing in the gastrointestinal tract, plays a crucial role in maintaining host health and homeostasis. Among the plethora of microorganisms residing in the gut, several key species have been identified, each contributing uniquely to the ecosystem's functionality. Bacteroides fragilis, a Gram-negative bacterium, produces polysaccharide A (psA), influencing immune responses in the large intestine. Similarly, Escherichia coli, another Gram-negative bacterium, harbors the lac Z gene, encoding for beta-galactosidase, essential for lactose metabolism in the colon. On the Gram-positive side, Faecalibacterium praunsitzii, known for its anti-inflammatory properties, synthesizes butyrate in the human gut, crucial for epithelial cell health.

Another significant player in the gut microbiota is Bifidobacterium longum, a Gram-positive bacterium that produces the chaperonin GroEL, aiding in protein folding and stress response. Akkermansia muciniphila, a Gram-negative bacterium, thrives on mucin and helps maintain gut barrier integrity. Meanwhile, Lactobacillus acidophilus, Streptococcus thermophilus, and Enterococcus faecalis, all Gram-positive bacteria, contribute to lactose metabolism and gastrointestinal health through various mechanisms such as lactose transport and colonization resistance against pathogens.

The gut microbiota also includes anaerobic bacteria like Clostridium difficile, which produces toxin A (tcdA), contributing to gastrointestinal infections. Ruminococcus bromii synthesizes glycosidase hydrolase (GH3 enzyme), essential for breaking down complex carbohydrates. Prevotella spp. and Roseburia spp., both Gram-negative and Gram-positive bacteria, respectively, produce carbohydrate-active enzymes (CAZymes), aiding in carbohydrate metabolism and fermentation.

Other notable members of the gut microbiota include Fusobacterium spp. and Parabacteroides spp., contributing to both oral and gut microbial communities, and Veillonella spp., predominantly found in the oral cavity and gastrointestinal tract. Eubacterium spp. and Blautia spp. are known for their roles in acetyl-CoA synthesis and phosphonate utilization, respectively, while Methanobrevibacter smithi produces methyl-coenzyme M reductase, contributing to methane production in the gut.

Klebsiella spp., a Gram-negative bacterium, synthesizes capsular polysaccharides, influencing mucosal immunity and gut colonization. Lastly, Lactobacillus rhamnosus, another Gram-positive bacterium, produces dihydrofolate reductase enzyme, essential for folate metabolism and gut health. Collectively, these diverse microorganisms form a complex ecosystem within the gut microbiota, orchestrating various metabolic, immune, and physiological functions essential for host well-being.

3.1 MSA and Phylogenetic tree

Multiple sequence alignment (MSA) is a method used in bioinformatics to align three or more biological sequences (protein, DNA, or RNA) simultaneously. This process is crucial for identifying regions of similarity and difference across the sequences, which can provide insights into structural, functional, and evolutionary relationships. For the purpose of a thesis, MSA can serve several key roles:

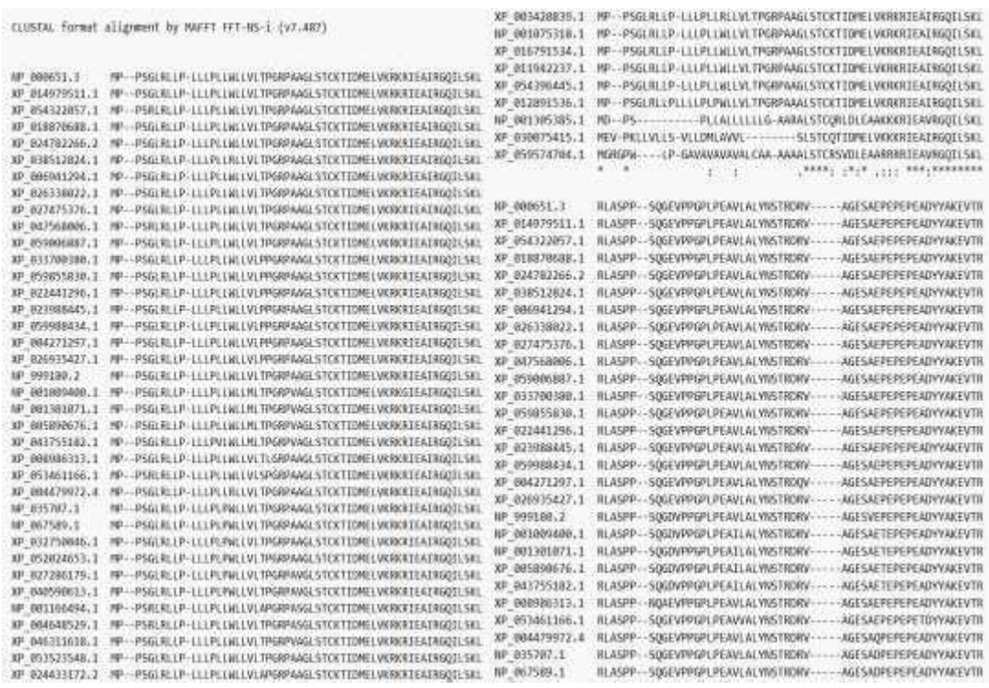
Functional Analysis: By aligning sequences, researchers can identify conserved regions that may be crucial for the biological function of the proteins or genes in question. Conserved sequences often indicate important structural or functional domains that have been preserved through evolution.

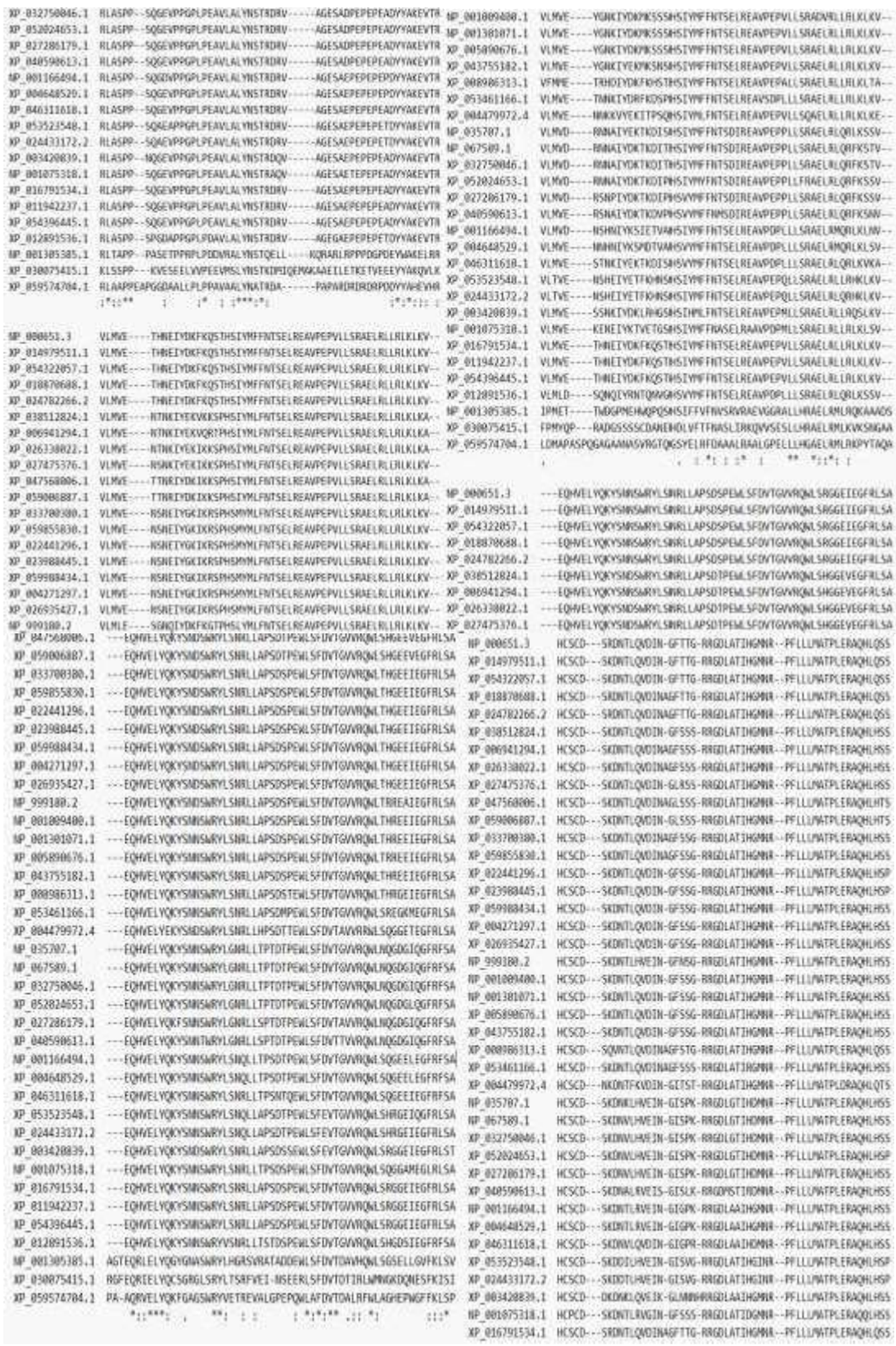
Evolutionary Studies: MSA can be used to infer phylogenetic relationships between sequences. By comparing aligned sequences, one can construct phylogenetic trees that illustrate the evolutionary pathways and common ancestors of the sequences.

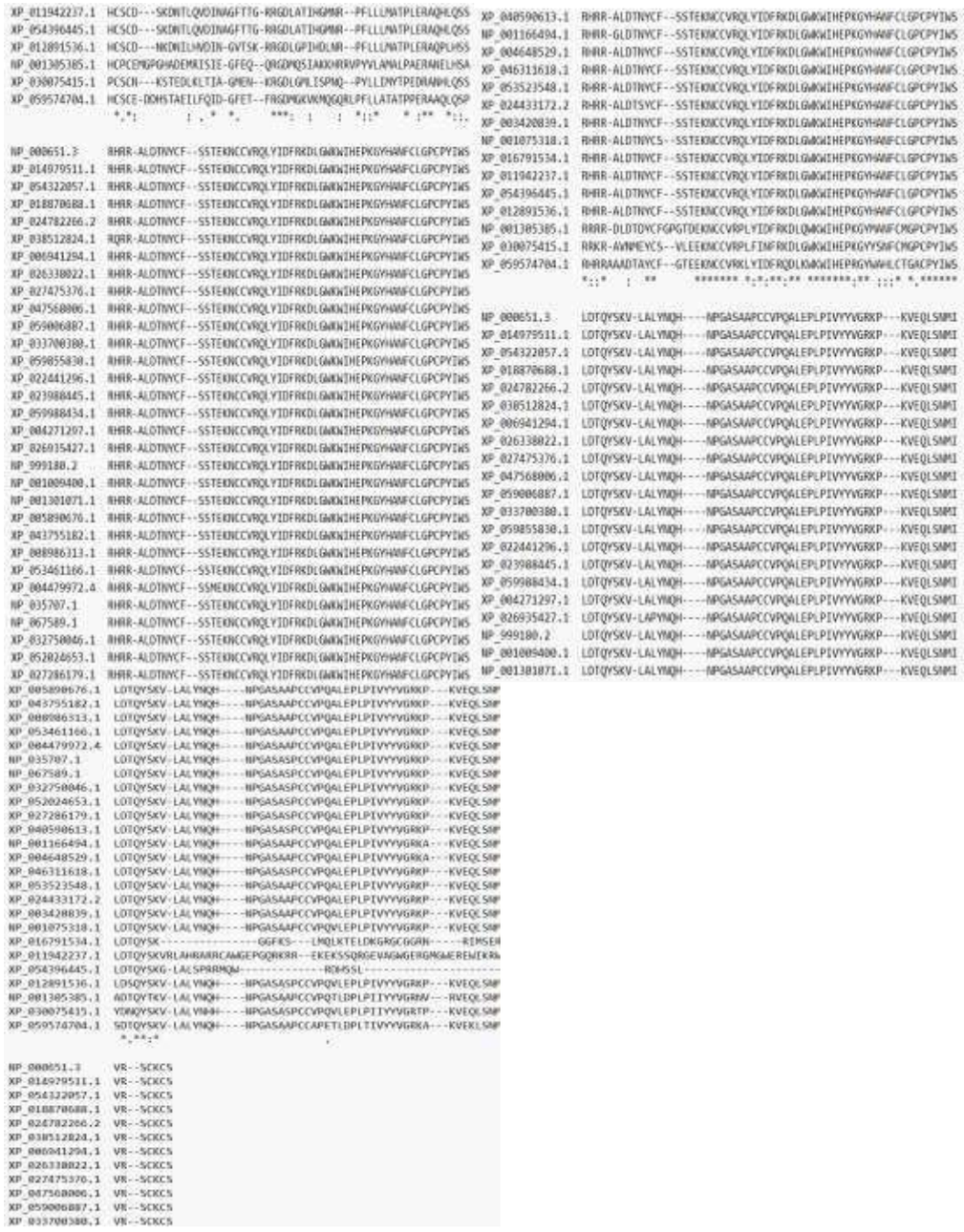
Structural Prediction: For proteins, conserved regions identified through MSA can help predict secondary and tertiary structures. Structural motifs that are conserved across different proteins often indicate similar folding patterns and functional properties.

Mutation Analysis: MSA allows for the identification of specific mutations, insertions, and deletions. This can be particularly useful in understanding genetic diseases, adaptation mechanisms, and the effects of evolutionary pressures on sequences.

Annotation of Genomes: When annotating new genomes, MSA can help transfer functional information from well-characterized sequences to newly sequenced genomes by identifying homologous regions.







3.3 Phylogenetic tree

A phylogenetic tree is a diagram that represents the evolutionary relationships among a group of organisms, typically based on genetic or genomic data. These trees illustrate the branching patterns of descent from a common ancestor, providing insights into the evolutionary history and relatedness of different species or groups of organisms.

Phylogenetic trees are constructed using various computational methods and algorithms that analyze molecular sequence data, such as DNA, RNA, or protein sequences. These methods include distance-based approaches, maximum likelihood estimation, and Bayesian inference, among others. By comparing sequences and identifying shared similarities and differences, bioinformaticians can infer the evolutionary distances between organisms and construct phylogenetic trees that depict their evolutionary relatedness.

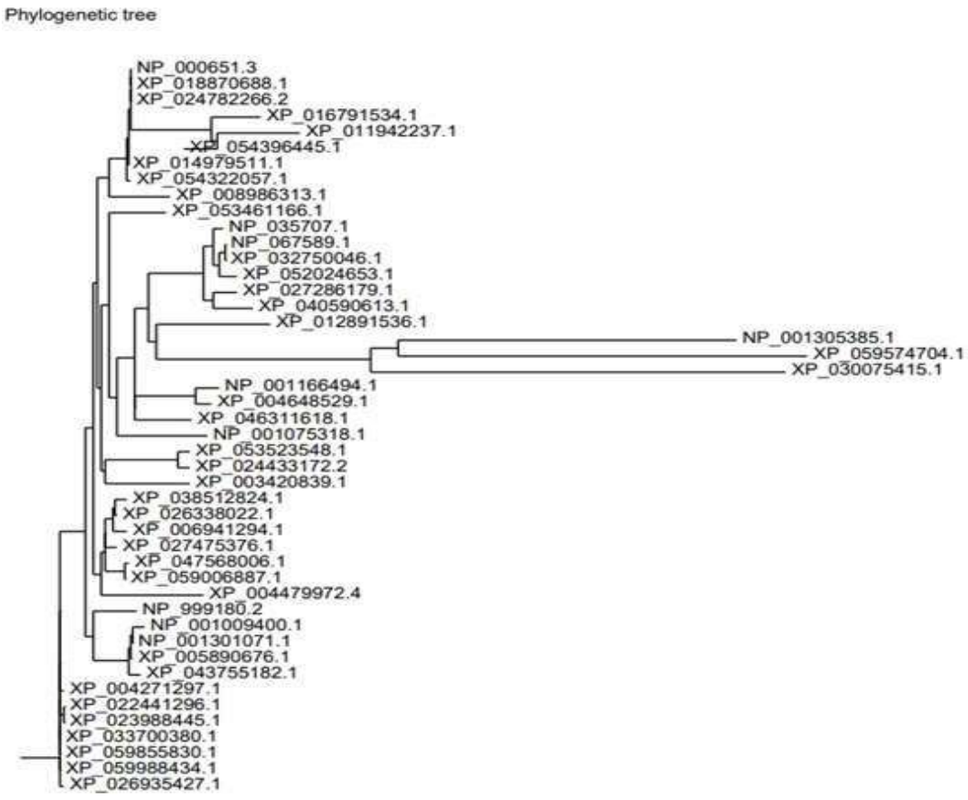


Fig.3.1: Phylogenetic tree of some microorganisms

Phylogenetic trees are commonly represented as branching diagrams, with branches representing evolutionary lineages and nodes indicating points of divergence or common ancestors. The length of branches often reflects the amount of evolutionary change or genetic divergence between taxa, with longer branches indicating greater evolutionary distance.

This phylogenetic tree organizes the microorganisms into two main branches based on their Gram staining characteristics (Gram-negative and Gram-positive). Within each branch, the organisms are further grouped based on their taxonomic relationships, with closely related species positioned closer together. This visual representation provides an overview of the evolutionary relationships among the listed microorganisms. Actual phylogenetic trees generated through molecular data analysis would provide more precise and detailed information about their evolutionary history

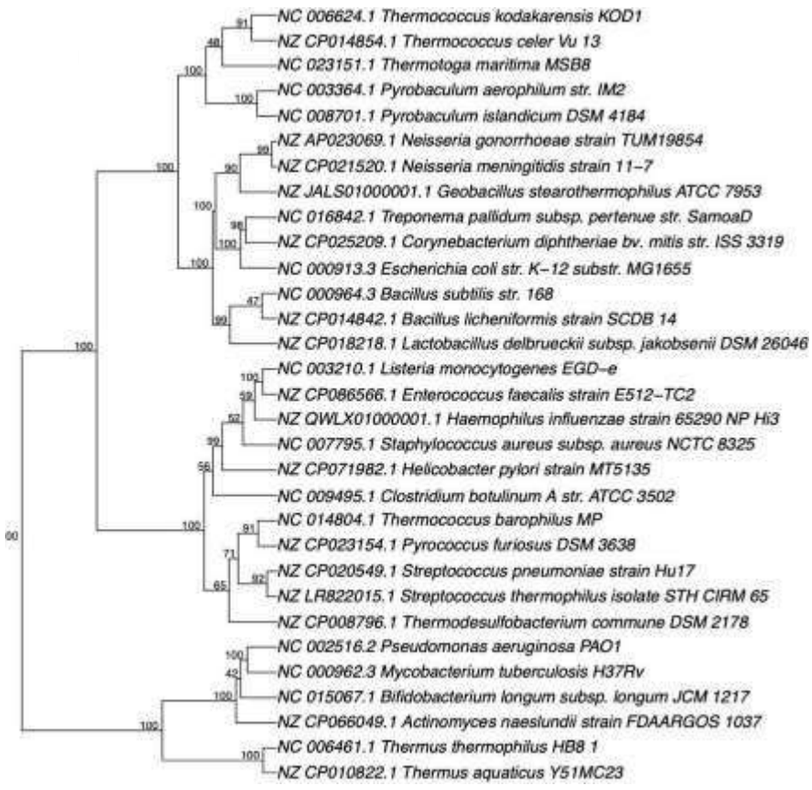


Fig.3.2: Phylogenetic tree shows some microorganisms present in gut microbiota

3.3 Diagnosis and current treatment

3.4.1 **Clinical Assessment**: A comprehensive medical history, cognitive tests, and a medical professional's assessment of the patient's symptoms are usually the first steps in the diagnosis process. Screening for memory loss, cognitive decline, behavioral abnormalities, and functional impairments may be part of these evaluations.

3.4.2 **Neuropsychological Testing**: To evaluate cognitive function, including memory, language, attention, and problem-solving abilities, neuropsychological tests are administered. These exams aid in the diagnosis of cognitive impairment and the monitoring of the course of illness.

3.4.3 **Imaging Studies:** The characteristic hallmarks of Alzheimer's disease (AD) that are most visible when using brain imaging methods like positron emission tomography (PET) scans and magnetic resonance imaging (MRI) are the presence of neurofibrillary tangles and amyloid plaques.

3.4.4 **Biomarker Analysis:** Cerebrospinal fluid (CSF) analysis and blood tests may be performed to measure levels of biomarkers associated with AD pathology, including amyloid-beta and tau proteins. These biomarkers aid in early detection and monitoring of disease progression.

3.4 Treatment :

Pharmacological Therapy: Memantine, an N-methyl-D-aspartate (NMDA) receptor antagonist, and cholinesterase inhibitors, such as donepezil, rivastigmine, and galantamine, are among the currently authorized treatments for AD that aim to relieve symptoms. For certain AD patients, these medications may aid with behavior, daily living tasks, and cognitive function.

Non-pharmacological Interventions: Non-pharmacological methods are essential for controlling AD symptoms and enhancing life quality. Cognitive stimulation treatment, physical activity, occupational therapy, speech therapy, and programs for the psychological support of patients and caregivers are a few examples of these therapies.

Lifestyle Modifications: A balanced diet high in fruits, vegetables, and omega-3 fatty acids, regular exercise, enough sleep, and social interaction are all components of a healthy lifestyle that may promote general brain health and may lower the risk of cognitive decline.

Experimental Therapies: Targeting many pathogenic pathways, including neuroinflammation, synaptic dysfunction, and the aggregation of tau and amyloid-beta proteins, a number of experimental medications and treatments are being developed to treat AD. These experimental treatments include monoclonal antibodies, immunotherapies, anti-inflammatory agents, and disease-modifying drugs.

Supportive Care: As AD progresses, individuals may require increasing levels of care and support to manage daily activities, behavioral changes, and medical needs. Caregiver support services,

respite care, day programs, and long-term care facilities can provide assistance and relief to caregivers while ensuring the safety and well-being of patients.

Treatment involves psychological therapies such as CBT, Life story work, CST, Cognitive rehabilitation, Music and creative arts, etc. along with prescribed medication including Donepezil, Rivastigmine, Galantamine and Memantine.

Two new drugs, named **“Aduhelm”** and **“Leqembi”** have been approved by FDA, to be used for Alzheimer’s treatment.

CHAPTER 4

CODON BAIS ANALYSIS

Through a complicated process involving codons, the language of DNA, which is made up of nucleotide base sequences, is translated into the twenty-letter language of amino acids in proteins. Three nucleotide triplets function as translation stop signals, while the remaining 61 nucleotide triplets encode the twenty standard amino acids. Because of the degeneracy of the genetic code, one amino acid can be encoded by more than one codon. However, the uneven use of these synonymous codons results in a condition called codon use bias (CUB) [1]. Natural selection, genetic drift, and mutational patterns all contribute to the variation of CUB among species, genes, and functional categories. Without changing the amino acid sequence, mutations in gene-coding areas, especially in the second or third nucleotide of codons, can lead to synonymous codon use biases. Furthermore, local recombination rates impact GC heterogeneity and GC-biased gene conversion, which both contribute to CUB. Comprehending CUB is essential as it affects a number of biological functions, including as transcription, the stability of mRNA, and the effectiveness of protein translation. Additionally, CUB analysis helps determine horizontally transmitted genes and clarify the evolutionary links between species [2]. Bias towards optimum codons is frequently observed in highly expressed proteins, underscoring the significance of CUB in genetic engineering and recombinant DNA technologies for augmenting protein production. The increasing amount of genomic data calls for a greater investigation and utilization of CUB in many species. CUB is an important component of gene expression, stress response, and environmental adaptability in plants. Understanding how codons are used by plants can help with crop development techniques and offer insights into their biological variability. The foundations of genetic code, variables impacting CUB, computer techniques for CUB analysis, and its applications to improving our knowledge of plant biology and crop breeding are covered in this study. [3].

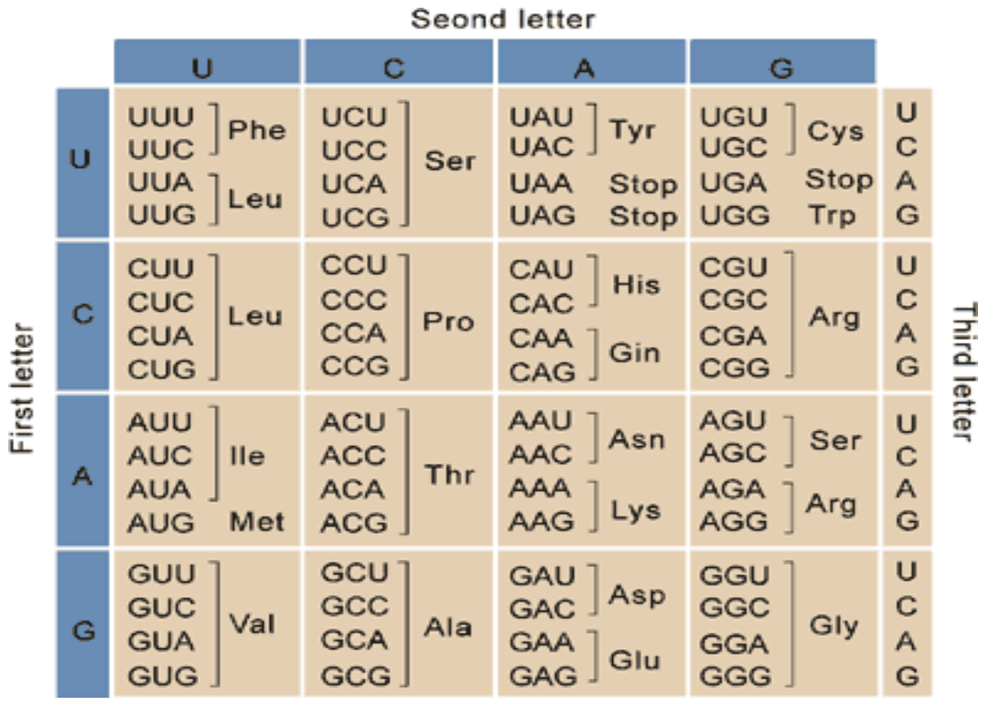


Fig.4.1: Genetic code

4.1 Factors affecting codon bias analysis:

Codon bias refers to the phenomenon where certain codons are preferred over others in the coding sequences of an organism, despite multiple codons coding for the same amino acid. Several factors influence codon bias, reflecting a complex interplay of molecular, evolutionary, and environmental influences[4].

1. Mutational Bias: The nucleotide composition of an organism's genome can influence codon bias. For example, organisms with a high GC content tend to prefer codons rich in G and C, while those with AT-rich genomes favor codons with more A and T nucleotides.
2. Natural Selection: Codon usage can be shaped by natural selection to optimize the efficiency and accuracy of protein translation. Highly expressed genes often use codons that match the most abundant tRNAs, facilitating faster and more accurate protein synthesis. This is particularly evident in highly expressed proteins like ribosomal proteins and heat shock proteins.
3. Genetic Drift: In small populations, random changes in codon usage can occur due to genetic drift. This can lead to variations in codon bias that are not necessarily advantageous but are fixed in the population by chance.
4. tRNA Availability and Adaptation: The abundance and availability of tRNAs corresponding to specific codons can drive codon bias. Organisms adapt their codon usage to match the tRNA pool, ensuring efficient translation. In some cases, the tRNA pool itself may evolve to accommodate preferred codons in highly expressed genes.
5. 5. Translational Accuracy and Efficiency: Codon bias can improve translation accuracy and efficiency. Genes that are highly expressed and necessary need to operate properly, and codons that match numerous tRNAs can translate more rapidly and with fewer mistakes.
6. 6. Expression Levels of Genes: Genes with high expression show more codon bias than those with low expression. This bias frequently favors codons that align with the most prevalent tRNAs, guaranteeing accurate and effective translation of widely required proteins.
7. 7. Secondary Structure and mRNA Stability: The secondary structure of mRNA is influenced by codon use, which in turn affects the translation initiation efficiency and stability of the mRNA. While too much stability might prevent ribosome access and lower translation efficiency, stable mRNA secondary structures can help a transcript last longer.
8. Evolutionary and Phylogenetic Constraints: Different evolutionary lineages exhibit distinct codon usage patterns due to their unique evolutionary histories and adaptations. Phylogenetic constraints can therefore play a significant role in shaping codon bias across different species.
9. Environmental Factors: Environmental pressures such as temperature, nutrient availability, and other stress conditions can influence codon bias. For instance, some organisms might adjust their codon usage in response to heat stress to favor codons that enhance the stability of the resulting proteins.[5,6]

4.2 Softwares for codon usage analysis:

Numerous indices for codon usage bias have been developed to estimate and comprehend codon usage preferences. However, computation and analysis are often complex and not straightforward. To simplify these processes, various software tools have been created, including INteractive Codon usage Analysis (INCA), JCat (Java Codon Adaptation Tool), COUSIN (COdon Usage Similarity INdex), Automated Codon Usage Analysis (ACUA) software, CAIcal, and others. These software tools primarily calculate the Codon Adaptation Index (CAI), Effective Number of Codons (ENC), and occasionally other indices. [7] COUSIN (http://cousin.ird.fr) encompasses COUSIN and seven other indices, offering additional functionalities such as statistical analyses, clustering, and optimization of codon usage for gene expression. Some software tools are regularly updated, while others are specifically designed for Linux or UNIX environments, like codonW. Unfortunately, some tools, such as Codon Explorer, are either obsolete or unavailable. [8].

4.3 Applications:

Codon usage analysis serves as a valuable tool for understanding various biological phenomena, encompassing evolution, phylogenetic relationships, host-pathogen co-evolution, and environmental adaptations. It also provides insights into the molecular evolution of individual genes, horizontal gene transfer between species, identification of protein-coding regions in uncharacterized genomic DNA, and translation studies. This knowledge is especially beneficial in designing degenerate oligonucleotides for PCR amplification of genes.

Furthermore, codon usage bias significantly impacts gene expression regulation, protein structure and function, co-translational protein folding, recombinant protein production, and the functional classification of proteins. Taking into account both GC content and codon usage is essential for achieving high levels of gene expression.[9]

A significant application of codon usage data lies in optimizing gene expression in heterologous systems. When introducing a foreign gene into a crop, its codons are modified to align with the codon usage pattern of the host plant, a procedure known as codon re-engineering. This optimization emphasizes the utilization of optimal codons rather than merely abundant ones. Numerous algorithms and software tools have been developed for codon optimization and synthetic gene design, such as Codon optimizer, OPTIMISER, Eugene, COStar, DNA-Tailor, COOL, CODA, ATGme, and CodonWizard.

The process involves designing candidate sequences by selecting codons based on their probabilities from codon usage tables, filtering sequences to meet other design criteria, eliminating unfavorable codon pairs, extreme GC content, repetitive sequences, and ensuring desirable mRNA structures. [10]

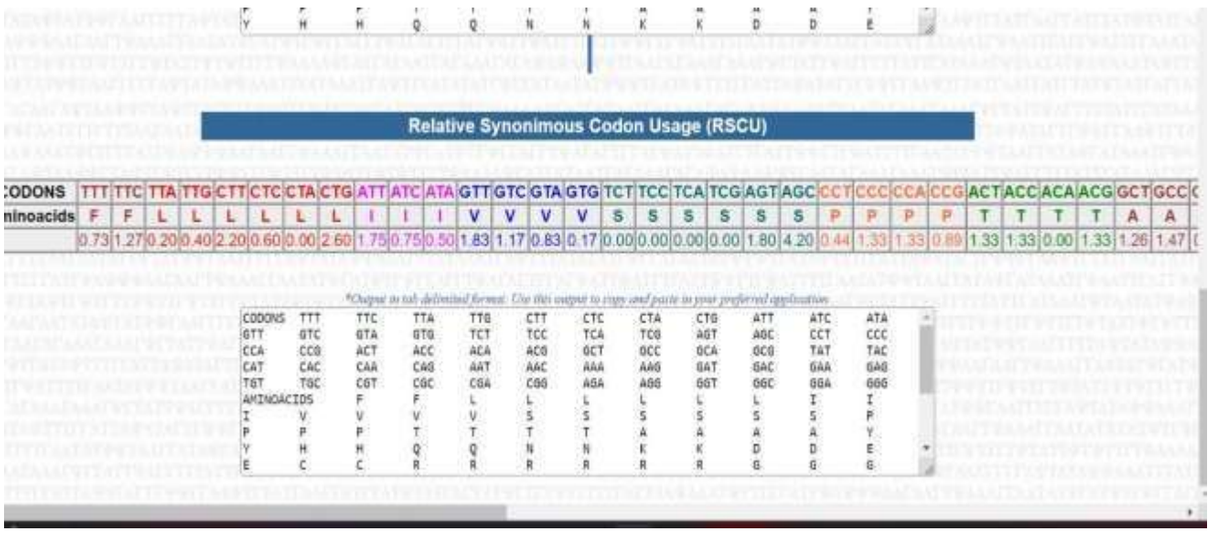
Efficient codons for endogenous genes may not be suitable for heterologous genes, as overexpression issues can lead to amino acid starvation and alter the abundances of charged tRNAs. Codon optimization has proven successful in plants for enhancing the expression of 'Cry' proteins for pest resistance and recombinant proteins for molecular pharming. For instance, synthetic cry genes from Bacillus thuringiensis, optimized for plant-preferred codons, have been introduced into crops like rice

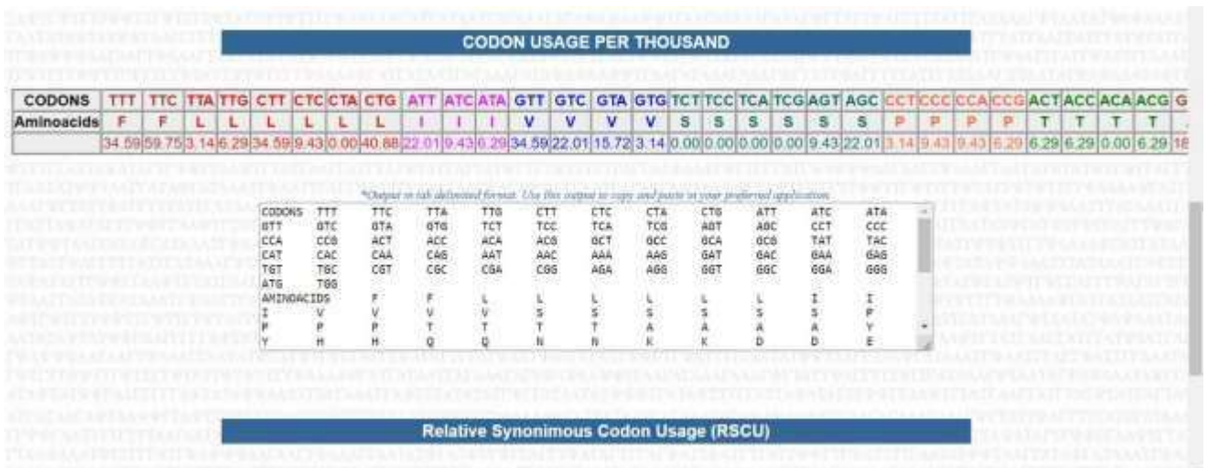
and cotton, providing significant protection against lepidopteran insects. Additionally, designing different codon-optimized cry genes for monocot and dicot plants is practiced to further enhance transgene expression levels.

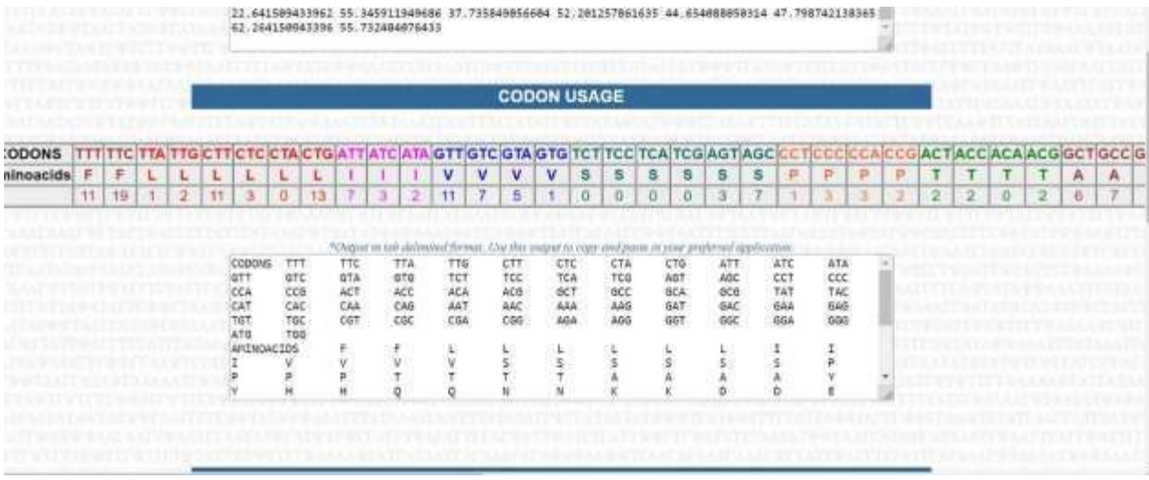
In another application, wheat cytochrome P450 genes, which exhibit high GC content and strong codon usage bias, were codon-optimized and introduced into yeast and tobacco for bioremediation purposes. Codon optimization has resulted in varying degrees of increased gene expression, often by enhancing translational efficiency rather than transcript abundance. This process can alleviate ribosome pausing at specific codons, particularly in chloroplasts. When expressing large genes, beyond codon optimization, understanding tRNA encoding by the genome and compatibility with regulatory sequences is crucial for optimal translation initiation and elongation.

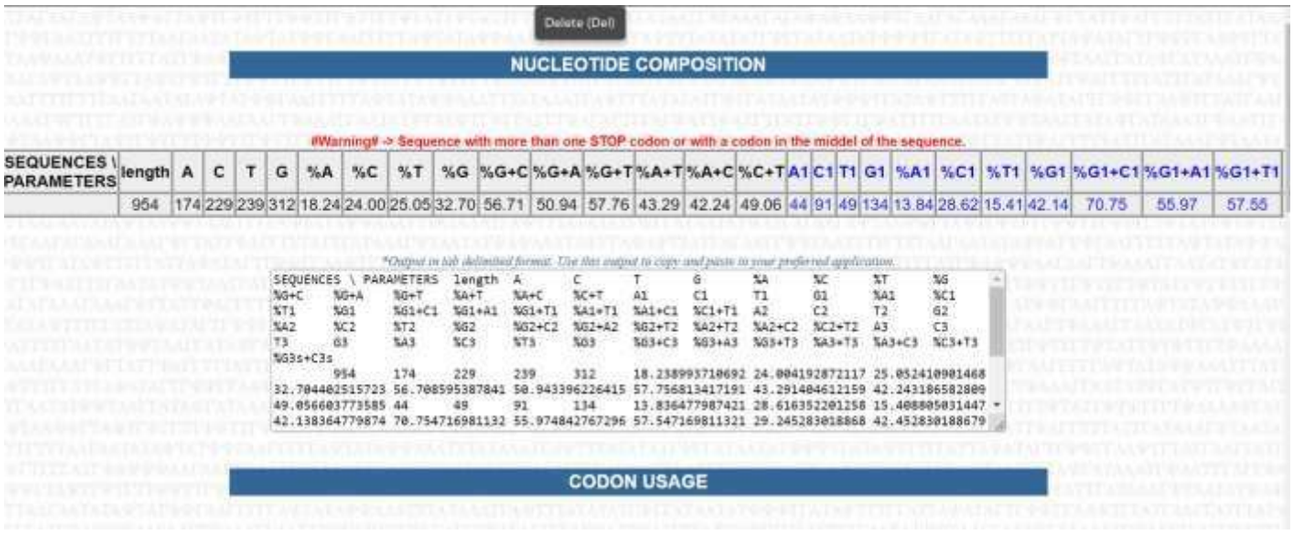
Furthermore, codon reassignment has been employed to incorporate non-canonical amino acids (ncAAs), enabling analysis of protein structure and interactions, introduction of post-translational modifications, production of constrained peptides, antibody-drug conjugates, and novel enzymes. Thus, the scope of applications for codon usage analysis is extensive. [11]

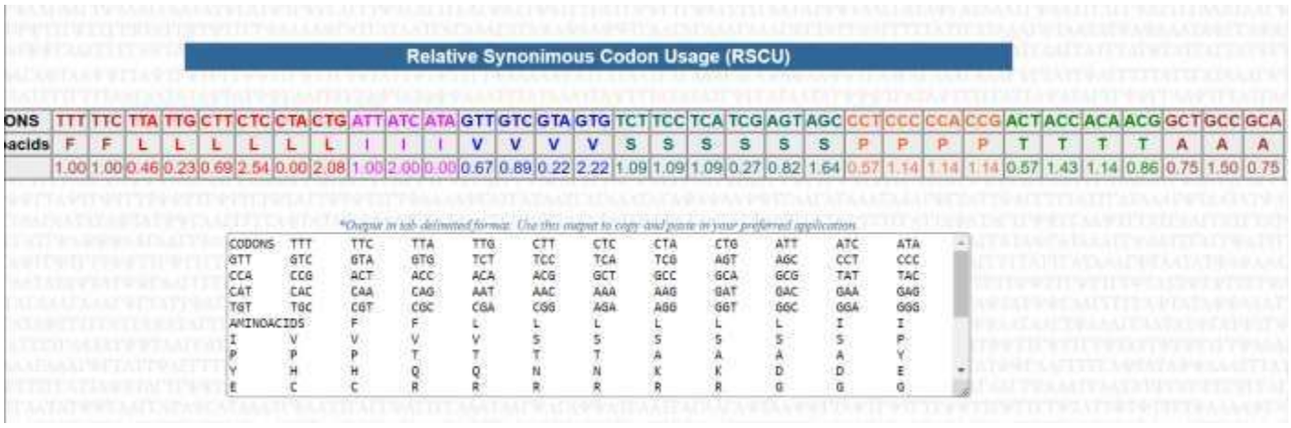
4.4 Results

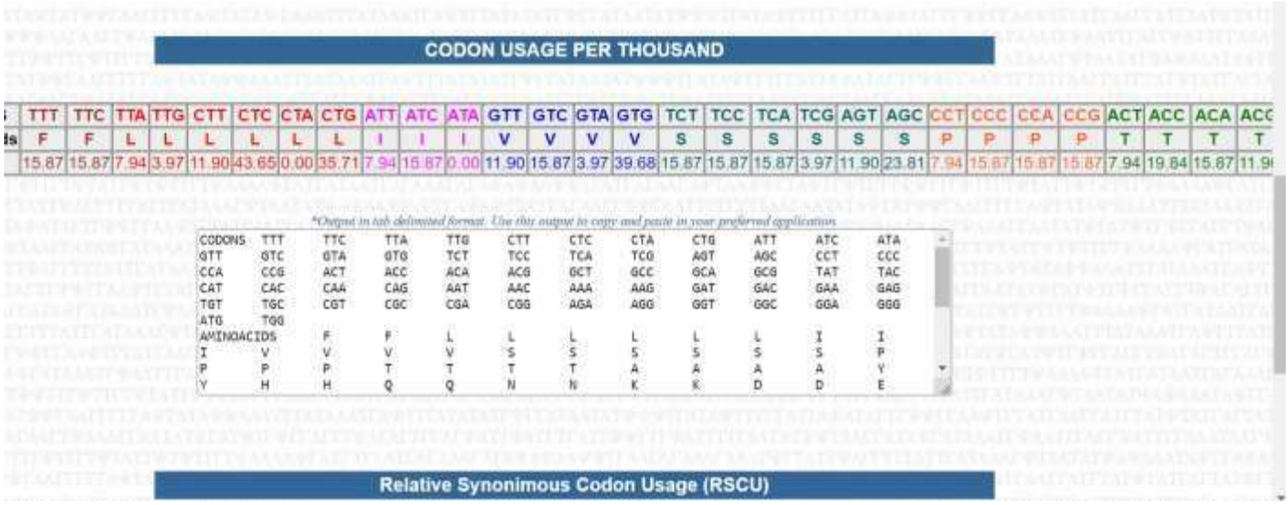





22.641589433962 55.345911949686 37.735849056604 52.201257861635 44.654088050314 47.798742138365
62.264150943396 55.732484076433
CODON USAGE
CODONS TTT TTC TTA TTG CTT CTC CTA CTG ATT ATC ATA GTT GTC GTA GTG TCT TCC TCA TCG AGT AGC CCT CCC CCA CCG ACT ACC ACA ACG GCT GCC G
Aminoacids F F L L L L L L I I I V V V V S S S S S S P P P P T T T T A A
11 19 1 2 11 3 0 13 7 3 2 11 7 5 1 0 0 0 0 3 7 1 3 3 3 2 2 0 2 6 7
*Output in tab delimited format. Use this output to copy and paste in your preferred application
CODONS TTT TTC TTA TTG CTT CTC CTA CTG ATT ATC ATA
GTT GTC GTA GTG TCT TCC TCA TCG AGT AGC CCT CCC
CCA CCG ACT ACC ACA ACG GCT GCC GCA GCG TAT TAC
CAT CAC CAA CAG AAT AAC AAA AAG GAT GAC GAA GAG
TGT TGC CGT CGC CGA CGG AGA AGG GGT GGC GGA GGG
ATG TGG
AMINOACIDS F F L L L L L L I I
I V V V V S S S S S P
P P P T T T T A A A A Y
Y H H Q Q N N K K D D E


Delete (Del)
NUCLEOTIDE COMPOSITION
#Warning# -> Sequence with more than one STOP codon or with a codon in the middel of the sequence.
SEQUENCES \ PARAMETERS length A C T G %A %C %T %G %G+C %G+A %G+T %A+T %A+C %C+T A1 C1 T1 G1 %A1 %C1 %T1 %G1 %G1+C1 %G1+A1 %G1+T1
954 174 229 239 312 18.24 24.00 25.05 32.70 56.71 50.94 57.76 43.29 42.24 49.06 44 91 49 134 13.84 28.62 15.41 42.14 70.75 55.97 57.55
*Output in tab delimited format. Use this output to copy and paste in your preferred application
SEQUENCES \ PARAMETERS length A C T G %A %C %T %G
%G+C %G+A %G+T %A+T %A+C %C+T A1 C1 T1 G1 %A1 %C1
%T1 %G1 %G1+C1 %G1+A1 %G1+T1 %A1+T1 %A1+C1 %C1+T1 A2 C2 T2 G2
%A2 %C2 %T2 %G2 %G2+C2 %G2+A2 %G2+T2 %A2+T2 %A2+C2 %C2+T2 A3 C3
T3 G3 %A3 %C3 %T3 %G3 %G3+C3 %G3+A3 %G3+T3 %A3+T3 %A3+C3 %C3+T3
%G3s+C3s
954 174 229 239 312 18.238993710692 24.004192872117 25.052410901468
32.704402515723 56.708595387841 50.943396226415 57.756813417191 43.291404611159 42.243186582809
49.056603773585 44 49 91 134 13.836477987421 28.616352201258 15.408805031447
42.138364779874 70.754716981132 55.974842767296 57.547169811321 29.245283018868 42.452830188679
CODON USAGE


Relative Synonimous Codon Usage (RSCU)
ONS TTT TTC TTA TTG CTT CTC CTA CTG ATT ATC ATA GTT GTC GTA GTG TCT TCC TCA TCG AGT AGC CCT CCC CCA CCG ACT ACC ACA ACG GCT GCC GCA
acids F F L L L L L L I I I V V V V S S S S S S P P P P T T T T A A A
1.00 1.00 0.46 0.23 0.69 2.54 0.00 2.08 1.00 2.00 0.00 0.67 0.89 0.22 2.22 1.09 1.09 1.09 0.27 0.82 1.64 0.57 1.14 1.14 1.14 0.57 1.43 1.14 0.86 0.75 1.50 0.75
*Output in tab delimited format. Use this output to copy and paste in your preferred application
CODONS TTT TTC TTA TTG CTT CTC CTA CTG ATT ATC ATA
GTT GTC GTA GTG TCT TCC TCA TCG AGT AGC CCT CCC
CCA CCG ACT ACC ACA ACG GCT GCC GCA GCG TAT TAC
CAT CAC CAA CAG AAT AAC AAA AAG GAT GAC GAA GAG
TGT TGC CGT CGC CGA CGG AGA AGG GGT GGC GGA GGG
AMINOACIDS F F L L L L L L I I
I V V V V S S S S S P
P P P T T T T A A A A Y
Y H H Q Q N N K K D D E
E C C R R R R R R G G G


CODON USAGE PER THOUSAND
TTT TTC TTA TTG CTT CTC CTA CTG ATT ATC ATA GTT GTC GTA GTG TCT TCC TCA TCG AGT AGC CCT CCC CCA CCG ACT ACC ACA ACC
F F L L L L L L I I I V V V V S S S S S S P P P P T T T T
15.87 15.87 7.94 3.97 11.90 43.65 0.00 35.71 7.94 15.87 0.00 11.90 15.87 3.97 39.68 15.87 15.87 15.87 3.97 11.90 23.81 7.94 15.87 15.87 15.87 7.94 19.84 15.87 11.9
*Output in tab delimited format. Use this output to copy and paste in your preferred application
CODONS TTT TTC TTA TTG CTT CTC CTA CTG ATT ATC ATA
GTT GTC GTA GTG TCT TCC TCA TCG AGT AGC CCT CCC
CCA CCG ACT ACC ACA ACG GCT GCC GCA GCG TAT TAC
CAT CAC CAA CAG AAT AAC AAA AAG GAT GAC GAA GAG
TGT TGC CGT CGC CGA CGG AGA AGG GGT GGC GGA GGG
AMINOACIDS F F L L L L L L I I
I V V V V S S S S S P
P P P T T T T A A A A Y
Y H H Q Q N N K K D D E
Relative Synonimous Codon Usage (RSCU)

NUCLEOTIDE COMPOSITION

SEQUENCES \ PARAMETERS	length	A	C	T	G	%A	%C	%T	%G	%G+C	%G+A	%G+T	%A+T	%A+C	%C+T	A1	C1	T1	G1	%A1	%C1	%T1	%G1	%G1+C1	%G1+A1	%G1+T1
	756	180	207	147	222	23.81	27.38	19.44	29.37	56.75	53.17	48.81	43.25	51.19	46.83	65	65	39	83	25.79	25.79	15.48	32.94	58.73	58.73	48.41

Output in tab delimited format. Use this output to copy and paste in your preferred application.

```
SEQUENCES \ PARAMETERS  length  A       C       T       G       %A      %C      %T      %G
%G+C    %G+A    %G+T    %A+T    %A+C    %C+T    A1      C1      T1      G1      %A1     %C1
%T1     %G1     %G1+C1  %G1+A1  %G1+T1  %A1+T1  %A1+C1  %C1+T1  A2      C2      T2      G2
%A2     %C2     %T2     %G2     %G2+C2  %G2+A2  %G2+T2  %A2+T2  %A2+C2  %C2+T2  A3      C3
T3      G3      %A3     %C3     %T3     %G3     %G3+C3  %G3+A3  %G3+T3  %A3+T3  %A3+C3  %C3+T3
%G3s+C3s
        756     180     207     147     222     23.809523809524 27.380952380952 19.444444444444
29.365079365079 56.746031746032 53.174603174603 48.809523809524 43.253968253968 51.190476190476
46.825396825397 65      39      65      83      25.793650793651 25.793650793651 15.476190476190
32.936507936508 58.730158730159 58.730158730159 48.412698412698 41.269841269841 51.587301587302
```

CODON USAGE

CODONS	TTT	TTC	TTA	TTG	CTT	CTC	CTA	CTG	ATT	ATC	ATA	GTT	GTC	GTA	GTG	TCT	TCC	TCA	TCG	AGT	AGC	CCT	CCC	CCA	CCG	ACT	ACC	ACA	ACG	GCT	GCC	G
Aminoacids	F	F	L	L	L	L	L	L	I	I	I	V	V	V	V	S	S	S	S	S	S	P	P	P	P	T	T	T	T	A	A	
	4	4	2	1	3	11	0	9	2	4	0	3	4	1	10	4	4	4	1	3	6	2	4	4	4	2	5	4	3	3	6	

Relative Synonimous Codon Usage (RSCU)

CODONS	TTT	TTC	TTA	TTG	CTT	CTC	CTA	CTG	ATT	ATC	ATA	GTT	GTC	GTA	GTG	TCT	TCC	TCA	TCG
Aminoacids	F	F	L	L	L	L	L	L	I	I	I	V	V	V	V	S	S	S	S
NC_000002.12:c135837169-135836530,c135833190-135833111,c135829676-135829593,c135824003-135823901,c135822098-135822020,c135818061-135817341,c135812956-135812311,c135809993-135808443,c135807396-135807128,c135805057-135804767,c135804128-135803930,c135800809-135800607,c135798138-135798029,c135794775-135794641,c135790881-135790658,c135789798-135789571,c135788544-135788324_Homo_sapiens_chromosome_2,_GRCh38.p14_Primary_Assembly	0.99	1.01	0.11	0.64	0.49	1.69	0.46	2.61	0.93	1.73	0.33	0.60	0.98	0.30	2.11	1.49	1.45	0.60	0.15

Output in tab delimited format. Use this output to copy and paste in your preferred application.

```
CODONS  TTT     TTC     TTA     TTG     CTT     CTC     CTA     CTG     ATT     ATC     ATA
GTT     GTC     GTA     GTG     TCT     TCC     TCA     TCG     AGT     AGC     CCT     CCC
CCA     CCG     ACT     ACC     ACA     ACG     GCT     GCC     GCA     GCG     TAT     TAC
CAT     CAC     CAA     CAG     AAT     AAC     AAA     AAG     GAT     GAC     GAA     GAG
TGT     TGC     CGT     CGC     CGA     CGG     AGA     AGG     GGT     GGC     GGA     GGG
AMINOACIDS      F       F       L       L       L       L       L       L       I       I
I       V       V       V       V       S       S       S       S       S       S       P
P       P       P       T       T       T       A       A       A       A       Y
Y       H       H       Q       Q       N       N       K       K       D       D       E
E       C       C       R       R       R       R       R       R       G       G       G
```

CODONS	TTT	TTC	TTA	TTG	CTT	CTC	CTA	CTG	ATT	ATC	ATA	GTT	GTC	GTA	GTG	TCT	TCC
Aminoacids	F	F	L	L	L	L	L	L	I	I	I	V	V	V	V	S	S
NC_000002.12:c135837169-135836530,c135833190-135833111,c135829676-135829593,c135824003-135823901,c135822098-135822020,c135818061-135817341,c135812956-135812311,c135809993-135808443,c135807396-135807128,c135805057-135804767,c135804128-135803930,c135800809-135800607,c135798138-135798029,c135794775-135794641,c135790881-135790658,c135789798-135789571,c135788544-135788324_Homo_sapiens_chromosome_2,_GRCh38.p14_Primary_Assembly	28.53	29.05	1.56	9.34	7.26	24.90	6.74	38.38	14.52	26.97	5.19	8.30	13.49	4.15	29.05	20.75	20.23

Output in tab delimited format. Use this output to copy and paste in your preferred application.

```
CODONS  TTT     TTC     TTA     TTG     CTT     CTC     CTA     CTG     ATT     ATC     ATA
GTT     GTC     GTA     GTG     TCT     TCC     TCA     TCG     AGT     AGC     CCT     CCC
CCA     CCG     ACT     ACC     ACA     ACG     GCT     GCC     GCA     GCG     TAT     TAC
CAT     CAC     CAA     CAG     AAT     AAC     AAA     AAG     GAT     GAC     GAA     GAG
TGT     TGC     CGT     CGC     CGA     CGG     AGA     AGG     GGT     GGC     GGA     GGG
ATG     TGG
AMINOACIDS      F       F       L       L       L       L       L       L       I       I
I       V       V       V       V       S       S       S       S       S       S       P
P       P       P       T       T       T       A       A       A       A       Y
Y       H       H       Q       Q       N       N       K       K       D       D       E
```

CODONS	TTT	TTC	TTA	TTG	CTT	CTC	CTA	CTG	ATT	ATC	ATA	GTT	GTC	GTA	GTG	TCT	TCC	TCA	TCG
Aminoacids	F	F	L	L	L	L	L	L	I	I	I	V	V	V	V	S	S	S	S
NC_000002.12:c135837169-135836530,c135833190-135833111,c135829676-135829593,c135824003-135823901,c135822098-135822020,c135818061-135817341,c135812956-135812311,c135809993-135808443,c135807396-135807128,c135805057-135804767,c135804128-135803930,c135800809-135800607,c135798138-135798029,c135794775-135794641,c135790881-135790658,c135789798-135789571,c135788544-135788324_Homo_sapiens_chromosome_2,_GRCh38.p14_Primary_Assembly	55	56	3	18	14	48	13	74	28	52	10	16	26	8	56	40	39	16	4

Output in tab delimited format. Use this output to copy and paste in your preferred application.

```
CODONS  TTT     TTC     TTA     TTG     CTT     CTC     CTA     CTG     ATT     ATC     ATA
GTT     GTC     GTA     GTG     TCT     TCC     TCA     TCG     AGT     AGC     CCT     CCC
CCA     CCG     ACT     ACC     ACA     ACG     GCT     GCC     GCA     GCG     TAT     TAC
CAT     CAC     CAA     CAG     AAT     AAC     AAA     AAG     GAT     GAC     GAA     GAG
TGT     TGC     CGT     CGC     CGA     CGG     AGA     AGG     GGT     GGC     GGA     GGG
ATG     TGG
AMINOACIDS      F       F       L       L       L       L       L       L       I       I
I       V       V       V       V       S       S       S       S       S       S       P
P       P       P       T       T       T       A       A       A       A       Y
Y       H       H       Q       Q       N       N       K       K       D       D       E
```

SEQUENCES \ PARAMETERS	length	A	C	T	G	%A	%C	%T	%G	%G+C	%G+A	%G+T	%A+T	%A+C	%C+T
NC_000002.12:c135837169-135836530,c135833190-135833111,c135829676-135829593,c135824003-135823901,c135822098-135822020,c135818061-135817341,c135812956-135812311,c135809993-135808443,c135807396-135807128,c135805057-135804767,c135804128-135803930,c135800809-135800607,c135798138-135798029,c135794775-135794641,c135790881-135790658,c135789798-135789571,c135788544-135788324_Homo_sapiens_chromosome_2,_GRCh38.p14_Primary_Assembly	5784	1397	1586	1312	1489	24.15	27.42	22.68	25.74	53.16	49.90	48.43	46.84	51.57	50.10

Output in tab delimited format. Use this output to copy and paste in your preferred application.

```
1397    1586    1312    1489    24.152835408022 27.420470262794 22.68326417704  25.740430152144
53.163900414938 49.896265560166 48.426694329184 46.836099585062 51.573305670816 50.103734439834
503     391     435     599     26.089211618257 22.56224066390  20.280082987552 31.06846473029
53.630705394191 57.157676348548 51.348547717842 46.369294605809 48.651452282158 42.842323651452
602     503     457     366     31.224066390041 23.703319502075 26.089211618257 18.983402489627
42.686721991701 50.207468879668 45.072614107884 57.313278008299 54.917385082116 49.792531120332
292     418     694     524     15.145228215768 35.995850622407 21.680497925311 27.178423236515
63.174273858921 42.323651452282 48.858921161826 36.829726141079 51.141078838174 57.676348547718
61.654199134199
```

CODONS	TTT	TTC	TTA	TTG	CTT	CTC	CTA	CTG	ATT	ATC	ATA	G
Aminoacids	F	F	L	L	L	L	L	L	I	I	I	
NC_079489.1:73799816-73801108_Spinacia_oleracea_cultivar_Varoflay_chromosome_3,_BTI_SOV_V1,_whole_genome_shotgun_sequence	1.12	0.68	0.00	1.62	2.25	1.50	0.12	0.50	0.94	1.29	0.77	2

Output in tab delimited format. Use this output to copy and paste in your preferred application.

```
CODONS  TTT     TTC     TTA     TTG     CTT     CTC     CTA     CTG     ATT     ATC     ATA
GTT     GTC     GTA     GTG     TCT     TCC     TCA     TCG     AGT     AGC     CCT     CCC
CCA     CCG     ACT     ACC     ACA     ACG     GCT     GCC     GCA     GCG     TAT     TAC
CAT     CAC     CAA     CAG     AAT     AAC     AAA     AAG     GAT     GAC     GAA     GAG
TGT     TGC     CGT     CGC     CGA     CGG     AGA     AGG     GGT     GGC     GGA     GGG
AMINOACIDS      F       F       L       L       L       L       L       L       I       I
I       V       V       V       V       S       S       S       S       S       S       P
P       P       P       T       T       T       T       A       A       A       A       Y
Y       H       H       Q       Q       N       N       K       K       D       D       E
E       C       C       R       R       R       R       R       R       G       G       G
```

CODON USAGE PER THOUSAND

CODONS	TTT	TTC	TTA	TTG	CTT	CTC	CTA	CTG	ATT	ATC
Aminoacids	F	F	L	L	L	L	L	L	I	I
_079489.1:73799816-01108_Spinacia_oleracea_cultivar_Varoflay_chromosome_3,_BTI_SOV_V1,_whole_genome_shotgun_sequence	32.48	25.52	0.00	30.16	41.76	27.84	2.32	9.28	25.52	34.80

Output in tab delimited format. Use this output to copy and paste in your preferred application.

```
CODONS  TTT     TTC     TTA     TTG     CTT     CTC     CTA     CTG     ATT     ATC     ATA
GTT     GTC     GTA     GTG     TCT     TCC     TCA     TCG     AGT     AGC     CCT     CCC
CCA     CCG     ACT     ACC     ACA     ACG     GCT     GCC     GCA     GCG     TAT     TAC
CAT     CAC     CAA     CAG     AAT     AAC     AAA     AAG     GAT     GAC     GAA     GAG
TGT     TGC     CGT     CGC     CGA     CGG     AGA     AGG     GGT     GGC     GGA     GGG
ATG     TGG
AMINOACIDS      F       F       L       L       L       L       L       L       I       I
I       V       V       V       V       S       S       S       S       S       S       P
P       P       P       T       T       T       T       A       A       A       A       Y
Y       H       H       Q       Q       N       N       K       K       D       D       E
```

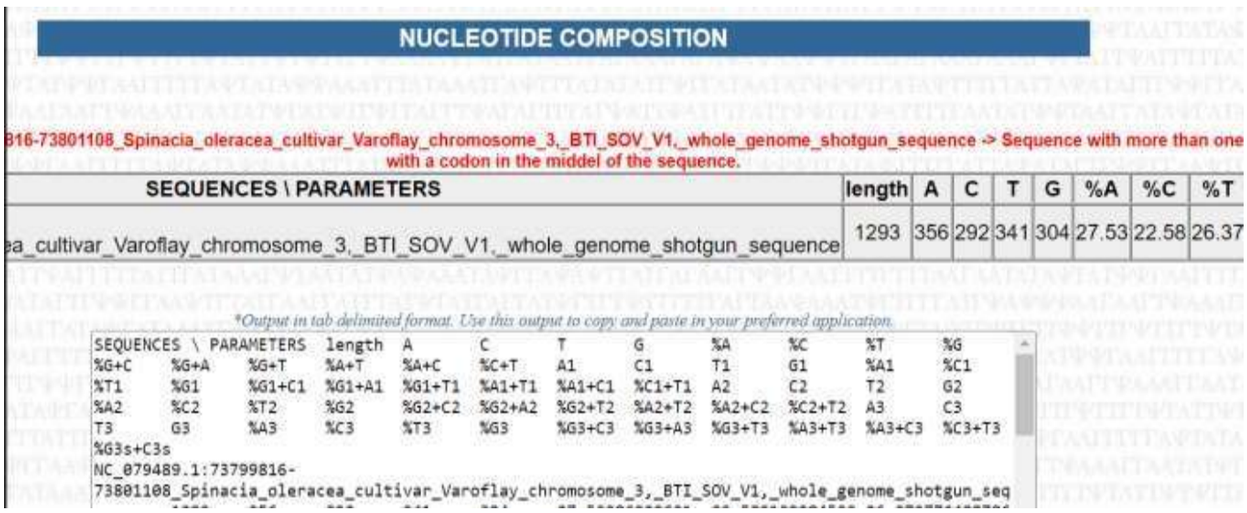
Relative Synonimous Codon Usage (RSCU)

CODON USAGE

CODONS	TTT	TTC	TTA	TTG	CTT	CTC	CTA	CTG	ATT	ATC	ATA	GTT
Aminoacids	F	F	L	L	L	L	L	L	I	I	I	V
79489.1:73799816-108_Spinacia_oleracea_cultivar_Varoflay_chromosome_3,_BTI_SOV_V1,_whole_genome_shotgun_sequence	14	11	0	13	18	12	1	4	11	15	9	15

Output in tab delimited format. Use this output to copy and paste in your preferred application.

```
CAT     CAC     CAA     CAG     AAT     AAC     AAA     AAG     GAT     GAC     GAA     GAG
TGT     TGC     CGT     CGC     CGA     CGG     AGA     AGG     GGT     GGC     GGA     GGG
ATG     TGG
AMINOACIDS      F       F       L       L       L       L       L       L       I       I
I       V       V       V       V       S       S       S       S       S       S       P
P       P       P       T       T       T       T       A       A       A       A       Y
Y       H       H       Q       Q       N       N       K       K       D       D       E
E       C       C       R       R       R       R       R       R       G       G       G
G       H       W
NC_079489.1:73799816-
```



NUCLEOTIDE COMPOSITION

816-73801108_Spinacia_oleracea_cultivar_Varoflay_chromosome_3,_BTI_SOV_V1,_whole_genome_shotgun_sequence -> Sequence with more than one with a codon in the middel of the sequence.

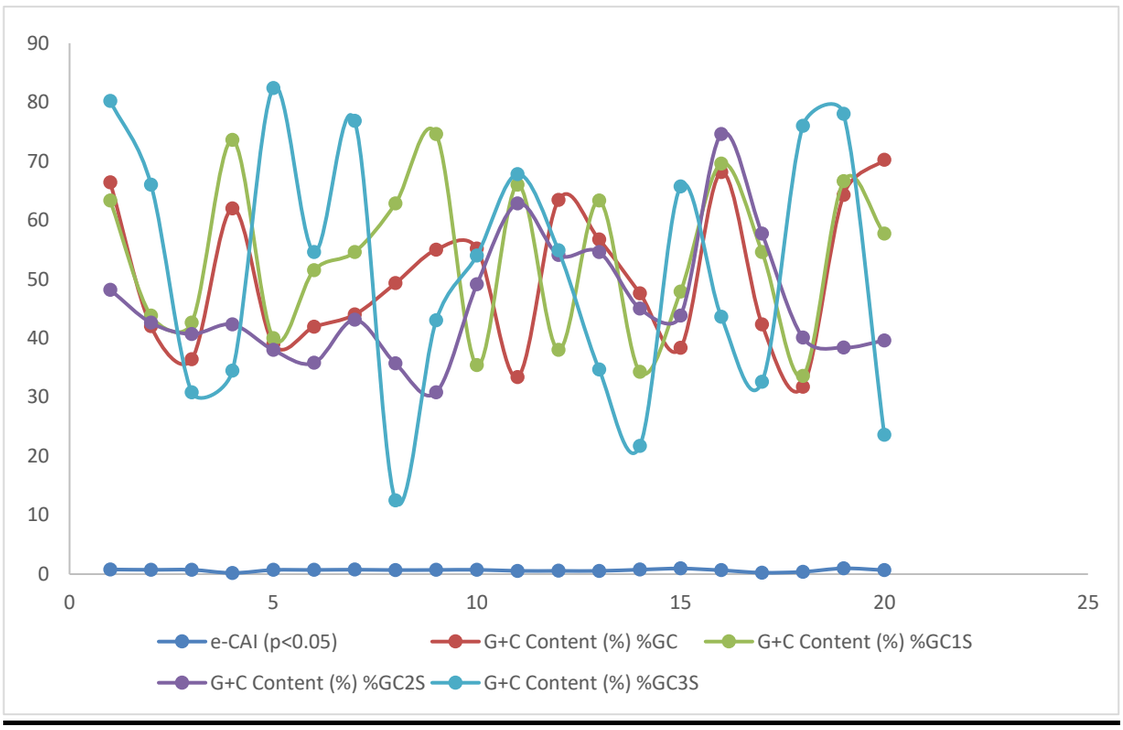
SEQUENCES \ PARAMETERS	length	A	C	T	G	%A	%C	%T
ea_cultivar_Varoflay_chromosome_3,_BTI_SOV_V1,_whole_genome_shotgun_sequence	1293	356	292	341	304	27.53	22.58	26.37

*Output in tab delimited format. Use this output to copy and paste in your preferred application.

SEQUENCES \ PARAMETERS	length	A	C	T	G	%A	%C	%T	%G
%G+C	%G+A	%G+T	%A+T	%A+C	%C+T	A1	C1	T1	G1	%A1	%C1
%T1	%G1	%G1+C1	%G1+A1	%G1+T1	%A1+T1	%A1+C1	%C1+T1	A2	C2	T2	G2
%A2	%C2	%T2	%G2	%G2+C2	%G2+A2	%G2+T2	%A2+T2	%A2+C2	%C2+T2	A3	C3
T3	G3	%A3	%C3	%T3	%G3	%G3+C3	%G3+A3	%G3+T3	%A3+T3	%A3+C3	%C3+T3
%G3s+C3s											
NC_079489.1:73799816-73801108_Spinacia_oleracea_cultivar_Varoflay_chromosome_3,_BTI_SOV_V1,_whole_genome_shotgun_seq											

Name of microorganisms :	e-CAI (p<0.05)	G+C Content (%)			
		%GC	%GC1S	%GC2S	%GC3S
Bacteroides fragilis	0.832(Average=0.786)	66.40	63.30	48.2	80.2
Escherichia coli	0.788(Average=0.731)	42.00	43.80	42.6	66
Faecalibacterium praunsitzii	0.637(Average=0.729)	36.40	42.60	40.7	30.8
Bifidobacterium longum	0.158(Average=0.195)	62.00	73.60	42.3	34.5
Akkermansia muciniphila	0.512(Average=0.701)	39.00	40.00	38	82.4
Lactobacillus acidophilus	0.577(Average=0.708)	41.90	51.50	35.8	54.6
Clostridium difficile	0.815(Average=0.754)	44.00	54.60	43.1	76.8
Ruminococcus bromii	0.761(Average=0.687)	49.30	62.80	35.7	12.5
Streptococcus thermophilus	0.782(Average=0.709)	55.00	74.60	30.8	43
Enterococcus faecalis	0.701(Average=0.738)	55.20	35.40	49.1	54
Roseburia spp.	0.687(Average=0.546)	33.40	66.00	62.8	67.8
Fusobacterium spp.	0.412(Average=0.540)	63.40	38.00	54.1	54.9
Parabacteriodes spp.	0.453(Average=0.541)	56.70	63.30	54.6	34.7
veillonella spp.	0.872(Average=0.759)	47.60	34.30	45	21.7
Eubacterium spp.	0.892(Average=0.949)	38.35	47.90	43.8	65.7
Blautia spp.	0.700(Average=0.649)	68.14	69.60	74.6	43.6
Methanobrevibacter smithi	0.234(Average=0.240)	42.30	54.60	57.7	32.6
Klebsiella spp.	0.382(Average=0.390)	31.76	33.60	40.1	76
Lactobacillus rhamnosus	0.958(Average=0.978)	64.28	66.60	38.4	78
Firmicutes	0.612(Average=0.670)	70.20	57.70	39.6	23.6

Table 4.1: Expected Codon Adaptive Index of different Micro-organisms.



The above provided graph have multiple lines representing different data series related to genetic content, specifically the G+C content at different positions in the genome, and an additional measure labelled "e-CAI (p<0.05)".

Here's what the borders in the graph represent:

1. **e-CAI (p<0.05)**:
    - This is represented by the blue line with circular markers.
    - e-CAI likely stands for a measure related to codon adaptation index, with the notation (p<0.05) indicating statistical significance.
- The values for e-CAI appear to be constant across the different positions or samples in the dataset.
2. **G+C Content (%) %GC**:
- This is represented by the red line with circular markers.
- It represents the overall G+C content percentage.
3. **G+C Content (%) %GC1S**:
- This is represented by the green line with square markers.
- It represents the G+C content percentage at the first codon position.
4. **G+C Content (%) %GC2S**:
- This is represented by the purple line with triangular markers.
- It represents the G+C content percentage at the second codon position.
5. **G+C Content (%) %GC3S**:
- This is represented by the cyan line with diamond markers.
- It represents the G+C content percentage at the third codon position.

Each line represents how the G+C content varies at different positions or under different conditions across the dataset. The borders of the markers help to differentiate between the different series of

data. The graph allows for a comparative visualization of these different measures of G+C content and the e-CAI across the samples or positions indicated on the x-axis.

CHAPTER 5

INTERACTION ANALYSIS

A database and online resource called STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) offers extensive data on protein-protein interactions (PPIs). Developed to facilitate the understanding of complex biological networks, STRING integrates known and predicted associations derived from multiple sources, including direct (physical) and indirect (functional) interactions. The interactions are sourced from curated databases, experimental data, computational prediction methods, text mining of scientific literature, and co-expression data.

The primary objective of STRING is to create a global perspective on PPIs, thereby enabling researchers to visualize and analyze the connectivity between proteins. By inputting a list of proteins, users can generate a network diagram that displays the interactions and relationships among these proteins. The resulting network is annotated with various metrics and confidence scores, which indicate the reliability of each interaction based on the source and type of evidence. Additionally, STRING offers functional enrichment analyses, which help identify biological pathways, processes, and molecular functions overrepresented in the network. These features make STRING an invaluable tool for hypothesis generation, functional annotation of proteins, and exploration of the molecular underpinnings of diseases.

Research:

Using STRING to study the association between gut microbiota and Alzheimer's disease involves several key steps, aimed at elucidating the molecular mechanisms underlying this complex relationship.

1. Data Collection and Preprocessing

To begin with, researchers collect gut microbiota samples from AD patients and healthy controls, followed by DNA extraction and metagenomic sequencing to identify the microbial genes and proteins present in these samples. After quality control and assembly, the sequences are annotated to predict protein-coding genes.

2. Identifying Differentially Expressed Proteins

Using functional genomics techniques, researchers identify differentially expressed proteins (DEPs) between AD and control samples. These DEPs represent the microbial proteins that may be implicated in the disease process.

3. STRING Database Analysis

The list of DEPs is then inputted into the STRING database. STRING generates a protein-protein interaction network, showing how these microbial proteins interact with each other and potentially with host proteins. The network includes known and predicted interactions, annotated with confidence

scores to reflect their reliability.

4. Functional Enrichment and Pathway Analysis

STRING's enrichment tools are used to identify biological pathways, processes, and molecular functions that are overrepresented among the DEPs. This helps in understanding the functional implications of the microbial proteins in the context of AD.

5. Comparative Analysis

Researchers compare the PPI networks between AD and control samples, looking for unique interactions or disrupted pathways in AD. This comparative analysis can highlight specific microbial proteins or interactions that might contribute to AD

6. Host-Microbiome Interactions

STRING can also be used to explore potential interactions between microbial proteins and host (human) proteins, providing insights into how gut microbiota might influence host cellular processes and contribute to neurodegeneration.

7. Visualization and Hypothesis Generation

STRING's visualization tools help create detailed maps of the interaction networks, allowing researchers to visually explore the connectivity and identify key nodes and interactions. These maps facilitate hypothesis generation regarding the roles of specific microbial proteins in AD.

5.1 Result of Interaction Analysis

- number of nodes:21
- number of edges:159
- average node degree:15.1
- avg. local clustering coefficient:0.902
- expected number of edges:26
- PPI enrichment p-value:< 1.0e-16

Fig5.1 (showing the nodes and edges)

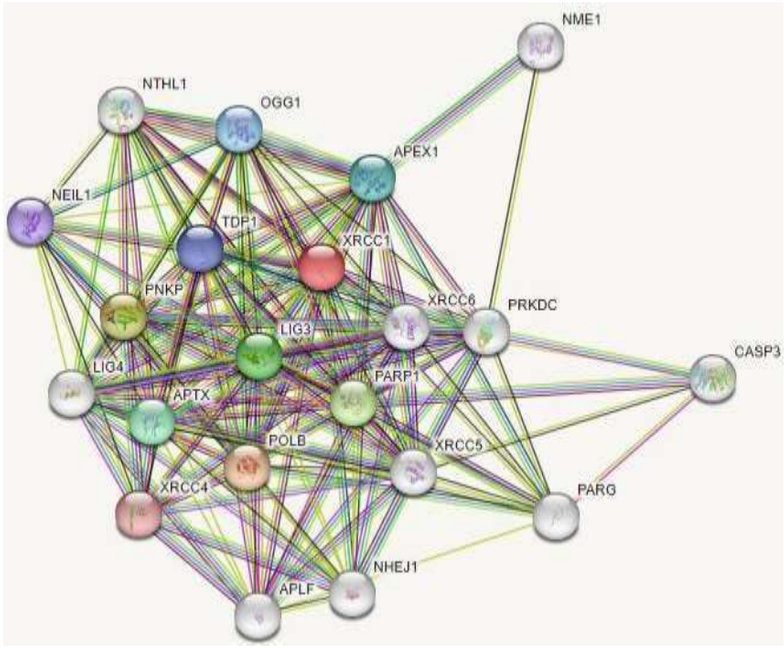


Fig.5.2 (Interacting network of XRCC1 protein obtaining through STRING database)

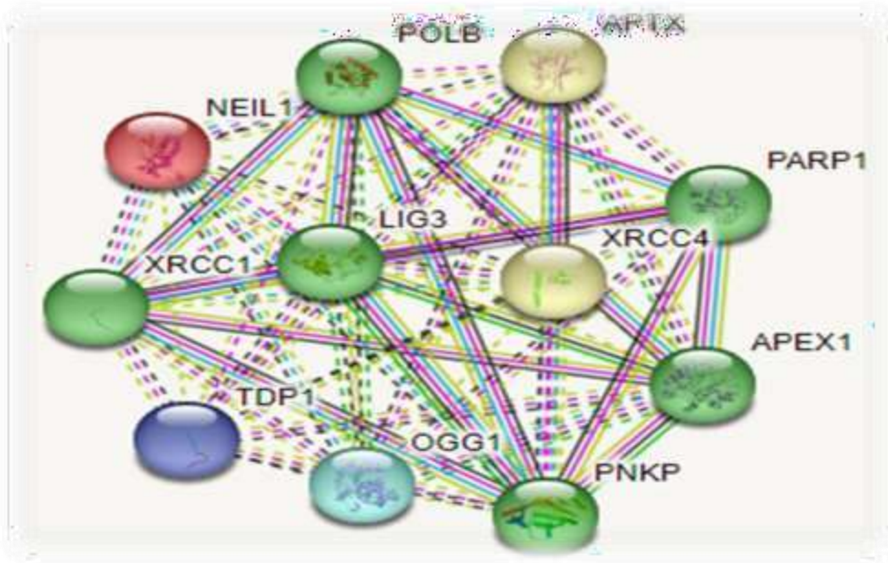


Fig.5.3 (Interacting network of XRCC1 gene)

bubble	cluster Id	gene count	protein names
	Cluster 1	1	NEIL1
	Cluster 2	2	APTX,XRCC4
	Cluster 3	6	APEX1,LIG3,PARP1,PNKP,POLB,XRCC1
	Cluster 4	1	OGG1
	Cluster 5	1	TDP1

Fig.5.4 (showing the clusters)

number of nodes:11
number of edges:55
average node degree:10
avg. local clustering coefficient:1
expected number of edges:11
PPI enrichment p-value:< 1.0e-16

Fig.5.5 (showing the nodes and edges)

CHAPTER 6

NETWORK MOTIFS AANALYSIS

FANMOD (Fast Network Motif Detection) is a bioinformatics tool used for identifying and analyzing network motifs in complex networks. Network motifs are small, recurring patterns of interconnections that occur significantly more often in a given network than in randomized networks. These motifs can represent fundamental building blocks of complex networks, such as biological, social, and technological networks. FANMOD is particularly efficient in detecting these motifs, making it valuable for understanding the underlying structural properties of large networks.

Compared to random networks, many biological networks include specific tiny subnetworks far more frequently. The term "network motifs" was created by Milo et al. (2002, 2004) who recommended using such an abundance of "topological modules" (Vespignanii, 2003) to reveal a structural design concepts of a biological networks. Three computationally intensive subtasks make up the task of finding network motifs: The second subtask has received some attention, while the other two have received far less up until lately. Kashtan et al. (2004) suggested a technique for sampling subgraphs to accelerate the first subtask. However, this technique has a number of shortcomings, including the fact that it only offers non-uniform sampling and performs badly as motif size grows; Wernicke (2005) gives a more thorough study of these issues [15].

1. Determining the number and kind of subgraphs present in the input network.

2. choosing the subsections that are isomorphic, or topologically identical, and dividing the resulting subsections in accordance.

3. figuring out whether subgraph classes are more common than random graphs in the context of a specific random graph model.

FANMOD is the network motif recognition tool which use the innovative RAND-ESUmethod (Wernicke, 2006) to sample & count subgraphs. The approach to orders magnitudes quicker than other one that has ever been used to perform this task, making it possible to find larger themes in larger networks than was previously feasible. In addition, FANMOD makes it feasible to recognise motifs in coloured networks, which is not achievable with other available tools. There are around 700 lines of non-library code in the C++ programme that creates the FANMOD utility. The wx WIDGETS framework (Smart et al., 2006) , and it is available for a wide range of platforms including Linux, Mac OS, and Windows, is used to create the GUI along with system dependent functionality. Up to eight vertices in size, FANMOD can recognise network patterns. Using the technique outlined by Wernicke (2005), all subgraphs of specified size can be either enumerated or evenly sampling in the input network. The subgraphs are categorised into isotropic subgraph classes using a rendition of the common graph-labeling technique NAUTY (McKay, 1981). FANMOD determines the frequency billion subgraph classes in a certain amount of random graphs.The user may select from a variety of switching techniques in order to retain specific graph attributes while generating random graphs by shifting edges between vertices in the original network. Depending on the amount of colours used for

edges and vertices, motifs of up to seven vertices in size can be discovered in coloured networks. Colours have no impact on how quickly the tool runs; in fact, because canonical graph labelling is made easier, it runs a tiny bit quicker overall. The amount of edges between vertices with various colours can be preserved by random networks at an optional step. You may export the determined relevance of each subgraph in the network in a number of different forms. A rapid review and sharing of results is made possible by an HTML export function with a number of filters [16].

6.1RESULTS OF FANMOD

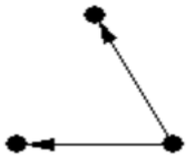
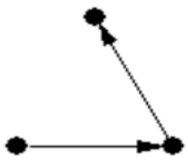
ID	Adj	Frequency [Original]	Mean-Freq [Random]	Standard-Dev [Random]	Z-Score	p-Value
6		31.933%	32.415%	0.0044633	-1.0805	0.789
36		16.807%	17.422%	0.0057237	-1.0743	0.768
12		49.58%	48.936%	0.010292	0.62581	0.464

Fig.6.1 Showing results of FANMOD tool

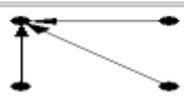
ID	Adj	Frequency [Original]	Mean-Freq [Random]	Standard-Dev [Random]	Z-Score	p-Value
204		2.439%	0.058383%	0.0013998	17.007	0
2182		2.8455%	0.1647%	0.0023097	11.607	0
2118		2.439%	0.48131%	0.0052171	3.7525	0.002
14		10.569%	7.7693%	0.007777	3.6001	0
28		11.382%	8.8461%	0.0075632	3.353	0
2184		2.8455%	1.9444%	0.0035636	2.5288	0
536		21.951%	17.457%	0.022211	2.0232	0.023

Fig. 6.2 Showing results of FANMOD tool

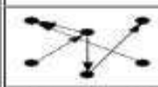
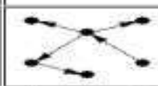
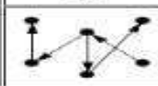
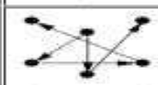
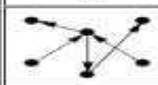
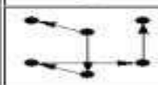
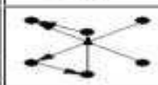
ID	Adj	Frequency [Original]	Mean Freq [Random]	Standard Dev [Random]	Z-Score	p-Value
4200488		2.2533%	1.0089%	0.0046279	2.6889	0.006
541130832		1.1655%	0.54203%	0.0023466	2.658	0.008
32884		1.2432%	0.66313%	0.0022192	2.6139	0.006
4325452		1.6317%	0.73836%	0.0035165	2.5404	0.018
4204556		3.4188%	2.0099%	0.0057187	2.4636	0.017
4198504		1.2432%	0.68039%	0.0023521	2.3928	0.009
1057932		3.4188%	2.0115%	0.0062201	2.2624	0.015
67405856		1.8648%	0.92903%	0.0042103	2.2249	0.028
8404016		0.6216%	0%	0	undefined	0
8402456		1.2432%	0%	0	undefined	0

Fig. 6.3 Showing results of FANMOD tool

CHAPTER 7

CONCLUSIONS

In conclusion, this thesis has undertaken a comprehensive exploration of computational studies and in vitro validations of medicinal compounds for Alzheimer's disease (AD), leveraging advanced tools and methodologies. The integration of computational approaches with experimental validations offers a powerful strategy for drug discovery and development in the context of AD.

The utilization of computational tools such as STRING and Fanmod has been instrumental in elucidating the complex molecular mechanisms underlying AD pathology. STRING, a database for protein-protein interactions, facilitated the investigation of the intricate networks of molecular interactions implicated in AD pathogenesis. By analyzing protein-protein interactions, STRING provided valuable insights into the molecular pathways involved in AD, aiding in the identification of potential therapeutic targets.

Additionally, the use of Fanmod, a tool for identifying functional modules within protein interaction networks, further refined our understanding of the interconnected pathways and biological processes associated with AD. Fanmod enabled the identification of cohesive groups of proteins within the AD interactome, shedding light on key regulatory mechanisms and potential intervention points for therapeutic development.

Furthermore, the incorporation of codon bias analysis into the study provided additional layers of insight into AD pathophysiology. Codon bias analysis offers a novel perspective on gene expression regulation, evolutionary adaptations, and functional genomics, which are critical aspects in the study of complex diseases like AD. Understanding codon usage patterns can inform gene expression strategies and optimize the design of therapeutic interventions, thereby enhancing the efficacy of medicinal compounds targeting AD-related pathways.

By integrating computational studies with in vitro validations, this thesis has advanced our understanding of AD and laid the groundwork for the development of novel therapeutic strategies. In the end, this could mean a great deal for the millions of people affected by Alzheimer's disease, as the synergistic combination of computational tools, experimental techniques, and insights from codon bias analysis holds great promise for speeding the discovery and development of effective treatments for the debilitating condition.

Reference

[1] Huang, Z., Liu, K., Ma, W., Li, D., Mo, T., & Liu, Q. (2020). The gut microbiome in human health and disease—Where are we and where are we going? A bibliometric analysis. *Corresponding Author*.

[2] A. Vijay and A. M. Valdes, "Role of the gut microbiome in chronic diseases: a narrative review," J. Clin. Invest.vol. 131, no. 1, pp. 1-12, Jan. 2021

[3] Y. Chen, J. Zhou, and L. Wang, "Role and Mechanism of Gut Microbiota in Human Disease," in *Front. Cell. Infect. Microbiol.*, vol. 10, p. 586, 2020.

[4] C. A. Lozupone, J. I. Stombaugh, J. I. Gordon, J. K. Jansson, and R. Knight, "Diversity, stability and resilience of the human gut microbiota," in *Nature*, vol. 489, no. 7415, pp. 220-230, 2012, doi: 10.1038/nature11550.

[5] H. J. Flint, K. P. Scott, P. Louis, and S. H. Duncan, "The role of the gut microbiota in nutrition and health," Nat. Rev. Gastroenterol. Hepatol., vol. 9, no. 10, pp. 577–589, Oct. 2012, doi: 10.1038/nrgastro.2012.156.

[6] Y. J. Kim, J. W. Park, and J. K. Lee, "Influence of diet on the gut microbiome and implications for human health," J. Microbiol. Biotechnol., vol. 30, no. 6, pp. 1–10, Jun. 2020, doi: 10.4014/jmb.2002.02024.

[7] T. V. S. Rinninella, E. Raoul, and P. Cintoni, "What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases," Microorganisms, vol. 7, no. 1, pp. 1–19, Jan. 2019, doi: 10.3390/microorganisms7010014.J. Hardy and D. J. Selkoe, "The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics," *Science*, vol. 297, no. 5580, pp. 353–356, 2002. https://doi.org/10.1126/science.1072994

[8] E. M. Reiman et al., "Declining brain activity in cognitively normal apolipoprotein E epsilon 4 heterozygotes: A foundation for using positron emission tomography to efficiently test treatments to prevent Alzheimer's disease," Proc. Natl. Acad. Sci. U.S.A., vol. 98, no. 6, pp. 3334–3339, 2001, doi: 10.1073/pnas.061509598.

[9] M. Cattaneo et al., "Association of brain amyloidosis with pro-inflammatory gut bacterial taxa and peripheral inflammation markers in cognitively impaired elderly," Neurobiol. Aging, vol. 49, pp. 60–68, 2017, doi: 10.1016/j.neurobiolaging.2016.08.019.

[10] T. E. Quax et al., "Codon bias as a means to fine-tune gene expression," Mol. Cell, vol. 59, pp. 149–161, 2015, doi: 10.1016/j.molcel.2015.05.035.

[11] Q. P. Ma et al., "Analysis of synonymous codon usage in FAD7 genes from different plant species," Genet. Mol. Res., vol. 14, pp. 1414–1422, Feb. 2015, doi: 10.4238/2015.February.13.20.Liu, Y., "A code within the genetic code: codon usage regulates co-translational protein folding," *Cell Commun. Signal.*, vol. 18, p. 145, 2020. https://doi.org/10.1186/s12964-020-00642-6

[12] S. A. Shabalina, N. A. Spiridonov, and A. Kashina, "Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity," Nucleic Acids Res., vol. 41, pp. 2073-2094, 2013, doi: 10.1093/nar/gks1205.

[13] H.-C. Wang and D. A. Hickey, "Rapid divergence of codon usage patterns within the rice

genome," BMC Evol. Biol., vol. 7, Suppl. 1, p. S6, 2007, doi: 10.1186/1471-2148-7-S1-S6.

[14] F. Chamani Mohasses, M. Solouki, B. Ghareyazie, L. Fahmideh, and M. Mohsenpour, "Correlation between gene expression levels under drought stress and synonymous codon usage in rice plant by in-silico study," PLoS ONE, vol. 15, e0237334, 2020, doi: 10.1371/journal.pone.0237334.Bourret, J., Alizon, S., and Bravo, I. G., "COUSIN (COdon usage similarity INdex): a normalized measure of codon usage preferences," *Genome Biol. Evol.*, vol. 11, pp. 3523–3528, 2019. https://doi.org/10.1093/gbe/evz262

[15] P. Mazumdar, R. Binti Othman, K. Mebus, N. Ramakrishnan, and J. Ann Harikrishna, "Codon usage and codon pair patterns in non-grass monocot genomes," Ann. Bot., vol. 120, pp. 893-909, 2017, doi: 10.1093/aob/mcx112.

[16] R. Manikandan, N. Balakrishnan, D. Sudhakar, and V. Udayasuriyan, "Transgenic rice plants expressing synthetic cry2AX1 gene exhibits resistance to rice leaf folder (Cnaphalocrocis medinalis)," 3 Biotech, vol. 6, p. 10, 2016, doi: 10.1007/s13205-015-0315-4.

[17] Y. Batard, A. Hehn, S. Nedelkina, M. Schalk, K. Pallett, H. Schaller, and D. Werck-Reichhart, "Increasing expression of P450 and P450-reductase proteins from monocots in heterologous systems," Arch. Biochem. Biophys., vol. 379, pp. 161-169, 2000.

[18] Z. Cui, W. A. Johnston, and A. Kirill, "Cell-free approach for non-canonical amino acids incorporation into polypeptides," Front. Bioeng. Biotechnol., vol. 8, p. 1031, 2020, doi: 10.3389/fbioe.2020.01031.D. Szklarczyk et al., “STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets,” Nucleic 31 Acids Res., vol. 47, no. D1, pp. D607–D613, 2019.

[19] D. Szklarczyk et al., "Correction to 'The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets,'" Nucleic Acids Research, vol. 49, no. 18, p. 10800, 2021.

[20] D. Szklarczyk et al., "The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets," Nucleic Acids Research, vol. 49, no. D1, pp. D605-D612, Jan. 2021.

[21] Wernicke and F. Rasche, “FANMOD: a tool for fast network motif detection,” Bioinformatics, vol. 22, no. 9, pp. 1152–1153, 2006.

[22] M. Megraw, S. Mukherjee, and U. Ohler, "Sustained-input switches for transcription factors and microRNAs are central building blocks of eukaryotic gene circuits," Genome Biol., vol. 14, no. 8, Art. no. R85, 2013.

[23] J. Gubatan, T. L. Boye, M. Temby, R. S. Sojwal, D. R. Holman, S. R. Sinha, S. R. Rogalla, and O. H. Nielsen, "Gut Microbiome in Inflammatory Bowel Disease: Role in Pathogenesis, Dietary Modulation, and Colitis-Associated Colon Cancer," [Journal Name], [Volume], [Issue], [Page range], [Year]. DOI or URL if available.

[24] Z. Breijyeh and R. Karaman, "Comprehensive Review on Alzheimer’s Disease: Causes and Treatment," Molecules, vol. 25, no. 24, Art. no. 5789, 2020, doi: 10.3390/molecules25245789.

[25] N. M. Nguyen, J. Cho, and C. Lee, "Gut Microbiota and Alzheimer’s Disease: How to Study and Apply Their Relationship," MDPI, vol. 24, no. 4, Art. no. 4047, 2024. [Online]. Available: https://www.mdpi.com/1422-0067/24/4/4047. [Accessed: 18-May-2024].

[26] M. T.-H. Truong, H.-P. Pham, T. T.-L. Nguyen, P.-N. Tran, D. M.-H. Nguyen, D. T.-C.

Nguyen, T. V.-H. Pham, H. Q. Nguyen, N. H.-T. Nguyen and L. Q.-S. Pham, "Insights into the Antimicrobial Resistance and Virulence of Edwardsiella ictaluri: A Systematic Analysis Using Genomic and Phenotypic Data," Microorganisms, vol. 10, no. 7, Article 1371, pp. 1-17, Jul. 2022. DOI: 10.3390/microorganisms10071371.

[27] R. Pourdarbani, S. S. Hosseini, E. Hosseinzadeh, M. Mehrara, S. Masoum and S. M. Nabavi, "Applications of Machine Learning in Targeting Protein-Protein Interactions with a Focus on Small Molecules," Molecules, vol. 25, no. 24, Article 5789, pp. 1-21, Dec. 2020. DOI: 10.3390/molecules25245789.

[28] S. Rodrigues, A. Vilela, A. Oliveira and J. M. S. Rocha, "Machine Learning in Yeast Industrial Biotechnology: Applications, Challenges, and Future Perspectives," International Journal of Molecular Sciences, vol. 24, no. 4, Article 4047, pp. 1-28, Feb. 2023. DOI: 10.3390/ijms24044047.