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

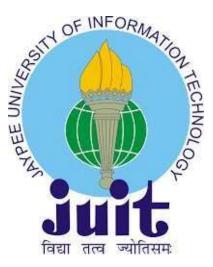
Dissertation submitted in partial fulfilment of the requirement for the degree of

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

CONTENT	Page Number
Declaration	iii
Supervisor's Certificate	iv
Acknowledgement	v
List of content	vi
List of Tables	vii
List of Abbreviation's	viii
Abstract	1
Aims and objectives	
Chapter 1	1-2
Introduction	
1.1 Background	2
1.2 Pathogenesis	3-4
Chapter 2	5-7
General Alzheimer's Disease	
2.1 Introduction	5
2.2 Association with Alzheimer's Disease	6-7
Chapter 3	8-17
Results	
3.1 MSA and Phylogenetic tree	11
3.2 Phylogenetic tree	14
3.3 Diagnosis and Treatment	16
3.4 Treatment	16
Chapter 4	18-27
Codon Bias Analysis	
4.1 Factor affecting Codon Bias analysis	19
4.2 Software's for Codon usage analysis	20
4.3 Applications	21
4.4 Results	22
Chapter 5	28-32
Interaction Analysis	
5.1 Result of interaction analysis	28
Chapter 6	33-35
Network motif's Analysis	
6.1 Results of Network motifs analysis	33
Chapter 7	35
Conclusion	36
References	37-39

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 (4th semester) Jaypee University of Information Technology Solan, HP

LIST OF FIGURES

Figure	Page Number
Fig.1.1 Pathogenesis of gut microbiota	4
Fig.2.1 Comparison between normal and abnormal Alzheimer's disease brain	5
Fig.2.2 Gut microbiota and its association with Alzheimer's disease	7
Fig3.1 Phylogenetic tree	15
Fig.3.2 Phylogenetic tree shows some microorganisms present in gut microbiota	16
Fig.4.1 Genetic code	19
Fig.5.1 Showing the nodes and edges	29
Fig.5.2 Interaction network of XRCC1 protein obtaining through string database	29
Fig.5.3 Interacting network of XRCC1 gene	30
Fig.5.4 showing the cluster	30
Fig.5.5 showing the nodes and edges	30
Fig. 6.1 Showing results of network motifs analysis	33
Fig 6.2 Showing results of network motifs analysis	33
Fig. 6.3 Showing results of network motifs analysis	34

LIST OF TABLES

Table	Page Number
Fig.3.1 20 microorganisms present in gut microbiota	9
Fig.4.1 Genetic Code	26

LIST OF ABBERATIONS

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].

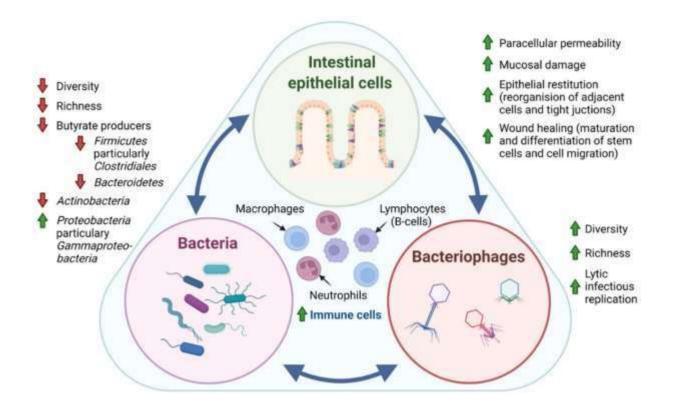


Fig 1.1: Pathogenesis of Gut Microbiota

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 (A β), 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].

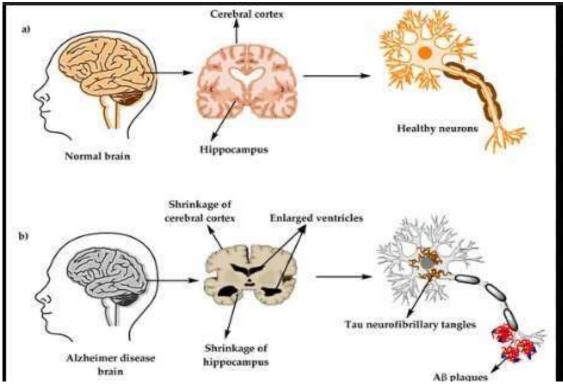


Fig.2.1 : comparison between normal and abnormal Alzheimer's disease brain

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: $A\beta$ deposition is a hallmark feature of AD pathology, contributing to synaptic dysfunction and neuronal toxicity. Recent studies have implicated gut dysbiosis in promoting A β accumulation in the brain, with alterations in gut microbial composition and function influencing A β production and clearance mechanisms. Dysbiotic gut microbiota may exacerbate A β 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].

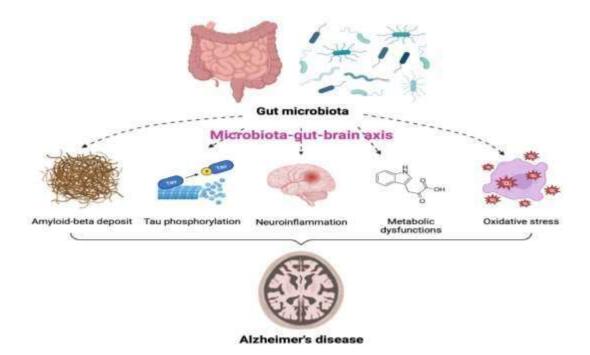


Fig.2.2 : Gut microbiota and its association with Alzheimer's disease

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.
- \checkmark Gene and protein level analysis (sequence and structure
- ✓ based studies)
- \checkmark Functional and evolutionary annotations

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	
Klebsiell <mark>a sp</mark> p.	"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	

20 microorganisms present in gut microbiota

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.

CUSTAL format	alignment by MAFFT FFT-HS-1 (V7.487)	XF_003428839.1	MP - PSGLALLP-LLLPLLRLLVLTPGRPAAGLSTCKTIDMELVKRRHJEATRGQTLSKI
C. 100010-10100100	and and the second second second	HP_001075318.1	MPPSGLILLEP-LLLPLLWLLVLTPGRPAAGLSTCKTTDHELWKRKRIEKINGQTLSKL
		XP_016791534.1	MPPSGLRLEP-LLEPLEWELVETPGRPAAGESTCKTIDPELVKRKBIEAERGQIESKE
MP 000651.3	HPPSGLRLLP-LLLPLLWLLVLTPGRPAAGLSTCKTICMELVKRKRTEATR601LSHL	XP_011042237.1	MPPSGURLEP-LLUPLENLEVETPGRPWAGESTCKTIDMELVKRKRIEATRGQTESKL
NP 814979511.1	MPPSGCRLLP-LLLPCLWLLVLTPGRPAAGESTOCTIONELWCRERTEATROOTESKL	XF_054390445.1	NPPSGLILLP-LLLPLLBLLVLTPGHPMAGLSTCKTIDPELVKRKHIEAERGQEL5KL
NF 854322857.1	#P - PS#LR: LP - LLLP: LNLLVLTPURPAGESTOCTTOMELVRRIVETEAT#601L581	XF_012891536.1	MPPSGLRLLPLLLLPLPWLLVLTPGRPAAALSTCKTIDHELVKRKRIEALRGQELSKI.
NP 019870698.1	NP PSGLRLLP . LLLP: LWLLVLTPGRPAAGESTCKTTOMELVKRKRTEAT8G01LSKL	IP_901305385.1	NDPSPLLALLILLG-AARALSTOQREDLEAAKKIRIEAVROQTESKE
WP 824782265.2	PPPSGLRLLP-LLLPLLWLLVLTPSRPAAGLSTOCTIONELVCRCRTEATROOTLSKL	XF_030075415.1	MEV-PROLIVELS-VELOMEAVVESLSTCQTIDMEEVROKKRIEAIR6QIESKE
XP #38512824.1	MP PSGERLEP-LEEPLERLEVETPMEPRAGESTERTTEMEEVKREKTEATROODESKE	XP_059574784.1	MGRGPW(P-GAVAVAVAVALCAA-AAAALSTCRSVDLEAARBRRTEAVRGQTLSK)
MP 005041294.1	NP PSGCREEP - LEEPELWEEVE TPSRPWAGESTOK TEOMELWORK REPARATE STE		a a 1 1 'assal Tala '115 assissment
MP 826338072.1	INP PSGLHLLP . LLLPLLWLLVI TPGRPAAGLSTOKTTONEL WORKTEATRIODUSEL		
NP 807875376.1	APPSGLRE1P-LLEPSENLLVLTPSRPARESTCKTTENELVCKKRTEATNOOLSNL	NP_000631.1	IILASPPSQGEVPPOPLPEAVLALVIISTRORVAGESAEPEPEPEADYYAKEVTII
NP 042568006.1	HPPSRLRLIP-LLLPCINCUVETPORPANOLSTOCTION/CVRRRETCATIGOTIS/L	XP 014079511.1	RLASPP - SQGEYPPGPLPEAVLALWISTRORY AGESAEPEPEPEADYYAKEVTR
XP 8559006487.1	MPPSGCRLIP-LLLPLINLL9LTPGRPAAGESTERTIONELVERKTITATIOODLSRL	XF 054323057.1	RLASPPSQGEVPPGPLPEAVLALVISTRORYAGESAEPEPEPEADHYAKEVTR
30 831700388.1	INPPSGEREEP-LEEP-LEEP-LIKELVEPNGRAMGESTOCTIONEEV/GROOTEATIND/01-SKE	XP 018370638.1	RLASPPSQGEVPPRPLPEAVLALWSTRDRYAGESAEPEPEPEADYYAKEVTR
MP 859851830.1	MPPSG-RLLP-LLLPLTHLLVLPPGRPANG STORTISMELVICRICITEATRICULSIC	XP 824782266.2	REASPP - SQGEVPPGPL/FEAVLAL WISTRORY AGESAEPEPEPEADYYAKEVTR
XP 822441296.1	IPPSGLRLLP-LLLPCLWLLVLPPGRPAWELSTCKTIONELVKPXRIEAT8600LSRL	XP #38512824.1	IILASPPSQGEVPPQPLPEAVLALWSTRORVAGESAEPEPEPEADYVAKEVTR
MP 0239886445.1	IP-PSGLRLIP-LLLPLINLLVEPPGRFANELSTORTIONELVRBORIEALRGOSESNE	XF 086941294.1	IILASPP SQGEVPPEPLPEAVLALWSTRDRV AGESAEFEFEFEFEADYYAKEVTR
XF 059988434,1	PPPSELRULP-LLLPLUNLLVLPPERPARELSTORTIONELVKRKRIEAIREQ3ESHL	XP 026338022.1	IILASPP-SOGEVPPOPLPEAVLALVIISTRORYAGESAEPEPEPEADIVYAKEVTR
NP_004271257.1	PPPSGLRLLP-LLLPULWLLVLPPGRPAAGLSTOCTIONELVKRKRTEATRGOTESKL	XP 827475376.1	RLASPPSOGEVPPGPLPEAVLALVINSTRORYAGESAEPEPEPEPERDVYAKEVTR
xF_826935427.1	PPPSGLRLLP-LLLPLLWLLVLPPGRFAAGLSTCKTIDMELWKROETEATRODUSKL	XP 847568806.1	RLASPPSOGEVPOGPLPEAVLALVNSTRDRVAGESAEPEPEPEADYVAKEVTR
MP 999130,2	MPP56LRLLP-LLLPLLWLLVLTPGRP4AGESTCKTIOMELVKRKRIEATROOSESKL	XF #55006887.1	III.ASPPSOGEVPPRPLPEAVLALYINTRONVAGESAEPEPEPEADYVAKEVTR
MP 001809400.1	MPPSGLRLLP-LLLPLINLINLTPGRPWAGLSTOCTIONELVKRKGIEATROQILSKL	XF 033700390.1	ILASPE SOGEVIPOPLPEAVLALVISTRORY AGESAEPEPEPEADYYAKEVTI
MP_001301071.1	HPPSGLRLLP-LLLPLLWLLHLTPGRPVAGLSTCKTEDHELVKRKRTEATROQDESKL	XP 050855838.1	ILLASOP SOCEVPPOPLPEAVLINLY/STRONV
30 605898676.1	MP - PSELFLLP - LLLPLINLIML TPGRPVAGESTCK TEDMELVKRKRTEAT86Q1LSRL	XF 022441296.1	IILASPP-SOGEVPPGPLPEAVLALVISTRORVAGESAEPEPEPERDVYAKEVTIL
NP_883755182.1	MPPS618LLP-LLLPV1WLLML1PGRPWAGLSTCKTTDMELVK9KRTEAE86QGLSKL	XP 023988845.1	RLASPPSOGEVPPOPLPEAVLALWSTRDAYAGESAEPEPEPEADYYAKEVTR
XP_000086313,1	MPPSGLRLIP-ILLPLINLLVLTLGRPANGLSTCKTTDMELVRRKRTEATROOLSAL	8P 059988434.1	ILASPPSOGEVPPOPLPEAVLALWSTRORVAGESACPEPCPEADVYAKEVTR
MP_053461166.1	MPPSRLRLP-LLLPCINLLVLSPGRPAAGLSTCKTIDMELVRIKKTEATRIQULSUL	XP 604271297.1	RLASPP - SQGEVPPGPLPEX/LALVINSTROOV AGESAEPEPEPEADYYAKEVTR
NP_0004679922.4	MPPS6LRL1P-LLLPLURLLVLTPG8PAAGLSTCKTTOMELVRRKKTEATR6QTLSRL	XP 026935427.1	ILLASPPSQLEVPROPLIPEAVLALVIISTRONVAGESAEPEPEPEADYYAKEVTI
MP_835702.3	MPPSGLRLIP-LLLPUPWLLVLTPSRPAWGLSTCKTIDMELVRRKRTEATAGQTLSRL	NP 999188.2	RLASPFSOGWPPGPLPEAVLALWISTRORYAGESVEPEPEPEADYVAKEVTR
MP_067589.5	MPPSGLALLP-LLLPUMILLVLTPGRPAAGLSTEXTIONELWORKTEADAG2LSKL	NP 081009400.1	ILASPP SOGUVPPOPLPEATLALVISTICALY AGESAETEPEPEADYYAKEVTR
XP_032750045.1	PPPS6LRLLP-LLLPLPw(LVLTPGRPAAGLSTOCTIDMELVORCRIEATROQUESE)	NF 001301071.1	RLASPPSOGOVPOSPLPEATLALVISTRORYAGESAETEPEPEADYXACEVTR
10/_052824653.1	mpPSGLRLLP-LLLPCPWLLVLTPGRPAAGLSTCXTTDMELVKRKRTEATROGSLSRL	XP 001301074.1	REASPP - SOGDVPPGPLPEATEALWSTRORY - AGESAETEPEPEADYYAKEVTR
MP_827286179.1	IPPSGLRLLP-LLLPUIWLLVLTPSRPAVGLSTOKTIONELVKRIKITEATRGESLSRL	XP 943755182.1	REASEPT SQUEEPERENT ACTIVITIES THAT AND A SALESSE EXPERSION AND A SALESSE EXPE
30,030590613.1	PPPSGLRLLP-LLLPCPWLLVLTPGRPAWGLSTCKTTDMELVKRKRTEACR6Q1L58L		ILASPP - NOAEVPPOPLPEALLALING TROPY AGESAE REPEPEADING TAKEN TH ILASPP - NOAEVPPOPLPEAVLALING TROPY AGESAE REPEPEADING AKEN TR
MP_003106494,1	MPPSHERLEP-LEEPCENELVEAPGRPASSESTORTIONELVERKREEAERGQUESKE	XP 003461166.1	ILISPE INVESTIGATION AND A CONTRACT
XP_004648529.1	MPPSBLRLLP-LLLPLLWLLVLTPGRPASGLSTCKTIOMELVKRKRIEATBGQDLSKL		
NP_046311618.1	MPPSGLRULP-LLLPLLNLLVLTPGRPVAGLSTCKTIDMELVKRKRJEATAGQTLSRL	XP_004479972.4	RLASPP SQGEVPPGPL/PEAVLALVMSTRORVAGESAQPEPEPEADVYAKEVTR
XP_003523548.1	MPPSGLRLLP-LLLPLLWLLVLTPGRP4AGLSTCKTIDMELVKRKRIEAT86Q1ESHL	WP_035707.1	RLASPP - SQGEVPPGPLPEAVLALYNSTRORY AGESADPEPEPEADYYAKEVTR
XF 024433172.2	HP-P561RLIP-LLEPETHELVERPERPARESTCKTEEMELVERPERFERENDED	IP_067589.1	IILASPPSQGEVPPGPLPEAVLALYNSTRORVAGESADPEPEPEADYYAKEVTR

XP_032750845.1	RLASPO-SOGEVPPOPLPEAVEALVINSTRURYAGESADPEPEPEADVYAREVTR am	001009430.1	VLMVEYGRKTYDK/KSSSR/STYNFFH/TSELREA//VEPVLLSRAD/HLIELKLKV
XP 852024653.1		The second s	VLMEVLNCTVDK/RSSSHSLTMEFRITSELREAVEPVLLSRAAULINLKLV VLMVEVLNCTVDK/RSSSHSLTMEFRITSELREAVEPVLLSRAELINLLNLKLV
NP 827285179.1			
NP 848598613.1	with the control of the sine repair of the second s	M. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.	VLMVEYGNKTYDIOWKSSSHSTYMFFNTSELREAVPEPVLLSBAELALLALKLKV
MP 001166494.1	and a second		VLHVEVGRKTYER/WSNSHSTYHFFRTSELBEAVPEPVLLSRAELALLALKLKV
NP 004648520.1		Marcold Andrews and	VFINETRHOEYDKFRHSTHISTWIFFRITSELREAMPERALLISRAELBLLIRLIKLTA
AP 840311018.1	and the second s		VLMVETMNKTYDRFKDSPHSTWWFTNTSELREAVSDPLLLSRAELRLUULKLXV
3P 053523548.1	In Action and approximation of source and the set of th	•• ***********************************	VLMVEIMAKKVYEKITPSQHSIVMLFNTSELREAVPEPVLLSQAELRLLALKLKE
V0+0020212222		035707.1	VLMVDRMMATYEKTKDISHSEVVFFNTSDIREAVPEPPLLSRAELELQBLXSSV
NP_024433172.2	HLASPPSQAEVPPGPLPEAVLALVIISTRORYAGESACPEPEPETDYVAKEVTA	067589.1	VLMVDRINATYOKTKOTTHSTYV#FNTSDTREAVPEPPLLSRAELBLQRFKSTV
3P_093429839.1	HLASPPNQCEVPPNOLPEAVLALINISTRDQVAGESAEPEPEPEADYVAKEVTA XP	032750046.1	VLMVDIMMATYOKTKOLTHISTYNF FIITSDIREAWPEPPLLSRAEURLORFKSTV
MP_001075318.1	HLASPP SQGEVPPGPLPEAVLALVHSTRAQV AGESAETEPEPEADYNRKEVTH XP	052024653-1	VLM/DRWNATYDKTKDTPH/STYWYFR/TSDTREAVPEPPLLFRAELRLQ8FKSSV
XP_816791534.1	RLASPP-SQGEVPPOPLPEAVLALYRSTRURVAGESAEPEPEPEADYYAKEVTR	027286179.1	VLMVDR5NPTYOKTKDEPH/SVYMFFN/TSDEREAVPEPPLLSRAELBLOBFKSSV
NP_811942237.1	RLASPPSXEVPPGPLPEAVLALVIISTRORVMESAEPEPEPEADYVIKEVTR 10	040590613.1	VLMVERSNATYOKTKOWPHSVVMFFRMSDTREAVPEPPLLSRAELRLORFKSAW
NP_854396445.1	HUNSPYSQREAMADPLPEAVLIETHISTRUMYARESAUPUPUPUPUPUPUPUPUP	************************************	VLMVDNSHNTYKSTETVAHSTYMETHTSELNEAVPDPLLLSRAELIMORUKUNV
NP_012891536.1	RLASPP SPGDAPPSPLFDAVLAL WISTRDRV MGEGAEPEPEPETOYVAKEVTR	Contraction of the last	VLMVE INNINEWSSMDTWARKSVYNFFRITSELREAVPOPELLISIAELINOREKUSV
HP_001305385,1	ALTANY - PASE INVERTIGATION OF A DATA AND AND A DATA AND A DATA AND A DATA AND AND AND AND AND AND AND AND AND AN	#450502237539#4.Clin.	VLMVE STIRETYEKTKOTSHISVYNFFINTSELREAMPOPLLLSRAELALORLKVKA
XP_030075415.1	ELSSPP EVESEEL VVPEE 0PS2 VIISTRIPPICEPOCKAAETLETRETVEEEVVARUVLR	MACOL STOCK COL	VLTVENSHETVETFIONSAVSEVVETRITSELREAVPEPOLLSRAELALLBIKLKV
NP_859574764.1	HERMOSERPOREMALLIPLIPPAMAALTINATADAPROMUMDRDRDRDRDRDRDRDRTMRETHR		VETVENSHELVETFK/NSMSTWFFHITSELNEAVPEPOLESRAELILORIKLKV
	UNIVERSITY AND	PERSONAL SCHOOL ST	
		#555/30165540	VLWVESSHICTVDKLAHIGSHSTHMLFHTSELREAVPEPHLLSRAELALLAQSLKV
SP 000051.3		H = 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	VLWVEKENETYX TVETGSHISTWIFFRWSELRAM/VPDPMLESRAELALLALKLSV
NP 014979511.1	VERYESS		VLMVETHREEVOKFRQSTHSEVPEFRTSEEREAVPEPVLLSRAELALLALKUKV
NP 854322857.1	VLINE THRETYDREROSTHSTYNEENTSELREAVEEPVLLSRAELRURIKLIKY XP	 Statistics 	VUNVETHNETYDKFKQSTH/STYNFFAITSELREAVPEPVLLSRAELRLLALKU.KV
XP 018879688.1	VUMETHNETYDRFROSTHSTYMFFRITSELREAWPEPVLLSRAELRLLRLKLKY	454396445.1	VENVETHNETYDKFKQSTH/STYNFFNTSELREAVPEPVLLSRAEURLALKUKV
NP 824782266.2	VLWETHNEDYDKFRQSTHSDYMFFNTSELREAVPEPVLLSRAELRLHLKLKV XP	012891536,1	VEHLDSQWQEYRNTQMVGARSVYHFTNTSELREAVPOPLLESBAELREQREXSSV
NP #38512824.1	VLMVENTIKETYEKVKKSPHSTYMLPATSELREAMPEPVLESKKELRELREKKK	001305385.1	1PMETTWDGPMEHWQPQSHSTFFVFNVSRVRAEVGGRALLHIVAELRMLRQKAAVOS
XP 906941294.1	VLMVE	630875415.1	FPMYQPRADQSSSSCDANEHOLVFTFINASLIBRQVVSESLLHRAELRIDKVKSIIGAA
XP 026330022.1		************************************	LONAPASPQGAGAMMASVRGTQGSYELHFDAAALRAALGPELLLHGAELRMLRKPYTAQA
	AFIAF		· · · · · · · · · · · · · · · · · · ·
MP_027475376.1	VUNVERSNKTYEKTEKSPHSTYMLENTSELREAVPEPVLLSRAELRUHLKKV UNVE		
XP_947558805.1	VLMVE TTIREYDKIKKSPHSEVHLERITSELREAVPERVLLSRAFLRLIRLKLKA	000651.3	EDHVELVOKYSIANSARYLSIIRLLAPSOSPEALSFOVTGVVRDALSRGGETEGERLSA
MP_059086887.1	TERME TO THE CONTRACT OF THE OTHER PROPERTY	The second se	 EQHYEL MAXYSMINI SINKLEAP SOSPERIES FOR TOWARDEL SMOLETED RESIN EQHYEL YOKYSMINI SINKLEAP SOSPERIES FOR TOWARDEL SMOLETED RESIN
NP_033700300.1	Tente Optica Grandia Francisco Practica Chemical Contractica Chemical Contractica Chemical Ch		 EQHYLLINGY INHOMYLSINKI LAPSOSPEM, SECONDARY MULTING SIGELEGERI, SA
XP_059855830.1	AFTAF, HOHETAGETAGETAGETAGETAGETAGETAGETAGETAGETAG	- CT-0100000000000000000000000000000000000	
NP_022441296.1	There is a substantial of the second s		EQHVELVQKYSNRSMKYLSNRLLAPSDSPEWLSFDVTGVVRQNLSRGGELEGFRLSA
NP_823988445.1		Control 010 (2010) (2010)	EQEVELYQKYSINSWRYLSINRLLAPSOSPEMLSFDVTGAVRQNLSRGGELEGFILSA
XP_859968434.1			EQHVELYQKYSNDSWRYLSNRLLAPSD1PEWLSFDV16VVNQNLSHGGEVEGFRLSA
XF_004271297.1	The contract of the second s		EQENELYQKYSNRSWRILSINRLLAPSDTPEMLSFDVTGVVRQMLSHGGEVEGFRLSA
MP_026935427.1	VLMVERSRETYGKEKRSPHSRMMLFINTSELREAVPEPVLLSRAELRLLRLKLKV XP	026338822.1	EQHVELYQKYSNDSMRYLSNRLLAPSD1PEMLSFDVTGVVRQNLSHGGEVEGFRLSA
MP_000189.7			FQHVELYQKYSIDSMYYLSIIRLLAPSDTPEMLSFGVTGVVRQNLSHGGEVEGFRLSA
10, 84126806011	EQNVELYQKYSMD50KYL500LLAPSD1PEWL5FDV16VVWQNE5H6EEVE6FHL5A	NP 000651.3	HESCOSRONTLONOIN-GETTG-RAGOLATINGNURPELLUNITPLERADHLOSS
NP_059006887.1	EQHVELYQKYSNDSMRYLSNHLLAPSOTPSHLSFDVTGVVRQNESHGEEVEGFRLSA	XP 014979511.	1 HESCOSKOWTLOVDEN-GETTG-RISOLATEHKINIKPELLIPATPLERAQHLOSS
MP 033700380.1	EQHVELVQKYSND5NRYLSNRLLAPSDSPENLSFDVTGWRQNLTHGEETEGFRLSA	XP 054322057.	
XF 859855830.1		Charles and the second second	일을 만난 것인을 다시면 것 같은 일부분에 다시면 절약을 해야 했다. 것은 것인 것은 것이 많이 것 같은 것 같은 것을 가지 않는 것 수요?
10.11 TO 0.00 CO	내는 그 사람이 수 있다. 정말 것이 많은 것이 많은 것이 많은 것이 없는 것이 같은 것이 같은 것이 집에 가지 않는 것이 없다. 것이 같이 많은 것이 없는 것이 없다. 것이 없는 것이 않는 것이 않는 것이 없는 것이 없는 것이 없다. 것이 않는 것이 없는 것이 않는 것이 없는 것이 않는 것이 없는 것이 않는 것이 없는 것이 않는 것이 않 않는 것이 않이 않는 것이 않이 않이 않는 것이 않이	xP_018870688.	
KP_822441296.1		XP_024782265.	2 HCSCDSRORTLONDINAGETTIG-BRIDLATIHGMBRPFLLLMATPLERAGHLOSS
NP_023988445.1	E — EQHVELYQKYSNDSkRYLSNRLLAPSDSPEWLSFDYTQWRQNLTHGEETEGFRLSA	XP_038512824.	1 HCSCDSKDNTLQNDIN-GFSSS-RRGDLATING/NRPFLLU/ATPLERAQHLHSS
JP_059588434.1	EQHVELYQKYSMDSNRYLSNRLLAPSDSPEWLSFDVTGVVRQNLTHGEE1EGFRLSA	XP 006941294.	1 HCSCDSKONTLONOINALESSS-REGULATIHGMURPPILLUPATECERACHLHSS
3F 004271297.1	EQHVELVQKYSNDSNRYLSNRLLAPSDSPENLSFDVTGVVRQNLTHGEETEGFRLSA	XP 026338022.	1 HESCOSIGNITLOVOINAGESSS-RRIDLATIHONIRPELLINATIN-ERAOHLHSS
XP 026935427.1	는 그는 이 가장 다양 것은 것은 것을 잘 하는 것을 것을 하는 것은 것은 것을 만들었다. 가장 가지 않는 것을 가 없는 것을 하는 것을 가 있는 것을 가 있는 것을 가 있다.	XP 027475376.	
집안 바람이 많은 것이 없다.	그는 그는 가장 승규가 정말한 것 같은 것은 것이다. 방법과 가장에서 가지 않는 것 같아요. 것은 것은 것이 같아요. 것이다.	C 10 17 C 20 C 10	이 같은 것은 것이 같은 것이 이 것은 것이 같은 것은 것은 것은 것은 것은 것은 것은 것을 가지면 것이 같이 가지만 것이 같이 같이 같이 같은 것이 같은 것이 같은 것이 같이 많이 많이 많이 많이 많이 없다.
MP_999188.2	EQNVELYQKYSNDSWRYLSHRLLAPSOSPEWLSTDVTGWVRQWLTBREATEGFRLSA	XP_047568806.	승규가 가장 것이 그 가장 이 것 같아요. 이 것 같아요. 이 것 않는 것 같아요. 이 것 같아요. 이 것 같아요. 한 것
MP_001009400.1	LEQWELVQKYSMISMRYLSHRLLAPSDSPEWLSFDVTGWRQNLTHREETEGFRLSA	XP_8599966887.	
MP_001301071.1	I EQHVELYQKYSINISMRYLSNRLLAPSOSPEWLSEDVTGV/RQNLTHREETEGERLSA	XP_033789380.	1 HCSCBSKONTLQVD3NAGESSG-BREDLAT3HGMMLPELLUMATPLEBAQHLHSS
XP 005890676.1	EQHVELYQKYSMISWRYLSNILLAPSDSPEWLSFDVTGV/RQNLTRIEEIEGFRLSA	XP_059855830.	1 HCSCDSKONTLQV01NAGESSG-RREDLATTHGNNRPELLUNATPLERAQHLHSS
AP 043755182.1		XP 022441296.	1 HCSCDSKDNTLONDIN-GESSG-RRGDLATING/WAPFILL/WATPLERACHLHSP
There is not a local to the second se	내 이는 것 것같은 전 300 명령이 있는 것 같아요. 것은 것 같아요. 한 것은 것 같아요. 한 것은 전 2010년 2010년 2011년 2011	XP 023988445.	
XP_088986313.1		XP 059988434.	
XP_853461166.1		VAR198-507-0005527	
XP_884479972.4	EQHVELYEKYSKDSNRYLSNRLLHPSOTTEWLSFDVTAVVRRNLSQGGETEGFRLSA	XP_004271297.	가는 말에서 사람이 잘 수가도 가지 않는 것 같아요. 이렇게 많은 것이 가지 않는 것이 같은 것이 같이 가지 않는 것이다.
MP_035707.1	EQHVELVQKYSMISMRVLGMRLLTPTDTPEWLSFDVTGVVRQNLNQSDGTQGFRF5A	XP_026935427.	그는 것 것 같아요. 이 집 것 같아요. 것은 것 같아요. 것 같아요. 집 것 같아요. 이 집 것 같아요. 이 집 것 같아요. ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
- 11 BOOM 11 1 1 1 1	EQHVELYQKYSINISNRYLGNILLTPTDTPENLSFDVTGV/RQMLHQSDG3QGF8F5A	IP 999188.2	HCSCDSKDWTLEWESN-GENSG-RRGDLATSHGMNEPFLLLMATPLERACHLRSS
MP 067589.1	EVENTICE TO A SOLUTION OF COMPLETE TO FEM STORE OF DRIVING TO AND A DRIVEN STORE S		
1		NP_001003480.	
MP_032750046.1	EQHVELYQKYSMISWKYLGHULLTPTOTPEWLSFUVTGV/HQNLHQGDGIQGFRFSA		1 HCSCDSKONTLOVDIN-GESSG-BREDLATJHORNAPELLL/WATPLERAQHLHSS
NP_032750046.1 NP_052824653.1	EQHVELYQKYSMISMRYLGHILLTPTOTPENLSFDYTGV/RQNLRQGDGLQGFRFSA EQHVELYQKYSMISMRYLGHILLTPTOTPENLSFDYTGV/RQNLRQGDGLQGFRFSA	NP_001003480. NP_001381072.	1 HCSCDSKONTLQNOIN-GESSG-BRODLATJHONNAPEILLINATPLERACHLHSS 1 HCSCDSKONTLQNOIN-GESSG-RRODLATJHONNAPEILLINATPLERACHLHSS
MP_032750046.1	EQWELVQCYSMISMRYLGHRLLTPTOTPENLSFDVTGVVRQNLIQGOGLQGFRFSA EQWELVQCYSMISMRYLGHRLLTPTOTPENLSFDVTGVVRQNLIQGOGLQGFRFSA EQWELVQCYSMISMRYLGHRLLSPTOTPENLSFDVTAVVRQNLIQGOGLQGFRFSA	NP_001009480. NP_001381072. XP_005890676.	1 HCSCDSKUNTLQXDD-GF556-BRIQLATDH948FF1LTMFPLEAQHURS5 1 HCSCDSKUNTLQXDD-GF556-BRIQLATDH948FF1LTMFPLEAQHURS5 1 HCSCDSKUNTLQXDD-GF556-BRIQLATDH948PF1LTMFPLEAQHURS5
NP_032750046.1 NP_052824653.1	EQWELVQCYSMISMRYLGIRULTPTOTPENLSFDVTGVVRQNLIQGOGLQGFRFSA EQWELVQCYSMISMRYLGIRULTPTOTPENLSFDVTGVVRQNLIQGOGLQGFRFSA EQWELVQRTSINISMRYLGIRULTSPTOTPENLSFDVTAVVRQNLIQGOGLQGFRFSA	NP_001003480. NP_001381071. XP_005890676. XP_043755182.	1 HCSCDSKUNTLQXDDI-GF55G-BRIQUATUHOPHIFFULINAFPLERACHURSS 1 HCSCDSKUNTLQXDDI-GF55G-BRIQUATUHOPHIFFULINAFPLERACHURSS 1 HCSCDSKUNTLQXDDI-GF55G-BRIQUATUHOPHIPFULINAFPLERACHURSS 1 HCSCDSKUNTLQXDDI-GF55G-BRIQUATUHOPHIPFULINAFPLERACHURSS
XP_032750046.1 XP_052824653.1 XP_027286179.1 XP_040590613.1	EQAVELYQCYSMISMRYLGIRLLTPTOTPENLSFDVTGVVRQNLIQGOGLQGFRFSA EQAVELYQCYSMISMRYLGIRLLTPTOTPENLSFDVTGVVRQNLIQGOGLQGFRFSA EQAVELYQCYSMISMRYLGIRLLSPTOTPENLSFDVTAVVRQNLIQGOGLQGFRFSA FQAVELYQCYSMISMRYLGIRLLSPTOTPENLSFDVTTVRQNLIQGOGLQGFRFSA	NP_001301071. NP_001301071. XP_005890676. XP_043755102. XP_000906313.	1 HCSCDSKUNTLQADIN-GF556-BRIQUATUHOPHIFFULLINATPLENQHURSS 1 HCSCDSKUNTLQADIN-GF556-BRIQUATUHOPHIFFULLINATPLENQHURSS 1 HCSCDSKUNTLQADIN-GF556-BRIQUATUHOWRPFULLINATPLENQHURSS 1 HCSCDSKUNTLQADIN-GF556-BRIQUATUHOWRPFULLINATPLENQHURSS 1 HCSCDSQWITLQADINAGF5T6-BRIQUATUHOWRPFULLINATPLENQHURSS
NP_032750046.1 NP_052824653.1 NP_027286179.1 NP_048590613.1 NP_001166494.1	EQAVELYQCYSMISMRYLGIRLLTPTOTPENLSFDVTGVVRQNLIQGOGLQGFRFSA EQAVELYQCYSMISMRYLGIRLLTPTOTPENLSFDVTGVVRQNLIQGOGLQGFRFSA EQAVELYQCYSMISMRYLGIRLLSPTOTPENLSFDVTAVVRQNLIQGOGLQGFRFSA EQAVELYQCYSMISMRYLGIRLLSPTOTPENLSFDVTTVRQNLIQGOGLQGFRFSA EQAVELYQCYSMISMRYLSIQLLSPTOTPENLSFDVTTVRQNLIQGOGLQGFRFSA EQAVELYQCYSMISMRYLSIQLLSPTOTPENLSFDVTGVRQNLIQGOGLQGFRFSA	NP_001003400. NP_001381071. XP_005890676. XP_043755182. XP_008986313. XP_053461166.	HCSDSKINTLQADIN-6F556-REQUATINGNEFF1LINEFPLENQHUSS HCSDSKINTLQADIN-6F556-REQUATINGNEFF1LINEFPLENQHUSS HCSDSKINTLQADIN-6F556-REQUATINGNEFF1LINEFPLENQHUSS HCSDSKINTLQADIN-6F556-REQUATINGNEFF1LINEFPLENQHUSS HCSDSKINTLQADIN-6F556-REQUATINGNEFF1LINEFPLENQHUSS HCSDSKINTLQADIN-6F555-REQUATINGNEFF1LINEFPLENQHUSS
NP_032750046.1 NP_052824653.1 NP_027286179.1 NP_040590613.1 NP_001166494.1 NP_004648529.1	EQAVELYQCYSMISMRYLGIRLLTPTOTPENLSFDVTGVVRQNLIQGOGLQGFRFSA EQAVELYQCYSMISMRYLGIRLLTPTOTPENLSFDVTGVVRQNLIQGOGLQGFRFSA EQAVELYQCYSMISMRYLGIRLLSPTOTPENLSFDVTAV/RQNLIQGOGLQGFRFSA EQAVELYQCYSMISMRYLSIQLLSPTOTPENLSFDVTTV/RQNLIQGOGLQFRFSA EQAVELYQCYSMISMRYLSIQLLTPSOTPENLSFDVTGVRQNLSQGEELEGFRFSA	NP_001003400. NP_001381071. XP_005890676. XP_043755182. XP_008986313. XP_053461166.	1 HCSCDSKUNTLQADIN-GF556-BRIQUATUHOPHIFFULLINATPLENQHURSS 1 HCSCDSKUNTLQADIN-GF556-BRIQUATUHOPHIFFULLINATPLENQHURSS 1 HCSCDSKUNTLQADIN-GF556-BRIQUATUHOWRPFULLINATPLENQHURSS 1 HCSCDSKUNTLQADIN-GF556-BRIQUATUHOWRPFULLINATPLENQHURSS 1 HCSCDSQWITLQADINAGF5T6-BRIQUATUHOWRPFULLINATPLENQHURSS
IP_832758046.1 IP_852824653.1 IP_027286179.1 IP_042590613.1 IP_040590613.1 IP_00166494.1 IP_001648529.1 IP_04648151618.1	EQHVELYQCYSINISARYLGHRLLTPTOTPENLSFIN/TGV/RQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLGHRLLTPTOTPENLSFIN/TGV/RQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLGHRLLSPTOTPENLSFIN/TWRQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLGHRLLSPTOTPENLSFIN/TWRQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLSINULLTPSOTPENLSFIN/TQV/RQKLIQGEELEFRFSA EQHVELYQCYSINISARYLSINULLTPSOTPENLSFIN/TQV/RQKLIQGEELEFRFSA EQHVELYQCYSINISARYLSINULTPSOTPENLSFIN/TQV/RQKLIQGEELEFRFSA EQHVELYQCYSINISARYLSINULTPSOTPENLSFIN/TQV/RQKLIQGEELEFRFSA	NP_001003400. NP_001381071. XP_005890676. XP_043755182. XP_008986313. XP_053461166.	HCSDSKINTLQADIN-6F556-REQUATINGNEFF1LINEFPLENQHUSS HCSDSKINTLQADIN-6F556-REQUATINGNEFF1LINEFPLENQHUSS HCSDSKINTLQADIN-6F556-REQUATINGNEFF1LINEFPLENQHUSS HCSDSKINTLQADIN-6F556-REQUATINGNEFF1LINEFPLENQHUSS HCSDSKINTLQADIN-6F556-REQUATINGNEFF1LINEFPLENQHUSS HCSDSKINTLQADIN-6F555-REQUATINGNEFF1LINEFPLENQHUSS
NP_032750046.1 NP_052824653.1 NP_027286179.1 NP_040590613.1 NP_001166494.1 NP_004648529.1	EQHVELYQCYSMISARYLORILLTPTOTPENLSFINTGVIRQLINGOGLQGFRFSA EQHVELYQCYSMISARYLORILLTPTOTPENLSFINTGVIRQNINQGOGLQGFRFSA EQHVELYQCYSMISARYLORILLTPTOTPENLSFINTGVIRQNINQGOGLQGFRFSA EQHVELYQCYSMISARYLORILLTPTOTPENLSFINTGVIRQNINQGOGLQGFRFSA EQHVELYQCYSMISARYLORILLSPTOTPENLSFINTGVIRQNINQGOGLQFRFSA EQHVELYQCYSMISARYLORILLSPTOTPENLSFINTGVIRQNINQGOGLQFRFSA EQHVELYQCYSMISARYLORILLSPTOTPENLSFINTGVIRQNINQGOGLQFRFSA EQHVELYQCYSMISARYLORILLSPTOTPENLSFINTGVIRQNINQGOGLQFRFSA EQHVELYQCYSMISARYLSNILLTPSNTQENLSFINTGVIRQNINQGISQEELEGFRFSA EQHVELYQCYSMISARYLSNILLTPSNTQENLSFINTGVIRQNINGGISQEELEGFRFSA EQHVELYQCYSMISARYLSNILLTPSNTQENLSFINTGVIRQNINGGISQEELEGFRFSA EQHVELYQCYSMISARYLSNILLSNILTPSNTQENLSFINTGVIRQNINGNISHGETQGFRFSA	NP 0012093400. NP 001381071. XP 005398676. XP 043755182. XP 008986313. XP 053461166. XP 064479972. NP 035787.1	HCSCD SKINTLQADIN-6F55G-REQUATINGNEFFTLLINETPLERQUERS HCSCD SKINTLQADIN-GF55G-REQUATINGNEFFTLLINETPLERQUERS HCSCD SKINTLQADIN-GF55G-REQUATINGNEPFTLLINETPLERQUERS HCSCD SKINTLQADIN-GF55G-REQUATINGNEPFTLLINETPLERQUERS HCSCD SKINTLQADINEF55S-REQUATINGNEPFTLLINETPLERQUERS HCSCD SKINTLQADINEF55S-REQUERTINGNEPFTLLINETPLERQUERS HCSCD SKINTLQADINEF55S-REQUERTINGNEPFTLLINETPLERQUERS HCSCD SKINTLQADINEF55S-REQUERTINGNEPFTLLINETPLERQUERS HCSCD SKINTLQADINEF55S-REQUERTINGNEPFTLLINETPLERQUERS HCSCD SKINTLQADINEF55S-REQUERTINGNEPFTLLINETPLERQUERS HCSCD SKINTLQADINEF555-REQUERTINGNEPFTLLINETPLERQUERS HCSCD SKINTLQADINEF555-REQUERTINGNEPFTLLINETPLERQUERS HCSCD SKINTLQADINEF555-REQUERTINGNEPFTLLINETPLERQUERS HCSCD SKINTLQADINEF555-REQUERTINGNEPFTLINETPLERQUERS HCSCD SKINTLQATFFTLINETPLERQUERS HCSCD SKINTLQATFFTLINETPLERQUERS HCSCD SKINTLQATFFTLINETPLERQUERS HCSCD SKINTLQATFFTLINETPLERQUERS HCSCD SKINTLQATFFTNETPLERQUERS HCSCD SKINTLQATFFTNETPLERQUERS HCSCD SKINTLQATFFTNETPLER
IP_832758046.1 IP_852824653.1 IP_027286179.1 IP_042590613.1 IP_040590613.1 IP_00166494.1 IP_001648529.1 IP_04648151618.1	EQHVELYQCYSMISARYLORILLTPTOTPENLSFINTGVIRQLINGOGLQGFRFSA EQHVELYQCYSMISARYLORILLTPTOTPENLSFINTGVIRQLINGOGLQGFRFSA EQHVELYQCYSMISARYLORILLTPTOTPENLSFINTGVIRQLINGOGLQGFRFSA EQHVELYQCYSMISARYLORILLTPTOTPENLSFINTTWRQLINGOGLQFRFSA FQHVELYQCYSMISARYLORILLSPTOTPENLSFINTGVIRQLINGOGLQFRFSA EQHVELYQCYSMISARYLORILLSPTOTPENLSFINTGVIRQLINGOGLQFRFSA EQHVELYQCYSMISARYLORILLSPTOTPENLSFINTGVIRQLINGOELGFRFSA EQHVELYQCYSMISARYLORILLSPTOTPENLSFINTGVIRQLINGOELGFRFSA EQHVELYQCYSMISARYLSNILLTPSMTQENLSFINTGVIRQLINGEELEFRFSA EQHVELYQCYSMISARYLSNILLTPSMTQENLSFINTGVIRQLINGEELFGFRFSA EQHVELYQCYSMISARYLSNILLTPSMTQENLSFINTGVIRQLINGEELEFRFSA EQHVELYQCYSMISARYLSNICLAPSOTPENLSFEVTONVIRQLINGEELFGFRFSA EQHVELYQCYSMISARYLSNICLAPSOTPENLSFEVTONVIRQLINGEELFGFRFSA	NP_001209400. NP_001381071. XP_005398676. XP_043755182. XP_008986313. XP_053461166. XP_064479972. NP_035787.1 NP_035787.1	 HCSCD SKUNTLQUDIN-GESSG-BRIQUATUHONIAPETLUMPPLENAQUUSS HCSCD SKUNTLQUDIN-GESSG-BRIQUATUHONIAPETLUMPPLENAQUUSS HCSCD SKUNTLQUDIN-GESSG-BRIQUATUHONIAPETLUMPPLENAQUUSS HCSCD SKUNTLQUDIN-GESSG-BRIQUATUHONIAPETLUMPPLENAQUUSS HCSCD SKUNTLQUDIN-GESSG-BRIQUATUHONIAPETLUMPPLENAQUUSS HCSCD SKUNTLQUDIN-GESSF-BRIQUATUHONIAPETLUMPPLENAQUUSS
NP_032750046.1 NP_052824653.1 NP_027286179.1 NP_001166494.1 NP_001166494.1 NP_04648529.1 NP_045311618.1 NP_053523548.1	EQWELYQCYSMISMRYLORILLTPTOTPENLSFINTGVIRQNLIQGOGLQGFRESA EQWELYQCYSMISMRYLORILLTPTOTPENLSFINTGVIRQNLIQGOGLQGFRESA EQWELYQCYSMISMRYLORILLSPTOTPENLSFINTGVIRQNLIQGOGLQGFRESA EQWELYQCYSMISMRYLSIQLLSPTOTPENLSFINTGVIRQNLIQGOGLQGFRESA EQWELYQCYSMISMRYLSIQLLTPSOTPENLSFINTGVIRQNLSQGEELEGFRESA EQWELYQCYSMISMRYLSIQLLTPSOTPENLSFINTGVIRQNLSQGEELEGFRESA	NP_001301071. NP_001301071. NP_005390676. NP_043755182. NP_003966313. NP_053461166. NP_06547972. NP_055787.1 NP_065787.1 NP_067589.1 NP_057289.1	HCSCD - SKINTLONDIN OF SSG. BRIDLATTHONH - FFILLINATPLERAPHUSS HCSCD - SKINTLONDIN OF SSG. BRI
IP 032750046.1 IP 052824653.1 IP 042530617.1 IP 042590613.1 IP 040590613.1 IP 040590613.1 IP 040590613.1 IP 040590613.1 IP 04051618.1 IP 040521523548.1 IP 052523548.1 IP 0424433171.2 IP 083420839.1	EQWELVQCYSMISMRYLGHRLLTPTOTPENLSFEVTGV/RQNLIQGDGLQGFRESA EQWELVQCYSMISMRYLGHRLLTPTOTPENLSFEVTGV/RQNLIQGDGLQGFRESA EQWELVQCYSMISMRYLGHRLLSPTOTPENLSFEVTGV/RQNLIQGDGLQGFRESA EQWELVQCYSMISMRYLSHQLLSPTOTPENLSFEVTGV/RQNLSQGEELGFRESA EQWELVQCYSMISMRYLSHQLLSPTOTPENLSFEVTGV/RQNLSQGEELGFRESA	107 001003400. 107 001301071. XP 005898676. XP 043755182. XP 000906313. XP 05461106. XP 054617972. 107 0578.1 XP 05583.1 XP 052824653.	HCSCD - SKINTLQADIN-GESSG-REQUATINGNE - FEILINETPLERAGUSS HCSCD - SKINTLQADIN-GESSG-REGUGTINGNE - FEILINETPLERAGUSS HCSCD - SKINTLQADIN-GESSG-REGUSTINGEP
xP 832758046.1 xP 852824653.1 xP 927286179.1 xP 948599613.1 xP 948599613.1 xP 946448259.1 xP 9465311618.1 xP 945311618.1 xP 94532548.1 xP 94433171.2 xP 943428839.1 xP 961075318.1	EQWELVQCYSMISMRYLGIRLLTPTOTPENLSFEVTGV/RQNLIQGDGLQGFRESA EQWELVQCYSMISMRYLGIRLLTPTOTPENLSFEVTGV/RQNLIQGDGLQGFRESA EQWELVQCYSMISMRYLGIRLLSPTOTPENLSFEVTGV/RQNLIQGDGLQGFRESA	107 001203400. 107 001301071. XP 065890570. XP 065890570. XP 055806513. XP 055707.1 107 057807.1 107 057807.1 XP 0525707.1 XP 05259056. XP 052240553. XP 052240553.	HCSDSUNTLQUDN-GFSSG-RIQUATHWANI -FFULINITPLENQUISS HCSDSUNTLQUDN-GFSSG-RIQUATHWANI - FFULINITPLENQUISS HCSDSUNTLQUDN-GFSSF-RIQUATHWANI - FFULINITPLENQUISS HCSDSUNTLQUDN-GFSSF-RIQUATHWANI - FFULINITPLENQUISS HCSDSUNTLQUDN-GFSSF-RIQUATHWANI - FFULINITPLENQUISS HCSDSUNTLQUTHWEN-GFSSF-RIQUATHWANI - FFULINITPLENQUISS HCSDSUNTLQUTHWEN-GFSSF-RIQUATHWANI - FFULINITPLENQUISS HCSDSUNTLQUTHWEN-GFSSF-RIQUATHWANI - FFULINITPLENQUISS HCSDSUNTUNEN-GFSSF-RIQUATHWANI - FFULINITPLENQUISS HCSDSUNTUNEN-GFSSF-RIQU
NP 832758046.1 NP 852824653.1 NP 84259631.1 NP 84259631.1 NP 84259631.1 NP 84259631.1 NP 84559631.1 NP 84559631.1 NP 845511618.1 NP 853525548.1 NP 863428839.1 NP 863428839.1 NP 863428839.1 NP 8634705318.1 NP 8634705318.1	EQHVELYQCYSMISARYLORILLTPTOTPENLSFITYTGVIRQUIQGOGLQGFRFSA EQHVELYQCYSMISARYLORILLTPTOTPENLSFITYTGVIRQUINQGOGLQGFRFSA EQHVELYQCYSMISARYLORILLTPTOTPENLSFITYTAVINQUINQGOGLQGFRFSA EQHVELYQCYSMISARYLORILLTPTOTPENLSFITYTAVINQUINQGOGLQGFRFSA EQHVELYQCYSMISARYLSMILLSPTOTPENLSFITYTAVINQUINQGOGLQFRFSA EQHVELYQCYSMISARYLSMILLSPTOTPENLSFITYTAVINQUINQGOGLQFRFSA EQHVELYQCYSMISARYLSMILLSPTOTPENLSFITYTAVINQUINQGOGLQFRFSA EQHVELYQCYSMISARYLSMILLSPTOTPENLSFITYTAVINQUISQGEELEGFRFSA EQHVELYQCYSMISARYLSMILLSPTOTPENLSFITYTAVINQUISQGEELEGFRFSA EQHVELYQCYSMISARYLSMILLSMILTPSOTPENLSFITYTAVINQUISMIELEGFRFSA EQHVELYQCYSMISARYLSMILLAPSOTPENLSFITYTAVINQUISMIELEFIEFISA EQHVELYQCYSMISARYLSMILLAPSOTPENLSFITYTAVINQUISMIGETEGFRISA EQHVELYQCYSMISARYLSMILLAPSOTPENLSFITYTAVINQUISMISARGETEGFRISA EQHVELYQCYSMISARYLSMILLAPSOTPENLSFITYTAVINQUISMIGLSMIELEFIFISA EQHVELYQCYSMISARYLSMILLAPSOTPENLSFITYTAVINQUISMICHGIAGGAFHELRIFL	NP 0012081400. NP 001301071. NP 065390676. NP 065390676. NP 0653906776. NP 0653906776. NP 065390677.1 NP 0557067.1 NP 0557067.1 NP 0557067.1 NP 055290.1 XP 052250006. XP 052250006. XP 0522500651. XP 0522500651. XP 042590613.	HCSDSIDITLQUDIN-GFSSG-RIGULATHKANI -FFULLINITPLENQULISS HCSDSIDITLQUDIN-GFSSG-RIGULATHKANI - FFULLINITPLENQULISS HCSDSIDINULAPIDI-GISPK-RIGULATHKANI - FFULLINITPLENQULISS HCSDSIDINULAPIDI-GISPK-RIGULATHKANI HCSD-SIDINULAPIDI-GISPK-RIGULATHKANI
NP 832758046.1 NP 852824653.1 NP 84259613.1 NP 84259613.1 NP 84459613.1 NP 84459613.1 NP 84559613.1 NP 84559613.1 NP 845511618.1 NP 85323548.1 NP 84423317.2 NP 804423317.3 NP 8043203531.3 NP 8014203513.1 NP 8149237.1	EQIVEL VQCYSINISMIYLGIRLLTPTOTPENLSFIVTGVIRQUINGOBLIQGFRESA EQIVEL VQCYSINISMIYLGIRLLTPTOTPENLSFIVTGVIRQUINGOBLIQGFRESA EQIVEL VQCYSINISMIYLGIRLLTPTOTPENLSFIVTGVIRQUINGOBLIQGFRESA EQIVEL VQCYSINISMIYLGIRLLTPTOTPENLSFIVTGVIRQUINGOBLIQGFRESA EQIVEL VQCYSINISMIYLSINILLISPTOTPENLSFIVTGVIRQUINGOBLIGFRESA EQIVEL VQCYSINISMIYLSINILLTPSOTPENLSFIVTGVIRQUINGOBLIGFRESA EQIVEL VQCYSINISMIYLSINILLTPSOTPENLSFIVTGVIRQUINGOBLIGFRESA EQIVEL VQCYSINISMIYLSINILLTPSOTPENLSFIVTGVIRQUINGOBLIGFRESA EQIVEL VQCYSINISMIYLSINILLTPSOTPENLSFIVTGVIRQUINGOBLIGFRESA EQIVEL VQCYSINISMIYLSINILLAPSOTPENLSFIVTGVIRQUINGBEETEGFRESA EQIVEL VQCYSINISMIYLSINILLTPSOSSINLSFIVTGVIRQUINGBEETEGFRESA EQIVEL VQCYSINISMIYLSINILLAPSOSSINLSFIVTGVIRQUINGBEETEGFRESA EQIVEL VQCYSINISMIYLSINILLAPSOSSINLSFIVTGVIRQUINGBEETEGFRESA EQIVEL VQCYSINISMIYLSINILLAPSOSFENLSFIVTGVIRQUINGBEETEGFRESA EQIVEL VQCYSINISMIYLSINILAPSOSFENLSFIVTGVIRQUINGBEETEGFRESA EQIVEL VQCYSINISMIYLSINILAPSOSFENLSFIVTGVIRQUINGBEETEGFRESA EQIVEL VQCYSINISMIYLSINILAPSOSFENLSFIVTGVIRQUINGBEETEGFRESA	107 001203400. 107 001301071. XP 065890570. XP 065890570. XP 055806513. XP 055707.1 107 057807.1 107 057807.1 XP 0525707.1 XP 05259056. XP 052240553. XP 052240553.	 HCSCD SEDITLOVIDIN-OFSSG-REGULATINOMIA - PETLUMATPLERACIUSS HCSCD SEDITLOVIDIN-OFSSG-REGULATINOMIA - PETLUMATPLERACIURSS HCSCD SEDITLOVIDIN-OFSSG-REGULATINOMIA - PETLUMATPLERACIURSS HCSCD SEDITLOVIDIN-OFSSG-REGULATINOMIA - PETLUMATPLERACIURSS HCSCD SEDITLOVIDIA/OFSSG-REGULATINOMIA - PETLUMATPLERACIURSS HCSCD SEDIMUN/NEIN-OFSPC-REGULATINOMIA - PETLUMATPLERACIURSS
NP 8327580465.1 NP 852824553.1 NP 827286179.1 NP 942559631.1 NP 942559631.1 NP 942659631.1 NP 96164494.1 NP 96164494.1 NP 9645311618.1 NP 963523548.1 NP 964433179.2 NP 963428383.1 NP 9634705318.1 NP 963791534.1		NP 0012081400. NP 001301071. NP 065390676. NP 065390676. NP 0653906776. NP 0653906776. NP 065390677.1 NP 0557067.1 NP 0557067.1 NP 0557067.1 NP 055290.1 XP 052250006. XP 052250006. XP 0522500651. XP 0522500651. XP 042590613.	 HCSCD SKUNTLQUOIN-GESSG-RIKULATUHONIA - FEILLINATPLERAQUURSS HCSCD SKUNTLQUOIN-GESSK-RIKULATUHONIA - FEILLINATPLERAQUURSS HCSCD SKUNTLANEIN-GESSK-RIKULATUHONIA - FEILLINATPLERAQUURSS
NP 832758046.1 NP 852824655.1 NP 84259613.1 NP 84859613.1 NP 84859613.1 NP 846511649.1 NP 846511649.1 NP 846511649.1 NP 846511618.1 NP 846311618.1 NP 84643171.2 NP 804423971.3 NP 804320839.1 NP 804320839.1 NP 804529513.1 NP 81942337.1 NP 811942237.1	EQHVELYQCYSINISMRYLGIRLLTPTOTPENLSFINTGVIRQNINGGOGLQGFRFSA EQHVELYQCYSINISMRYLGIRLLTPTOTPENLSFINTGVIRQNINGGOGLQGFRFSA EQHVELYQCYSINISMRYLGIRLLTPTOTPENLSFINTGVIRQNINGGOGLQGFRFSA EQHVELYQCYSINISMRYLGIRLLTPTOTPENLSFINTGVIRQNINGLSQGELGFRFSA EQHVELYQCYSINISMRYLSINGLLTPSOTPENLSFINTGVIRQNISGGEELGFRFSA EQHVELYQCYSINISMRYLSINGLLTPSOTPENLSFINTGVIRQNISGGEELGFRFSA EQHVELYQCYSINISMRYLSINGLLTPSOTPENLSFINTGVIRQNISGGEELGFRFSA EQHVELYQCYSINISMRYLSINGLLTPSOTPENLSFINTGVIRQNISGGEELGFRFSA EQHVELYQCYSINISMRYLSINGLLAPSOTPENLSFINTGVIRQNISMRGEELEGFRFSA EQHVELYQCYSINISMRYLSINGLLAPSOTPENLSFEVTGVIRQNISMRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLLAPSOTPENLSFEVTGVIRQNISMRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLLAPSOTPENLSFEVTGVIRQNISMRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLLAPSOTPENLSFEVTGVIRQNISSKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLLAPSOTPENLSFEVTGVIRQNISSKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLLAPSOSPENLSFEVTGVIRQNISKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLLAPSOSPENLSFEVTGVIRQNISKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLLAPSOSPENLSFEVTGVIRQNISKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLLAPSOSPENLSFEVTGVIRQNISKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLLAPSOSPENLSFEVTGVIRQNISKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLLAPSOSPENLSFEVTGVIRQNISKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLLAPSOSPENLSFEVTGVIRQNISKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLAPSOSPENLSFEVTGVIRQNISKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLAPSOSPENLSFEVTGVIRQNISKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLAPSOSPENLSFEVTGVIRQNISKRGETGGFRLSA EQHVELYQCYSINISMRYLSINGLAPSOSPENLSFEVTGVIRQNISKRGETGGFRLSA	107_001083400. 107_001301071. XP_008500676. XP_008500676. XP_008500113. XP_0550011 XP_055011 XP_055011 XP_052020651. XP_052020651. XP_052020651. XP_05220066. XP_052200679. XP_004500513. XP_004500513. XP_004500513.	HCSCD - SKINTLOVOIN-GESSG-BRIDLATDHONIE - FEILLINATPLERACHUSS HCSCD - SKINTLOVOIN-GESSG-BRIDLATDHONIE - FEILLINATPLERACHUSS HCSCD - SKINTLOVOIN-GESSG-BRIDLATDHONIE - PEILLINATPLERACHUSS HCSCD - SKINTLOVOIN-GESSG-BRIDLATHONIE - PEILLINATPLERACHUSS HCSCD - SKINTLOVOIN-GESSG-BRIDLATHADHONIE - PEILLINATPLERACHUSS HCSCD - SKINTLOVOIN-GESSG-BRIDLATHADHONIE - PEILLINATPLERACHUSS HCSCD - SKINTLOVOIN-GESSG-BRIDLATHADHONIE - PEILLINATPLERACHUSS HCSCD - SKINTLOVOIN - GESSG-BRIDLATHADHUSAHUE - PEILLINATPLERACHUSE HCSCD - SKINTLOVOIN - GESSG-BRIDLATHADHUSAHUE - PE
NP 832758046.1 NP 852824653.1 NP 84259613.1 NP 90116494.1 NP 90116494.1 NP 90116494.1 NP 9045101.1 NP 9045101.1 NP 9045101.6 NP 9045101.6 NP 9045101.6 NP 9045120.2 NP 9045120.2 NP 9045120.2 NP 9045120.2 NP 901075134.1 NP 81542237.1 NP 81542237.1 NP 9154342237.1 NP 9154342237.1 NP 9154342237.1 NP 912981536.1		107 001203400. 107 001301071. XP 005300753102. XP 045300753102. XP 045300753102. XP 04530031108. XP 045479572. 107 055787.1 107 055787	HCSCD - SKINTLQADIN-GESSG-REQUATINGNE -FEILINATPLERAGNISS HCSCD - SKINTLQADIN-GESSG-REQUATINGNE -FEILINATPLERAGNISS HCSCD - SKINTLQADIN-GESSG-REQUATINGNE - PEILINATPLERAGNISS HCSCD - SKINTLQADIN-GESST-REQUATINGNE - PEILINATPLERAGNISS HCSCD - SKINTLQADIN-GESST-REQUATINGNE - PEILINATPLERAGNISS HCSCD - SKINTLQADIN-GESST-REQUATINGNE - PEILINATPLERAGNISS HCSCD - SKINTLQADIN-GESSF-REQUATINGNE - PEILINATPLERAGNISS HCSCD - SKINTLANDIN-GESSF-REQUATINGNE - PEILINATPLERAGNISS HCSCD - SKINTLANDIN-GESSF-REQUATINGNE - PEILINATPLERAGNISS HCSCD - SKINTLANDIN-GESSF-REQUARIANGNE - PEILINATPLERAGNISS HCSCD - SKINTLA
NP 832758046.1 NP 852824653.1 NP 842596613.1 NP 84311648.1 NP 84311648.1 NP 84312648.1 NP 843426839.1 NP 843426839.1 NP 843426839.1 NP 841571548.1 NP 84154273.1 NP 84154231.1 NP 841542431.1 NP 841542431.1 NP 841542431.1 NP 841542431.1 NP 841542431.1 NP 8415424431.1 NP 8415424431.1 NP 8415424431.1 NP 8415424431.1 NP 8415424431.1 NP 8415424431.1 NP 84154244431.1		NP 001.0004400. NP 001.301071. XP 065580676. XP 065580676. XP 0557872. NP 055787.1 NP 052024053. XP 052024053. XP 052024053. XP 052024053. XP 050116404. XP 040509513. NP 052024053. XP 040504134. XP 040544524. XP 04553523548.	HCSDSIDITLQUDIN-GESSG-RIQUATURQUIN-FEULINATPLERQUINSS HCSDSIDITLQUDIN-GESSG-RIQUATURQUIN-FEULINATPLERQUINSS HCSDSIDITLQUDIN-GESSG-RIQUATURQUIN-PEULINATPLERQUINSS HCSDSIDITLQUDIN-GESSG-RIQUATURQUIN-PEULINATPLERQUINSS HCSDSIDITLQUDIN-GESSG-RIQUATURQUIN-PEULINATPLERQUINSS HCSDSIDITLQUDIN-GESSG-RIQUATURQUIN-FEULINATPLERQUINSS HCSDSIDITLQUDIN-GESSG-RIQUATURQUIN-FEULINATPLERQUINSS HCSDSIDITLQUDIN-GESSG-RIQUATURQUIN-FEULINATPLERQUINSS HCSDSIDITLQUDIN-GESSG-RIQUATURQUIN-FEULINATPLERQUINSS HCSDSIDITLQUDIN-GESSF-RIQUATURQUIN-FEULINATPLERQUINSS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSDSIDITLQUDIN-GESSF-RIQUATURQUINS HCSD-
IP 932758045.1 IP 952824553.1 IP 927286175.1 IP 942590613.1 IP 9445590613.1 IP 9445590613.1 IP 944511618.1 IP 94352364.1 IP 953523544.1 IP 963420839.1 IP 963490431.1 IP 961990536.1 IP 961990381.1 IP 98075415.1	EQHVELYQCYSINISARYLGHRLLTPTOTPENLSFIDYTGWRQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLGHRLLTPTOTPENLSFIDYTGWRQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLGHRLLTPTOTPENLSFIDYTAWNQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLGHRLLTPTOTPENLSFIDYTWRQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLSHLLISPTOTPENLSFIDYTWRQKLIQGOELQGFRFSA EQHVELYQCYSINISARYLSHLLISPTOTPENLSFIDYTWRQKLIQGOELQGFRFSA EQHVELYQCYSINISARYLSHLLISPTOTPENLSFIDYTWRQKLIQGOELQFRFSA EQHVELYQCYSINISARYLSHLLISPTOTPENLSFIDYTQWRQKLSQGEELEGFRFSA EQHVELYQCYSINISARYLSHLLIPSOTPENLSFIDYTQWRQKLSRGEELEGFRFSA EQHVELYQCYSINISARYLSHLLIPSOTPENLSFIDYTQWRQKLSRGEELEGFRFSA EQHVELYQCYSINISARYLSHLLIPSOTPENLSFIDYTQWRQKLSRGEELEGFRFSA EQHVELYQCYSINISARYLSHLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSFSDYGFRLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSFDSFRLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSFDSFRLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSFDSFRLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSHCSHCYCHARDISFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSFDSFRLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHCSHCYCHARDISFRCYCHARDISFRCYHARDISFRCYHRXISHRUSHRYLSHRLSFDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHCYCHARDISFRCYHRXINGKARDISFRCYHRXINGKARDISFRCYHRXIXIII STDYFGHYTGWRQKLSRGEELEGFRLSA	NP 0012081400. NP 0013818071. NP 043755182. NP 043755182. NP 043755182. NP 043755182. NP 043755182. NP 0453461104. NP 0457967.1 NP 055707.1 NP 055707.1 NP 055707.1 NP 055707.1 NP 055707.1 NP 04559046. XP 042240653. NP 045590613. NP 040590613. NP 040590513. NP 040590513. NP 053523548. NP 053523548. NP 024433177.	HCSDSIDITLQUDIN-GESSG-BRIDLATHKANI -FELLINIFPLENQURSS HCSDSIDITLQUARTH-GESSG-BRIDLATHKANI -FELLINIFPLENQURSS HCSDSIDITLUNGIN-GESSG-BRIDLATHKANI -FELLINIFPLENQURSS HCSDSIDITLUNGIN-GESSG-BRIDLATHKANI -FELLINIFPLENQURSS HCSDSIDITLUNGIN-GESSG-BRIDLATHKANI -FELLINIFPLENQURSS HCSDSIDITLATHCIN-GESSG-BRIDLATHKANI -FELLINIFPLENQURSS HCSDSIDITLATHCIN-GESSG-BRIDLATHKANI -FELLINIFPLENQURSS HCSDSIDITLATHCIN-GESSG-BRIDLATHKANI -FELLINIFPLENQURSS
NP 832758046.1 NP 852824653.1 NP 842596613.1 NP 8642596613.1 NP 8641166484.1 NP 864116648.1 NP 864116648.1 NP 864116648.1 NP 86312618.1 NP 863420839.1 NP 863420839.1 NP 8645171.2 NP 8613627518.1 NP 813420839.1 NP 816791534.1 NP 813420837.1 NP 813420837.3 NP 813420837.3 NP 813420837.3 NP 813420837.3 NP 813420837.3 <td>EQHVELYQCYSINISARYLGIRLLTPTOTPENLSFIVTGVIRQUIQGOGLQGFRFSA EQHVELYQCYSINISARYLGIRLLTPTOTPENLSFIVTGVIRQUIQGOGLQGFRFSA EQHVELYQCYSINISARYLGIRLLTPTOTPENLSFIVTGVIRQUIQGOGLQGFRFSA EQHVELYQCYSINISARYLGIRLLTPTOTPENLSFIVTGVIRQUIQGOGLQGFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUIQGOGLQFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUIQGOGLQFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLLAPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLLAPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLIAPSOTPENLSFIVTGVIRQUISRGGETEGFRFSA EQHVELYQCYSINISARYLSINILLTPSOSFENLSFIVTGVIRQUISRGGETEGFRLSA EQHVELYQCYSINISARYLSINILLAPSOSFENLSFIVTGVIRQUISRGGETEGFRLSA EQHVELYQCYSINISARYLSINILAPSOSFENLSFIVTGVIRQUISRGGETEGFRLSA EQHVELYQCYSINISARYLSINILAPSOSFENLSFIVTGVIRQUISRGETEGFRLSA EQHVELYQCYSINISARYLSINILAPSOSFENLSFIVTGVIRQUISRGSETEGFRLSA EQHVELYQCYSINISARYLSINILAPSOSFENLSFIVTGVIRQUISRGSETEGFRLSA EQHVELYQCYGNISARYLSINILAPSOSFENLSFIVTGVIRQUISRGSETEGFRLSA EQHVELYQCYGNISARYLSINILAPSOSFENLSFIVTGVIRQUISRGSETEGFRLSA</td> <td>NP 001200400. NP 001301071. NP 005500676. NP 005500676. NP 005500676. NP 0057061116. NP 0057061116. NP 005706111. NP 055707.1 NP 055707.1 NP 052024053.1 NP 052024053.1 NP 052024053.1 NP 052024053.1 NP 00116404. NP 00454959513. NP 00116404. NP 00343172. NP 003423823.</td> <td>HCSDSONTAQUIN-6F595-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAQUIN-6F595-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F595-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-</td>	EQHVELYQCYSINISARYLGIRLLTPTOTPENLSFIVTGVIRQUIQGOGLQGFRFSA EQHVELYQCYSINISARYLGIRLLTPTOTPENLSFIVTGVIRQUIQGOGLQGFRFSA EQHVELYQCYSINISARYLGIRLLTPTOTPENLSFIVTGVIRQUIQGOGLQGFRFSA EQHVELYQCYSINISARYLGIRLLTPTOTPENLSFIVTGVIRQUIQGOGLQGFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUIQGOGLQFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUIQGOGLQFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLLTPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLLAPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLLAPSOTPENLSFIVTGVIRQUISQGETEGFRFSA EQHVELYQCYSINISARYLSIQLIAPSOTPENLSFIVTGVIRQUISRGGETEGFRFSA EQHVELYQCYSINISARYLSINILLTPSOSFENLSFIVTGVIRQUISRGGETEGFRLSA EQHVELYQCYSINISARYLSINILLAPSOSFENLSFIVTGVIRQUISRGGETEGFRLSA EQHVELYQCYSINISARYLSINILAPSOSFENLSFIVTGVIRQUISRGGETEGFRLSA EQHVELYQCYSINISARYLSINILAPSOSFENLSFIVTGVIRQUISRGETEGFRLSA EQHVELYQCYSINISARYLSINILAPSOSFENLSFIVTGVIRQUISRGSETEGFRLSA EQHVELYQCYSINISARYLSINILAPSOSFENLSFIVTGVIRQUISRGSETEGFRLSA EQHVELYQCYGNISARYLSINILAPSOSFENLSFIVTGVIRQUISRGSETEGFRLSA EQHVELYQCYGNISARYLSINILAPSOSFENLSFIVTGVIRQUISRGSETEGFRLSA	NP 001200400. NP 001301071. NP 005500676. NP 005500676. NP 005500676. NP 0057061116. NP 0057061116. NP 005706111. NP 055707.1 NP 055707.1 NP 052024053.1 NP 052024053.1 NP 052024053.1 NP 052024053.1 NP 00116404. NP 00454959513. NP 00116404. NP 00343172. NP 003423823.	HCSDSONTAQUIN-6F595-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F595-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-BROLATHONIA-FFILINATPLEAQUESS HCSDSONTAUN-1-6F596-
NP 832758045.1 NP 852824553.1 NP 84259631.3 NP 84259631.4 NP 84259631.4 NP 84259631.4 NP 84551618.1 NP 84551618.1 NP 84551618.1 NP 85352554.1 NP 89342833.3 NP 801443171.2 NP 8014791534.1 NP 8154257.1 NP 81542554.3 NP 81542537.3 NP 81542537.3 NP 81542537.3 NP 912891536.3 NP 91390381.1 NP 813903451.3	EQHVELYQCYSINISARYLGHRLLTPTOTPENLSFIDYTGWRQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLGHRLLTPTOTPENLSFIDYTGWRQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLGHRLLTPTOTPENLSFIDYTAWNQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLGHRLLTPTOTPENLSFIDYTWRQKLIQGOGLQGFRFSA EQHVELYQCYSINISARYLSHLLISPTOTPENLSFIDYTWRQKLIQGOELQGFRFSA EQHVELYQCYSINISARYLSHLLISPTOTPENLSFIDYTWRQKLIQGOELQGFRFSA EQHVELYQCYSINISARYLSHLLISPTOTPENLSFIDYTWRQKLIQGOELQFRFSA EQHVELYQCYSINISARYLSHLLISPTOTPENLSFIDYTQWRQKLSQGEELEGFRFSA EQHVELYQCYSINISARYLSHLLIPSOTPENLSFIDYTQWRQKLSRGEELEGFRFSA EQHVELYQCYSINISARYLSHLLIPSOTPENLSFIDYTQWRQKLSRGEELEGFRFSA EQHVELYQCYSINISARYLSHLLIPSOTPENLSFIDYTQWRQKLSRGEELEGFRFSA EQHVELYQCYSINISARYLSHLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLLIPSOSPENLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSFSDYGFRLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSFDSFRLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSFDSFRLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSFDSFRLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSHCSHCYCHARDISFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHLSFDSFRLSFIDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHCSHCYCHARDISFRCYCHARDISFRCYHARDISFRCYHRXISHRUSHRYLSHRLSFDYTQWRQKLSRGEELEGFRLSA EQHVELYQCYSINISARYLSHRLISHCYCHARDISFRCYHRXINGKARDISFRCYHRXINGKARDISFRCYHRXIXIII STDYFGHYTGWRQKLSRGEELEGFRLSA	NP 0012081400. NP 0013818071. NP 043755182. NP 043755182. NP 043755182. NP 043755182. NP 043755182. NP 0453461104. NP 0457967.1 NP 055707.1 NP 055707.1 NP 055707.1 NP 055707.1 NP 055707.1 NP 04559046. XP 042240653. NP 045590613. NP 040590613. NP 040590513. NP 040590513. NP 053523548. NP 053523548. NP 024433177.	HCSDSONTAQUDI-6FSS-BROLATIKANI-FFLLINATPLENQUARS HCSDSONTAQUDI-6FSF-BROLATIKANI-FFLLINATPLENQUARS HCSDSONTAQUARS-SS-BROLATIKANI-FFLLINATPLENQUARS HCSDSONTANA-IN-6FSF-BROLATIKANI-FFLLINATPLENQUARS HCSDSONTANA-IN-6FSF-BROLATIKANI-FFLLINATPLENQUARS HCSDSONTANA-IN-6FSF-BROLATIKANI-FFLLINATPLENQUARS HCSDSONTANA-IN-6FSF-BROLATIKANI-F

XP #11942277.1	HCSCDSKDWTLQV03NvGFTTG-RRGDLATIRGMRPFLLUMATPLERAQHLQSS	The Construction of the	NAME AND ADDRESS OF A DESCRIPTION OF A D
XP 054396445.1		XP_048598613.1	BHRR-ALDTRYCESSTERNCCVRQLYTDERKDLGWWIEHEPKGYHANFCLGPCPYINS
		NP_001166494.1	RHR-GLDTNYCFSSTEKNCCVRQLYIDFRKDLGWKWIHEPKGYNWNFCLGPCPYIWS
XP_012091536.1		XP_004648529.1	RHRR-ALDTWYCFSSTERNCCVRQLYIDFRKDLGMKWIHEPRGYHWNFCLGPCPYINS
NP_001305385.1	HCPCEM6PGHADEMRISIE-GFEQQRGDMQSIAKKHRRVPYVLAMALPAERANELHSA	XP 046311618.1	RHRR-ALDTRIVCFSSTEKNCCVROLYTOFRKDLGWMITHEPKGYHWIFCLGPCPYINS
XP 030075415.1	PCSCHKSTEDLKETEA-GMENKRGDLGPLESPNQPVLLEPVTPEDHAIPILQSS	XP 053523548.1	RHRR-ALDTINYCFSSTERNCCVRQLYIDFRKDLGWWIHEPKGYHANFCLGPCPYINS
XP 059574704.1	HCSCE-DOHSTAETLFOID-GFETFREDMOKWONOGORLPFLLATATPPERAAOLOSP		
an account of the	*** 1.*** ***** 1 1 *15* * 1** *1**	XP_024433172.2	RHRR-ALDTSYCFSSTEKNCCVRQLYIDFRKDLGM/WIHEPKGYHM/FCLGPCPYINS
	· 영상 · 영상 영 · 영상 · 영상 · 영상 · 영상	XP_003420839.1	RHRR-ALDINYCFSSITEKNCCVRQLYIDFRKDLGWWLHEPKGYHANFCLGPCPYINS
		NF 001075318.1	RHR-ALDTNYC5SSTEKNCCVRQLYIDFRKDLGMXIJHEPKGYHANFCLGPCPYIN5
NP_000651.3	8HRR-ALOTIVICESSTEKNECVRQLYIDERKDLGNKWTHEPK6YHANECLGPCPYTHS	XP 016791534.1	RHRR-ALDTWYCFSSTEKNCCVRQLYIDFRKDLGMWIHEPKGYHWIFCLGPCPYINS
XP_014979511.1	RHER-ALDTWYCFSSTEKNICCVRQLYIDFREDLEMINTHEPKEYHANFCLGPCPYINS	XP 011942237.1	RHRR-ALDINYCFSSTEKNCCVROLYIDFRKDLGWKWIHEPRGYHAMECLGPCPYINS
XP 054322057.1	RHRR-ALDTINYCFSSTERNCCVRQLYIDFRKDLGMKWIHEPKGYHWAFCLGPCPYINS		
XP 018870688.1	RHRR ALOTHIYCF - SSTERNCCVROLYIDERRDLGWKNTHEPKGYHWIECLGPCPYINS	XP_054396445.1	RHRR-ALDINYCESSTEKNCCVRQLYIDERKDLGWKWEHEPKGYHWNECLGPCPYINS
XP 024782266.2	8HRR-ALOTWYCESSTERNCCVROLYIDERKDLENKWIHEPKGYHWNECLGPCPYINS	XP_012891536.1	RHIR-ALDTNYCFSSTEKNCCVRQCYTDFRKDLGMKWIHEPKGYHANFCLGPCPYTNS
10 10 m 10 10 10 10 10 10 10 10 10 10 10 10 10		NP_001305385.1	BRRR-DLDTDYCFGPGTDEKNCCVRPLYIDFRKDLQMKKIHEPKGYPWNFCRGPCPYINS
XP_038512824.1	RQRR-ALOTNYCFSSTERNCCVRQLYIDFRKDLGWONTHEPRGYHANFCLGPCPYINS	XP 030075415.1	RRKR-AVMMEYCSVLEEKNCCVRPLFINFRXDLGAXAUHEPKGYYSNFCMGPCPYINS
XP_000941294.1	RHRR-ALDTRYCESSTEKNCCVRQLYIDERKDLGNKWIHEPKGYHANECLGPCPYINS	XP 059574704.1	RHRRAAADTAYCFGTEEKNCCVRKLYTDFRODLKMAATHEPRGYMHLCTGACPYTMS
XP 826338822.1	RHIR-ALOTIVICESSTERNECVRQLYIDEREDLGWRNTHEPKGYHWIFELGPCPYTHS	W_00000400411	
XP 827475376.1	IHRR-ALDTWYCFSSTEKNCCVROLYIDFRRDLGMKWTHEPKGYHMNFCLGPCPYTHS		COND. Diez Dissessionandere construction MRD diseases
XP 847568886.1	INNR-ALOTINICESSTERNCCVROLVIDFRKDLGWNINEPKSYNANFCLGPCPYINS		
		MP_000651.3	LOTOYSKV-LALYNOHNPGASAAPCCVPQALEPLPIVYYVGRKPKVEQLSMMI
XP_059006887.1	IHRR-ALDTRYCFSSTERNCCVRQLYIDFRKDLGARNIHEPKGYRAMFCLGPCPYIAS	XP 014979511.1	LOTOYSKV-LALYNOHNPGASAAPCCVPOALEPLPTVYYVGRKPKVEDLSIMT
XP_003700380.1	INAR-ALOTINYCFSSTERNCCVRQLYIDFRKDLGWINITHEPKGVIWNFCLGPCPYINS	XP 054322857.1	LOTOYSKV-LALVNOHNPGASAAPCCVPOALEPLPLVVVVGRKP
XP_059855830.1	AHHR-ALDTINICFSSTERNCCVRQLVIDFRRDLGMRNIHEPRGYHWNECLGPCPVIWS		
XP 022441295.1	INNR-ALOTWYCFSSTERRCVRQLYIDFRRDLGMRNTHEPROVHANECLGPCPYINS	XP_018870688.1	LDTQYSKV-LALYNQHNPGASAAPCCVPQALEPLPTVYYVGRKPKVEQLSMMT
XP 023988445.1	RHRR-ALOTINYCFSSTEKINCCVROLYIDFRKDLOWKWIHEPKKYHWAFCLGPCPYTWS	XP_024782266.2	LDTQYSKV-LALYNQHNPGASAAPCCVPQALEPLPTVYYVGRKPKVEQLSIMT
	IHRR-ALDINYCFSSTEKNCCVR0LYIDFRKDLGMXIHEPKGYWAFCLGPCPYIWS	XP 038512824.1	LDTQYSKV-LALYNOHNPGASAAPCCVPQALEPLPEVYYVGRKPKVEQLSMMI
XP_059988434.1		XP 006941294.1	LDTQYSKV-LALWIGHNPGASAWPCCVPQALEPLPEVYYVGRKPKVEQLSIMT
XP_004271297.1	RHRR-ALOTINYCFSSTERNCCVBQLYIDFRKDLUMINIHEPKSYHWNECLGPCPYTWS	XP 026338022.1	LDTQYSKV-LALYNQHNPGASAAPCCVPQALEPLPIVYYVGRKPKVEQLSMMI
XP_826935427.1	RHR-ALDTWYCFSSTERNCCVRQLVIDFREDLENKNIHEPKGVHARFCLGPCPVINS		
NP 999188.2	INHIR-ALDTINYCF SSTERDICCVRQLYIDFREDLGMRNTHEPREVHARECLGPCPYINS	XP_027475376.1	LDTQYSRV-LALYNQHNPGASNAPCCVPQALEPLPTVYYVGRKPKVEQLSNMT
NP 081009400.1	INRR-ALDINIVCE SSTERNICCVRQLYIDERRDLGWRWIHEPKGYHWIECLGPCPVEWS	XP_047568006.1	LOTQYSKV-LALVNQHNPGASAAPCCVPQALEPLPIVYYVGRKPKVEQLSMPI
NP 081301071.1	INNR-ALDTINCESSTERNCCVROLVIDEREDLOWINIHEPKGYHANECLOPCPVINS	XF 059006887.1	LOTQVSKV-LALVNOHNPGASAAPCCVPQALEPLPIVYYVGRKPKVEQLSMMI
CONCERNING CONTRACTOR FOR		XP 033700380.1	LOTOYSKV-LALYNOHNPGASAAPCCVPQALEPLPEVYYVGRXPKVEQLSIMT
XP_005890676.1	HHIR-ALDTINYCESSTERNCCVRQLYIDERKDLGWKWIHEPKUYHANECLGPCPYIWS	XP 059855830.1	LDTDYSKV-LALYNDHNPGASAAPCCVPCALEPLPTVYVVGRKPKVEDLSIMT
XP_843755182.1	HHRR-ALDINYCFSSTERNCCVRQLYIDFRKDLGNKNIHEPKGYHANFCLGPCPYINS	100.000 C = 500 C = 11.	
XP 008986313.1	IHRR-ACOTIVICESSTEKINCCVRQLYIDERKDEGWKNTHERKGYHWIECLGPCPYIWS	XP_022441296.1	LOTQYSKV-LALVNQHNPSASAAPCCVPQALEPLPTVYYVGRXPKVEQLSAMT
XP 453461166.1	IHAR-ALDTWYCFSSTEKNCCVRQLYIDFRKDLGWINHEPKGYHWNFCLGPCPYINS	XP_023988445.1	LOTQYSKV-LALYNQHNPGASAAPCCVPQALEPLPTVYYVGRKPKVEQLSMMT
XP 884479972.4		XP_059988434.1	LDTQY5KV-LALYNOHNPGASAAPCCVPQALEPLPTVYYVGRXPKVEQLSMMT
NP 035707.1	IHRR-ALDTINCE SSTERNCCYRQLYIDFRRDLGWKNIHEPKGYWWFCLGPCPYIWS	XP 004271297.1	LOTOYSKV-LALVNOHNPGASAAPCCVPOALEPLPTVYYVGRKPKVEQLSM/I
1715-000 C C C C C C C C C C C C C C C C C C		XP 026935427.1	
NP_867589.1	RHRR-ALDTINYCFSSTERNCCVRQLYIDFRKDLGNKNTHEPKGVHMIFCLGPCPYINS		
XP_832758846.1	INNR-ALDTWYCESSTERNCCVRQLVIDERKDLGMINITHEPKGVHWAFCLGPCPYTHS	NP_999180.2	LDTQYSKV-LALYNQHNPGASAAPCCVPQALEPLPIVYYVGRKPKVEQLSMMI
XP 852824653.1	INNER-ALDTWYCESSTERNCCVRQLVIDERRDLGWINTHEPKGYWWFCLGPCPVTW5	NP_001005400.1	LDTQYSKV-LALYNQHNPGASAAPCCVPQALEPLPIVYYVGRKPKVEQLSMMI
XP 027286179.1	INHER-ALIDTINYCESSTEKNOCVROLYIDERKOLOMINTHEPKOVHANECLOPOPYTIKS	NP_001383871.1	LOTQYSKV-LALYNOHNPGASAAPCCVPQALEPLPTVYYVGRKPKVEQLSNMI
XP 005890676.1		0.538	
XP 043755182.1			
XP 080286313.1	LOTQY5KV LALY9QH INPGASAAPCCVPQALEPLPTVYYVQRKP KVEQ	6.50P	
XP_053461166.1		6.508	
XP 084479972.4			
HP_035707.1	LOTQY5KV-LALY9QH		
MP_067589.1	LOTQY5KV - LALYNQH NPSASASPCCVPQALEPLPTVYYVGRKP KVEQ		
XP_032750046.1			
XP_852024653.1			
XP_027286179.1			
XP_040590613.1			
NP_901166494.1			
XP_004648529.1 XP_046311618.1			
XP_046311618.1 XP_053523548.1			
XP_024433172.2			
XP 803428839.1			
NP 001075318.1			
XP_016791534.1			
XP_011942237.1			
XP_054396445.1	LOTQV5KG · LAESPARMQM · · · · · · · · · · · · · · · · · ·		
XP_012891536.1			
MP_001305385.1			
XP_030075415.1			
XP_859574704.1		1.54	
	Ay8424 i		
In contract	Children .		
NP_00051.3 XP 014979511.1	VRSCKCS		
XP_016979511.1 XP_054322957.1			
XP 918878688.1			

0.0000000

VRSCICS
VRSCKCS
VRSCKCS
VRSCICCS
VRSCICCS
VRSEKCS
VRSCKCS
VRSCICS
VRSCICCS
VR-SCKC5
VIC-SEKCS

010040072.4

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 distancebased 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.

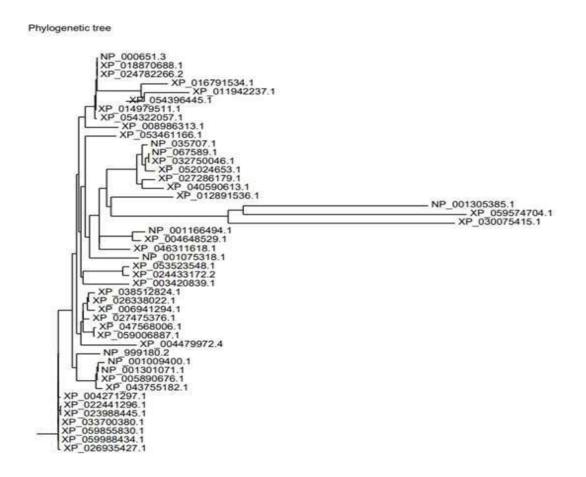


Fig.3.1: Phylogenetic tree of some microorganisms

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

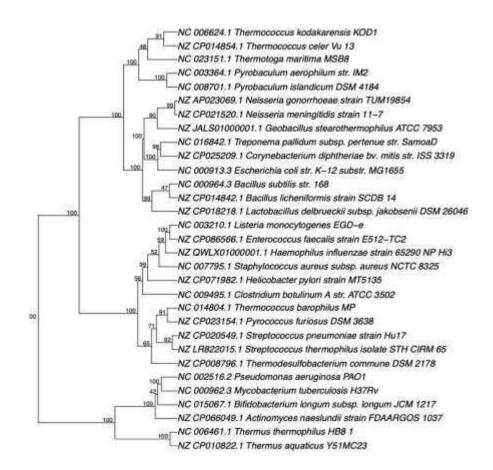


Fig.3.2: Phylogenetic tree shows some microorganisms present in gut microbiota

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].

			Seon	d letter			
		U	с	A	G		
	U	UUU]Phe UUC] UUA UUG]Leu	UCU UCC UCA UCG	UAU UAC] Tyr UAA Stop UAG Stop	UGU UGC UGA Stop UGG Trp	U C A G	
First letter	с	CUU CUC CUA CUG	CCU CCC CCA CCG	CAU CAC] His CAA CAA] Gin	CGU CGC CGA CGG	U C A G	Third letter
First	A	AUU AUC AUA AUG Met	ACU ACC ACA ACG	AAU AAC AAA AAG] Lys	AGU AGC] Ser AGA AGG] Arg	U C A G	tter
	G	GUU GUC GUA GUG	GCU GCC GCA GCG	GAU GAC GAA GAA GAG GIu	GGU GGC GGA GGG	U C A G	

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].

- 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 (<u>http://cousin.ird.fr</u>) 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

	_	-	_	The second second	WAR OF	COLOR AND AND	North Colorest	ALC: NO.	Trent and	and the second second	-			-				
	y star star		1	Relati	ive Syr	ionimo	us Cod	on Usa	ige (RS	CU)			12.44	_				
IS TTT TTC TTA TTG	CTTICTC	TACT	GATT		Constanting start	and the second second	And in case of the local division of the loc	- Contraction of the	forestand strength	and second s	distant in the local distant	-	CAC	-	-	CCAC	AACO	GCT
ids F F L L 0.731.270200.40	L L	L L	and the second second	1 1	-	/ V 17 0.83	V 5	5	S S	S 1.80 4	5 P	P	-	-	and the second	T T 33 0.0	T 1.33	A 1.26
ALL PROPERTY.																		
	cobons	TTT	TTC	TTA .	TTE	CTT	to inter CTC	CTA	CTG	ATT	ATC	ATA	- 23					
	GTT CCA CAT	GTC CCB CAC	GTA ACT CAA	GTO ACC CAS	TCT ACA AAT	ACG AAC	TCA OCT AAA	TCG OCC AAG	AGT GCA GAT	40C 0C0 6AC	CCT TAT GAA	CCC TAC 6A6						
	TGT AMINOAC		CGT F	CGC F	664 L	C00 L	A6A L	466 L	66T L	66C L	66A I	666 I	- 8					
		V.	V	V	V	5	5	5	5	S A	5 A	P						
	P	P H	P H	T Q	T Q	TN	TN	A K	ŝ	ŝ	D	E						
		250			T Q R				- 112				•					
		H	H.				N		- 112	D	D	E	•					
		H	H.				N		- 112	D	D	E	•					
		H	H.		9	N R	N	K R	K R	D	D	E	•					
	P	H C	H C	Q R	CODO	N USA	N R	K R	K R	D S	D	8	•			lace		
DONS TIT TIC TTA coacids F F L	P Y E TTG CTT L	H C C C C C C C C C C	H C TA CTG L L	Q R		N USA	R GE PER TC GTA V V	K R THOU GTG 1	K R USAND CT TCC S S	D B TCA TC	D G G AGT S S	E B AGC C	p p	P	ACCC	T	ACC A	т
	P Y E TTG CTT L	H C C C C C C C C C C	H C TA CTG L L	Q R		N USA	R GE PER TC GTA V V	K R THOU GTG 1	K R USAND CT TCC S S	D B TCA TC	D G G AGT S S	E B AGC C	p p	P		T	T	ACA A T 2.00 6
oacids F F L	P YE TTG CTT L L 6 29 34 59	H C C C C C C C C C C C C C C C C C C	H C TA CTG L L 00-40.88	Q R ATT A 1 22:010 TTA	8 CODC ITC ATA 1 1 43 0 29 TT0	N R ON USA GTT G V 34 59 22	H R GE PEI TC GTA V V 2.01 15.7	THOU GTG T V 3 14 0	K R ISAND CT TCC S IS .00 0.00 CTS	D 8 TCA TC 5 S 0.00 0	D G G AGT 8 S 00 9 43	AGC C	p p	P		T	T	т
oacids F F L	P Y E TTG CTT L L L 6.2934.59 StT CCA CCA	H C C L 9.430	H C TA CTG L L L L OO 40 88 Original TTC GTA ACT CAA	Q B ATT A 1 22.010 TTA 0TG ACC CAG	CODC TTC ATA I I 43 0 29	N USA GTT G V 34 59 22 Corport TCC Acco	N 8 GE PEI TC GTA V V V V COI 15.7 CC CC CC CC CC CC CC CC	K R THOL GTG T 3 14 0 CTA TCS 0CC AA6	K R SAND CT TCC S S S 000 0 00 CTS ABT SCAT	0 8 TCA TC S 5 0 00 0 4 TT 400 6640	CGAGT 8 S 00 9.43	AGC C	p p	P		T	T	т
oacids F F L	P Y E TTG CTT L L C CODO 871 CCT CCT C CODO 871 CCT CCT C TG T TG TG TG TG TG TG	H C C L I 9 43 0 NS TTT arc cac tac tac cac TBC	H C TA CTG L L L 00 40 88 100 40 88	Q R ATT A 1 22.010 TTA GTG ACC	CODC TTC ATA 1 1 43 0 29	N USA GTT G V 34 59 22 CTT TCC	H R GE PEI TC GT/ V V R01 15.7	K R THOL GTG T V 2 3 14 0 CTA TCS GCC	K R ISAND CT TCC S S 0000000 CTS ABT	0 8 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	CGAGT 8 S 00 9 43	AGC C S 22.01 1	p p	P		T	T	т
oacids F F L	P Y E TTG CTT L L C CODO 871 CCT CCT C CODO 871 CCT CCT C TG T TG TG TG TG TG TG	H C C L 9.43 0	H C TA CTG L L L L OO 40 88 Original TTC GTA ACT CAA	Q B ATT A 1 22.010 TTA 0TG ACC CAG	CODC TTC ATA I I 43 0 29	N USA GTT G V 34 59 22 Corport TCC Acco	N 8 GE PEI TC GTA V V V V COI 15.7 CC CC CC CC CC CC CC CC	K R THOL GTG T 3 14 0 CTA TCS 0CC AA6	K R SAND CT TCC S S S 000 0 00 CTS ABT SCAT	0 8 TCA TC S 5 0 00 0 4 TT 400 6640	CGAGT 8 S 00 9.43	AGC C	p p	P		T	T	т

		150943396 415094339				49956684	52,201257	861635 44	65488858	314 47.7	987421383	65				1111	
							N USAGI										
ODONS TIT TTC TTA TTG inoacids F F L L 111 19 1 2	L L 11 3	CTACTO L L 0 13	1	TC ATA 1 1 3 2	GTT GTC V V 11 7	April 1997 State Street	1 5	C TCA T 5 5 0 0	CGAGT	AGC CC S P 7 1	P 1	CAICCG P P 3 2	ACT ACC T T 2 2	ACA ACC	A A	1	
	CODON GTT CCA CAT TGT AFG APLINC I P Y	S TIT OTC CCS CAC TOO HCLDS V P H	YCAGUET III TTC QTA ACT CAA CGT F Y P H	TTA STO ACC CAG CSC F V T Q	TTG TCT ACA AAT CGA	CTT TCC AC6 AAC COS L S T N	TCA 1 GCT G AAA A	TA CTI CQ AS CC BC AS GA GG GG L S	ATT AGC 608 640	ATC CCT TAT GAA GGA I S A D	ATA CCC TAC 546 605 I P V E						
						1.1	Nelace (Del)	E COM	POSITIC	ол							
EQUENCES ARAMETERS 954 174	с т	G %/	4 %C	%Т	%G %	G+C %(T%A+T	%A+C %	C+TA	1 C1 T1	G1 %A	1 %C1			C1 %G1+	
		49.8566	NG+A NG1 NC2 G3 S 954 8251572 8377358	RAMETERS NG+T NG1+C1 NT2 NA3 174 3 56.700 5 44	1 lengtl %A+T %61+A: %62 %C3 229 35953878- 49	1 A NA+C 1 VG1+T NG2+C NT3 239 41 50.94 91		T A1 %A1+C1 %G2+T2 %G3+C3 18.238 5 57.756 13.836	6 C1 %C1+T1 %A2+T2 %03+A3 993718692 \$13417191 \$77987421	NA T1 A2 NA2+C2 NG3+T3 24.004) 43.2914 28.616	%C 61 C2 %C2+T2 %A3+T3 19287211 48461215 35229125	\$T \$A1 T2 A3 \$A3+C3 25.052 42.243 15.408	186582889 985831447				
						1112	CODO	ON USA	GE								
		1172	1217							0.011		11100	12741				
NS TTT TTC TTA TTG		-	TG ATT			GTC G	TA GTG			GAGT	and the second second		CCACC				
1.00 1.00 0.46 0.23	and a second	and the second second	08 1.00		00 0.67	0.89 0	and a second second	1.09 1.09	And and a state of the state of	27 0.82	1.64 0.	1			All all and a second	a submitted and a	and the second se
	CODOR GTT CCA CAT TGT	GTC CCG CAC TGC	TTC GTA ACT CAA CGT F	TTA GTG ACC CAG CGC F	TCT ACA AAT CGA	TCI ACI AAI CGI	C TCA 5 GCT 5 AAA	GCC AAG AGG	CTG AGT GCA BAT GGT	ATT AGC GCG 0AC 66C	ATC CCT TAT GAA GGA I	ATA CCC TAC GAG GGG					
	I P Y E	V P H C	F V P H C	F V T Q R	L V T Q R	LSTNR	5 T N R	L S A K R	L S A K R	L S A D G	S A D G	I P Y E G	•				
	TT CT	CTA	CTG A	1000		100			AND	TCA	TCG AG	TAGO	CCTIC	CCICCA	CCG	CTACC	ACA AC
TTT TTC TTA TTG C	Contraction of the local division of the loc	L	L	1	11	VN	/ V	V s	S	S	S S	S	P	P P	P	T T 94 19 84	TI
And the second s	90-43.6	00.001-			1400	Une this au	tput to capy						100				
F F L L 15.87 15.87 7 94 3.97 11	90 43 6	40 TTT 1 STC 0 SCG 4 CAC 0	TTC STA ACT CA4	TTA GTG ACC CAG	TTG TCT ACA AAT	CTT TCC ACG AAC	CTC TCA GCT AAA	CTA TCO GCC AAG	CTG AGT GCA GAT	ATT AGC GCG GAC	ATC CCT TAT GAA	ATA CCC TAC GAG					
F F L L 15.87 15.87 7 94 3 97 11	90 43 6	50 570 0 570 0 506 4 506 0 506 0 506 0 50	TTC STA ACT	TTA GTG ACC	TTG TCT ACA	TCC ACG	TCA GCT	TCO GCC	AGT GCA	AGC GCG	CCT TAT	CCC TAC					

								1.4.2.4			NUCL	EOTID	ECOM	POSIT	ON	LALKE	1.1.2	λų p						
EQUENCES \	length	A	c	т	G	%A	%c	%т	%G	%G4	+C %G+	+A %G-	T %A+1	%A+C	%C+TA	1 C1 T1	G1 %A	1 %0	1 %	T1 %G1	%G1+(1 %G1+A	1 %0	31+T
	756	180	207	147	222	23.81	27.38	19.44	29.3	7 56.7	75 53.1	17 48.8	1 43.25	51.19	46.83 6	5 65 39	83 25.	79 25,7	79 15	48 32.94	58.73	58.73	4	8.41
					1			Chapter	ter toth d	whether the	format 2	he that an	treat he copy	and paste	DI SUNE JPE	formed appla	when:							
					SEQU NG+0	UENCES	S \ PAR	AHETER NG+T	S le	ngth	A SAHC	C %C+T	T A1	G C1	%A T1	%C 61	ST SA1	%G %C1	1					
					ST1		VG1	%G1+0			%91+71	SA1+T1		SC1+T1		C2	TZ	G2						
					\$42		\$C2	\$72	36		\$62+62	%62+A2		%A2+T2			A3	C3						
					13	5+C3s	13	NA3	\$0	3	\$73	\$63	\$63+03	%63+A3	NG3+T3	%A3+T3	MA3+C	NC3-	eta 🛛	1.1.1.1				
							756	150	28		147	222	23.899	\$2300952	4 27.388	95238895	2 19.44	644444	4444	1				
							365079		68317		53.1746	0317468		52388952 65879365		95825396		8476196	C 200 (10)					
									39 01587		99 58.7381	5873815				65079365 84126984		198474 7301587						
					(Sector)		1111	-			and the second		Carl Color		- All Carry		100000	-		1111				
			12		100	200	1177	11.07	12.00	1.1.1	(Cont)		-	11110	12.5.5.5	111000	1000	Price.	19.00	1.000				
			12									CODO	DN USA	GE										
														CU.U						11141				
CODONE T	-	TTO		-	FIGT	-	Alexe	larr	arela	TALA	-	- TA	melnes	izeelz	a meet	ACTIAC	Clean	enel	reak	reelact	Incel	calace	COY	nee
alterization and a second second	and the second se	-	and the second second	distant in the	-	and so the	ACTO	ATT	ATC /	كالمشارقة ستنجارهم	فيهددنك وتصليحا	And the owner where the party is not the party of the par	stands in the second	and the second states in the	a hard a standard of the standard of the	AGTAG	and minderside	ccc	CCA	and in case of the local division of	ACC	CAACG	in the local division of	distant sectors
CODONS TT minoacids F	F	TTA L 2	TTG L	CT L 3	L	. L	L	ATT	ATC 4	1 1	TT GTC V V 3 4	GTA	V S	and the second states in the	CA TCG	AGT AG	and minderside	P	P	P T	T	T T 4 3	GCT A 3	GCC A 6

relative o	ynonimou	s codon c	rsage (Nou	u)

	CODONS					TIT	TTC	TTA	TTG	CTT	CTC	CTA	CTG	ATT	ATC	ATA	GTT	GTC	GTA	GTG	TCT	TCC	TCA	TCG
	Aminoaci	ls.				1.	F	L	L	L	L	L	L	040	1.1	1.1	V	V	V	V	S	5	5	5
IC_000002.12:c135837169- 35829593.c135824003.135 35817341.c135812956-135 35807128.c135805057-135 35800607.c135788138-135 35790658.c135789798-135	823901.c13582 812311.c13580 804767.c13580 798029.c13579 789571.c13578	2098-1 9993-1 4128-1 4775-1 8544-	358220 358084- 358039 357946-	20,c1358 43,c1358 30,c1358 41,c1357	318061- 107396- 300809- 790881-	0.99	1.01	0.11	0.64)	D.491	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
35788324_Homo_sapiens_	chromosome_2	GRO	n.36.p14	Zictimen)	Alaseum	91	1																	-
35788324_Homo_sapiens_	chromosome_2	GRC			und format. D		iquet to	CONN. D	oni pari	e in pos	or regi	berned a	pyshare	20019										
35788324_Homo_sapiens_	chromosome_2 copons	111	"Output A	tab dicion TTA	und format. D	e tên m CTT	CTC		CTA .	CT	FG	ATT	-	ATC	AT		2							
35788324_Homo_sapiens_	CODONS BTT	TTT QTC	"Ovent i	tub oklow TTA 070	and format. D TTG TCT	e Min ei CTT TCC	CTC TCA		CTA TCO	CT AG	TG ST	ATT AGC		ATC	00	¢	1							
35788324_Homo_sapiens_	CODDHS 017 CCA	TTT OTC CCG	*Cumul A TTC ATA ACT	TTA 070 ACC	ted format. D TTG TCT ACA	e Min in CTT TCC ACG	CTC TCA GCT		CTA TCO BCC	CT 40 60	76 37 54	ATT AGC GCG		ATC CCT TAT	CC TA	ć C	1							
35788324_Homo_sapiens_	CODONS BTT	TTT QTC	"Output A	TTA 070 ACC CA3	tud format. D TTG TCT ACA AAT	e Min er CTT TCC ACG AAC	CTC TCA		CTA TCO BCC	CT AG GC	FG 57 54	ATT AGC GCG		ATC CCT TAT	CC TA QA	¢ C G	i							
35788324_Homo_sapiens_	CODORIS GTT CCA CAT TGT	TTT OTC CCG CAC TGC	*Cumul A TTC ATA ACT	TTA 070 ACC	ted format. D TTG TCT ACA	e Min in CTT TCC ACG	CTC TCA GCT		CTA TCO	CT 40 60	FG 57 54	ATT AGC		ATC	CC TA	¢ C G	-							
35788324_Homo_sapiens_	CODDHS 017 CCA	TTT OTC CCG CAC TGC	*Cumul A TTC ATA ACT	TTA 070 ACC CA3	tud format. D TTG TCT ACA AAT	e Min er CTT TCC ACG AAC	CTC TCA GCT		CTA TCO BCC	CT AG GC	FG 57 54	ATT AGC GCG		ATC CCT TAT	CC TA QA	¢ C G	1						1222222222	
35788324_Homo_sapiens_	CODORIS GTT CCA CAT TGT	TTT OTC CCG CAC TGC	*Cumul A TTC ATA ACT	TTA 070 ACC CA3	tud format. D TTG TCT ACA AAT	e Min er CTT TCC ACG AAC	CTC TCA GCT		CTA TCO BCC	CT AG GC	FG 57 54	ATT AGC GCG		ATC CCT TAT	CC TA QA	¢ C G	•							
35788324_Homo_saplens_	CODORIS GTT CCA CAT TGT	TTT OTC CCG CAC TGC	*Cumul A TTC ATA ACT	TTA 070 ACC CA3	tud format. D TTG TCT ACA AAT	e Min er CTT TCC ACG AAC	CTC TCA GCT		CTA TCO BCC	CT AG GC	FG 57 54	ATT AGC GCG		ATC CCT TAT	CC TA QA	¢ C G								

CODONS	TTT	TTC	TTA	TTO	CTT	CTC	CTA	CTG	ATT	ATC	ATA	GTT	GTC	GTA	GTG	TCT	TCC
Aminoacids	F	F	L	L	L	L	L	L	1	1	1	V	V	V	V	5	5
NC_000002 12:c135837169-135836530,c135833190-135833111,c135829676- 135829593,c135824003-135823901,c135822098-135822020,c135818061- 135817341,c135812966-135812311,c135809993-135808443,c135807396- 135807128,c135805057-135804767,c135804128-135803930,c135800809- 135800607,c135798138-135798029,c135794775-135794641,c135790881- 13579058,c135789798-135789571,c135788544- 135786324 Homo sapiens chromosome 2, GRCh38.p14 Primary Assembly	28.63	29.05	1 56	9.34	47.26	24 90	6,74	38.08	14,52	26.97	5 19	8.30	13,49	4.15	29.05	20.75	20,23

CODONS	TIT	TTC	TTA	TTO	CTT	CTC	CTA	CT0	ATT	ATC	ATA	1.04-3
GTT	GTC	GTA.	GTG.	TCT	TCC	TCA	TCG	AGT	AGC	CCT	CCC	101
iccA	cca	ACT	304	ACA	Aco.	OCT.	0CC	OCA	000	TAT	TAC	- 10.
CAT	CAC	CAA	CAG	AAT	AAC	AAA:	AAG	GAT	GAC	GAA	GAG	180
TOT	TOC	COT	COC	CGA.	C00	AGA	A00.	OGT	000	-90A	000	100
ATG	TGG											
ANINCA	CIDS			- E	L.	L.	£.	1	6	I	1	
I	× .	V.	V	V	5	s	s	5	s	5	P	
p	P	P	Τ.	T :	T	7	A	A	A	A	14	
Y.	H	H	9	Q.	N	N	×	×.	D	D	E	- 12

	CODONS	5				111	TTC	TTA	TTG	CTT	CTC	CTA	CTG	ATT	ATC	ATA	GTT	GTC	GTA	GTG	TCT	TCC	TCA	TCO
	Aminoacio	İs	_			F	F	L	L	L	L	L	L	T	1	1	۷	٧	V	V	5	s	S	5
VC_000002 12:c135837169-1356 35829593.c135824003-1358230 35817341.c135812956-1358123 35807128.c135805057-1358047 35800607.c135798138-1357980 35790658.c135789798-1357895	01.c13582 111.c13580 767.c13580 929.c13579 971.c13578	2098-1 9993-1 4128-1 4775-1 8544-	358220 358084 358039 357946	20,c1358 43,c1358 30,c1358 41,c1357	318061- 107396- 300809- 790881-	55	56	3	18	14	48	13	74	28	52	10	16	26	8	56	40	39	16	4
35788324_Homo_sapiens_chro	mosome_2	GRC	h38.p14	Primary	Y_Assembly	1											_							-
35788324_Homo_sapiens_chro	mosome_2	GRC	*Corport 2	_Primary	ind format. Un	shin m	011110	citer i	and pai	h. 700) %	uner bre	(erend)	iggilica	VILLA.										
35788324_Homo_sapiens_chro	copovs	GRC	*Chugowia TTC	n nah airitawa TTA	ied format the	Shir m	geve to CTC	ciapy i	and par	<i>la inc)</i> is C	ner jeg	ATT		ATC		TA	-							
35788324_Homo_sapiens_chro		TTT GTC	*Chuguer a TTC GTA	TTA TTA GTG	ted format the TTO TCT	MAR IN	TCA	£.	TCG	- A	GT	ATT		ATC	Ð	22	- 7							
35788324_Homo_sapiens_chro	CODONS GTT CCA	TTT GTC CCG	*Chugner a TTC GTA ACT	TTA GTG ACC	ted format the TTO TCT	MAR IN	TC/	1. F	TCG	A O	IGT ICA	ATT		ATC CCT TAT	C) T	CC AC	-							
35788324_Homo_sapiens_chro	CODONS GTT CCA CAT	TTT GTC CC0 CAC	*Chugner a TTC GTA ACT	TTA TTA GTG	ted format the TTO TCT	MAR IN	TC/	1. F	TCG	4 0 6	IGT ICA IAT	ATT		ATC	C) T	22	-							
35788324_Homo_sapiens_chro	CODONS GTT CCA	TTT GTC CCG	*Chuguer a TTC GTA	TTA GTG ACC	ted format the TTO TCT	Shite and	TCA	1. F	TCG	4 0 6	IGT ICA			ATC CCT TAT	C) T	CC AC	-							
35788324_Homo_sapiens_chro	CODONS GTT CCA CAT	TTT GTC CC0 CAC TBC	*Chugner a TTC GTA ACT	TTA GTG ACC	ted format the TTO TCT	MAR IN	TC/	1. F	TCG	4 0 6	IGT ICA IAT	ATT		ATC CCT TAT	C) T	CC AC	-							
35788324_Homo_sapiens_chro	CODONS GTT CCA CAT	TTT GTC CC0 CAC TBC	*Chugner a TTC GTA ACT	TTA GTG ACC	ted format the TTO TCT	MAR IN	TC/	1. F	TCG	4 0 6	IGT ICA IAT	ATT		ATC CCT TAT	C) T	CC AC	-							
35788324_Homo_sapiens_chro	CODONS GTT CCA CAT	TTT GTC CC0 CAC TBC	*Chugner a TTC GTA ACT	TTA GTG ACC	ted format the TTO TCT	MAR IN	TC/	1. F	TCG	4 0 6	IGT ICA IAT	ATT		ATC CCT TAT	C) T	CC AC								

SEQUENCES \ PARAMETERS	len	ngth	A	C	T	G	%A	%C	%T	%G	%G+C	%G+A	%G+T	%A+T	%A+C	%C+1
NC_000002.12.c135837169-135836530.c135833190-13 135829593.c135824003.135823901.c135822098-13582 135807341.c135812956-135812311.c135809993-13580 135807128.c135805057-135804767.c135804128-13580 135800607.c135798138-135798029.c135784775-13579 135790658.c135789798-135789571.c135788544- 135786324_Homo_sapiens_chromosome_2_GRCh38.p	2020.c135818061- 8443.c135807396- 3930.c135800809- 4641.c135790881-	5784	1397	1586	1312 14	489	24,15	27.42	22.68	25.74	53.16	49.90	48.43	46.84	51.57	50.10
1397 1586 131 53.165900314938 49 503 201 435 53.658785394191 57 602 503 457	896265560166 48.426694329 5 599 26.089211618 157676348548 51.348547717 7 366 31.2240663990				794 22 662 51 9 29 809 48 975 26 1299 54				. 74343 . 10373 . 06846 . 64232 1.98346 . 79253	015214 443983 473029 365145 248962 112033	4 *					
292 418 694	Adventation before the fact was expected.			850622 72614			497925 978938	311 27 174 37	.17842	323651 854773	5					
292 418 694 65.174273858921 42. 61.654199134199	4 524 15,145228215							311 27 174 37			1	TTICTO	CTA	CTGA	TTATC	ATA
252 418 694 65.174273858921 42. 61.654199134199 CC Ami	4 524 16.145228215 323651452282 48.858921161							Ţ			1	TTICTO L L	CTA C	CTG A	TT ATC	ATA
292 418 694 63.174273858921 42. 61.654199134199 CC_079489.1.73799816- 73801108_Spinacia_oleracea_cultivar_Varoflay_chromo	4 524 15.145228215 323651452282 48.859311431 DDONS inoacids	whole	16.825 e_ger	nome_	ete si	1414 IT_S	equer	T I ICE	TT TTC F 12 0 8	TTA	TTG C	TTICTO L L 25 1.50	L	L	1 1	1

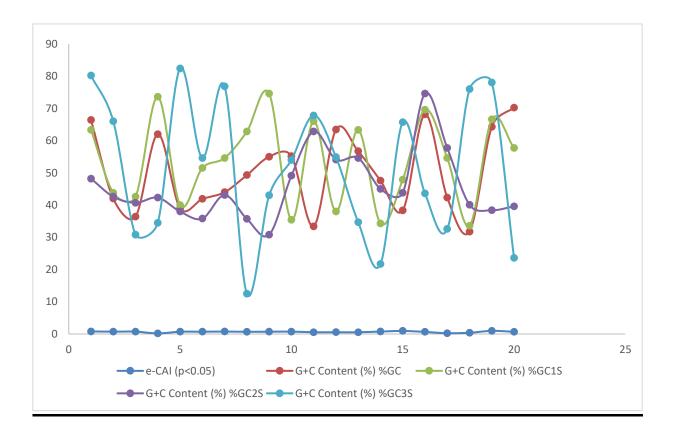
			CODO	ONS							TTT	TTC	TTA	TTG	CTT	CTC	CTA	CTG	ATT	AT
			Amino	acids							F	F	L	L	L	L	L	L	1	1
79489.1:73799816- 108_Spinacia_oleracea	_cultivar_Varo	flay_ch	romosor	ne_3,_B	TI_SOV	_V1wt	nole_ger	ome_sh	otgun_se	equence	32.48	25.52	0.00	30.16	41,76	27.84	2.32	9.28	25.52	34,
			*Channel II	a trade che from	ned format	The flor o	and the first	e and mark	In your pre	German I man	-									
	CODONS	TTT	TTC	TTA	TTG	CTT	CTC	CTA	CTG.	ATT	ATC	AT	A	-						
	GTT	GTC CCG CAC	GTA ACT	GTG ACC CAG	TCT ACA AAT	TCC ACG AAC	TCA	TCG	AGT GCA	AGC	CCT	AT CC TA	c							
	CCA	CAC	CAA	CAG	ALA	ALG	GCT AAA	GCC AAG A00	GAT	GCG GAC	TAT	GA	с. а							
	TOT	TOC	COT	COC	CBA	000	AGA	A05	OGT	660	00A	00								
	ATG	T66	-	0.000		215	a-	-14	-		-									
	AMINGA	105	E U	F		8	L.	1.4	-	2	I S A	I F		100						
	P	p	P	V T	Ť.	Ť	T	A	Â	Ā	A	Q.								
	Y	н	H	9	Q	N	N		*	D	D	E		120						

			COD	ONS							TTT	TTC	TTA	TTG	CTT	CTC	CTA	CTG	ATT	ATC	ATA
			Amino	acids							F	F	L	L	L	L	L	L	1	T	L
9489.1:73799816- 108_Spinacia_olerace	a_cultivar_Var	oflay_ch	romoso	me_3,_B	TI_SOV	_V1wt	nole_ger	iome_sh	otgun_s	equence	-14	11	0	13	18	12	3	4	31	15	9
	CAT	CAC TOC TOC	*Chepur I CAA CGT	w tub deluw CAO	AAT CGA	Che that of AAC COS	AAA	n and park AAO ADO	QAT GAT GGT	derred app GAC GGC	dA	4	640 600		1						
	TGT ATG AMINOA		F	CGC F	L	.coo.	AGA L	ADO L	L	60C	00. I		I DOG	1	117						
	I P	V P	V P	v T	¥ T	S T	S T	S A	s A	s A	5 A		р V		úi:						
	E G NC 975	H C H	H C W 3799816-	Q R	Q R	R	R	К R	R	D G	G		E G								

				_	NUCL	EOTIDE	COM	POSITIO	DN						-			
316-738()1108_Sp	inacia_ole	racea_cult		flay_chro with a cod					ome_sh	otgun_s	equei	nce -	> Sec	uend	e with	more th	ian on
		SEQUEN	ICES \ PA	ARAME	TERS						length	A	C	T	G	%A	%C	%T
ea culti	var Var	oflay chr	omosom	e 3, BT	I SOV	V1. who	le geno	me shot	gun sec	uence	1293	356	292	341	304	27.53	22.58	26.3
11110	nim	NTIN D	WALNU		12.55	L.W.C.F.	191197	NO GM	ARCE	POA IO	1101	(Try				1479	81.8	100
			*Output in t	tab delimite	d format. L	se this out	nit to copy i	and paste is	i your prefe	rred appl	ication.							
	SEQUE	ICES \ PA	RAMETERS	length	A	с	T	G	%A	%C	%Т		G		1			
	%G+C	%G+A	%G+T	%A+T	%A+C	%C+T	A1	C1	T1	G1	%A1		C1	- 11				
	%T1	%61	%G1+C1	%G1+A1	%G1+T1 %G2+C2	%A1+T1 %G2+A2	%A1+C1 %G2+T2	%C1+T1	A2 %A2+C2	C2 %C2+T2	T2		3	- 11				
	%A2 T3	%C2 G3	%T2 %A3	%G2 %C3	%G2+C2 %T3	%G3	%62+12	%A2+T2 %G3+A3	%A2+C2	%A34T3	2 A3		.5 IC3+T	÷.				
	%G3s+0		SAMPLE .	ACS	ALS.	203	MOSTES	AUSTAS	203713	AM3713	AMOTO		ic off i	2				

Name of microorganisms	e-CAI (p<0.05)		G+C C	ontent (%)	
:		%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

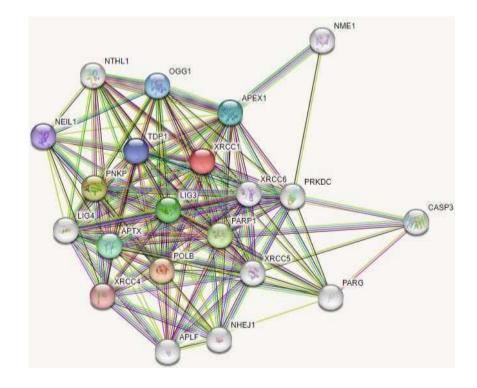


Fig5.1 (showing the nodes and edges)

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

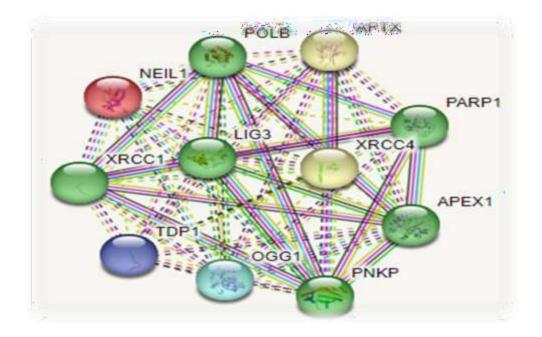


Fig.5.3 (Interacting network of XRCC1 gene)

bubble	cluster Id	gene count	protein names
0	Cluster 1	Ť	NEIL1
۲	Cluster 2	2	APTX,XRCC4
	Cluster 3	6	APEX1,LIG3,PARP1,PNKP,POLB,XRCC1
۲	Cluster 4	1	OGG1
٢	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	o
2182		2.8455%	0.1647%	0.0023097	11.607	o
2118		2.439%	0.48131%	0.0052171	3.7525	0.002
14		10.569%	7.7693%	0.007777	3.6001	o
28		11.382%	8.8461%	0.0075632	3.353	o
2184		2.8455%	1.9444%	0.0035636	2.5288	o
536	\geq	21.951%	17.457%	0.022211	2.0232	0.023

OI	Adj	[Original]	[Random]	[Random]	Z-Score	p-Value
4200488	\gg	2.2533%	1.0089%	0.0046279	2.6889	0.006
541130832	A.	1.1005%	0.64203%	0.0023466	2,668	0.008
32884	\geq	1.2432%	0.00313%	0.0022192	2.0139	0.008
4325452	LX.	1.0317%	0.73830%	0.0035165	2.5404	0.018
4204556	$\geq \sim$	3 4 199%	2.0099%	0.0057187	2 4636	0.017
4198504	\geq	1.2432%	0.68039%	0.0023521	2.3928	0.009
1067832	11	3.4188%	2.0115%	0.0062201	2.2624	0.015
67405856	X	1.8648%	0.9280396	0.0042103	2 2240	0.028
8404016	X	0.6216%	0%	D	undefined	0
8402456	XI	1,243296	0%	o	undefined	.0

Fig. 6.3 Showing results of FANMOD tool

<u>CHAPTER 7</u> <u>CONCLUSIONS</u>

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. <u>https://doi.org/10.1126/science.1072994</u>

[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.

Q. P. Ma et al., "Analysis of synonymous codon usage in FAD7 genes from different plant [11] 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 cotranslational protein folding," *Cell Commun.* Signal., 18, vol. 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 PLoS rice plant by in-silico study," 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 proteinprotein 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.

38

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.