MOLECULAR EVOLUTION AND POPULATION GENETICS OF ACID RESISTANT PATHWAY GLUTAMATE DECARBOXYLASE IN LACTIC ACID BACTERIA

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Biotechnology

by
Burcu TEKİN

June 2022
İZMİR

ACKNOWLEDGMENTS

ABSTRACT

MOLECULAR EVOLUTION AND POPULATION GENETICS OF ACID RESISTANT PATHWAY GLUTAMATE DECARBOXYLASE IN LACTIC ACID BACTERIA

The Glutamate Decarboxylase(GAD) Pathway (GDP) is a major acid resistance mechanism that allows Lactic acid bacteria (LABs) to survive in low pH food environments.

In the thesis, we aimed to study the molecular evolution and population genetics of GDP genes in LABs to understand evolutionary processes shaping adaptation to high acid environments by contrasting species where the GDP genes are organized as an operon structure (*Levilactobacillus brevis*) versus lack of an operon structure (L*actiplantibacillus plantarum*).

Intraspecies molecular population genetic analyses with GDP genes of *L. brevis* and *L. plantarum* from various environments revealed that synonymous and non-synonymous nucleotide diversity is driven mainly by low-frequency changes. Neutrality tests revealed mostly negative values indicating negative selection against replacement changes. Similarly, molecular structure and amino acid characteristic analyses showed that none of the replacement changes on the GDP genes alter the important residues of the proteins supporting negative selection against non-conservative amino acid changes.

Interspecies analyses were used to identify the closely related LABs. Moreover, phylogenetic analyses showed that the GDP gene tree topologies differed from the LAB species tree, indicating divergent evolutionary histories. The functionally preserved two *gad* copies of the *L. brevis* grouped separate phylogenetic clades, showing that the origin of the second *gad* gene might be via horizontal gene transfer from a phylogenetically distant LAB species rather than gene duplication.

In conclusion, GDP in LABs exhibits a dynamic molecular evolutionary history that enables organisms to thrive in high acid environments.

ÖZET

LAKTİK ASİT BAKTERİLERİNDE ASİDE DİRENÇLİ YOL GLUTAMAT DEKARBOKSİLAZIN MOLEKÜLER EVRİMİ VE POPULASYON GENETİĞİ

Glutamat dekarboksilaz (GAD) yolağı (GDY) Laktik asit bakterilerinin (LAB) düşük pH'lı ortamlara adaptasyonunda oldukça önemli olan bir mekanizmadır.

Tezde LAB'ların GDY genlerinin yüksek asitli ortamlara adaptasyonunu şekillendiren evrimsel süreçleri anlamak için, GDY genlerinin organizasyonu açısından farklı iki LAB türünün (*L. brevis* and *L. plantarum*) moleküler evrimi ve populasyon genetiği incelenmiştir.

Çeşitli ortamlardan alınmış *L. brevis* ve *L. plantarum* GDY genleri ile yapılan tür içi moleküler popülasyon genetiği analizlerinde, nükleotit çeşitliliğinin esas olarak düşük frekanslı değişiklikler tarafından yönlendirildiği bulundu. Tarafsızlık teslerinden negatif seçilimi işaret eden sıfırdan küçük değerler elde edilmiştir. Benzer şekilde, GDY genlerinde tespit edilen amino asit değişimine yol açan mutasyonların negatif seçilimi işaret eder şekilde proteinlerin aktivitesini ve yapısını etkilemediği gözlenmiştir.

Evrimsel açıdan yakın LAB türlerini belirlemek için türler arası filogenetik analizler kullanılmıştır. GDY genleri için oluşturlan gen ağaçlarının LAB tür ağacından topolojik olarak farklı olduğu görülmüştür.GAD geni gen ağacında *L. brevis* genomunda bulunan ve fonksiyonal olarak aktif iki GAD geninin farklı dallarda bulunmaktadır.Bu durum *L. brevis*'in ikinci GAD geninin gen duplikasyonu yerine yakın farklı bir türden gen aktarımı yolu ile alınmış olabileceğini göstermiştir.

Sonuç olarak, LAB ların sahip olduğu GDY genlerinin organizmaların yüksek asitli ortamlara uyumunu sağlayan dinamik bir moleküler evrimsel geçmişi olduğu ortaya çıkarılmıştır.

To my MOTHER…

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ADI Arginine Deiminase.

AY Achaz Y test statistic

AgDI Agmatine Deiminase.

B Potential side-chain H-bonds.

C.Amino-acids Correspond Amino-acids.

CAI Codon Adaptation ındex.

CBI Codon Bias Index.

CUB Codon Usage Bias.

Cod. Coding Sites.

DoS Direction of Selection.

Entrez Entrez Global Query Cross-Database Search System.

Eta Number of mutations.

FDA Food and Drug Administration.

FL-D Fu and Li's D test statistic

FL-F Fu and Li's F test statistic

FW-Hn Normalized Fay and Wu's Hn test statistic

GABA Gamma-Aminobutyric Acid.

GAD Glutamate Decarboxylase.

GDP Glutamate Decarboxylase Pathway.

GIT Gastrointestinal Tract.

GRAS Generally Recognized as Safe.

H Hydrophobicity.

HDC Histidine Decarboxylase.

InL Log-likelihood Score.

JC Jukes-Cantor correction applied estimates.

LAB Lactic Acid Bacteria.

MK McDonald–Kreitman test.

N Sample Size.

N.No. Nucleotide Number.

NCBI National Center for Biotechnology Information.

NI Neutrality Index.

No. Rep. Sequence with no replacement sites.

Nonsyn. Non-synonymous.

Np Number of parameters.

Num. RP Number of replacement polymorphisms.

ODC Ornithine Decarboxylase.

Par. Parsimony Informative.

pI Isoelectric Point.

PLP Pyridoxal phosphate.

R R programming language.

R.No. Residue Number.

Rep.Pol. Replacement Polymorphism.

rRNA Ribosomal RNA.

S Number of segregating sites.

Sing. Singleton.

Syn. Synonymous.

Syn. Pol. Synonymous Polymorphism.

TD Tajima's D.

TDC Tyrosine Decarboxylase.

ZE Zeng E tes statistic

* $p < 0.05$

** $p < 0.01$

*** $p < 0.0001$

\# $0.10 < p < 0.05$

CHAPTER 1

INTRODUCTION

1.1 General Features of Lactic Acid Bacteria

The bacteria that produce lactic acid while its metabolic processes are called Lactic acid bacteria (LAB)[1]. In the Lactic acid bacteria, lactic acid is generated through several pathways of carbohydrate metabolism, mainly by the glycolysis pathway[2]. In detail, the carbohydrate metabolism divides into two groups referred to as homofermentative and heterofermentative[3]. Homofermentative groups of LABs use the two pathways, which are glycolysis and pentose phosphate pathway, to produce lactic acid from sugars hexose and pentose. Two lactic acid molecules are given to the system for each glucose molecule during the process[4]. Besides that, Unlike the homofermentative groups, the heterofermentative groups of LABs use the more complex pathways for producing lactic acids like phosphoketolase and phosphogluconate pathways. Also, shown in Figure 1.1, the heterofermentative pathways generate CO_2, ethanol, or acetic acid besides lactic acid[5].

In addition to features of LABs, they are gram-positive bacteria that have the shape of cocci, rods with non-spore-forming, and can grow absence of O_2, but the presence of O_2 does not affect the growth of LABs[6].They are named aerotolerant anaerobes that can boost their numbers even under the existence of O_2[7]. Moreover, In most articles, it is mentioned that lactic acid bacteria do not show catalase activity; however, this claim is not primarily valid. Several reports indicate that numerous lactic acid bacteria show catalase-positive activity[8].
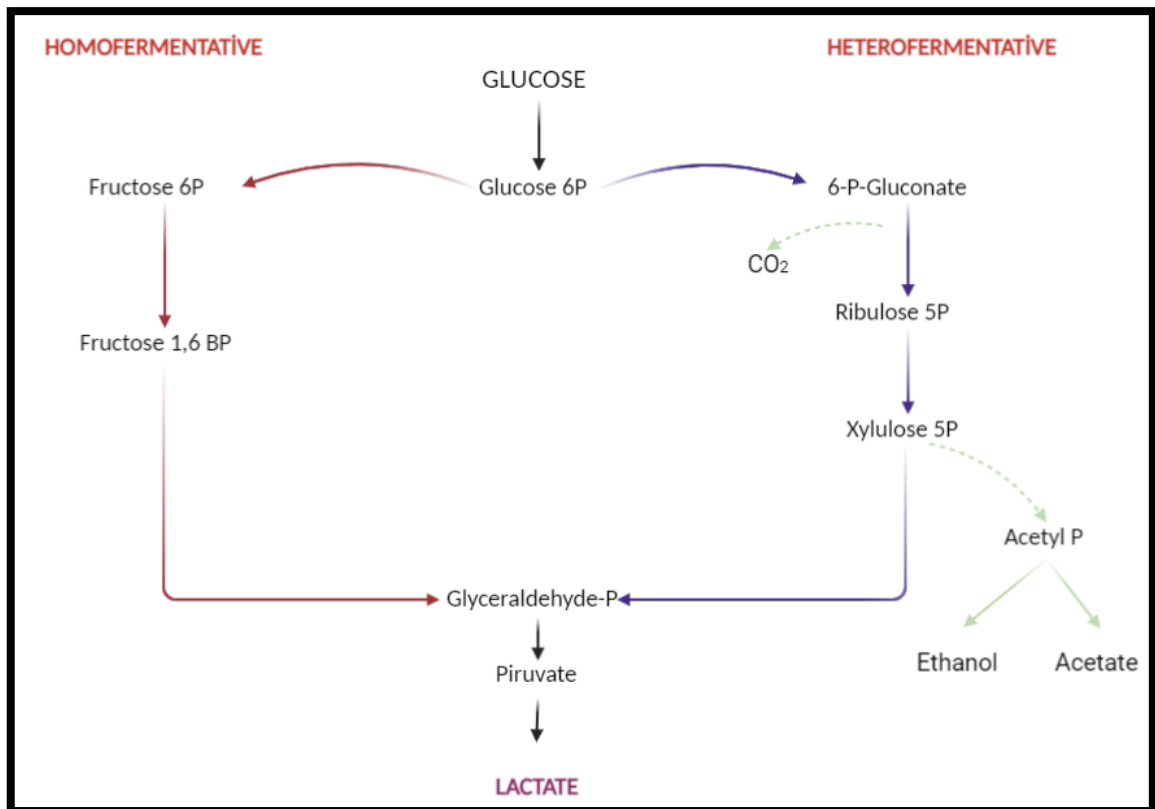
Figure 1.1: A brief illustration of carbohydrate metabolism of LABs.
(Created with BioRender.com)

1.1.1 Taxonomy of LABs

Lactic acid bacteria groups were first classified by their features like cell types and their growth temperatures. Also, their metabolic properties, like fermenting glucose, were considered a classification point of LABs. With these primary properties, the LABs were grouped into four bacterial genera. Thus, Lactobacillus, Leuconostoc, Pedicoccus, and Streptococcus were accepted as classified LAB genera[9]. After years of developments in the classification, physiological and biochemical typing is not preferred much. Recently, the improvements in molecular biology have given new insight into the taxonomical studies of bacteria. Nowadays, several molecular tools are used to identify prokaryotes, like classification with the 16S rRNA by generating phylogenetic trees[10], specific PCR assays[11], some protein patterns, and other techniques[12].

With lots of literature studies, the lactic acid bacteria group represents highly heterogeneous taxonomic genera. LABs group members could be taxonomically diverse from each other in terms of their adaptation to different environmental niches and conditions[13]. The species accepted as a LAB are genera of Lactobacilli, Lactococci, Streptococci, Pediococci ,Enterococci and also, Tetragenococci, Vagococci, Oenococci, Leuconostocs ,Carnobacteria ,Weissella[14]. Besides that, the bacteria with similar properties to the lactic acid bacteria are misleadingly defined as a member of LABs. For instance, Gram-positive Bifidobacterium living mainly in the GIT and vaginal environment is sometimes suggested as LAB in terms of using probiotics and founding in the fermented products[15]. Furthermore, Lactobacillaceae is one of the lactic acid bacteria families(type genus Lactobacillus). Commonly, Lactobacillaceae is inaccurately considered to be regarded as all LABs despite the taxonomic heterogeneity of the LABs group. Although *Lactobacillus* does not cover the entire LABs, It is the most studied and most extensive type of the genus.

Several studies tried to identify and taxonomically classify *Lactobacillus* species in the literature. One of the studies used the 16S rRNA sequences to identify *Lactobacillus* species. They applied comparative phylogenetic analysis and divided the *Lactobacillus* genus into 15 subgroups[16]. On the other hand, they observed that the species do not always show the same features, even in the same *Lactobacillus* subgroup[16, 17]. Phylogenetic trees based on specific genes, such as the 16S rRNA gene, allow species to be classified into phylogenetic clades; however, they lack sufficient resolution to reveal the phylogenetic connections between the various clades. The updated number of *Lactobacillus* species is more than 250 (https://www.bacterio.net/genus/lactobacillus), and It has proved that only 16S rRNA phylogenetic analyses will not be sufficient for taxonomic classification[18, 19]. Also, being a highly diverse group makes *Lactobacillus* compulsory to classify with different approaches. Hence, in most similar studies, analyses such as other metabolic genes, ribosomal proteins, and average nucleotide or amino acid similarity were used in addition to 16S rRNA analyses[20]. Moreover, it has become widespread that the core genome and pan-genome are used to classify *Lactobacillus* species[21].

One of the research groups, which foresees the acceptance and use of the usual *Lactobacillus* classification as an obstacle to researching the diversity in this *Lactobacillus*, has recently reclassified the *Lactobacillus* group[22]. For the reclassification of the *Lactobacillus* genus, They took into account average nucleotide identity, the

average amino acid identity and phylogeny of the core genome, signature genes, and metabolic and ecological features. Thus, The Lactobacillus genus has been reclassified into 25 genera[22], and their names have changed. Table 1.1 shows the species used in this thesis and their changed names.

Table 1.1: *Lactobacillus* species used in the thesis and their new names[22].

Common Names	Reclassified Names
Lactobacillus brevis	*Levilactobacillus brevis*
Lactobacillus plantarum	*Lactiplantibacillus plantarum*
Lactobacillus reuteri	*Limosilactobacillus reuteri*
Lactobacillus parakefiri	*Lentilactobacillus parakefiri*
Lactobacillus buchneri	*Lentilactobacillus buchneri*
Lactobacillus argentoratensis	*Lactiplantibacillus argentoratensis*
Lactobacillus paraplanrarum	*Lactiplantibacillus paraplantarum*
Lactobacillus nuruki	*Companilactobacillus nuruki*
Lactobacillus fermentum	*Limosilactobacillus fermentum*
Lactobacillus futsaii	*Companilactobacillus futsaii*
Lactobacillus oris	*Limosilactobacillus oris*
Lactobacillus senmaizukei	*Levilactobacillus senmaizukei*
Lactobacillus coleohominis	*Limosilactobacillus coleohominis*
Lactobacillus suebicus	*Paucilactobacillus suebicus*
Lactobacillus rossiae	*Furfurilactobacillus rossiae*
Lactobacillus sakei	*Latilactobacillus sakei*
Lactobacillus antri	*Limosilactobacillus antri*
Lactobacillus curvatus	*Latilactobacillus curvatus*
Lactobacillus paracollinoides	*Secundilactobacillus paracollinoides*
Lactobacillus zymae	*Levilactobacillus zymae*
Lactobacillus gastricus	*Limosilactobacillus gastricus*
Lactobacillus paracasei	*Lacticaseibacillus paracasei*
Lactobacillus tangyuanensis	*Levilactobacillus tangyuanensis*
Lactobacillus angrenensis	*Levilactobacillus angrenensis*
Lactobacillus herbarum	*Lactiplantibacillus herbarum*
Lactobacillus cerevisiae	*Levilactobacillus cerevisiae*
Lactobacillus spicheri	*Levilactobacillus spicheri*
Lactobacillus rennini	*Loigolactobacillus rennini*

1.1.2 Ecological Niches and Dynamic Metabolic Processes of LABs

The environments where lactic acid bacteria are found and their abundance in these niches are pretty diverse. The number of LABs may increase or decrease depending on the characteristics of the environment in which they are located. Lactic acid bacteria, which have a great diversity within and between species, show genomic diversity according to their environments[23]. LABs can be isolated from natural environments such as water and soil[24]. However, their abundance in these environments has not been found at high levels. In addition, researchers found that the number of LABs increased in habitats where they were in close contact with plants and showed an antimicrobial effect beneficial to the plant[25].

Although human-specific, a healthy female vagina is characterized as an environment dominated by LABs[26, 27]. There are some significant benefits that the LABs dominating the vaginal environment provide to the individual. One of these advantages is that LABs prevent diseases caused by harmful microorganisms such as sexually transmitted viruses, bacterial pathogens, and fungi in the vagina[28]. Due to the lactic acid produced by LABs, other microorganisms are easily eliminated in the vagina environment with low pH levels[28]. Another positive impact is the release of hydrogen peroxide, which prevents vaginosis, into the environment by LABs[29]. The species diversity of LABs, which dominate the vaginal flora, is quite restricted, and it varies depending on ethnicity, pregnancy, and menopause[30, 31]. The LABs are also found in the human stomach[32] and gastrointestinal tract[33]. Nevertheless, their presence highly differs according to individuals' diets and health conditions. Moreover, the considerable heterogeneity of LABs in the human intestine makes it difficult to identify and compare their benefits or pathogenicity[34]. LABs can be found in the human lung[35]. However, similar to the GIT environment, their identification is challenging in terms of their variability in different ages, diets, and unique living environments. LABs have been demonstrated to reduce asthma symptoms[36] and prevent influenza infection[37] in several trials involving the administration of certain LABs species. As an absolute fact that LABs are major members of the lung environment, their applications might be used to prevent several diseases[38].

The utilization of lactic acid bacteria in food products, which has numerous positive features in terms of metabolism, has become extremely popular nowadays[39]. Characteristics such as organic acid and aroma production in foods, hydrolysis of

proteins, and elimination of spoilage-causing microorganisms are among the prominent reasons for using LABs in foods[40]. Most importantly, as their use in terms of health gains importance, the production of foods containing LABs, defined as functional foods, is increasing[41]. As a result, the food-based habitats of LABs are rather varied. Fermented products are the most significant food environments for LABs. Simply put, fermentation is when organic molecules are converted into various products such as alcohol and acids due to microbial metabolism by microbial enzymes[42, 43]. Fermentation is a centuries-old process that is widely used in the production of various foods by human societies all over the world[44]. Fermented foods have an important place in maintaining human life because the fermentation process plays a significant role in preserving the food taste, quality, and enrichment of the food content[45]. In detail, bacteriocins, small proteins generated by LABs in fermented foods, are critical in protecting food from other hazardous microorganisms and providing food safe for consumption[46, 47]. The fermentation process can be natural[48] or with the addition of bacterial culture to the medium. Microorganisms found naturally in raw foods play a role in natural fermentation. However, unlike natural fermentation, specially selected bacterial cultures are added to the food medium in culture-dependent fermentation[49]. Figure 1.2 shows the fermented food environments where lactic acid bacteria are most commonly found and used. Among these examples, raw plant-based foods such as kimchi[50] and pickles[51] are formed by natural fermentation, but foods such as cheese[52], kefir[53], and fermented beverages[54] are foods created by adding a selected bacterial culture to the medium[55].

The numbers and types of LABs found in fermented foods vary substantially depending on the kind of product, the nutrients provided to the environment, and the metabolic dynamics used by LABs[56, 57]. Along with the diversity in foods, LABs vary in the compounds they provide to the food environment. One of the reasons for this diversity is the nutrients provided by the environment to the bacteria and, accordingly, the different metabolic processes used by the LABs. In traditional and industrial manufacturing, many healthy foods are developed using these different metabolic processes of LABs, and their application is growing more common[58].

Figure1.2: The main fermented products that LABs found.

1.1.3 Safety and Applications of LABs in Food Products

LABs are often considered safe or defined by FDA approval as GRAS (Generally Recognized as Safe) because LABs are not only widely used in food but also play a unique role in health applications[59]. In addition, the European Food Safety Agency has recognized many types of LAB as safe[60].

The human GIT provides a habitat for many LAB. In return, LABs prevent conditions that can cause many health problems[61]. For instance, LABs help regulate insulin levels in the blood[62] and facilitate fatty acid absorption from the intestines[63]. Most importantly, they do not permit harmful organisms to survive[64]. Furthermore, many studies have demonstrated that LABs have a favorable effect on colon cancer patients and slow the proliferation of cancer cells[65, 66].

Nowadays, fortified foods for health are made and used widely. Several studies have shown that fermented foods containing LABs can help avoid lactose intolerance[67] and diarrhea problems[68]. Since especially children are in the high-risk group for diarrhea, the use of foods containing LABs, such as fermented milk products[69], comes to the preliminary. Consequently, the use of LABs in food has increased. Biotechnologically, good metabolic properties provided by LABs are examined, and their usage in foods is investigated as additives. Traditional fermented foods have been recognized to provide several health benefits for millennia[70]. Many studies, for example, have demonstrated that fermented foods like kimchi, which have been around for a long time, offer anti-aging and anti-obesity properties[71]. Despite the fact that the health benefits of culturally fermented foods are well recognized and widely consumed, scientific research into the metabolic and genetic features of LABs in these foods are still needed.

Food becomes high in nutrients, bioactive substances, and amino acids during fermentation, depending on the metabolic activities of the bacterial populations it contains. Foods whose composition changes as a result of bacterial fermentation have significant health implications. As proof, fermented milk products have been found in the literature to reduce the risk of heart disease[72], lower cholesterol levels[73], and avoid insulin resistance[74]. Furthermore, foods containing LABs, such as kefir and kimchi, have been demonstrated to prevent allergic responses[75] in certain people. In particular, several studies have indicated that consuming fermented foods containing LABs might effectively prevent stomach[76], colon, and cervical cancer types[77].

Some LAB species, on the other hand, can be pathogenic or spoilage-causing microbes. Pathogenic species are found in the genera enterococci and streptococci. However, depending on how they evolve, they may be mutualistic, pathogenic, or clinically harmful. Although certain *Enterococcus* species cause the meat to deteriorate, they have many beneficial impacts on the ripening and flavor of traditional foods like cheese[78, 79]. Most notably, ingesting fermented food containing biogenic amines produced by LABs' decarboxylation activities might result in various health issues[80]. Biogenic amines, depending on their types and the person's allergic sensitivity, might induce headaches, vomiting, allergy symptoms, diarrhea, and certain cardiac issues when consumed[81]. There are several biogenic amines that can be produced by LABs that might be present in fermented foods. The four forms of biogenic amines most typically detected in fermented foods generated by LABs can be listed as putrescine, agmatine, and the most

harmful, tyramine and histamine, are among them[82, 83]. The presence of other biogenic amines together with histamine and tyramine, which cause severe health problems, worsens the severity of health problems by inhibiting the activity of enzymes that will eliminate the toxicity of histamine and tyramine[84]. Decarboxylase enzymes in LABs play a role in the formation of biogenic amines. LABs have a unique metabolism that enables the formation of each biogenic amine and various proteins involved in these metabolic pathways. Depending on the nutritional content of the environment and the metabolic capabilities of LABs in the environment, the synthesis of biogenic amine types in foods might change. Even different species strains might differ in their capability to produce biogenic amines. In general, decarboxylation metabolic pathways are used by LABs, and the production of biogenic amines has tremendous importance in adapting the bacteria to the environment and maintaining their viability in that environment. Simply put, the decarboxylation processes of amino acids are the mechanism by which LABs protect themselves against the environment whose pH decreases during fermentation[85].

LABs have several stress reactions against the low pH of the environment as a result of living in varied conditions. One of these mechanisms, decarboxylation, will be discussed in detail in the following sections.

1.2 Acid Resistance Pathways in LABs

LABs expose to a variety of stress factors in their surroundings. Among many stresses, acid stress is the one that LABs mainly encounter, and the mechanisms against acid stress are essential for the survival of the bacteria[86]. Lactic acid is released into the environment due to LABs metabolism during the fermentation of LABs in a nutrient-rich environment, and this process causes acidification of the extracellular environment. Owing to the apparent charge difference between the extracellular and intracellular environments, lactic acid in the acidified extracellular environment eventually passes into the cytoplasm by simple diffusion[87]. Acidification of the intracellular environment has effects that may result in bacterial mortality. Changes in intracellular ion charge disrupt the integrity of the bases in protein structures, and then the structural disturbance of protein structures prevents enzymatic reactions required for metabolic activities[88]. Many acid tolerance mechanisms have evolved in LABs to recognize and resist their self-created acid environment.

Intracellular pH maintains at an average level with the working of acid-resistant metabolic processes. One of the first defensive steps for bacteria against high-level extracellular acidity is increasing the amount of unsaturated fatty acids proportion in the bacterial membrane[89]. In addition, proton pumps, like F1-F0-ATPase, which transport protons and have a catalytic effect on ATP synthesis or hydrolysis, are positively regulated by acid stress in LABs[86]. Several studies have found that lower pH levels increase the transcription of the F1-F0 complex gene[90]. Likewise, various repair processes utilize to protect central material DNA and metabolic enzymes due to acid adaptation in bacteria[91]. The protection of genetic material is critical for LABs adapting to acidic environments.

Living in low pH conditions results in a variety of selection pressures, including the evolution or loss of metabolic pathways. Many metabolic pathways in LABs play an active part in acid tolerance, and these pathways might vary depending on the nutritional qualities of the habitats in which they exist. The metabolic pathways involved in acid tolerance can divide into two groups that neutralize the environment and consume ions like proton pumps[92]. The effective pathways in the neutralization process help neutralize the acidity by producing alkaline molecules in the intracellular environment through a series of reactions. Arginine and agmatine deiminase metabolic pathways are effective systems in intracellular neutralization processes[93]. Homologous deiminase pathways have similar operon structures. Genes in the operon create the transporter protein, the regulator protein that regulates the system, and three major enzymes. Figure 1.3 illustrates the operon configuration and metabolic activities of deiminase pathways. The deiminase catalyzes the conversion of the arginine molecule, which is the substrate of the arginine deiminase (ADI) pathway[94], to citrulline. The citrulline molecule is further converted to ornithine by ornithine carbamoyltransferase, and the carbamate kinase enzyme produces ATP. Ammonia is formed and given to the intracellular environment in the metabolic process's citrulline and ATP formation steps[95].

On the other hand, in the agmatine deiminase (AgDI) system, different from the ADI system, the product produced is putrescine since the substrate taken into the cell is agmatine[93]. Other metabolic steps of the AgDI pathway[96] create ATP and ammonia in the intracellular environment through enzymatic activities similar to the ADI pathway. In conclusion, deiminase pathways generate energy for the bacterial cell; meanwhile, ammonia regulates intracellular pH[97].
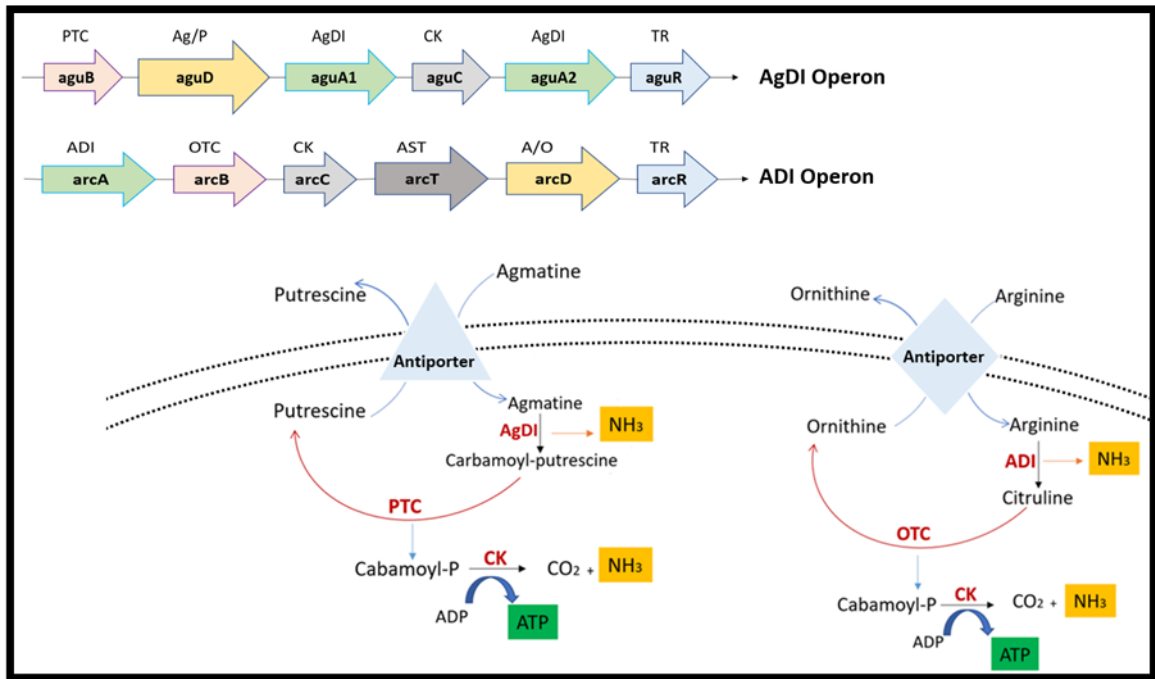
Figure 1.3: Operons and metabolic pathways of deiminase systems. The arrow-shaped illustrations and the names on the images show the genes and their positions in the operon. Short names represent proteins. Enzymes; ADI and AgDI = Deiminases, CK= Carbamate kinase, OTC= Ornithine carbamoyltransferase, PTC= Putrescine carbamoyltransferase. TR stands for regulator protein, whereas Ag/P and A/O represent antiporter proteins.

Proton pump-like metabolic acid tolerance pathways are proton-depleting decarboxylase processes. The basic structure of these pathways that decarboxylation of amino acids is the decarboxylase enzyme and the antiporter that ensures substrate/product transitions. The decarboxylation process consumes the hydrogen ions in the intracellular environment, preventing the pH value from decreasing. Four different pathways are named according to the amino acid type involved in the decarboxylation process and are effective in acid tolerance. Moreover, these pathways are essential in terms of molecule types they release into the environment as a result of reactions alongside maintaining intracellular pH levels[98]. The first decarboxylase system is the tyrosine decarboxylase (TDC) pathway (TDCP), which performs tyrosine decarboxylation[99]. In addition to the

enzyme and antiporter genes, there are tRNA synthetase and Na/H+ antiporter genes in the TDCP operon structure[100]. While decarboxylation of the tyrosine molecule, the tyramine molecule is formed by hydrogen consumption and released into the extracellular environment via the antiporter protein. Meanwhile, the decarboxylase process also provides carbon dioxide to the environment.

The second decarboxylase pathway is the histidine decarboxylase[101] (HDC) pathway (HDCP), which produces histidine by consuming ions from the cytoplasm. In the HDCP operon structure, there are enzyme and antiporter genes located side by side, similar to the TDCP operon. In addition, unlike the TDCP operon, a gene synthesizes the protein that helps the maturation of the histidine decarboxylase enzyme[102]. The structurally simpler ornithine decarboxylase (ODC) pathway is another decarboxylase process. The pathway's operon structure consists of only two genes (enzyme and antiporter)[103]. The decarboxylation of ornithine which is the substrate of the ODC produces putrescine, and the antiporter protein transports putrescine to the extracellular environment[104].

Figure 1.4 describes the three decarboxylase pathways' operon structures and chemical transformations. Although the effect of these three decarboxylase systems on bacterial survival and adaptability is beneficial, consuming the compounds produced by the reactions through food may cause harmful effects on humans[105]. As previously stated in the preceding sections, the metabolic products of arginine/agmatine deiminase, TDC, HDC, and ODC pathways: putrescine, ornithine, tyramine, and histidine, are toxic compounds known as biogenic amines. These molecules, which can be taken into the human body in high amounts through foods, have profound effects such as causing migraine, vomiting, neurological and gastrointestinal diseases[106]. Legal limits, which are accepted and prohibited, determine for biogenic amines, which are highly toxic, such as histamine, due to their effects on the quality of life[107]. Moreover, since the presence of acid tolerance pathways in bacterial species and their genomic organization may differ according to the species and the type of nutrients the environment provides, the diversity of biogenic amines in foods also changes.

Unlike the acid tolerance pathways that create biogenic amines, the other decarboxylase pathway, the glutamate decarboxylase (GAD) system, produces gamma-aminobutyric acid (GABA) that has several beneficial effects on the human body[108]. Due to the beneficial effects provided by the GABA molecule, the bacterial species that have

the GAD system and the foods containing these bacteria are the subject of many studies. Similarly, the subject of the thesis has been chosen as the GAD system and will discuss in detail.
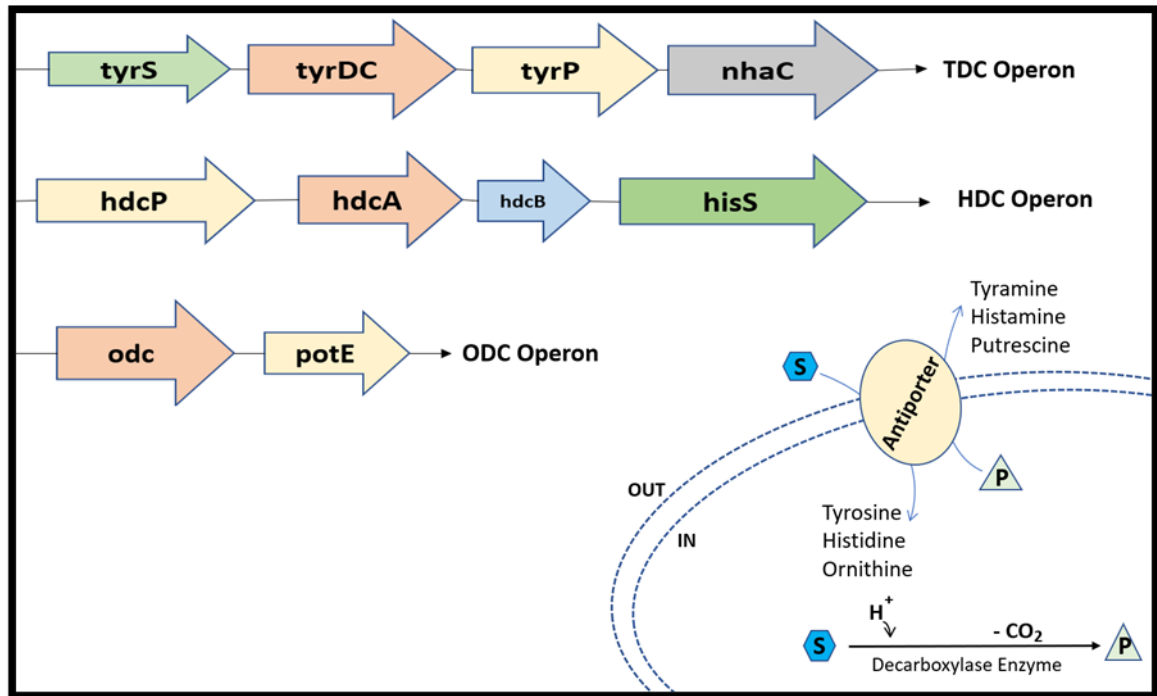
Figure 1.4: Biogenic amine producing decarboxylase pathway operons and schematic image of decarboxylation process. The arrow-shaped illustrations and the names on the images show the genes and their positions in the operon. Enzyme genes : *tyrDC* =Tyrosine decarboxylase, *hdcA*= Histidine decarboxylase, *odc* = Ornithine decarboxylase.Antiporter genes: *tyrP*= Tyrosine/Tyramine antiporter, *hdcP*= Histidine/Histamine antiporter, *potE*= Ornithine/Putrecine antiporter. tRNA synthetase genes are *tyrs* (Tyrosyl-tRNA Synthetase) and *hisS* (histidyl-tRNA synthetase). Respectively, *nhac* and *hdcB* genes represent the NA/H+ antiporter gene and the protein gene that catalyzes the maturation of the HDC enzyme. In addition, the bold S symbolizes the substrate, while the bold P symbolizes the product.

1.2.1 Glutamate Decarboxylase System in LABs

Glutamate plays a role in critical metabolic processes such as glycolysis, gluconeogenesis, and the citric acid cycle, which participate in protein synthesis; therefore, it is a molecule that is found and plays an essential role in all living organisms[109]. Moreover, glutamate acts as a link between nitrogen and carbon metabolism. There are two enzymes actively involved in the catabolism process of the glutamate molecule. The first is the dehydrogenase enzyme, which accomplishes ammonia amino acid assimilation[110]. The second is the glutamate decarboxylase (GAD) enzyme, which bacteria use as an acid tolerance mechanism, among other stress responses[111]. Because acid resistance and the GAD pathway (GDP) are central to the thesis, this part will focus on the genomics of the GAD system, as well as the structure and evolution of genes that encode proteins involved in the GDP.

The bacterial glutamate-dependent GAD system transports extracellular glutamate into the cell and converts it to gamma-aminobutyric acid (GABA). During this process, as in all other decarboxylase processes, the protons present in the intracellular environment participate in the reaction. Thus, the bacterial cell gains the advantage of regulating the decreasing pH level[112]. Since the GABA molecule, a byproduct of glutamate decarboxylation, is a very stable molecule, it does not re-ionize and does not alter the ion balance in the cell. Then, the GABA molecule produced in the cell can be transported out of the cell by the antiporter of the GAD system, or it can remain in the cell[113].

Nowadays, the expanding genomic knowledge and the accessibility of genomes from various organisms demonstrate that the GAD system is widespread in a wide range of bacterial species. Because the GAD system allows bacteria to successfully adapt to low acid conditions, different bacterial species from distinct habitats contain GDP genes in their genome. The structure and genomic organization of the GAD system, on the other hand, may change between species and even strains. Some species, for instance, may have more than one decarboxylase and antiporter gene, while others may not have the antiporter gene together with the enzyme gene[114]. Moreover, as in *Listeria monocytogenes*, different strains of the same species may differ from each other in enzyme and antiporter gene numbers[115].

According to the literature, *E.coli*, which adapts well to high acid conditions, is the most studied species for the genetic and structural information of the GAD system. Many studies revealed the genomic organization of the GAD system found in *E.coli* and the three-dimensional structures of the system proteins. These studies discovered two glutamate decarboxylase enzyme genes (*gadA* and *gadB*) formed due to duplication in the *E.coli* genome, which are 95% identical[116]. Additionally, one of the *E.coli* GAD system decarboxylase enzyme genes is near the antiporter gene (*gadC*), while the other is further away, and the multiple regulator proteins complexly control the system[117].

Many raw or fermented foods may have high glutamate levels. LABs, which live in various glutamate-rich environments, may maintain acid balance by utilizing glutamate to neutralize the generated acid environment. Hence, numerous LAB species have been examined in LAB genera, such as *Lactobacillus*[118], *Lactococcus*[119], and *Streptococcus*[120], which produce the GABA molecule by activating GAD while adjusting intracellular pH. As shown in numerous species, the GAD system and system's operon organization in the LAB genome is highly variable. For example, like other bacterial species, LAB species may contain one or two GAD enzymes, and the antiporter gene (*gadC*) may not be present in all LAB species. Among many examined LAB genera, the *Lactobacillus* genus has the most significant GABA production potential, with species such as *L. brevis*[121], *L. plantarum*[122], *L. buchneri*[123], *L. paracasei*[124], and others.

Levilactobacillus brevis is the best-examined species in the genus and has a high GABA-producing capability. The *L. brevis* genome has two GAD genes, and the genes' products are isozymes with each other[125]. One of the GAD enzyme genes is located in the *L. brevis* genome as a specific operon, as illustrated schematically in Figure 1.5. The second GAD gene lies in a separate genome region, away from the operon arrangement. The *L. brevis* GAD cluster structure consists of four genes. The first of these genes is the *gadR* gene, which synthesizes the protein that will regulate the system, and the second is the *gadC* gene, which synthesizes the glutamate/GABA antiporter protein located downstream of *gadR*. In most studies, the glutamate decarboxylase gene, called *gadA*, is located just upstream of the *gadC* gene. Since no promoter or terminator structure could be found between *gadA* and *gadC* genes[125], it suggests that these two genes are transcribed together in a synchronized manner by forming an operon structure. Furthermore, the *gltX* (glutamate-tRNA ligase) gene, positioned upstream of the *gadA* gene, is part of the GAD cluster structure[126]. The second GAD gene of *L. brevis* is usually called *gadB* and locate

in a different part of the bacterial genome without any operon formation. Some studies suggest that since *gadA* and *gadB* genes are found far from each other, there may be different gene regulations within these two genes.
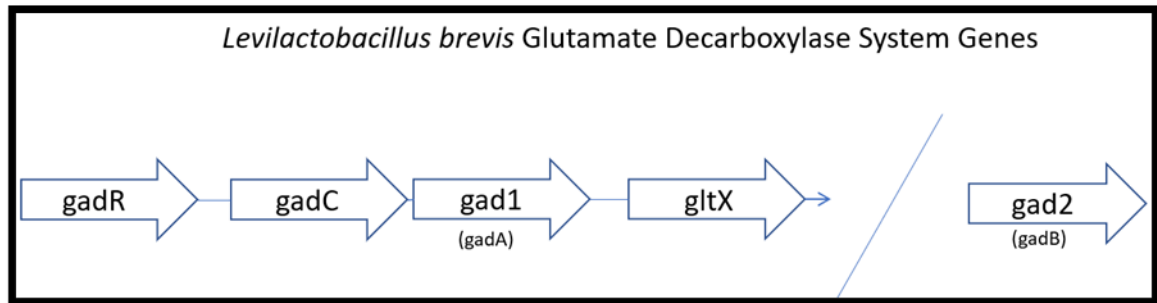
Figure 1.5: Schematic explanation of *L. brevis* GAD system genes.

It has been observed that the nomenclature for *L. brevis* GAD genes can vary in the literature, and some review publications have mentioned this[126]. Some articles used the *gadA* nomenclature for the GAD gene in the operon structure[127], while others named the GAD gene in the operon structure as *gadB*[125]. Therefore, to avoid any confusion in the names of the studies and explanations within the scope of the thesis, the nomenclature *gad1* for the GAD gene in the *L. brevis* GAD operon and *gad2* for the GAD gene outside the operon has been used.

Only *L. brevis* possesses two GAD genes among the LABs examined thus far[128]. As mentioned before, the distribution and presence of GAD system genes on the LAB's genome vary according to species. In species such as *L. brevis*, *L. reuteri*, *L. buchneri*, and *L. lactis*, the antiporter gene creates an operon structure with the GAD gene, while in species such as *L. plantarum*, *L. fermentum,* there is only a single GAD gene, and there is no antiporter gene[127]. Additionally, the *gadB* nomenclature is used in the literature for LAB species with only one GAD gene in their genome. The fact that the *gadC* gene is found only in certain types of LABs may define the specific potential of those species to produce GABA. Although *Lactiplantibacillus plantarum* species, one of the thesis materials, does not have the *gadC* gene[127], many studies show that this species produces GABA. However, a study using *L. plantarum WCFS1* strain observed that the amount of glutamate added to the medium did not significantly change the bacterium's GAD

synthesis, GABA formation, and acid resistance status[127]. Such inferences indicate that the *gadC* gene and its product, the antiporter protein, are fundamental for GABA synthesis[129].

Enzyme assay tests with *L. brevis*, two GAD proteins, and *L. plantarum* GadB protein revealed that the activity features of these proteins differed from one another. When the activities of enzymes were evaluated against changing temperature values, it was discovered that the enzyme produced by the *L. brevis gad1* gene displayed limited activity at varied temperature values; *L. brevis* Gad2 and *L. plantarum* GadB proteins were determined to be heat-stable enzymes[130]. More notably, *L. brevis* Gad2 protein and *L. plantarum GadB* protein displayed activity in similar pH ranges in enzyme activity assays done with various pH levels (pH levels from 3 to 5.5). The GAD enzyme, the gene product of *L. brevis gad1*, on the other hand, has activity in weaker acid pH ranges (pH 5.5-6.6)[127]. Moreover, *L. brevis gad2* gene expression levels are lower than other GAD genes, and it was determined by a study that there was no change in the expression of the *gad2* gene at acid (pH 5.2) or neutral (pH 6.8) levels[125]. Also, they suggested that although *gad2* gene expression is not affected by pH, the expression levels of *L. brevis gadC* and *gad1* genes, which form an operon, are pH-dependent. Another study supporting enzyme assay results is mutant studies with *L. brevis* GAD system genes. As a result of the study, they found that the *L. brevis gadC* and *gad1* gene mutants are acid sensitive. The findings show that the *gad1* and *gadC* genes, which constitute the operon structure, are required for acid tolerance. Consequently, it can be concluded that *gad1* and *gadC* are necessary for acid tolerance in *L. brevis*, while *gad2* plays a minimal role[125].

1.2.1.1 The Glutamate Decarboxylase Enzyme (GAD, EC 4.1.1.15)

Glutamate decarboxylase is a PLP-dependent intracellular enzyme found in both eukaryotes and prokaryotes, and its presence varies according to organisms. When GAD is classified, it is in the 1st type in PLP-dependent enzymes and the 2nd group among decarboxylase enzymes[131]. During the decarboxylation reaction inside the cell, L-Glu, the GAD enzyme's substrate, reacts by binding close to the PLP molecule in the enzyme's active site. As a result of the reaction, GAD converts the L-Glu to the GABA molecule by consuming H+ from the cytoplasm[132]. The three-dimensional structure of the GAD enzyme in LAB and the critical properties of the amino acids that make up the protein

have been studied recently by isolating them from the *L. brevis CGMCC 1306* [133] strain. The structure observed as a result of the study is stored in the PDB, which is the protein database, under the code 5GP4. The 468 amino acid long GAD protein isolated from *L. brevis* consists of three domains. In Figure 1.6, the domains are shown both in the three-dimensional structure of the GAD enzyme and on the amino acid sequence using different colors. Compared to the N-terminal and Small-domains, the PLP-binding domain that covers the most significant portion of protein includes the active site amino acids to which the cofactor PLP and the substrate L-Glu molecule interact [133].
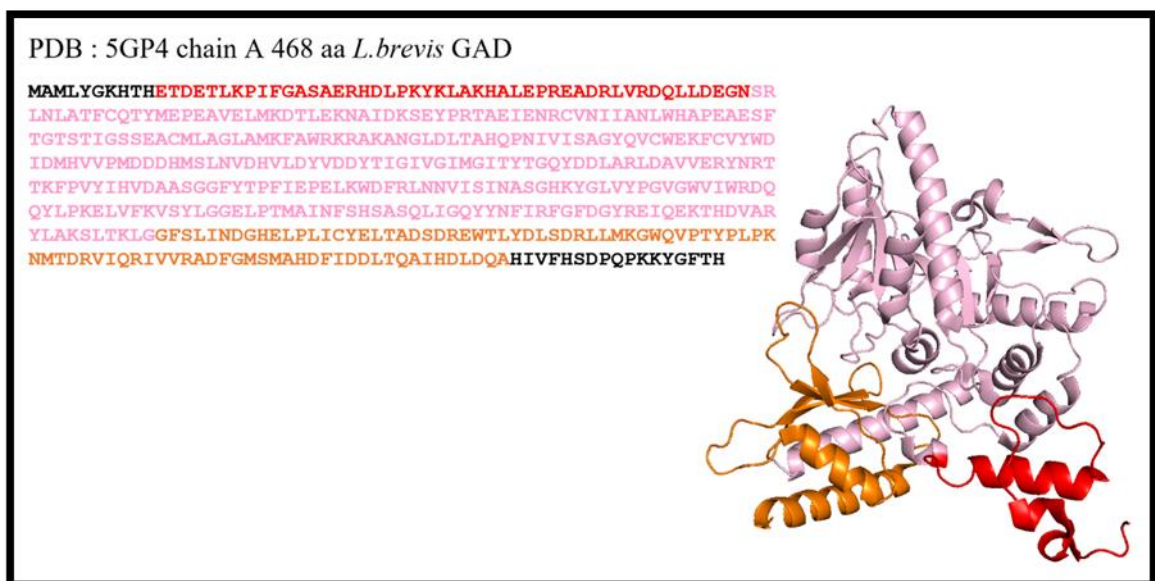
Figure 1.6: Schematic representation of *L. brevis* homodimeric GAD enzyme domains (PDB: 5GP4). The parts shown in red in the image indicate the N-terminal domain and the parts shown in orange indicate the Small-domain. The large area shown in pink is the PLP-binding domain containing the residues that make up the active site of the GAD enzyme. The black parts on the amino acid sequence are those whose structure could not determine in the study [133]. Coloring on the 3D structure was added using PyMOL.

In the *L. brevis,* GAD enzyme called the active site, PLP and L-Glu molecules are located close to each other, and they form bonds with specific amino acid residues. One of the most important bonds made by the cofactor PLP molecule is the covalent bond with

the amino acid Lysine at the 279th site of the *L. brevis* GAD protein[132, 133]. Moreover, PLP forms hydrogen bonds and hydrophobic bonds with many amino acids in the active site. The residues with which PLP establishes hydrogen bonds with the GAD enzyme are the amino acids Serine (sites at 126,127,276,321), Histidine (site 278), and Lysine (site 279). Also, Hydrophobic bonds form between the PLP-pyridine ring and the Ala248, Cys168, and Ile211 residues (Figure 1.7). On the other hand, L-Glu, which is the substrate of the GAD enzyme, forms hydrogen bonds with Thr64, Cys66, Gln166, and Ser321 in the active site; It also forms salt bridges with amino acids His278, Lys279, and Arg422[133].

Another important region in the GAD enzyme structure is the area between the 308 and 312 residues, which is called the flexible loop. The flexible loop region, which is close to the active and substrate-binding part, is effective in the enzyme's catalytic activity. Importantly, the amino acid Tyrosine at position 308 is involved in the decarboxylation of L-glutamate[133]. Many studies have investigated the forms in which GAD enzymes are active, and it has been found that they may differ according to species. For instance, *L. brevis* Gad1 protein is active in tetramer form[134], *L. brevis* Gad2 protein is functional as a monomer[135], and *L. plantarum* GadB protein is active as a dimer[136].

As explained in the previous section, the pH ranges of active L.brevis Gad1 and Gad2 proteins differ. In comparison, the Gad1 enzyme is active at a pH closer to neutral, and the Gad2 enzyme functions at a more acidic pH. Using the GAD enzymes, which have different effective pHs, is very advantageous for the *L. brevis* strains to achieve acid tolerance in a wide range of pH. Furthermore, since the ability of GAD enzymes to be active in a wide pH range is a situation that increases GABA synthesis, it is the subject of many studies to investigate the parts of GAD enzymes that can be effective in pH control. In mutagenesis studies with both *L. brevis* and *L. plantarum,* GAD enzymes concluded that the C-terminal parts of the enzymes affected pH. One of the studies observed that the enzyme activity increased to neutral pH when 14 amino acids were removed from the *L. brevis* GAD enzyme's C-terminal[136]. Similar results have also been shown by deletion studies in the C-terminal part of the GAD enzyme of *L. plantarum.* In addition, studies discovered that the C-terminal region of the *L. plantarum* GAD enzyme blocked the active site by undergoing a conformational change in response to pH changes[137].
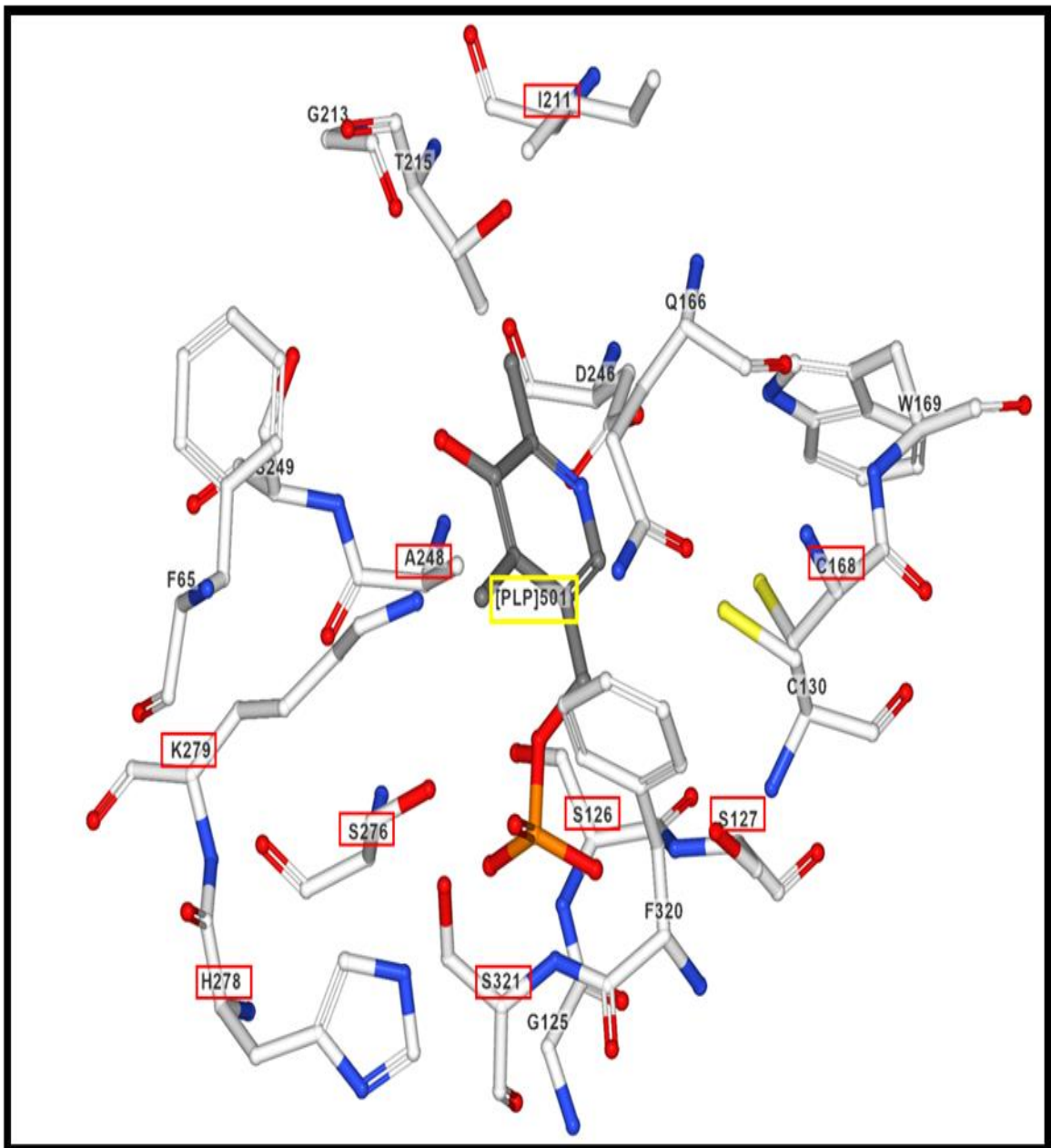
Figure 1.7: Active PLP binding sites of *L. brevis* GAD enzyme. The figure shows the residues with the potential to bond with PLP in the active site of GAD (surroundings 5Å). Essential amino acids around the PLP molecule marked with a yellow square are marked with red boxes. The figure was created with the NGL Viewer over the PDB 5GP4 code.

1.2.1.2 Glutamate/GABA Antiporter and Regulation of GAD System

The *gadC* gene in the GAD system synthesizes the transporter gene, and the antiporter protein, which is the product of *gadC*, is positioned in the bacterial cell membrane, allowing glutamate to be taken from the extracellular environment and given the formed GABA molecule out of the cell[138]. The presence of the *gadC* gene in LABs differs according to the LAB species. For example, in *L. brevis* strains, there is the *gadC* gene, which synthesizes the 501 aa antiporter protein, right next to the *gad1* gene[139], while there is no strain containing any *gadC* gene in *L. plantarum*. Moreover, there is no scientific paper that analyzes the structure of GadC protein by isolating it from LAB.

Like those on other GAD systems proteins, many investigations have revealed the biochemical and crystal structural features of the GadC antiporter protein in the *E. coli* (PDB codes 4DJI and 4DJK). The crystal structure of *E. coli* GadC protein has 12 transmembranes (TM) helix configurations. Also, the C-terminal region of the 511 amino acid *E .coli* antiporter protein forms the C-plug, which changes its conformation in response to pH changes (Figure 1.8)[140]. The C-plug structure has an essential role in ensuring the passage of glutamate into the cell through the antiporter protein. The transport of substrate glutamate with the GadC antiporter is pH-dependent and decreases towards neutral, according to proteoliposome transport test results employing *E. coli* GadC protein. Similarly, the study observed glutamate accumulation at near-neutral pH in GadC variants formed by the deletion of the C-plug part[140]. Additionally, the basic amino acids (especially H491 and R499) in the GadC protein's C-terminus enable the C-plug's conformational changes.

Regulation of the GAD system is critical in bacteria living in high acid environments. Many regulatory proteins regulate the GAD system in a species adapted to highly acidic environments, such as *E. coli*[141]. In contrast, limited information is available on the regulation of the GAD system in LABs. In LABs, the *gadR* gene typically found upstream of the *gadC* gene in the bacterial genome regulates the GAD system. Mutation studies of the *L. brevis gadR* gene have shown that the *gadR* gene is a positive regulator of the GAD system, increasing the expression of the *gadC1* operon (*gadC* and *gad1*) and GABA synthesis[142]. Therefore, the GAD system, positively regulated by the GadR protein, gives LABs an advantage in acid environments. Although the presence of a second gene for the GAD enzyme in the *L. brevis* genome implies that both genes may

have distinct regulatory proteins, there are no different genes other than *gadR* found in *L. brevis*. In addition, some of the other LAB species (*Latilactobacillus sakei*, *Lactococcus lactis*)[143] other than *L. brevis* may also have the *gadR* gene, but these genes have minimal similarity to each other.
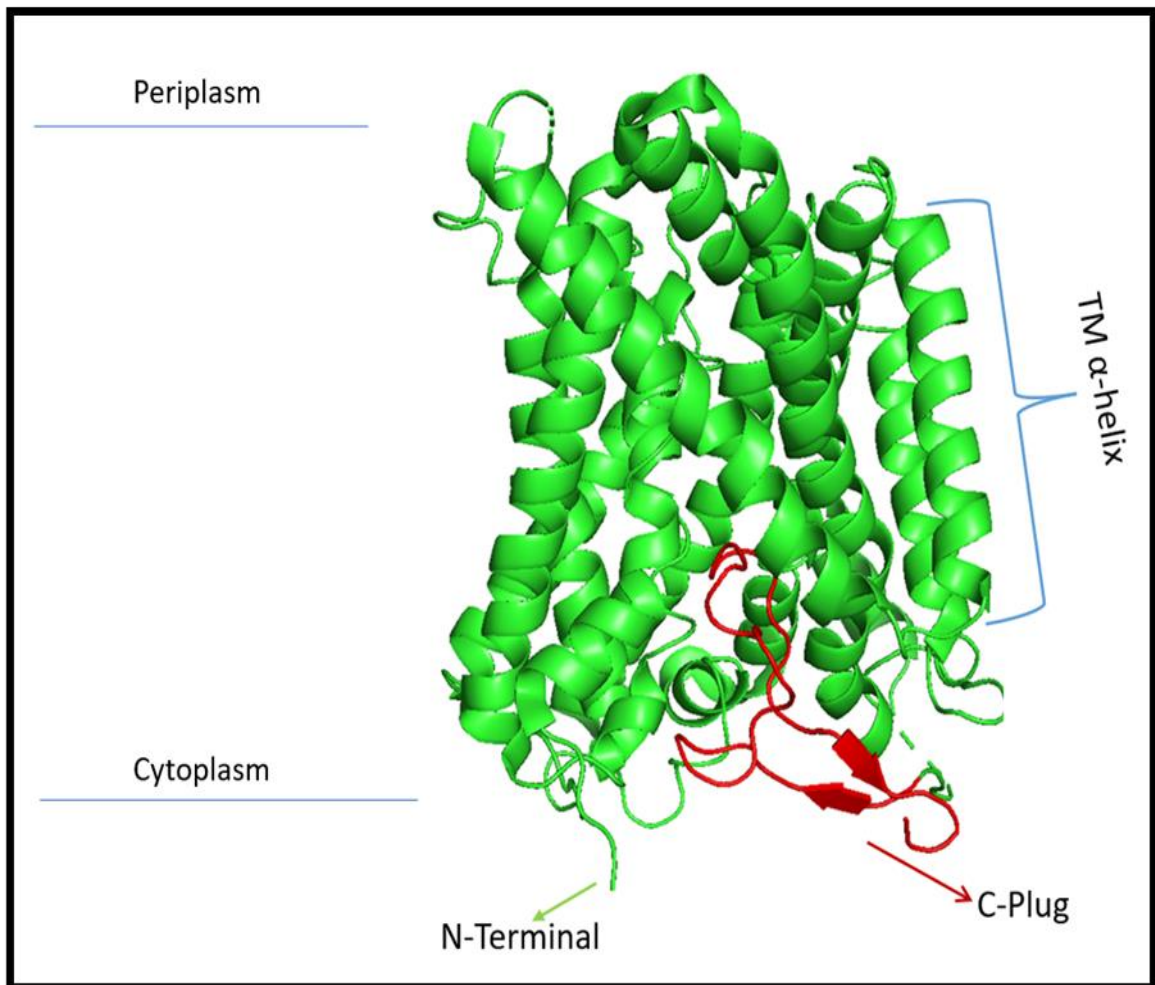
Figure 1.8: Structure of *E. coli* GadC (PDB: 4DJI). The red part of the protein represents the pH-dependent C-plug. TM α-helix structures between periplasm and cytoplasm are schematically indicated. The structure is visualized, and residues are colored by PyMOL.

1.2.2 GABA Features and Applications in Human Health

In addition to the advantages of the GAD system to LABs, the GABA molecule, which is the product of the system, has many benefits for human health (Figure 1.9). The GABA molecule, formed by the decarboxylation of L-Glu through the GAD system, is a non-protein amino acid that can also synthesize in plants and mammals[144]. GABA was identified in plants first[145], then in mammals, and found to work as an inhibitory neurotransmitter in the human brain[146]. Later, as research on the GABA molecule increased, it was shown to have antidepressant[147] and blood pressure-lowering properties[148]. Furthermore, research on mice fed a high-fat diet found that the GABA molecule could help prevent diabetes and obesity[149]. The benefits of the GABA molecule have increased the demand for functional foods containing GABA in recent years, and this demand continues to grow today. The ability of LABs to increase their numbers in various fermented products and produce GABA has proved efficient in obtaining GABA via microorganisms rather than chemically synthesizing it[150]. As a result, various studies in the literature on the GABA production processes in LABs and the system genes that comprise these pathways. In fermented foods, LAB species with high GABA production potential and their GABA production pathways are popularly studied.

1.3 Hypothesis and Aims of Thesis

LAB species with high GABA synthesis have an essential role in producing beneficial functional fermented foods. To increase GABA synthesis, it is critical to discover the species of LAB that produce GABA from food-derived environments and to understand the metabolic pathways involved in GABA production, both genetically and metabolically. Although several studies have identified LAB species capable of GABA synthesis, there is little knowledge of the genes involved in GABA synthesis and the system's control. Furthermore, there is no research in the literature on the evolutionary processes of the genes that comprise the GAD system.

The hypothesis of the thesis is that GAD pathway genes in LAB show molecular adaptation and selection to diverse high acidity ecological niches. Therefore, the thesis aimed to investigate two distinct LAB species (*L. brevis* and *L. plantarum*) with different GAD system organizations in order to better understand their molecular evolutionary

23

processes in adaptation to high acid environments by conducting molecular population genetic analysis with the GAD pathway genes. For this reason, genome sequences of *L. brevis* and *L. plantarum* strains isolated from various food and non-food environments are used.

In addition, the GAD system gene organization in diverse LAB species is compared via phylogenetic analyses. Based on phylogenetic relationships, divergence patterns of *L. brevis* and *L. plantarum* GAD system genes are compared with respect to other LAB species. Finally, the functional consequences of amino acid changes observed in *L. brevis* and *L. plantarum* genes have been investigated with molecular structural analyses.
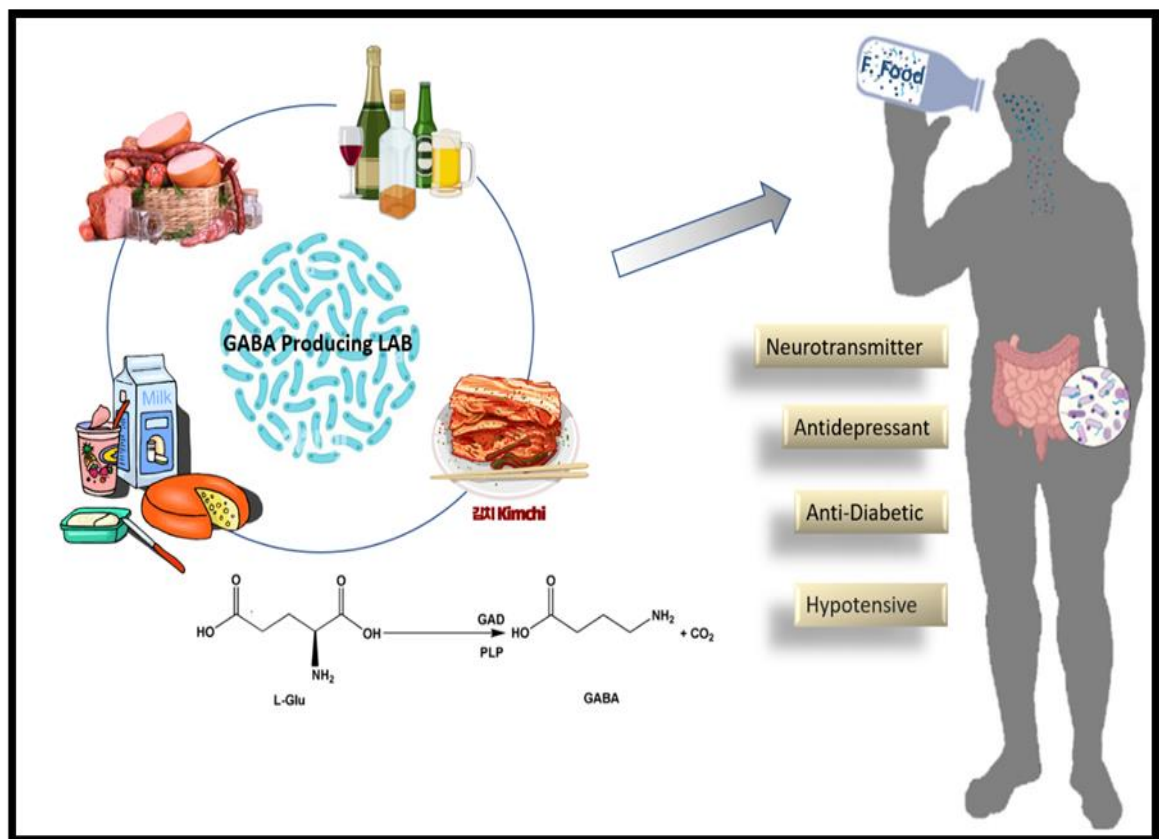
Figure 1.9: Schematic overview of GABA-containing foods and their beneficial effect on humans. F.Food: Functional food.

CHAPTER 2

MATERIALS AND METHODS

The study carried out within the scope of the thesis is computational analysis. In this context, all of the materials used in the study are data obtained from specific databases. There are two lactic acid bacteria species(*L. brevis* and *L. plantarum*) selected as the main species within the content of the thesis. The National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) databases such as nucleotide, gene, protein, and genome are used the obtain information about the GAD system of *L. brevis* and *L. plantarum*. From the information found in the databases, suitable gene and protein data for the study were selected and used as thesis material. In addition to the species selected as the main species, we used data of the species given in Table 1.1, and additionally, *Streptococcus thermophilus*, *Lactococcus lactis*, and *Escherichia coli*.

2.1. Data Collection

Entrez (http://www.ncbi.nlm.nih.gov/Entrez/) is a web interface that contains many databases and is located on the NCBI homepage[151]. When users enter the gene names or gene tags together with the organism name into the Entrez search box, the relevant information is retrieved from specified databases. Therefore, we used the NCBI Entrez searching system to access the necessary data for the thesis. We saved the information about the GAD system genes, which we found via NCBI, in FASTA (https://www.ncbi.nlm.nih.gov/genbank/fastaformat/) format.

2.1.1 *Levilactobacillus brevis* GAD Pathway Data

The GAD system genes in *L. brevis* and the positioning of these genes in the bacterial genome were mentioned in section 1.2.1. Accordingly, while searching on

NCBI, one of the selection criteria was that the genes in the operon form were located on the same gene cluster. When a search was done in the Entrez nucleotide database by typing (glutamate decarboxylase) AND "Lactobacillus brevis," 106 (checked on 17/03/2022), data were obtained for the *L. brevis* glutamate decarboxylase gene. Among the 106 *L. brevis* data found, strains containing the entire GAD system(*gadR-gadC-gad1* and separated *gad2*) were separated. Out of the 106 data, 36 strain data with known isolation sources were found, and the most suitable 30 *L. brevis* strain data were selected among these. Incomplete and pseudogenes were not included when determining appropriate data. Then, the 30 strains' data were divided into five groups according to where they were isolated. Table 2.1 shows the *L .brevis* strains selected for the study and their accession numbers on NCBI, grouped by isolation source.

There are two genes encoding the glutamate decarboxylase enzyme on the *L. brevis* genome, and one of these genes(*gad1*) creates an operon structure with the antiporter gene. As mentioned in the previous chapter, there is the *gadR* gene upstream of the *gadC* gene, which is included in the operon structure. In addition, *gad2*, which encodes the second GAD enzyme, is found in a different part of the genome. Considering previous information, the GAD operon structure and the *gad2* gene located outside the operon were found on the genomes of the selected *L. brevis* stains and saved in FASTA format. A 3785 nucleotide (nt) gene cluster including three genes(*gadR-gadC-gad1*) for the GAD operon structure was determined on the *L. brevis* strain genomes. There are protein-coding genes and non-coding regions between the genes in this cluster. In addition, the 1407 nucleotide *gad2* gene is located separately from the operon in each *L. brevis* strain (Figure 2.1).

In order to clearly distinguish the genes in the 3785 nt gene cluster, which is shown schematically in Figure 2.1, in addition to the information provided on NCBI, the regions that encode and do not code for the genes in the cluster were determined using the GeneMark.hmm prokaryotic (http://exon.gatech.edu/GeneMark/gmhmmp.cgi) web base. For analysis, gene clusters of strains were loaded in FASTA format, and *L.brevis ATCC 367* strain in the system was selected as a model organism. The purpose of using the GeneMark.hmm prokaryotic web base is that when individual gene data are retrieved via NCBI, sometimes the beginning or end of genes can be under-identified due to sequencing. Therefore, to obtain precise information, the GAD operon structure of *L.*

brevis strains was recorded as a gene cluster, and GeneMark determined the positions of related genes.
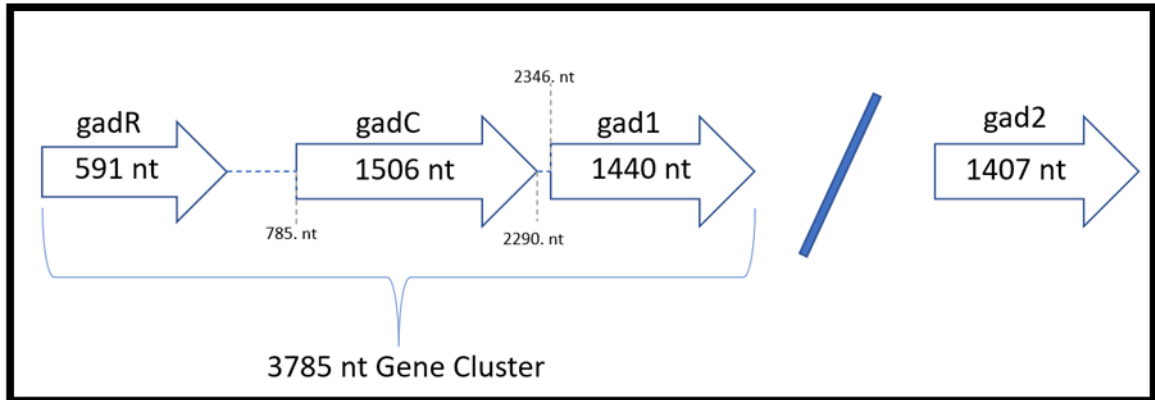
Figure 2.1: Schematic representation of the lengths and organizations of *L. brevis* GAD pathway genes in selected strains.

2.1.2 *Lactiplantibacillus plantarum* GAD Pathway Data

The same strategy used to gather *L. plantarum* GAD system data via NCBI was used to collect *L. brevis* GAD system data. The Entrez search engine was used to find the *L. plantarum* glutamate decarboxylase enzyme, and suitable strains were chosen from the findings. When the glutamate decarboxylase enzyme data of *L. plantarum* was searched on NCBI, 457 (checked on 17/03/2022) strains' data were retrieved. A total of 92 strains were saved from the findings, with a preference for those whose entire genome was complete and whose isolated environments were defined. In the study, 88 of the chosen strains were analyzed.

There is only one GAD gene in *L. plantarum*, and in addition, there is no studies have been published defining the *L. plantarum* antiporter or regulator gene. Hence, a single GAD gene found in *L. plantarum* was used in the analyses. Moreover, when the *L. plantarum* strains with the whole genome were examined within the scope of the research, only one GAD gene (*gadB*-1410nt) was found, and no defined *gadC* or *gadR* genes were found. However, despite the absence of a specific *gadC* gene detected in the *L. plantarum* genome, general transporter protein data under the YjeM tag was found when protein

BLAST(https://blast.ncbi.nlm.nih.gov/Blast.cgi) was applied to *L. plantarum* genomes using *L. brevis* GadC protein sequence.

The gene product under the YjeM tag in the *L. plantarum* genome is named glutamate/gamma-aminobutyrate transporter family protein. Although no particular studies on the *yjeM* gene have been published, the identical tag and protein nomenclature define in the *E. coli* genome (UniProtKB - P39282 (YJEM_ECOLI)). Following these findings, it was hypothesized that the *L. plantarum yjeM* (1485nt) gene might be a general transporter for GABA and glutamate, and the *yjeM* gene data from the 88 strains were obtained.

As seen in Table 2.2, the number of *L. plantarum* strains collected from NCBI is high, and the sources from which the strains are isolated are quite diverse. Nine specific groups were formed considering the diversity in isolation sources of *L .plantarum* strains. Eight of the nine created groups contain strains isolated from food-based environments, and the other group includes *L. plantarum* strains isolated entirely from fecal samples.

2.1.3 GAD Pathway and 16S rRNA Information of 32 Bacterial Species

GAD enzyme gene data and *gadC* gene data of 31 bacterial species were obtained from LAB genomes available on NCBI to perform interspecific analysis. In addition, two GAD enzyme genes (*gadA* and *gadB*) and *gadC* gene information of *E. coli*, the organism in which the GAD system is most studied, were used. All 31 selected LAB species do not have the *gadC* gene. The *yjeM* gene was taken for analysis from those LAB species that do not have the *gadC* gene in their genomes.

Table 2.3 and Table 2.4 show the selected species' *gad* genes and *gadC*(or *yjeM*) genes. While collecting data on the species, attention was paid that the isolation environments were especially food-based. However, for some of the species given in the tables, *gadC* (or *yjeM*) and isolation source information are not available on NCBI.

One of the methods used to determine taxonomic similarities between bacteria is to construct phylogenetic trees using 16S rRNA sequences.16S rRNA is found in all bacterial species and has been highly conserved species-specific throughout evolution[152]. Therefore, to determine the taxonomic relationships of 32 selected species, 16S rRNA data of 32 selected species were found using the NCBI Entrez system. In Table 2.5, the 16S rRNA accession numbers of 32 species, the length of the gene, and the information

about the isolation source of the species are shared. Some species' isolation sources were not included in the table because the NCBI did not contain their source information.

2.2 Sequence Alignments and Phylogenetic Trees

Before beginning analyses to be made in the thesis, it is essential to check the accuracy of the gene data recorded in FASTA format and make the necessary arrangements on the data. Therefore, UGENE[153] and MEGA-X[154] bioinformatics tools were used to visualize and control the data obtained from NCBI about the GAD system genes of *L. brevis* and *L .plantarum* strains. In the control of the collected data with bioinformatics tools, the presence of any stop codons, insertions, or deletions that could prevent proper protein synthesis, especially in the nucleotide sequences of protein-coding genes, was investigated. Furthermore, when the *L .brevis* and *L. plantarum* genes are collected using the NCBI Entrez system, a text-based system, the similarities of the strain genes to each other must be controlled using sequence alignment to prove accurate data were collected. MEGA-X software was used to align the GAD system genes. The GAD pathway genes(*gadR-gadC-gad1/gad2*) of 30 *L. brevis* strains were checked by aligning each gene to each other. Similarly, the genes were controlled by making separate alignments for the *gadB* and *yjeM* gene data of 88 *L. plantarum* strains.

MEGA-X is a tool that allows multiple sequence alignment with many different algorithms. In the thesis, the MUSCLE[155] algorithm was used while aligning the genes with MEGA-X. MUSCLE is a popular alignment algorithm with fast and high accuracy, which is mostly used in studies with bacterial genes too1[156-158]. The MUSCLE alignment algorithm also was used to align gad enzyme, *gadC*(or *yjeM*), and 16S rRNA[159] genes taken from selected 32 bacterial species for interspecies analysis. In particular, while homologous genes belonging to different bacterial species were aligned with each other, nucleotide data was first converted to protein data on MEGA-X and aligned with MUSCLE, and then aligned protein data was converted to nucleotide sequences. Thus, a more uniform alignment was obtained according to the encoded amino acid types.

The outputs generated by the alignment of GAD genes were visualized using MEGA-X and Jalview[160]. From the visualized alignment files, sites that are conserved between strains, observed in a single (non-informative/singleton site) strain, or observed in many strains(informative sites) were identified. Also, the sites found to be conserved

in all species were determined from the alignments made with the genes of 32 species, and the properties of the conserved residues were examined. Moreover, gene trees were constructed using alignment files of the *gad* gene and *gadC*(or *yjeM*) genes of 32 species to identify interspecific taxonomic similarities. Gene trees were created using single-gene data of 32 species with MEGA-X.

Phylogenetic trees are made by calculating the genetic distances of the gene sequences of the species. There are many models defined in the literature to calculate the distances of the sequences from each other, and these models differ according to whether the sequence is a nucleotide or protein sequence. For this reason, before creating phylogenetic trees, the most suitable model with MEGA-X was analyzed according to the type of substitutions (nucleotide or protein) to be used. With the MEGA-X/Models section, the estimation of the most suitable model can be done quickly by using the alignments of the sequences. The model combinations with the lowest Bayesian Information Criterion (BIC) score are chosen as the best acceptable model among the various models tried for substitutions with MEGA-X. After selecting the appropriate model, gene trees were created using the Maximum Likelihood (ML)[161] method with MEGA-X. The models obtained for each gene alignment were employed when the phylogenetic tree on MEGA-X was created using the ML approach. In addition, the bootstrap method[162] (1000 replication) was employed to evaluate the phylogeny, and the complete deletion option was used to treat the provided data.

Phylogenetic trees consisting of data belonging to a single gene often do not show the taxonomy of the species from which the gene was taken. Therefore, a species tree was created with 16S rRNA data of 32 bacterial species to determine their taxonomic relationship. The 16S rRNA species tree was constructed similarly to gene trees with MEGA-X. First, the 16S rRNA sequences of 32 species were aligned with MUSCLE, and then, using this alignment output, the model suitable for the species tree was defined. A phylogenetic tree was created with the ML method by using the appropriate model, 1000 bootstrap replication values, and the complete deletion option.

As a result, phylogenetic trees generated for 32 bacterial species and GAD pathway genes were recorded as the rooted and unrooted trees.

2.3 Molecular Population Genetic Analyses

This section aims to perform intraspecific and interspecies population genetic analyses by using the alignments created with the GAD system genes of *L. brevis* and *L. plantarum* strains. DNA Sequence Polymorphism (DnaSP version 6)[163] software was used for the necessary analyses, and the results were saved in tables.

2.3.1 Intraspecies Analyses

Firstly, in intraspecies analysis, 3785 nucleotide cluster(operon) alignments of *L. brevis* strains containing three genes(*gadR-gadC-gad1*) were loaded into the DnaSP program. The gene regions determined in section 2.1.1 are defined for each gene as coding regions. Afterward, 30 *L. brevis* strains were divided into five groups on DnaSP according to the sources where they were isolated. Similar procedures were performed for *L. brevis gad2* and *L. plantarum* GAD system genes(*gadB* and *yjeM*), which are not in operon form. Sequences of 88 *L. plantarum* strains were identified on DnasSP as nine different isolation groups. Analyzes were made both for the whole population of *L. brevis* and *L. plantarum* and for each isolation group separately.

The alignments of *L. brevis* and *L. plantarum* strains defined on DnaSP were used for the polymorphism analysis. According to this study, polymorphic sites were identified in all populations and isolation groups. Monomorphic and polymorphic sites were determined through polymorphism analysis, and singletons, synonymous, and replacement (non-synonymous) sites were identified from polymorphic sites. Moreover, DNA polymorphism calculations were made simultaneously with the data on which polymorphism analyses were applied. For each gene, polymorphic(segregating) sites(S) and total mutation numbers(Eta) were determined for *L. brevis/L. plantarum* populations and also for isolation groups. In addition, nucleotide diversity(Pi,π) and Theta-W (Theta(Θ) (per site) from S) calculations were performed with DnaSP.

2.3.1.1 Neutrality Evolutionary Tests

The analyzes made in the previous sections are not sufficient to explain the evolutionary scenarios of the populations. In order to understand what kind of

evolutionary processes the populations are in, some statistical calculations are needed. With statistical analyses called neutrality tests it is tested whether the data of a certain population deviates from the neutral scenario, which is accepted as the null hypothesis. The neutrality analysis applied in the study tested which evolutionary scenarios the genes of the GAD system shown in different isolation environments of the same bacterial species.

The Pi(π) and Theta(Θ) values calculated in the previous section form the basis of neutrality tests. Pi(π), defined as nucleotide diversity, refers to the average pairwise difference between individuals of a population. On the other hand, Theta(Θ) is the calculation of the expected Pi(π) value for a population that is in a completely neutral process(Figure 2.2).

$$\theta = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}} \qquad \begin{array}{l} \text{S : Segregating Sites} \\ \text{n : Number of Sequences} \end{array}$$

Figure 2.2: Theta(Θ) Equation

Two different neutrality tests were used in the intraspecies analysis. Both tests are tests performed with a single population. The first of these is Tajima's D (TD)[164] statistics test. The TD test is based on calculating the difference between nucleotide diversity(π) and Theta(Θ). In a neutrally evolving population, the two values expect to be equivalent under the test. Because mutations in the DNA sequence occur randomly in neutrally evolving populations and do not affect the population's adaptation to its environment, different evolutionary scenarios may arise if the mutations in the genetic sequence are not random. Therefore, the TD test is an instrumental analysis for identifying sequences that deviate from neutral theory[164]. For TD analysis on DnaSP, four or more aligned

homologous gene sequences from a species are required. In the thesis, each GAD system gene of *L. brevis* and *L. plantarum* species was tested individually for different isolation environments.

The values obtained from the TD analysis represent three different interpretations[165]. The first is the interpretation that accepts the neutral null hypothesis that Pi(π) and Theta(Θ) values are equal. The other two interpretations are cases where the null hypothesis is rejected, and differences are observed between the Pi(π) and Theta(Θ) values. If the calculated TD value is negative, it means that the observed Pi(π) value is less than the expected Theta(Θ) value. In this case, called negative selection, rare alleles are present in the genetic sequence. This is observed when the variation on the gene is deleted by the sudden growth of a population that has been small for a long time[166]. Unlike negative selection, a positive TD value is observed in the case of positive selection, and this is when the observed nucleotide diversity(π) is greater than the expected Theta(Θ) value. The positive selection indicates balancing selection in which variation is kept in the population, and rare alleles are found less. Balancing selection[167] is observed in populations that have experienced sudden contraction recently.

Fu and Li's tests[168] (D* and F*) are another of the intraspecies neutrality tests. Fu and Li tried to explain the selection by determining that the mutations occurring in a population are in the external or internal branches of the gene tree with their tests based on singletons that affect one individual in the population. When gene trees of a population are created, old mutations are found in the internal branches, while newly formed mutations tend to be found in the external branches. Mutations in the negative selection(or purifying) are observed in the external branches because the frequency of deleterious mutations occurring in the population is very low. Similarly, a new mutation fixed in the population tends to be found in the external brach[168]. However, when there is a balancing selection, the alleles formed in the population tend to be found in the internal branches because they are old[168].

DnaSP can perform Fu-Li statistic tests quickly. The interpretations of the Fu-Li test results are similar to those of the TD test. If the numbers returned from the Fu-Li statistical analysis are significantly negative, it implies that this population contains singletons. On the contrary, there are old mutations in the population in the positive selection scenario, and the values of Fu-Li statistics are pretty positive.

2.3.1.2 Sliding Window Analysis

The visualization of both neutrality tests applied to gene sequences of populations and nucleotide diversity observed in these sequences are mostly used in evolutionary biology. The sliding window analysis is one in which the statistics tested on the gene sequence are shown as plots of varying widths[169]. Sliding windows can be created in desired lengths for statistics of genetic data TD, Fu-Li, and nucleotide diversity analysis on DnaSP[170]. In this way, the statistical values corresponding to each nucleotide on the gene sequence and their distribution in the sequence can be observed.

Sliding window analyses in the thesis were performed for all GAD system genes of *L. brevis* and *L. plantarum* species and mainly focused on non-synonymous sites. Sliding window images for nucleotide diversity in non-synonymous sites on GAD system genes were easily created with DnaSP. However, there is no separate tool to create sliding window images of TD values non-synonymous sites on DnaSP. Therefore, the R[171] language was used to create sliding window images showing the TD values of non-synonymous sites. Simple plot graphics were created with the non-synonymous site TD values of each GAD system gene defined on R.

2.3.2 Interspecies Analyses

The 32 bacterial species from which *gad* enzyme gene and *gadC*(or *yjeM*) gene information were obtained were used in the interspecies analysis. Three genes(*gadC-gad1*and *gad2*) from *L.brevis* and two(*gadB* and *yjeM*) from *L. plantarum* were analyzed one by one with other bacterial species. One of the statistical tests used in the interspecies analysis is the McDonald and Kreitman test(MK)[172]. The MK test is one of the tests used to measure the presence and level of adaptive evolution. The MK test compares the diversity within the population of a species(polymorphism) with the observed diversity among species(fixed). The MK test evaluates whether a mutation is neutral, deleterious, or favorable, as well as whether it is synonymous(silent mutation) or non-synonymous in a protein-coding gene[173]. In the MK analysis, a table is created, as shown in Figure 2.3. The polymorphism observed within the species, and the number of fixed changes between

species, as well as the information on whether the sites are synonymous or non-synonymous, are revealed.

In MK analysis, the null hypothesis assumes that the non-synonymous to synonymous ratios within and between species are equal, and events that deviate from this equality are defined as positive(Dn/Ds > Pn/Ps) or negative(Dn/Ds < Pn/Ps) selection[173]. On top of the general definition of the MK test, new statistical calculations such as alpha(α)[174] and neutrality index(NI)[175] are derived(Figure 2.4). Alpha(α) is an indicator for positive selection, while the NI determines the degree and direction of deviation from the MK neutral null hypothesis.

	Fixed Sites (Between Species)	Polymorphic Sites (Within Species)
Synonymous	Ds	Ps
Non-synonymous	Dn	Pn

Dn/Ds : the ratio of non-synonymous to synonymous variation between species

Pn/Ps : the ratio of non-synonymous to synonymous variation within species

Figure 2.3: Schematic representation of the MK test table (2 x 2 table)

$$\alpha = 1 - \frac{D_s P_n}{D_n P_s} \qquad NI = \frac{P_n/P_s}{D_n/D_s}$$

$\alpha = 0$ Neutral $\qquad$ NI = 1 Neutral

Figure 2.4: The equations of α and NI

In addition to the MK analysis done in DnaSP, the direction of selection(DoS)[175] calculations were also made using the 2x2 MK table. The reason for making DoS calculations is that NI is undefined due to zero values in the MK table. With DoS calculation, the direction and extent of the evolution of a gene can be determined[175]. Although DoS and NI are similar equations, they are not the same. Genes with equal NI values may not have the same DoS values. Figure 2.5 shows the DoS equation and interpretations of NI and DoS values.

NI > 1	Excess of amino acid polymorphism. (There are slightly deleterious mutations/Negative selection)
NI < 1	Excess of non-silent divergence. (Positive selection)
NI =1,Dos =0	Neutral
DoS > 0	Positive selection. (Evidence of adaptive evolution)
DoS < 0	There are slightly deleterious mutations. (Negative selection)

$$DoS = \frac{D_n}{D_n + D_s} - \frac{P_n}{P_n + P_s}$$

Figure 2.5: DoS equation and interpretations of DoS and NI values[175].

Another addition to the MK test on DnaSP for interspecies comparisons, the Fu-Li (D and F)[168] test, was done using species close to *L. brevis* and *L. plantarum* as outgroups for each GAD system genes of *L. brevis* and *L. plantarum*. The species close to *L. brevis* and *L. plantarum* in terms of GAD system genes were determined according to the results of MK analysis.

2.3.3 Codon-based Model: Site Model Analyses

There are codon-based models used to determine how a protein-coding gene evolved through time. These models allow the identification of genes that are under selection pressure while also allowing them to be identified in positively selected sites of a protein-coding gene[176]. One of the methods of detecting protein-coding genes under selection is omega(ω = dN/dS) values using non-synonymous(dN) and synonymous(dS)

rates[177]. The neutral value for omega(ω) is considered to be one; cases, where the omega is greater than one, indicate the positive selection, while cases, where it is less than one indicates negative selection. When the omega value is greater than one, it means that non-synonymous changes provide an advantage to the population and are fixed. The rate(dN) is higher than the rate of synonymous changes.

The site model has been developed to identify sites in the protein sequence under selection using varying omega values[178]. Lactic acid bacteria are also exposed to different environmental pressures in various environments in which they live. Therefore, site model analysis was used to investigate whether there are positively selected sites in the *L. brevis* and *L. plantarum* GAD system genes. The PAML package(codeml)[179] and its graphical user interface PAMLX[180], which performs the calculation using maximum likelihood scores, were used in the site model analysis. Aligned data for each GAD system gene of strains belonging to *L. brevis* and *L. plantarum* species and Newick(.nwk) formats of phylogenetic trees created from these alignments were used in the analysis. Analysis tables were created by testing each GAD system genes of *L. brevis*(*gadR-gadC-gad1* and *gad2*) and *L. plantarum*(*gadB* and *yjeM*) on PAMLX with models using different omega(ω) values.

2.4 Protein Sequence-Based Analyses

Non-synonymous polymorphisms occurring on the nucleotide sequence of a gene cause changes in the amino acid sequence of the protein molecule to be formed. Amino acids that change the protein sequence structure can affect the normal function of the protein[181]. If amino acids that change in protein structure due to non-synonymous polymorphisms have similar properties to previous amino acids, they have little effect on the protein's function. However, if the changed amino acids have different properties than the previous ones, they can significantly impact the function of the protein. Therefore, the non-synonymous sites determined by polymorphism analyses on DnaSP have been studied in the GAD pathway gene alignments of *L. brevis* and *L. plantarum* strains. It has been defined in which regions of the proteins the amino acid changes occur and which amino acid replaces the previous one. In addition, the distribution of non-synonymous sites in GAD system genes according to defined isolation groups of *L.brevis* and *L. plantarum* was also studied. Replacement changes were identified for each of the six

genes(*L. brevis gadR-gadC-gad1/gad2* and *L. plantarum gadB/yjeM*), and it was determined which amino acid codons these changes constituted. Much information from the literature and NCBI(www.ncbi.nlm.nih.gov/Class/Structure/) about the new amino acids formed due to non-synonymous changes and the amino acids found in that site before were collected, and detailed tables were created.

The isoelectric points(pI)[182] of protein data from *L. brevis* and *L. plantarum* strains with replacement changes were analyzed to assess the influence of replacement changes on the protein. The factors affecting the pI gain importance as the isoelectric point indicates the pH value of the protein at which the net charge is zero. Because proteins have ionizable types of amino acids and the charges of these amino acids change according to pH[182]. The pI values were calculated with a web-based application Prot Pi (www.protpi.ch), and the pKa values to be used in the pI calculation was set to be the values taken from ExPASy. Furthermore, the charges of the strains containing replacement changes were calculated using Prot Pi web-based application at varying pH values.

Another tool used to describe the effects of replacement changes on protein functions is PROVEAN[183]. The web-based application predicts the effects of amino acid changes on protein function. Protein sequences from each of the *L. brevis* and *L. plantarum* GAD genes that do not contain replacement changes were used for PROVEAN analysis, and the amino acid changes identified in these protein sequences were processed as variants on the program. PROVEAN performs BLAST analysis on NCBI using the given data and variant information and determines scores for each variant according to other data it finds homologous to the provided data[183]. PROVEAN defines either the variants as neutral or deleterious according to analysis scores.

2.4.1 Domain Analyses of *gad* Genes from *L. brevis* and *L. plantarum*

Among the LAB species, the only species whose GAD enzyme structure is known chemically is *L. brevis*. The *L.brevis* GAD enzyme contains three major domains, as shown by its PDB structure (5DP4).In the domain analysis, the sequence of the protein whose structure is known in the database was taken and aligned with the *L. brevis gad* genes. After finding which *gad* gene of *L. brevis* represents the structure in PDB, the domains were determined by aligning the *L. brevis* (*gad1* and *gad2*) and *L. plantarum*

(*gadB*) *gad* genes. The π and TD values of the domain regions defined in the three genes were calculated separately. Moreover, the residues, which are essential in terms of their bonding potential with PLP and L-Glu in the structure study, were determined for both *L. brevis* and *L. plantarum gad* genes as a result of alignment.

2.4.2 Codon Usage Bias (CUB)

CUB[184] is based on the codon composition of a gene and expresses the frequency of synonymous codons found for particular amino acids in the gene sequence. Besides, CUB-related analyses are correlated with gene selection and expression levels. Since CUB indices are strain-specific values, they are frequently used for gene prediction or for detecting a gene that has been transferred to the microorganism. Moreover, it is stated that the CUB value of a gene is closely related to the expression level of the gene in prokaryotic organisms. Because in a rapidly growing bacterial population, too much tRNA recognizes a small number of codes for genes that are expressed too much[185]. There are many CUB-related analyzes. Two of these analyzes, codon bias index(CBI) and codon adaptation index(CAI), were used in the study.

CBI measures whether synonymous codons used in a gene are preferred or not. It uses the ratio between preferred codons in a gene and the total number of codons to measure the usage of preferred codons. CBI calculations are generally used to identify a foreign gene in the bacterial genome. The CBI measurement can have values ranging from 1 to -1[184]. The exact mean value of zero implies that codons are chosen at random, value 1 shows that preferred codons are utilized, and values around -1 suggest that codons are not favored.

On the other hand, CAI measures the frequency of usage of all codons for a gene by using predetermined reference values. CAI values are utilized to determine the expression level of the gene of interest since the codon use tables generated for the organism examined are employed in the computation. Unlike CBI, the values of CAI measurements vary between 1 and zero[184]. The closer the score obtained from the analysis result is to 1, the higher the frequency of the codons used in the analyzed gene and the higher the expression level of the gene.

CBI analysis can be easily performed using aligned sequences on DnaSP. CBI values were calculated for each *L. brevis* and *L. plantarum* GAD gene in the thesis.

Besides this, CAIcal(http://genomes.urv.es/CAIcal/), a web-based application, was used to calculate the CAI values. The codon usage tables required for CAI calculation were obtained from a database(http://www.kazusa.or.jp/codon/).In addition to the genes for which CBI analysis was performed, CAI analysis also was calculated for *gad* genes and *gadC*(or *yjeM*) genes from 32 species.

2.4.3 GAD System Protein Structure Prediction

As discussed in Chapter 1, the three-dimensional structure of only one of the *L. brevis gad* genes among the *L. brevis* and *L. plantarum* GAD system genes studied in the thesis is known. No other study examines and reveals the three-dimensional structure of the GAD system genes in LAB species. Only the non-LAB type *E.coli gadC* transporter structure has been determined. Therefore, servers that can predict the structure of many proteins have been used to understand the protein structure of the GAD system genes and observe where important amino acids are located in the three-dimensional structure. Since the information provided by each protein structure prediction server is different, a few of these servers were used. From the servers, the secondary structure of the protein, its three-dimensional structure, and the predictions of regions with transmembrane helix formations, especially for transporter proteins, were examined. The protein structures obtained from the protein structure prediction servers shown in Figure 2.6 were examined by visualizing with PyMOL[186].

2.5 Statistics and Graphs

The results of intraspecies polymorphism and neutrality test values were gathered together by preparing a table specifying the isolation group identified for each *L. brevis* and *L. plantarum* GAD gene. Then, this table was compared for each variable, gene, and group, and their statistical significance was calculated using the Kruskal and Shapiro tests using the R programming language[187]. Comparisons with statistically significant results were graphed using the R ggplo2[188] package.

SERVERS		Links
SWISS-MODEL[189]	Template-based structure prediction Target-template alignment Different interaction modes	swissmodel.expasy.org/
I-TASSER[190]	Ranked the best method in 7th CASP experiment Full-length 3D protein structural prediction Output includes C-score, TM-score, and RMSD	zhanggroup.org/I-TASSER/
Phyre2[191]	Homology detection of 3D protein structure Analyzes protein function and mutations	www.sbg.bio.ic.ac.uk/phyre2/
AlphaFold[192]	Neural network-based model Can predict protein structures even no homologous structure The best method in 14th CASP	AlphaFold Colab (AlphaFold.ipynb)

Figure 2.6: Servers for 3D protein structure prediction[189-192].

Table 2.1: GAD pathway genes (*gadR+gadC+gad1/gad2*) information of *L. brevis* strains.

Sequence Mode	Name	Accession Number	Source	Group
Complete	*Levilactobacillus brevis NPS-QW-145*	CP015398	kimchi(Hong Kong)	**Fermented Vegetable Group**
Contig	*Levilactobacillus brevis D7*	NVYO01000001	kimchi(South Korea)	
Contig	*Levilactobacillus brevis G101*	MULL01000001	kimchi	
Contig	*Levilactobacillus brevis SRCM103306*	SBJN01000006	Food(South Korea)	
Complete	*Levilactobacillus brevis 100D8*	CP015338	rye silage(South K)	
Complete	*Levilactobacillus brevis KB290*	AP012167	Suguki(Fermented vegetable)	
Contig	*Levilactobacillus brevis CRL2013*	MZMW01000001	sourdough(Argentina)	**Sourdough Group**
Contig	*Levilactobacillus brevis TR055*	QFDK01000006/03	sourdough(Ireland)	
Contig	*Levilactobacillus brevis TR052*	QFDL01000013/03	sourdough(Ireland)	
Contig	*Levilactobacillus brevis TR169*	QFDG01000024/16	sourdough(Ireland)	
Complete	*Levilactobacillus brevis UCCLBBS124*	CP031169	Beer Keg	**Fermented Beverage Group**
Complete	*Levilactobacillus brevis TMW 1.2108*	CP019734	wheat beer(Germany)	
Complete	*Levilactobacillus brevis TMW 1.2112*	CP016797	wheat beer(Germany)	
Complate	*Levilactobacillus brevis NSMJ23*	CP050541	K.T.Alcoholic Beverage	
Complate	*Levilactobacillus brevis UCCLB521*	CP031208	brewery environment	
Contig	*Levilactobacillus brevis TMW 1.465*	JXUG01000003	brewery environment	
Contig	*Levilactobacillus brevis TMW 1.313*	JXUF01000002	beer	
Complete	*Levilactobacillus brevis UCCLBBS449*	CP031198	beer	
Contig	*Levilactobacillus brevis BSO 310*	LGIX01000030/17	brewery	
Complete	*Levilactobacillus brevis BSO 464*	CP005977	Brewery	
Contig	*Levilactobacillus brevis TUCO-5E*	QMCB01000017/01	milk	**Fermented Dairy product Group**
Complete	*Levilactobacillus brevis HQ1-1*	CP046631	T. dairy products	
Contig	*Levilactobacillus brevis D6*	LQNG01000043/24	Cheese	
Contig	*Levilactobacillus brevis KMB_620*	QMJT01000025/48	bryndza cheese	
Contig	*Levilactobacillus brevis KMB_615*	QMJO01000051/32	bryndza cheese	
Contig	*Levilactobacillus brevis 47f*	LBHR01000030/01	feces(Homo sapiens)	**Feces Group**
Contig	*Levilactobacillus brevis 15f*	JXCD01000017/02	feces(Homo sapiens)	
Contig	*Levilactobacillus brevis TMW 1.6*	JXUE01000015	feces	
Complate	*Levilactobacillus brevis ZLB004*	CP021456	feces (pig)	
Contig	*Levilactobacillus brevis DPC 6108*	MDUA01000001	feces(Homo sapiens)	

Table 2.2: GAD pathway genes (*gadB/yjeM*) of *L. plantarum* strains.

Sequence Mode	Name	Source	Accession Number	
Complete	*Lactiplantibacillus plantarum CAUH2*	Sichuan pickled vegetables	CP015126	
Complete	*Lactiplantibacillus plantarum JBE245*	fermented soybean (South Korea)	CP014780	
Complete	*Lactiplantibacillus plantarum CNEI-KCA5*	Fermented Okpei-Nsukka	CP059294	
Complete	*Lactiplantibacillus plantarum BK-021*	Fermented onions (South Korea)	CP044233	
Complete	*Lactiplantibacillus plantarum CNEI-KCA4*	Fermented Okpei-Onitsha	CP053571	
Complete	*Lactiplantibacillus plantarum AMT74419*	kimchi (South Korea)	CP052869	
Complete	*Lactiplantibacillus plantarum KCCP11226*	kimchi (South Korea)	CP046262	
Complete	*Lactiplantibacillus plantarum EM*	kimchi (South Korea)	CP037429	
Complete	*Lactiplantibacillus plantarum ATG-K6*	kimchi (South Korea)	CP032464	
Complete	*Lactiplantibacillus plantarum ATG-K2*	kimchi (South Korea)	CP032460	
Complete	*Lactiplantibacillus plantarum IDCC3501*	kimchi (South Korea)	CP031702	
Complete	*Lactiplantibacillus plantarum b-2*	pickle (China)	CP027349	
Complete	*Lactiplantibacillus plantarum DSR_M2*	kimchi (South Korea)	CP022294	**Kimchi**
Complete	*Lactiplantibacillus plantarum HAC01*	White Kimchi (South Korea)	CP029349	**Group**
Complete	*Lactiplantibacillus plantarum KC28*	kimchi (South Korea)	CP026743	
Complete	*Lactiplantibacillus plantarum LM1004*	cabbage kimchi (South Korea)	CP025988	
Complete	*Lactiplantibacillus plantarum PC520*	Chinese fermented food-pickles	CP023772	
Complete	*Lactiplantibacillus plantarum NCU116*	Chinese pickle	CP016071	
Complete	*Lactiplantibacillus plantarum DSM 20174*	pickled cabbage	CP039121	
Complete	*Lactiplantibacillus plantarum SRCM101511*	Radish Kimchi (South Korea)	CP028235	
Complete	*Lactiplantibacillus plantarum SRCM102737*	Soybean paste (South Korea)	CP028261	
Complete	*Lactiplantibacillus plantarum SRCM101518*	Radish Kimchi (South Korea)	CP028241	
Complete	*Lactiplantibacillus plantarum SRCM101222*	Radish Kimchi (South Korea)	CP028229	
Complete	*Lactiplantibacillus plantarum SRCM101187*	White Kimchi (South Korea)	CP028226	
Complete	*Lactiplantibacillus plantarum SRCM101167*	Water kimchi (South Korea)	CP028334	
Complete	*Lactiplantibacillus plantarum SRCM101105*	kimchi (South Korea)	CP028222	
Complete	*Lactiplantibacillus plantarum SRCM100995*	Pickled Green Chili Peppers	CP028275	
Complete	*Lactiplantibacillus plantarum PMO08*	kimchi (South Korea)	CP062059	

Table 2.2: (cont.).

Sequence Mode	Name	Source	Accession Number	
Complete	*Lactiplantibacillus plantarum SRCM103297*	Food (South Korea)	CP035556	
Complete	*Lactiplantibacillus plantarum SRCM103473*	Food (South Korea)	CP035224	
Complete	*Lactiplantibacillus plantarum SRCM103472*	Food (South Korea)	CP035223	
Complete	*Lactiplantibacillus plantarum SRCM103426*	Food (South Korea)	CP035174	**Food**
Complete	*Lactiplantibacillus plantarum SRCM103418*	Food (South Korea)	CP035168	**Group**
Complete	*Lactiplantibacillus plantarum SRCM103357*	Food (South Korea)	CP035143	
Complete	*Lactiplantibacillus plantarum SRCM103362*	Food (South Korea)	CP035156	
Complete	*Lactiplantibacillus plantarum SRCM103295*	Food (South Korea)	CP035113	
Complete	*Lactiplantibacillus plantarum AS-9*	Fruits and Vegetables	CP028421	
Complete	*Lactiplantibacillus plantarum AS-6*	Fruits and Vegetables	CP028424	**Raw**
Complete	*Lactiplantibacillus plantarum AS-8*	Fruits and Vegetables	CP028422	**Foods**
Complete	*Lactiplantibacillus plantarum AS-10*	Fruits and Vegetables	CP028420	
Complete	*Lactiplantibacillus plantarum 12_3*	Tibet kefir (China)	CP035012	
Complete	*Lactiplantibacillus plantarum YW11*	Tibet kefir (China)	CP035031	
Complete	*Lactiplantibacillus plantarum 13_3*	Tibet kefir (China)	CP035020	
Complete	*Lactiplantibacillus plantarum Q7*	Yak fer. milk(China)	CP019712	**Kefir**
Contig	*Lactiplantibacillus plantarum YW32*	Tibet kefir (China)	SDKA01000009/15	**Group**
Contig	*Lactiplantibacillus plantarum B-1*	Tibet kefir (China)	SDJY01000003/04	
Contig	*Lactiplantibacillus plantarum XZ3303*	Tibet kefir (China)	SDJU01000003/13	
Contig	*Lactiplantibacillus plantarum C4*	Kefir (Spain)	PVNN01000002/06	
Contig	*Lactiplantibacillus plantarum SKT109*	Tibet kefir (China)	SAZE01000001/05	

Table 2.2: (cont.).

Sequence Mode	Name	Source	Accession Number	
Complete	*Lactiplantibacillus plantarum 10CH*	cheese (UK)	CP023728	
Complete	*Lactiplantibacillus plantarum NCIMB 700965*	New Zealand cheese	CP023490	
Contig	*Lactobacillus plantarum UCMA 3037*	cheese (France)	APHP01000021/05	
Contig	*Lactiplantibacillus plantarum YLBGNL-S7*	traditional cheese (China)	VHJC01000067/05	**Cheese**
Contig	*Lactiplantibacillus plantarum LL441*	traditional cheese (Spain)	LWKN01000058/80	**Group**
Contig	*Lactiplantibacillus plantarum KMB_597*	ovine cheese (Slovakia)	QMJE01000007/02	
Contig	*Lactiplantibacillus plantarum KMB_614*	bryndza cheese (Slovakia)	QMJN01000031/08	
Contig	*Lactiplantibacillus plantarum KMB_619*	bryndza cheese (Slovakia)	QMJS01000088/108	
Contig	*Lactiplantibacillus plantarum KMB_618*	bryndza cheese (Slovakia)	QMJR01000039/45	
Contig	*Lactiplantibacillus plantarum KMB_621*	bryndza cheese (Slovakia)	QMJU01000031/28	
Contig	*Lactiplantibacillus plantarum CECT 9435*	Chicha	CAADHQ010000026/09	
Contig	*Lactiplantibacillus plantarum 9434*	Chicha	CAADEV010000026/09	
Contig	*Lactiplantibacillus plantarum 8965*	Chicha	OMOO01000005/04	**Beverage**
Complete	*Lactiplantibacillus plantarum TMW 1.277*	palm wine (Germany)	CP017363	**Group**
Complete	*Lactiplantibacillus plantarum UNQLp11*	Pinot noir wine (Argentina)	CP031140	
Complete	*Lactiplantibacillus plantarum JBE490*	Nuruk(Korea)	CP020861	
Complete	*Lactiplantibacillus plantarum MF1298*	Fermented sausage (Norway)	CP013149	
Complete	*Lactiplantibacillus plantarum TMW 1.708*	raw sausage (Germany)	CP017374	
Complete	*Lactiplantibacillus plantarum TMW 1.25*	raw sausage (Germany)	CP017354	**Meat**
Contig	*Lactiplantibacillus plantarum CRL 681*	Fermented sausage (Argentina)	QOSF01000014/13	**Group**
Contig	*Lactiplantibacillus plantarum A6*	Vietnamese fer. Sausage	LRUO01000066/03	
Complete	*Lactiplantibacillus plantarum LPL-1*	fermented fish (China)	CP021997	

Table 2.2: (cont.).

Sequence Mode	Name	Source	Accession Number	Group
Complete	*Lactiplantibacillus plantarum pc-26*	healthy adult fecal sample (China)	CP023301	Feces Group
Complete	*Lactiplantibacillus plantarum Heal19*	Human GI tract (Sweden)	CP055123	
Complete	*Lactiplantibacillus plantarum FBL-3a*	Fecal sample (China)	CP034694	
Complete	*Lactiplantibacillus plantarum 83-18*	Human feces (Russia)	CP046669	
Complete	*Lactiplantibacillus plantarum 123-17*	Human feces (Russia)	CP046656	
Complete	*Lactiplantibacillus plantarum DR7*	Milk (Malaysia)	CP031318	Other Milk-Based Fermented Products
Complete	*Lactiplantibacillus plantarum LZ227*	raw cow milk (China)	CP015857	
Complete	*Lactiplantibacillus plantarum LZ206*	raw cow milk (China)	CP015966	
Complete	*Lactiplantibacillus plantarum TS12*	Stinky Tofu (Malaysia)	CP018324	
Complete	*Lactiplantibacillus plantarum X7021*	a brine of stinky tofu (China)	CP025412	
Contig	*Lactiplantibacillus plantarum T9*	milk tofu (China)	RPOJ01000071/05	
Complete	*Lactiplantibacillus plantarum 2025*	milk products (Russia)	CP076824	
Complete	*Lactiplantibacillus plantarum TK-P2A*	probitic product (China)	CP045593	
Complete	*Lactiplantibacillus plantarum 8P-A3*	probiotic preparation (Russia)	CP046726	
Contig	*Lactiplantibacillus plantarum CCFM605*	yogurt-like dairy product (China)	QPQR01000004/03	
Contig	*Lactiplantibacillus plantarum CGMCC12436*	yogurt-like dairy product (China)	QPQS01000007/02	
Complete	*Lactiplantibacillus plantarum J26*	T. fermented dairy products (China)	CP033616	

Table 2.3: Thirty-two bacterial species *gad* gene information.

Organisms	Protein Accession	Nucleotide Accession	Source
Levilactobacillus brevis	ANN50035/ ANN49747	CP015398	Korean kimchi
Levilactobacillus zymae	AHF72525	KF690143	kimchi
Levilactobacillus senmaizukei	KRN01061	AYZH01000037	pickles
Levilactobacillus tangyuanensis	WP_125641322	NZ_RHOA01000007	pickle
Levilactobacillus angrenensis	WP_125574762	NZ_RHOB01000001	yogurt
Levilactobacillus cerevisiae	WP_125583210	NZ_RHNN01000018	Spoiled beer
Levilactobacillus spicheri	KJW12820	JZCR01000015	food
Lactiplantibacillus plantarum	AWY49471	CP022294	kimchi
Lactiplantibacillus argentoratensis	AYJ36890	CP032751	Fermented cassava roots
Lactiplantibacillus paraplantarum	AYJ39960	CP032744	beer
Lactiplantibacillus herbarum	WP_048001054	NZ_LFEE01000043	fermented radish
Limosilactobacillus reuteri	KEQ19823	JOKX02000003	Type II sourdough
Limosilactobacillus fermentum	BAO00911	AB856984	pickle juice
Limosilactobacillus oris	KRM16609	AZGE01000002	saliva
Limosilactobacillus antri	EEW54005	ACLL01000020	-
Limosilactobacillus gastricus	EHS85882	NZ_AICN01000052	human milk
Limosilactobacillus coleohominis	KRM84158	AZEW01000005	vagina
Latilactobacillus sakei	AJR27923	KM982734	fermented sea-food
Latilactobacillus curvatus	ASN62676	CP022475	salami
Lentilactobacillus parakefiri	GAW73186	BDGB01000146	kefir grain
Lentilactobacillus buchneri	AEB72391	CP002652	-
Companilactobacillus nuruki	PMD69826	NIPR01000025	beverage starter
Companilactobacillus futsaii	BBA26472	AB986192	Fermented shrimp
Lacticaseibacillus paracasei	BAG12190	AB295641	fermented fish
Loigolactobacillus rennini	WP_057874537	NZ_AYYI01000075	rennin
Paucilactobacillus suebicus	KRM13113	AZGF01000003	apple mash
Furfurilactobacillus rossiae	KRL56399	AZFF01000004	wheat sourdough
Secundilactobacillus paracollinoides	ANZ65502	CP014915	brewery
Lactococcus lactis	QEA60644	CP042408	kimchi
Enterococcus faecium	KST45897	LKPH01000077	cheese
Streptococcus thermophilus	QTA50001	CP061019	Cheese starter
Escherichia coli	CAA0191038/ CAA0194968	CACSHO010000110/ CACSHO010000111	Urine

Table 2.4: Thirty-two bacterial species *gadC* or *yjeM* gene information.

Organisms	Protein Accession	Nucleotide Accession	Source	Gene
Levilactobacillus brevis	ANN48031	CP015398	Korean kimchi	*gadC*
Levilactobacillus zymae	AHF72526	KF690144	kimchi	*gadC*
Levilactobacillus senmaizukei	KRN01060	AYZH01000037	pickles	*gadC*
Levilactobacillus tangyuanensis	WP_125641325	NZ_RHOA01000007	pickle	*gadC*
Levilactobacillus angrenensis	WP_125574761	NZ_RHOB01000001	yogurt	*gadC*
Levilactobacillus cerevisiae	WP_125583211	NZ_RHNN01000018	Spoiled beer	*gadC*
Levilactobacillus spicheri	KJW12731	JZCR01000015	food	*gadC*
Lactiplantibacillus plantarum	AWY49189	CP022294	kimchi	*yjeM*
Lactiplantibacillus argentoratensis	AYJ36618	CP032751	fermented cassava roots	*yjeM*
Lactiplantibacillus paraplantarum	AYJ39671	CP032744	beer contaminant	*yjeM*
Lactiplantibacillus herbarum	-	-	-	-
Limosilactobacillus reuteri	KEQ19824	JOKX02000003	Type II sourdough	*gadC*
Limosilactobacillus fermentum	WP_104878675	NZ_LT906621	sourdough	*yjeM*
Limosilactobacillus oris	KRM16610	AZGE01000002	saliva	*gadC*
Limosilactobacillus antri	EEW54006	ACLL01000020	-	*gadC*
Limosilactobacillus gastricus	WP_007122298	NZ_AICN01000047	human milk	*yjeM*
Limosilactobacillus coleohominis	KRM81387	AZEW01000121	vagina	*gadC*
Latilactobacillus sakei	AKE47364	KP310071	fermented sea-food	*gadC*
Latilactobacillus curvatus	ASN62677	CP022475	salami	*gadC*
Lentilactobacillus parakefiri	GAW73185	BDGB01000146	kefir grain	*gadC*
Lentilactobacillus buchneri	AEB72390	CP002652	-	*gadC*
Companilactobacillus nuruki	PMD67803	NIPR01000064	beverage starter	*yjeM*
Companilactobacillus futsaii	QCX25054	CP040736	-	*yjeM*
Lacticaseibacillus paracasei	-	-	-	-
Loigolactobacillus rennini	WP_057874538	NZ_AYYI01000075	rennin	*gadC*
Paucilactobacillus suebicus	-	-	-	-
Furfurilactobacillus rossiae	KRL56398	AZFF01000004	wheat sourdough	*gadC*
Secundilactobacillus paracollinoides	ANZ62949	CP014915	brewery environment	*yjeM*
Lactococcus lactis	QEA60645	CP042408	kimchi	*gadC*
Enterococcus faecium	KST45898	LKPH01000077	cheese	*gadC*
Streptococcus thermophilus	QTA50002	CP061019	cheese starter	*gadC*
Escherichia coli	CAA0194932	CACSHO010000111	urine	*gadC*

Table 2.5: Thirty-two bacterial species 16S rRNA gene information.

16S rRNA Accession	Organisms	bp	Source
KX458105.1	*Levilactobacillus brevis*	1508 bp	kimchi
KJ607887.1	*Levilactobacillus zymae*	1543 bp	sourdough
MT898562.1	*Levilactobacillus senmaizukei*	1593 bp	Cucumber kimchi
MK110861.1	*Levilactobacillus tangyuanensis*	1432 bp	Pickle
MK110858.1	*Levilactobacillus angrenensis*	1445 bp	yogurt
MT211345.1	*Levilactobacillus cerevisiae*	1489 bp	Wufeng pickle water
KT757220.1	*Levilactobacillus spicheri*	1476 bp	fermented bamboo shoot
MT898568.1	*Lactiplantibacillus plantarum*	1522 bp	Cabbage kimchi
MZ959460.1	*Lactiplantibacillus argentoratensis*	1490 bp	From koji
MZ365311.1	*Lactiplantibacillus paraplantarum*	1520 bp	kimchi
MH548359.1	*Lactiplantibacillus herbarum*	1505 bp	milk
L23507.1	*Limosilactobacillus reuteri*	1535 bp	-
AB856983.1	*Limosilactobacillus fermentum*	1491 bp	Chinese pickle juice
X94229.1	*Limosilactobacillus oris*	1512 bp	sourdough
AY253659.1	*Limosilactobacillus antri*	1520 bp	-
AY253658.1	*Limosilactobacillus gastricus*	1550 bp	-
AM113776.1	*Limosilactobacillus coleohominis*	1564 bp	-
KX886806.1	*Latilactobacillus sakei*	1496 bp	kimchi
MT898647.1	*Latilactobacillus curvatus*	1590 bp	Cabbage kimchi
LC096211.1	*Lentilactobacillus parakefiri*	1514 bp	-
OK135530.1	*Lentilactobacillus buchneri*	1500 bp	Zha-chili
MT786425.1	*Companilactobacillus nuruki*	1419 bp	bioproduct
AB839950.1	*Companilactobacillus futsaii*	1528 bp	Fermented shrimp
OK559730.1	*Lacticaseibacillus paracasei*	1498 bp	Dairy product
LC258150.1	*Loigolactobacillus rennini*	1514 bp	-
KU761840.1	*Paucilactobacillus suebicus*	1573 bp	Sichuan pickle
MZ749583.1	*Furfurilactobacillus rossiae*	1496 bp	From acidic gruel
NR_112846.1	*Secundilactobacillus paracollinoides*	1525 bp	brewery environments
MN749817.1	*Lactococcus lactis*	1546 bp	kimchi
KM495946.1	*Enterococcus faecium*	1495 bp	Siahmazgi cheese
HM218518.1	*Streptococcus thermophilus*	1472 bp	fermented dairy products
MW059027.1	*Escherichia coli*	1464 bp	-

CHAPTER 3

RESULTS AND DISCUSSION

3.1 DNA Polymorphism in *L. brevis* and *L. plantarum* Populations

The polymorphism analyses were performed on *L. brevis* and *L. plantarum* populations in the thesis. Polymorphism analysis results of the *L. brevis* population are shown in Table 3.1. The *L. brevis* population contains the data of the GAD pathway from 30 different strains, and these strains were divided into five other groups according to the ones from which they were isolated. One of the five identified groups represents the non-food group. According to polymorphism analysis, 214 polymorphic sites in the operon and 76 in the *gad2* gene that was found outside the operon were detected for the whole *L. brevis* population (N=30). The operon contains three different genes as a cluster of genes, and the non-coding regions among these genes were included in the analysis. Therefore, each gene in the operon was analyzed one by one, and the number of specific segregating sites(S) was found. The most S was found in the *gadC* gene among the four analyzed genes. When the polymorphic sites were partitioned as synonymous and polymorphic replacement sites, which led to amino acid changes, it was found that again *gadC* gene had the highest number of synonymous(non-amino acid changes) polymorphisms (Figure 3.1). However, the number of replacement(non-synonymous) segregating sites in Table 3.1 shows that the *gad2* gene outside the operon has the highest number of non-synonymous segregating sites.

The five groups defined for the *L. brevis* population were evaluated separately, and scenarios different from the observed situation for the whole population were encountered. According to Table 3.1, the *gad2* gene has the highest number of S and replacement polymorphisms(RP) in the fermented vegetable (N=6) group. The same situation is observed in the sourdough (N=4) and fermented dairy product (N=5) groups. In the fermented beverage (N=10) group, the highest number of S is found in the *gadC*,

and the highest number of RP is found in the *gad2* gene. A different scenario is observed in the feces (N=5) group. In the feces group, the ratios of S number to other genes in the *gadC* gene are quite high, and the RP is most high in the *gadC* gene. In addition, no RP is found in the *gadR* gene in the fermented vegetable and fermented beverage groups and the *gadC* gene in the fermented dairy product group (Figure 3.3-A).
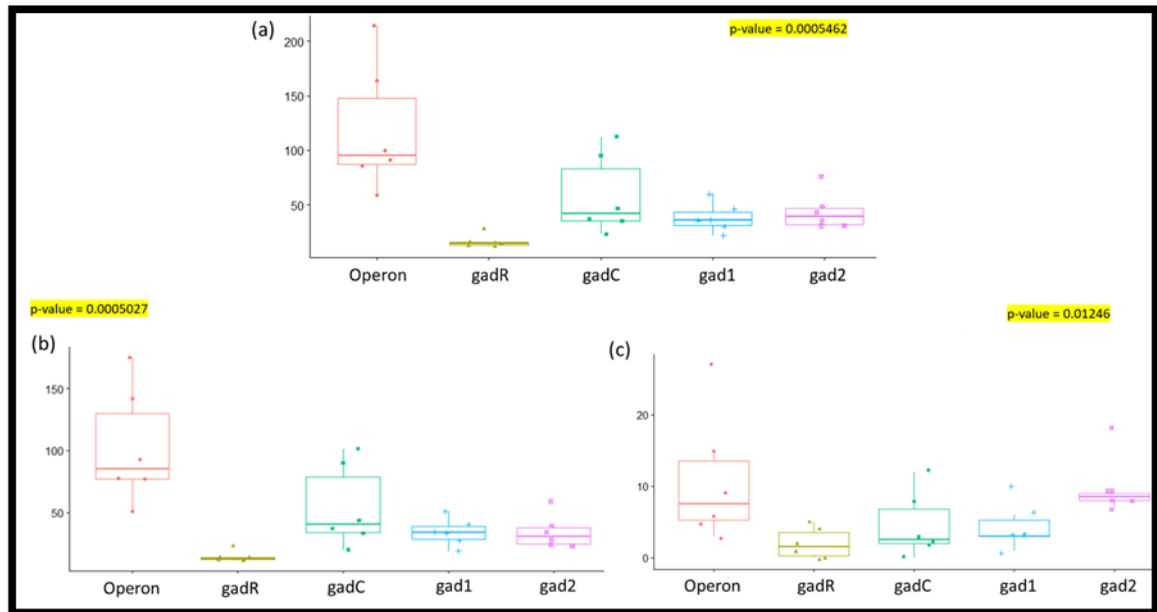
Figure 3.1: Box-plots of *L. brevis* GAD system genes to compare the number of segregating sites. (a) the number of overall segregating sites, (b) the number of synonymous polymorphisms, and (c) the number of non-synonymous(replacement) polymorphisms. The p-values represent the non-parametric ANOVA tests. (Graphs were created with ggplot2 in R)

Since the total base numbers of the four genes and groups' sample sizes (N) examined differed, nucleotide diversity (Pi,π) of the genes was calculated to compare the genes. As shown in Table 3.1, the gene with the highest π value in the *L. brevis* population (N=30) is *gadC* (129.6×10^{-4}), and this situation varies within isolation groups. The *gadR* (92.5×10^{-4}) gene in the fermented vegetable group, the *gad1* (138.9×10^{-4}) gene in the sourdough group, the *gadC* (293.5×10^{-4}) gene in the feces group, also the *gad2* (147×10^{-4} and 99.1×10^{-4}) gene in fermented dairy product and fermented beverage groups have the

highest π value. Nevertheless, when the π values of non-synonymous sites are examined, it is seen that the *gad2* gene π non-synonymous value is the highest in all groups except the sourdough group. In the sourdough group, the highest π non-synonymous value is in the *gadR* (44.1x10-4) gene (Figure 3.3-B). The high rate of non-synonymous nucleotide diversity in the *gad2* gene in all *L.brevis* isolation groups indicates many amino acid changes in this gene.

In the following Figure 3.2, the feces population has the highest nucleotide diversity, followed by sourdough and fermented dairy products considering all sites(a). The partitioning graphs for the nucleotide diversity as synonymous(b) and non-synonymous(b) show a similar trend. The feces population has the highest number of synonymous and non-synonymous diversity, followed by sourdough and fermented dairy products.

Polymorphism analyses were carried out in the *L. plantarum* population using *gadB* and the *yjeM* gene, which is assumed to be involved with the GAD system. The *L. plantarum* population used for the analysis contains data from 88 different strains. Since the amount of data for the *L. plantarum* population is high, nine isolation area-specific groups were formed. Like *L. brevis*, one of the *L. plantarum* groups is the feces group. Table 3.2 shows the results of *L. plantarum* polymorphism analysis both for the whole population and for each group separately. According to the table, in the examination of the total *L. plantarum* population (N=88), there are 121 segregating sites (S) in the *gadB* gene and 65 S in the *yjeM* gene. While the number of RP in the *gadB* gene is 38 for the whole *L.plantarum* population, this number is 14 in the *yjeM* gene. Considering the S, the RP, and the total base numbers of genes, it is seen that π all sites and π non-synonymous sites values are highest for the *gadB* gene (Table 3.2). For the *gadB* gene, the kimchi (N=28) group has the greatest S of all of the isolation groups. Kimchi is followed by cheese (N=10), kefir (N=9), and other milk-based (N=12) groups, all of which are dairy sources. The food (N=8) group comes in fifth, followed by the meat (N=6) and beverage (N=6) groups. The feces (N=5) group, which is a non-food source, ranks eighth in terms of S number, and the raw-food (N=4) group has the lowest S number (Figure 3.4). However, the same ordering is not correct for π values because of the groups' sample sizes (N) (Figure 3.5).
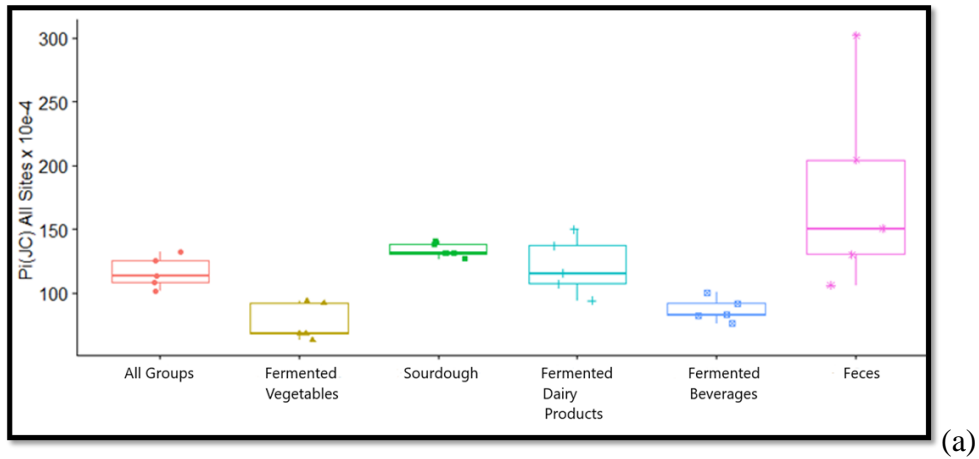
(a)
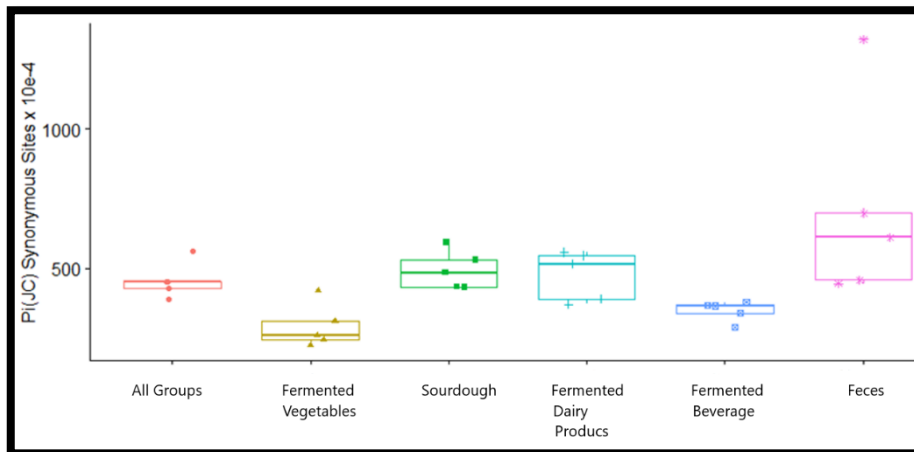
(b)
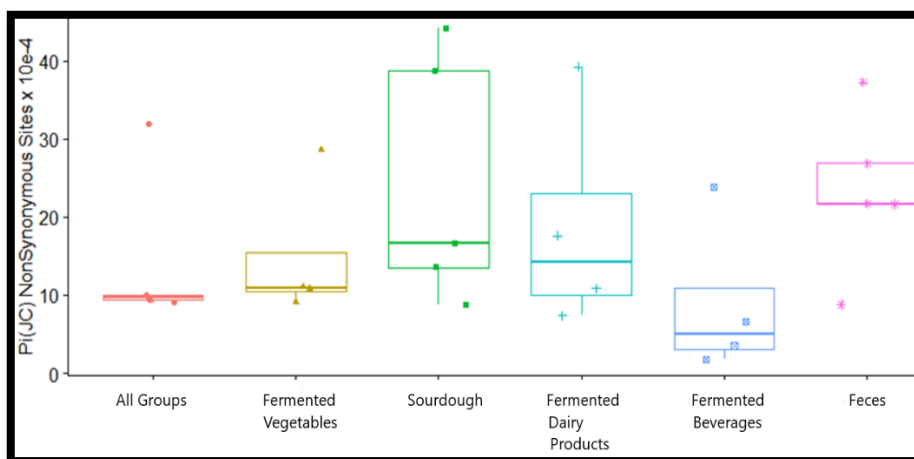
(c)

Figure 3.2: Box-plots of Jukes-Cantor correction applied nucleotide diversity estimates of *L. brevis* isolation groups. The P-values of the non-parametric ANOVA test for graphs a,b, and c are 0.0006984,0.001248 and 0.3065, respectively. (Graphs were created with ggplot2 in R)
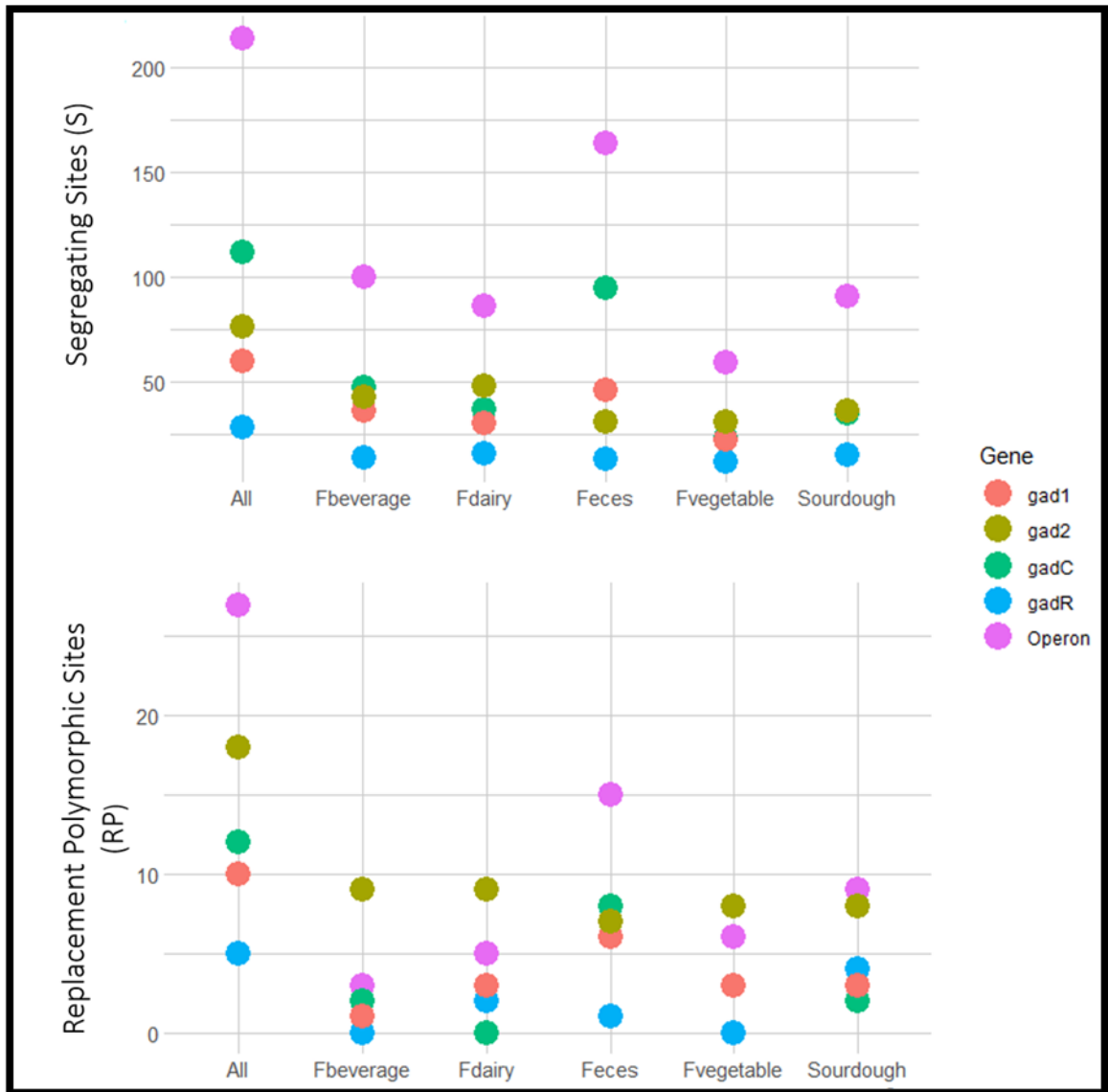
(A)

Figure 3.3: Summary graphics of *L. brevis* DNA polymorphism analyses. Two-way plots: (A) represents the number of S and RP by genes in each isolation group, and (B) shows the general nucleotide diversity(π) and non-synonymous nucleotide diversity graphs calculated for the genes in each group. In part (B), the graph values are values multiplied by 10^{-4}. (Graphs were created with ggplot2 in R)
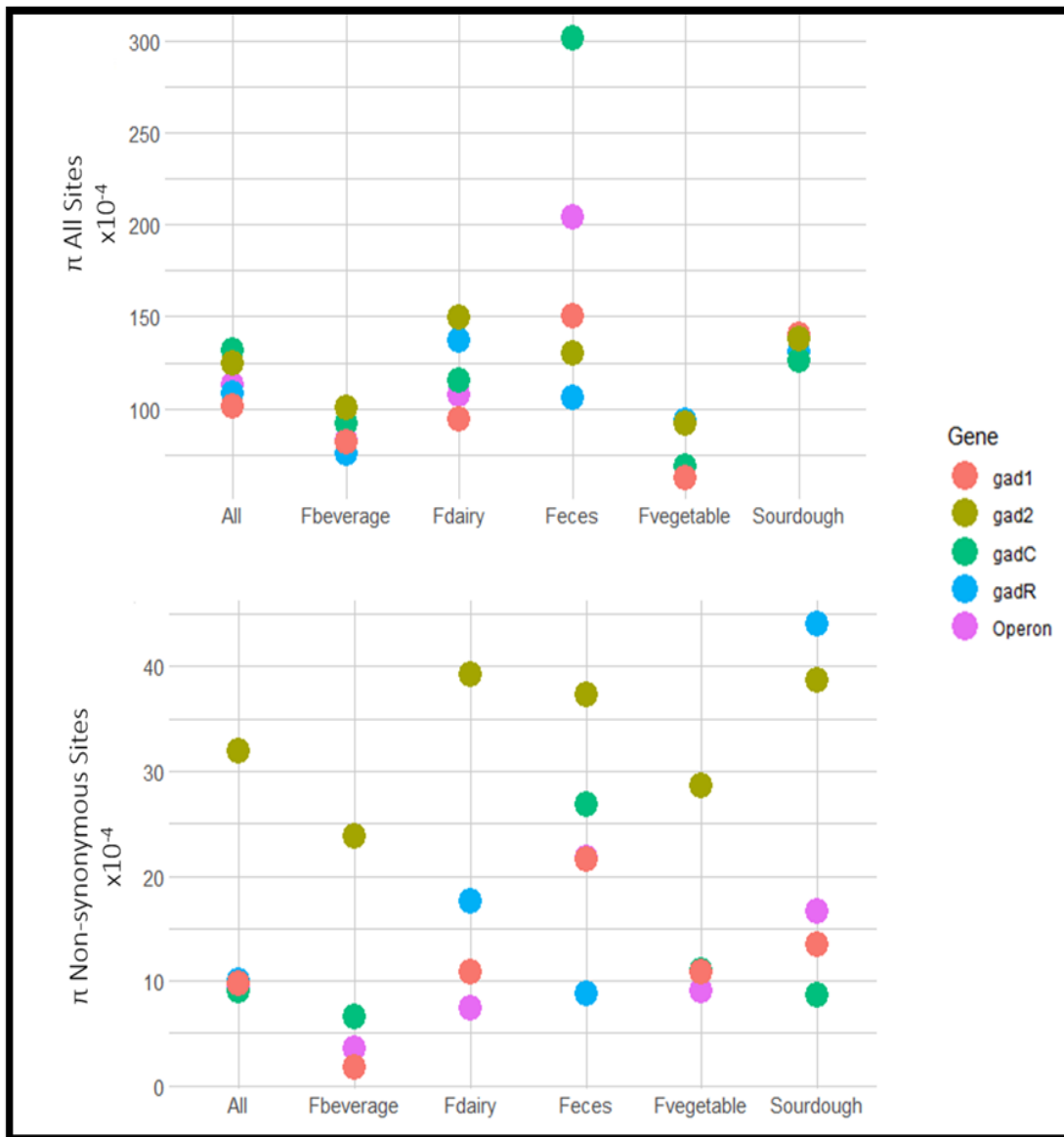
(B)

Figure 3.3: (cont.)

The group with the highest *L. plantarum gadB* gene nucleotide diversity is the cheese group, with a value of 136×10^{-4}. It is also seen that π values of the *gadB* gene are high in other dairy isolation groups (Kefir (111.9×10^{-4}) and milk-based (80.3×10^{-4})). In addition, the cheese group is the isolation source with the highest RP number and π non-synonymous value among other groups.

The order of the groups according to the S number for the *yjeM* gene starts with kimchi, cheese, and other milk-based products, followed by meat, feces, food, beverage, kefir, and raw food groups (Figure 3.4).
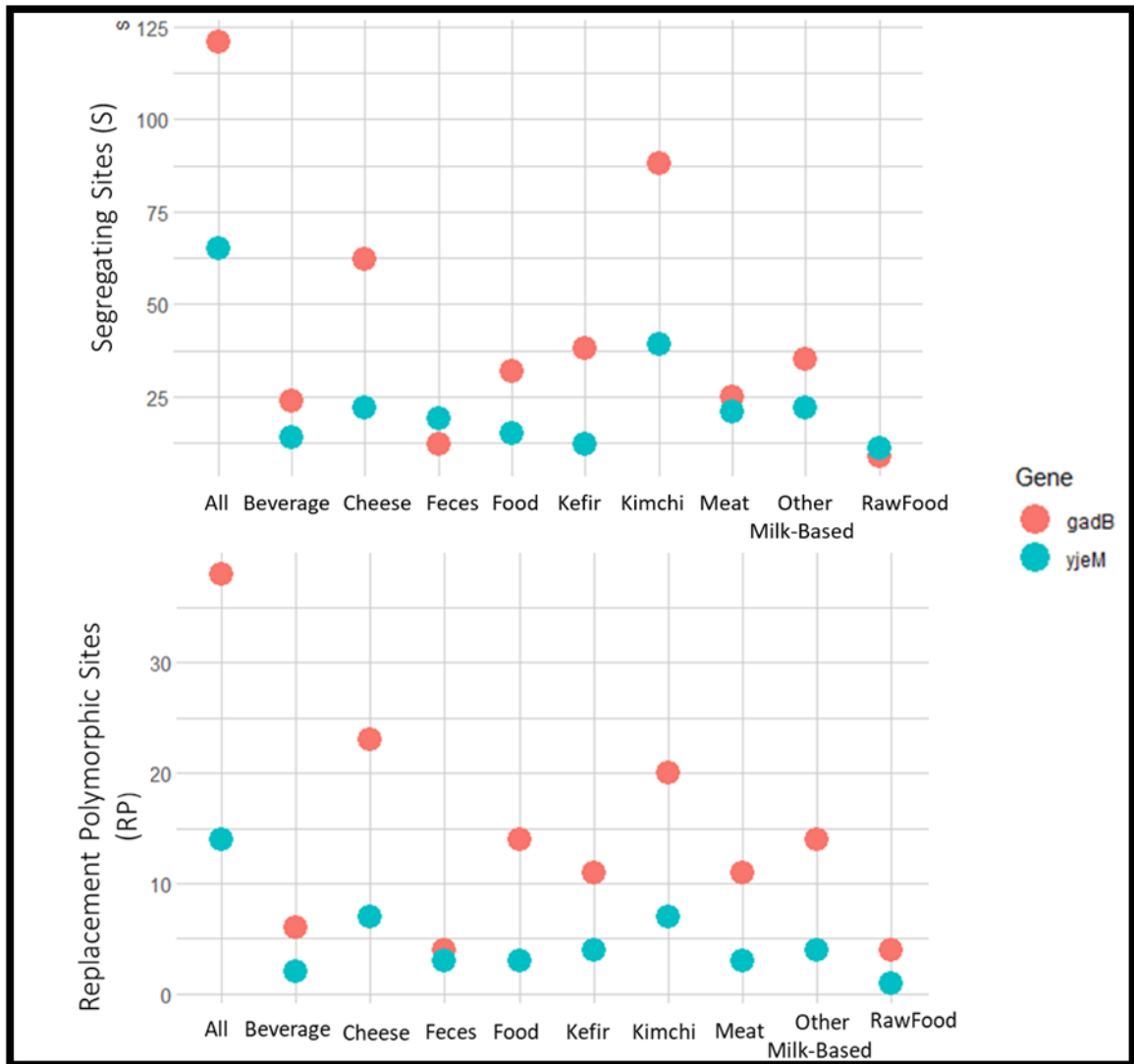
Figure 3.4: Summary graphics of *L. plantarum* DNA polymorphism analyses. Two-way graphs show the S and RP number of the genes *gadB* and *yjeM* for different isolation groups. Nine different groups are represented in the figures, and "All" mean the entire *L. plantarum* population (N=88). (Graphs were created with ggplot2 in R)

When the groups are examined according to the *yjeM* gene π value, it is seen that the group with the highest π value of the *yjeM* ($61.1\text{x}10^{-4}$) gene is the meat group (Figure 3.5). According to table 3.2 table, the number of polymorphisms that cause replacement change in the *yjeM* gene is indicated equally in the kimchi and cheese groups. However, although the RP numbers are the same, the non-synonymous nucleotide diversity value of these two groups in the *yjeM* gene is not the same due to the difference in the groups' sample size (N). Therefore, the cheese group with a smaller sample size has a higher π non-synonymous value (16.5x10-4) (Figure 3.5).
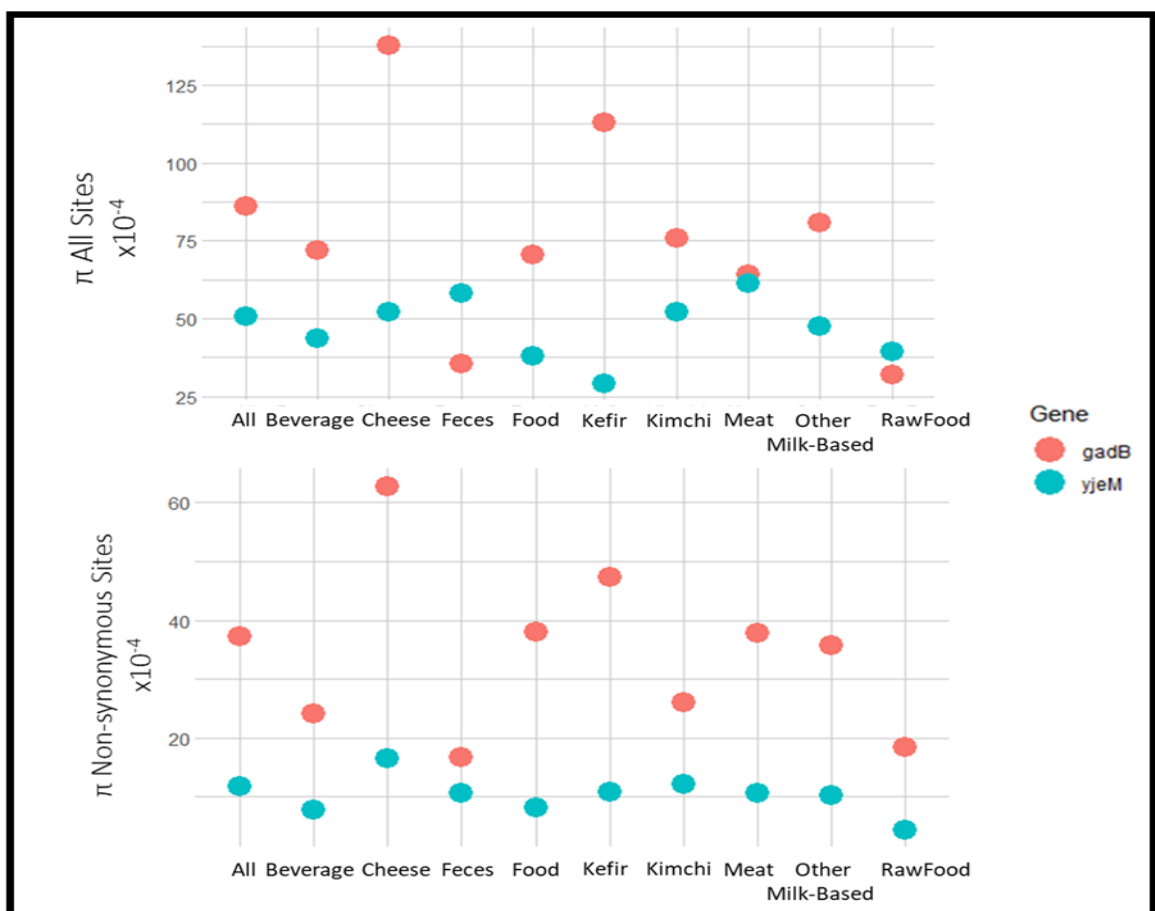
Figure 3.5: Summary graphics of *L. plantarum gadB* and *yjeM* genes nucleotide diversity. The calculated π values for the entire *L. plantarum* population and nine isolation groups are shown in the figure. (Graphs were created with ggplot2 in R)

Table 3.1: Population genetic summary statistics for nucleotide diversity of GAD pathway genes among *L. brevis* population.

Population L.brevis		N	Total bases	Syn. sites	Nonsyn. sites	S(Eta)	Sing.	Par.	Syn. Pol.	Rep. Pol.	π (Pi) All Sites	Theta-W All Sites (θ)	π (JC) All Sites	π (JC) Syn. sites	π (JC) Nonsyn. sites
All Samples	Operon	30	3771 (3516)	1058.32 (807.32)	2708.68	214(217)	104	110	175	27	112	143.2	113.3	387.2 (485.1)	9.4
	gadR		591	135.36	455.64	28(28)	9	19	23	5	107.4	119.6	108.4	452.9	10
	gadC		1492	346.71	1141.29	112(114)	65	47	101	12	129.6	189.5	131.8	561.2	9
	gad1		1440	325.26	1111.75	60(61)	20	40	51	10	100.4	105.2	101.3	427.4	9.7
	gad2		1401	321,24	1076,76	76(77)	32	44	59	18	123.5	136.9	125	451.7	31.9
Fermented Vegetable	Operon	6	3785 (3534)	1061.89 (810.89)	2723.11	59(59)	32	27	51	6	67.8	68.3	68.2	222.6 (279.4)	9.1
	gadR		591	135,06	455,94	12(12)	5	7	12	0	92.5	88.9	93.2	419	-
	gadC		1506	350.53	1155,47	23(23)	13	10	20	3	67.7	66.9	68.1	260.4	11
	gad1		1440	325.31	1111.69	22(22)	14	8	19	3	62	66.9	62.5	245	10.8
	gad2		1401	320,81	1077,19	31(31)	19	12	23	8	90.9	96.9	91.8	311.8	28.6
Sourdough	Operon	4	3785 (3534)	1062.13 (811.13)	2722.88	91(91)	70	21	77	9	129.5	131.1	130.7	431.9 (537.9)	16.6
	gadR		591	135.38	455.63	15(15)	14	1	11	4	129.7	138.4	131	432.6	44.1
	gadC		1506	350.58	1155.42	35(35)	27	8	33	2	125.1	126.8	126.3	530.4	8.7
	gad1		1440	325.17	1111.83	36(36)	24	12	33	3	138.9	136.4	140.3	593	13.5
	gad2		1407	322,63	1081,38	36(36)	29	7	28	8	136.2	139.6	137.8	484.2	38.7
Fermente Dairy P.	Operon	5	3771 (3516)	1058.03 (807.03)	2708.97	86(86)	58	28	78	5	106.1	109.5	107.1	371.2 (470.8)	7.4
	gadR		591	135.03	455.97	16(16)	8	8	14	2	135.4	129.9	136.9	557.9	17.6
	gadC		1492	346.67	1141.33	37(37)	26	11	37	0	113.9	119	115.2	515.2	-
	gad1		1440	325.33	1111.67	30(30)	23	7	27	3	93.1	100	93.9	389.9	10.8
	gad2		1401	320,73	1077,27	48(48)	41	7	39	9	147	164.5	149.6	545.7	39.2

(cont. on next page)

Table 3.1: (cont.).

Population *L.brevis*		N	Total bases	Syn. sites	Nonsyn. sites	S(Eta)	Sing.	Par.	Syn. Pol.	Rep. Pol.	π (Pi) All Sites x 10⁴	Theta-W All Sites (θ) x 10⁴	π (JC) All Sites x 10⁴	π (JC) Syn. Sites x 10⁴	π (JC) Nonsyn. sites x 10⁴
Fermented Beverage		10	3771 (3516)	1058.95 (807.95)	2708.05	100(100)	52	48	93	3	82.1	93.7	82.8	289.8 (367.3)	3.5
	Operon														
	gadR		591	135.67	455.33	14(14)	6	8	14	0	75.2	83.7	75.8	339.5	-
	gadC		1492	346.83	1141.17	47(47)	27	20	44	2	90.9	111.4	91.7	379	6.6
	gad1		1440	325.45	1111.55	36(36)	18	18	35	1	81.5	88.4	82.1	367.5	1.8
	gad2		1401	321.43	1076.57	43(43)	23	20	34	9	99.1	108.5	100.1	364.8	23.8
Feces		5	3785 (3534)	1062.03 (811.03)	2722.97	164(167)	116	48	142	15	200.8	208	204.2	698.7 (871.2)	21.7
	Operon														
	gadR		591	135.43	455..7	13(13)	8	5	12	1	104.9	105.6	105.9	445.6	8.8
	gadC		1506	350.80	1155.20	95(98)	68	27	90	8	293.5	302.8	301.9	1321.7	26.9
	gad1		1440	324.80	1112.20	46(46)	31	15	40	6	148.6	153.3	150.2	611.1	21.6
	gad2		1401	321,60	1076,40	31(31)	3	28	24	7	128.5	106.2	130.1	457.8	37.3

Table 3.2: Population genetic summary statistics for nucleotide diversity of GAD pathway genes among *L. plantarum* population.

Population *L.plantarum*	N	Total bases	Syn. sites	Nonsyn. sites	S(Eta)	Sing.	Par.	Syn. Pol.	Rep. Pol.	π (Pi) All Sites x 10^4	Theta-W All Sites (θ) x 10^4	π (JC) All Sites x 10^4	π (JC) Syn. Sites x 10^4	π (JC) Nonsyn. sites x 10^4
Total	88													
gadB		1410	323,35	1077,65	121	39	82	82	38	85,3	170,7	86,2	247,4	37,2
yjeM		1485	351,76	1130,24	65	30	35	51	14	50,4	86,7	50,6	178,2	11,7
Kimchi Group	28													
gadB		1410	323,31	1077,69	88	48	40	67	20	74,8	161,1	76	252,4	26,1
yjeM		1485	351,43	1130,57	39	19	20	32	7	51,8	67,5	52	183,3	12,1
Food Group	8													
gadB		1410	324,42	1082,58	32	22	10	17	14	70,2	87,5	70,7	174,2	38,1
yjeM		1485	351,71	1130,29	15	6	9	12	3	37,8	39	37,9	134,7	8,2
Raw Foods	4													
gadB		1410	324,17	1082,83	9	9	0	4	4	31,9	34,8	32	62,1	18,5
yjeM		1485	352	1130	11	9	2	10	1	39,3	40,4	39,4	153,5	4,4
Kefir Group	9													
gadB		1410	324,02	1082,98	38	14	24	26	11	111,9	99,2	113,1	327,7	47,4
yjeM		1485	352,24	1129,76	12	2	10	8	4	29,2	29,7	29,3	89,4	10,8

(cont. on next page)

60

Table 3.2: (cont.).

Population / L.plantarum	N	Total bases	Syn. sites	Nonsyn. sites	S(Eta)	Sing.	Par.	Syn. Pol.	Rep. Pol.	π (Pi) All Sites x 10⁴	Theta-W All Sites (θ) x 10⁴	π (JC) All Sites x 10⁴	π (JC) Syn. Sites x 10⁴	π (JC) Nonsyn. sites x 10⁴
Cheese Group	10													
gadB		1410	323,30	1077,70	62	38	24	38	23	136	156	137,8	382,8	62,8
yjeM		1485	351,77	1130,23	22	8	14	15	7	51,9	52,4	52,2	168,7	16,5
Beverage Group	6													
gadB		1410	323,64	1077,36	24	14	10	17	6	71,7	74,9	72,1	219	24,2
yjeM		1485	352,08	1129,92	14	5	9	12	2	43,5	41,3	43,7	161,6	7,7
Meat Group	6													
gadB		1410	323,56	1077,44	25	22	3	14	11	63,6	78	64,2	154,7	37,9
yjeM		1485	351,83	1130,17	21	11	10	18	3	61,1	61,9	61,4	228	10,6
Feces Group	5													
gadB		1410	323,87	1077,13	12	11	1	7	4	35,6	41	35,7	87	16,7
yjeM		1485	351,73	1130,27	19	14	5	16	3	57,9	61,4	58,2	214	10,6
Other Milk-Based Fermented Products	12													
gadB		1410	323,31	1077,69	35	8	27	20	14	80,3	82,5	80,9	217,9	35,8
yjeM		1485	351,94	1130,06	22	11	11	18	4	47,2	49,1	47,4	168,7	10,3

3.2 Neutrality Tests Summary Statistics

The neutrality tests were applied in order to see how GAD system genes of *L. brevis* and *L. plantarum* depart from the neutral theory expectations. For the *L. brevis* species, the tests were performed on the whole *L. brevis* population and each isolation group individually for every GAD pathway gene. Table 3.3 shows the results of *L. brevis* TD and Fu-Li's(D* and F*) analyses. The table shows that the results of *L. brevis* GAD pathway genes for all site TD analyses are negative values for each gene. The TD all-site value calculated for the *gadC* gene is more negative than other genes. However, these negative values are not statistically significant. A similar scenario is observed when GAD system genes are calculated for *L. brevis* isolation groups. Negative values that are not statistically significant are also found in isolation groups. The only difference is in the *gad2* gene in the feces population. Although it is not statistically significant, the calculated TD value for all sites of the *gad2* gene in the feces population is positive.

The partitioned TD values for the synonymous and non-synonymous sites are somewhat similar to the TD all site values of the *L. brevis* population. In terms of statistical significance, TD non-synonymous values of all *L. brevis* population are significantly negative values for operon (p < 0.01), *gadC* (p < 0.05), and *gad1* (p < 0.05). The TD non-synonymous values of the *L. brevis* isolation groups are mostly negative values that are not statistically significant. However, the feces group is quite different from the other groups. According to table 3.3, the TD non-synonymous value calculated for the operon (p < 0.001) in the feces group is a statistically significant negative value. In contrast, the TD non-synonymous value calculated for the *gadC* (p < 0.001) gene in the feces group is a statistically significant positive value. Also, the TD non-synonymous value of the *gad2* gene in the feces population is positive, although it is not statistically significant.

Negative TD values in rapidly growing populations in nutrient-rich habitats, such as the *L. brevis* population, indicate that low-frequency alleles are slightly more common than intermediate-frequency alleles. As a result, in the *L. brevis* population, negative TD values found for GAD pathway genes imply the presence of rare alleles in the genes. Whenever the feces group from the total *L. brevis* population is analyzed, the positive TD values in the *gadC* and *gad2* genes imply that the non-food feces group may be in a somewhat different evolutionary process.

Moreover, In table 3.3, Fu-Li's D* and F* analyses for the whole *L. brevis* population is similar to TD estimates, whereas the estimates are negative for all GAD pathway genes. The negative values for Fu-Li's estimations for GAD system genes indicate that most of the mutations observed in these genes are relatively recent in the gene tree of the genes. In contrast, the *gad2* gene Fu-Li's D* and F* values are positive in the feces group. The positive Fu-Li's analysis results suggest that in *L.brevis* species living in the fecal environment, the *gad2* gene has a different evolutionary process than the *gad1* gene in the operon. Although, Fu-Li's values calculated for the *gad2* gene are greater than one but not statistically significant.

Similar to the *L. brevis*, negative TD values were found from the neutrality analysis performed with the *L. plantarum gadB* and *yjeM* genes. As shown in Table 3.4, the TD values obtained using data from all *L. plantarum* strains for the *gadB,* and *yjeM* genes are negative values without any statistical significance. Moreover, when the *L. plantarum* population is divided into isolation groups and the TD values are calculated, it is seen that each TD values of the *gadB* (p < 0.05) gene in the kimchi group are statistically significant negative value. Except for the kefir group, TD values for *gadB* and *yjeM* genes in other *L. plantarum* isolation groups were close to zero or negative values without statistical significance. In the kefir group, the calculated TD values for the *gadB* gene are positive, and the TD non-synonymous value is greater than one. However, the *gadB* TD values calculated for the kefir group are not statistically significant. According to Table 3.4, Fu-Li's D* and F* values for *L. plantarum gadB* and *yjeM* genes are mainly close to zero or negative values. Also, İt is seen that Fu-Li's D* and F* values of the *yjeM* (p < 0.05) gene are statistically significant negative values in the total *L. plantarum* population.

As stated for *L. brevis*, negative TD and Fu-Li's values calculated for *L. plantarum gadB* and *yjeM* genes indicate the presence of rare alleles in these genes and show that the mutations are mostly found in external branches of *gadB* /*yjeM* gene trees.

Table 3.3: Neutrality tests summary statistics for GAD pathway genes among *L. brevis* populations.

Population *L.brevis*	TD	TD - Cod.	TD - Syn.	TD - Nonsyn.	TD - Silent	Fu-Li's D*	Fu-Li's F*
All Samples (N=30)							
Operon	-0.89	-0.84	-0.60	-2.24**	-0.67	-1.74	-1.72
gadR	-0.37	-0.37	0.06	-1.75#	0.06	-0.40	-0.45
gadC	-1.26	-1.27	-1.11	-2.16*	-1.11	-2.40#	-2.39#
gad1	-0.23	-0.23	0.15	-1.81*	0.15	-0.59	-0.56
gad2	-0.42	-0.42	-0.25	-0.85	-0.25	-1.21	-1.12
Fermented Vegetable (N=6)							
Operon	-0.04	-0.09	-0.06	-0.35	0.00	-0.06	-0.06
gadR	0.24	0.24	0.24	n.a.	0.24	0.34	0.35
gadC	0.08	0.08	0.12	-0.19	0.12	-0.13	-0.09
gad1	-0.46	-0.46	-0.44	-0.45	-0.44	-0.36	-0.41
gad2	-0.39	-0.39	-0.25	-0.74	-0.25	-0.29	-0.34
Sourdough (N=4)							
Operon	-0.13	-0.09	0.00	-0.83	-0.05	-0.13	-0.14
gadR	-0.64	-0.64	-0.56	-0.78	-0.56	-0.64	-0.66
gadC	-0.14	-0.14	-0.10	-0.71	-0.10	-0.14	-0.15
gad1	0.19	0.19	0.29	-0.75	0.29	0.19	0.20
gad2	-0.25	-0.25	-0.18	-0.45	-0.18	-0.25	-0.26
Fermented Dairy Product (N=5)							
Operon	-0.24	-0.27	-0.21	-1.12	-0.17	-0.24	-0.26
gadR	0.31	0.31	0.52	-0.97	0.52	0.31	0.33
gadC	-0.32	-0.32	-0.32	n.a.	-0.32	-0.32	-0.35
gad1	-0.52	-0.52	-0.44	-1.05	-0.44	-0.52	-0.56
gad2	-0.80	-0.80	-0.93	-0.20	-0.93	-0.80	-0.86
Fermented Beverage (N=10)							
Operon	-0.62	-0.64	-0.65	-0.36	-0.60	-0.54	-0.63
gadR	-0.47	-0.47	-0.47	n.a.	-0.47	-0.14	-0.25
gadC	-0.90	-0.87	-0.92	0.22	-0.92	-0.75	-0.89
gad1	-0.38	-0.38	-0.33	-1.11	-0.33	-0.44	-0.48
gad2	-0.42	-0.42	-0.28	-0.86	-0.28	-0.59	-0.62
Feces (N=5)							
Operon	-0.40	-0.36	-0.25	-1.32***	-0.30	-0.36	-0.40
gadR	-0.05	-0.05	0.05	-0.82	0.05	-0.05	-0.05
gadC	-0.46	-0.46	-0.37	1.36***	-0.37	-0.39	-0.44
gad1	-0.23	-0.23	-0.08	-1.15	-0.08	-0.23	-0.25
gad2	1.57	1.57	1.60	1.33	1.60	1.57	1.69

Table 3.4: Neutrality tests summary statistics for GAD pathway genes among *L. plantarum* populations.

Population *L.plantarum*	TD	TD - Cod.	TD - Syn.	TD - Nonsyn.	TD - Silent	Fu-Li's D*	Fu-Li's F*
Total (N=88)							
gadB	-1.68#	-1.71#	-1.75#	-1.48	-1.70#	-1.46	-1.86
yjeM	-1.37	-1.37	-1.26	-1.46	-1.26	-2.91*	-2.74*
Kimchi Group (N=28)							
gadB	-2.06*	-2.05*	-2.13*	-1.60#	-2.14*	-1.96#	-2.36#
yjeM	-0.87	-0.87	-0.85	-0.73	-0.85	-1.46	-1.50
Food Group (N=8)							
gadB	-1.05	-1.03	-0.83	-1.19	-0.89	-0.91	-1.05
yjeM	-0.16	-0.16	0.05	-0.81	0.05	0.14	0.08
Raw Foods (N=4)							
gadB	-0.83	-0.82	-0.78	-0.78	-0.80	-0.83	-0.82
yjeM	-0.28	-0.28	-0.22	-0.61	-0.22	-0.28	-0.28
Kefir Group (N=9)							
gadB	0.65	0.66	0.37	1.24	0.37	0.18	0.33
yjeM	-0.09	-0.09	0.26	-0.69	0.26	0.90	0.74
Cheese Group (N=10)							
gadB	-0.63	-0.68	-0.58	-0.81	-0.50	-0.92	-0.96
yjeM	-0.04	-0.04	0.47	-1.04	0.47	0.12	0.09
Beverage Group (N=6)							
gadB	-0.27	-0.34	-0.43	-0.06	-0.33	-0.19	-0.22
yjeM	0.34	0.34	0.40	-0.05	0.40	0.53	0.53
Meat Group (N=6)							
gadB	-1.16	-1.16	-1.27	-0.94	-1.27	-1.15	-1.26
yjeM	-0.09	-0.09	-0.01	-0.45	-0.01	0.01	-0.02
Feces Group(N=5)							
gadB	-0.95	-0.93	-1.16	-0.41	-1.17	-0.95	-1.01
yjeM	-0.42	-0.42	-0.27	-1.05	-0.27	-0.42	-0.45
Other Milk-Based Fermented Products (N=12)							
gadB	-0.12	-0.21	0.18	-0.72	0.30	0.59	0.46
yjeM	-0.16	-0.16	-0.08	-0.42	-0.08	-0.60	-0.55

3.3 Structural Domains of *L. brevis* and *L. plantarum* GAD Enzymes

As explained in Chapter 1, the *L. brevis* gad gene's three-dimensional structure was identified in the literature and stored in the PDB under accession number 5GP4. When the protein's amino acid sequence in PDB(5GP4) was taken and aligned with the amino acid sequences of the *L. brevis gad1* and *gad2* genes, it was found that the gene whose structure was defined was the *gad2* gene. Afterward, the *L .brevis* and *L. plantarum* GAD protein sequences were aligned, and the parts of the domains defined for the GAD enzyme in the literature were shown in the alignment (Figure 3.6).

Table 3.5: *L. brevis gad1 /gad2* and *L. plantarum gadB* genes nucleotide ranges of enzyme domain regions.

Gene	N-terminal Domain	PLP-binding Domain	Small (C-terminal) Domain
gad1 (479 aa)	1-59 aa (1-177 nt)	60-375 aa (180-1125 nt)	375-479 aa (1125-1437 nt)
gad2 (468 aa)	1-57 aa (1-171 nt)	58-364 aa (174-1092 nt)	364-468 aa (1092-1404 nt)
gadB (469 aa)	1-57 aa (1-171 nt)	58-365 aa (174-1095 nt)	365-469 aa (1095-1407 nt)

As a result of the alignment, it was determined that the domain regions defined for each gad gene correspond to which nucleotide ranges in the DNA sequence of the gad genes (Table3.5). Nucleotide diversity and TD analyses were performed for each domain using the data of all *L. brevis*(N=30) and *L. plantarum* (N=88) populations (Table 3.6). According to the calculated π values of the domains, it was found that the most nucleotide diversity was in the PLP-binding domain of the *L. brevis gad1* and *L. plantarum gadB* gene, while the nucleotide diversity was the most in the N-terminal domain of the *L.brevis gad2* gene. Besides, while TD values of gad genes were calculated as close to zero or negative values in domains, only *L. brevis gad2* gene small domain TD values were found positive.
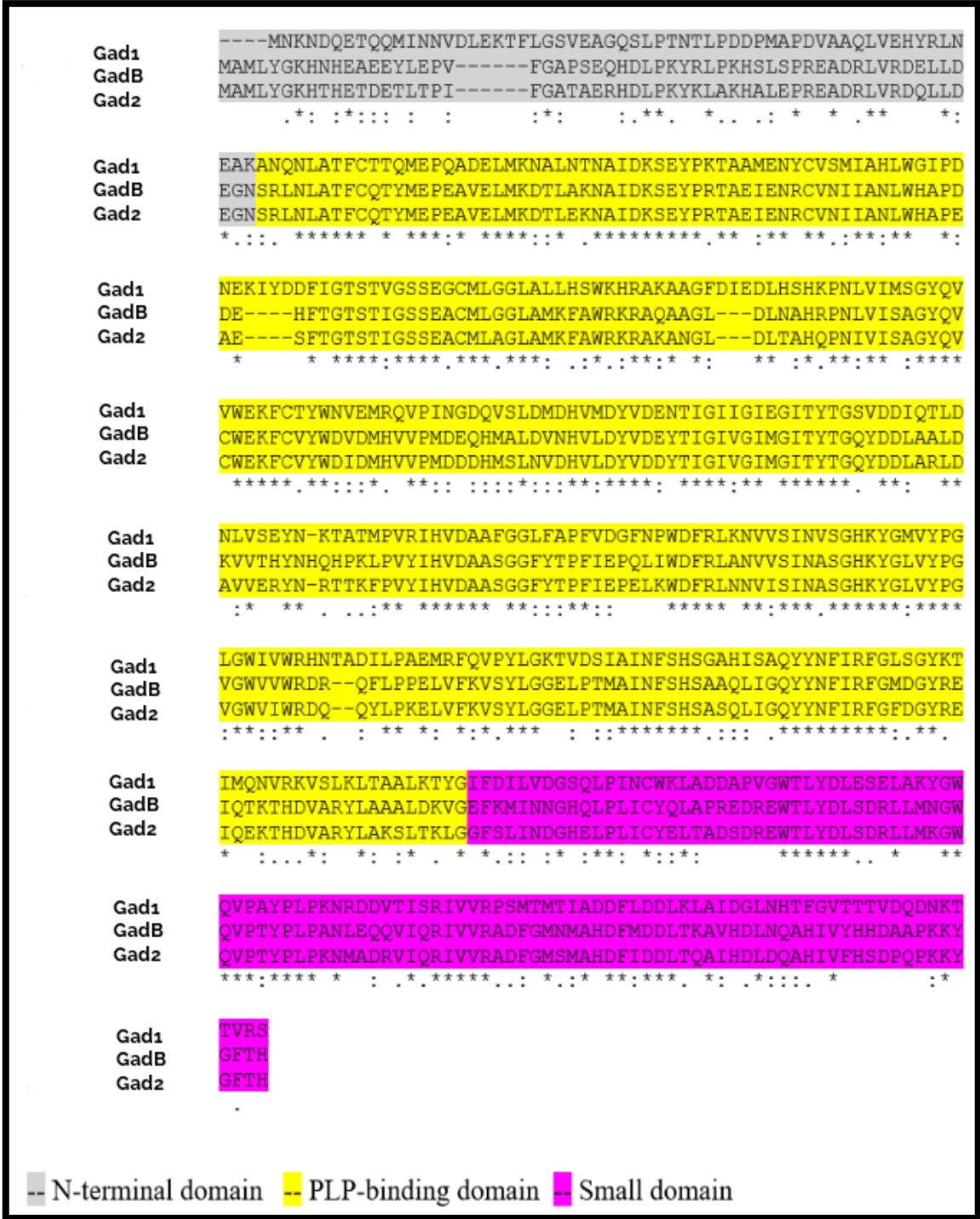
Figure 3.6: The alignment of *L. brevis* (Gad1 and Gad2) and *L. plantarum* (GadB) GAD enzymes protein sequences by MUSCLE algorithm. Coloring on the alignment indicates domain regions of GAD enzymes.

Table 3.6: Nucleotide diversity and TD calculations of *L. brevis* and *L. plantarum gad* genes enzyme domains.

N-terminal Domain	π-Total	π-Nonsyn	TD	TD-Nonsyn
gad1	0.0055	0.0014	-1.57	-1.73
gad2	0.0210	0.0094	-0.30	-0.14
gadB	0.0027	0.0009	-1.61	-1.74
PLP-binding Domain	π-Total	π-Nonsyn	TD	TD-Nonsyn
gad1	0.0113	0.0007	0.24	-1.28
gad2	0.0130	0.0024	-0.62	-1.42
gadB	0.0093	0.0042	-1.66	-1.50
Small (C-terminal) Domain	π-Total	π-Nonsyn	TD	TD-Nonsyn
gad1	0.0087	0.0015	-0.55	-1.20
gad2	0.0065	0.0021	1.01	1.63
gadB	0.0082	0.0037	-1.52	-0.19

Furthermore, the similarity ratios of the *L. brevis* and *L. plantarum gad* genes to each other were established based on their alignment. Interestingly, the amino acid sequences of the *L. brevis gad1* and *gad2* genes were discovered to be 51.2 percent identical, whereas the nucleotide sequences were 57 percent similar. Additionally, the *gad2* gene of *L. brevis*, which is positioned outside the GAD operon, was found to be 83.12% identical to the *L. plantarum gadB* gene in terms of amino acid sequences.

3.4 Analysis of Replacement Sites on The GAD Pathway Genes

The non-synonymous (replacement) mutations identified in the *L. brevis* and *L. plantarum* GAD system genes due to polymorphism analyses were analyzed one by one to determine which nucleotides were changed. Afterward, it was determined which nucleotides that changed in the genes formed which amino acids in the protein sequence. The biochemical characteristics of the changed and previous amino acids in the GAD system genes were compared. When comparing amino acids with each other, their polarities, charge states, pI, bonding potentials, and whether they are singletons or not

were taken into account. Moreover, whether replacement changes are neutral or deleterious was also tested by PROVEAN analysis.

3.4.1 Replacement Sites of *L.brevis gadR* Gene

Table 3.7 shows which nucleotides the replacement changes detected in the *L. brevis gadR* gene occur and which isolation groups these changes are seen. Also, it was determined how the nucleotide changes in the *gadR* gene changed which amino acid in the protein sequence of *gadR*. As indicated in the table, the change on nucleotide 571 of the *L. brevis gadR* gene creates a protein synthesis-stopping codon instead of a different amino acid.

Among the *L. brevis gadR* gene amino acid changes, it was determined that all changes are found only in a single strain(singleton) except for the 93rd amino acid. In addition, PROVEAN analysis determined that the mutation from arginine to cysteine in the 81st amino acid of the *gadR* gene could be deleterious (Table 3.8).

Table 3.7: Replacement(Non-synonymous) sites in the *L. brevis gadR* gene and their corresponding amino acids. The nucleotides changed in the triple codon structure, and the new amino acids formed by the change are shown in red.

L.brevis gadR 591 nt	Isolation Groups	Num. RP	Nucleotide Numbers of Mutations	
	Fermented Vegetable	0		
	Sourdough	4	164-210-241-277	
	Fermented Dairy P.	2	277-571	
	Fermented Beverage	0		
	Feces	1	277	
Amino-acid Changes on *L.brevis gadR*				
N.No	**Nucleotide Changes**	**Codons**	**C. Amino-acids**	**R.No**
164. nt	A change to G	CAG/CGG	Glutamine/Arginine	55. aa
210. nt	C change to A	GAC/GAA	Aspartic acid/Glutamic acid	70. aa
241 .nt	C change to T	CGC/TGC	Arginine/Cysteine	81. aa
277. nt	G change to A	GTC/ATC	Valine/Isoleucine	93. aa
571. nt	C change to T	CAA/TAA	Glutamine/Stop	191. aa

Table 3.8: Biochemical features of amino acids mutated in *L. brevis gadR* gene.

R.No	Amino Acid 1-Letter Code		Class	Charge Features			Polarity		Interaction Modes	B	pI	H	Sing.	PROVEAN
				Positive	Negative	Neutral	Polar	Nonpolar						
55	Glutamine	Q	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.7	0.430	YES	Neutral
	Arginine	R	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000		
70	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417	YES	Neutral
	Glutamic acid	E	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.2	0.458		
81	Arginine	R	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000	YES	Deleterious
	Cysteine	C	Sulfuric	-	-	YES	YES	-	Covalent disulfide bonds, van der Waals	0	5.0	0.721		
93	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	NO	Neutral
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1.000		
191	Glutamine	Q	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.7	0.430	YES	
	Stop	*	-	-	-	-	-	-	-	-	-	-		-

3.4.2 Replacement Sites *of L. brevis gadC* Gene

Table 3.9 below shows the distribution of replacement changes on the *L. brevis gadC* gene in isolation groups and the amino acid changes they cause. It was discovered that there are two different alterations in the 1105th nucleotide region of the *L.brevis gadC* gene, which result in two distinct amino acid mutations. The table shows that only the feces group has two different modifications in the 1105th nucleotide of the *gadC* gene. Among the 15 replacement amino acid changes detected in the *L. brevis gadC* gene, only two(171 and 367) were observed in more than one strain. In addition, four of the non-synonymous changes in the *gadC* gene were deleterious mutations. However, only the deleterious change in amino acid 171 of the *gadC* gene was observed in more than one strain. Other deleterious changes detected were singleton changes (Table 3.10).

Table 3.9: Replacement(Non-synonymous) sites in the *L. brevis gadC* gene and their corresponding amino acids. The nucleotides changed in the triple codon structure, and the new amino acids formed by the change are shown in red.

L.brevis gadC 1506 nt	Isolation Groups	Num. RP	Nucleotide Numbers of Mutations	
	Fermented Vegetable	3	99-1028-1099	
	Sourdough	2	80-1313	
	Fermented Dairy P.	0		
	Fermented Beverage	3	278-512-1099	
	Feces	8	20-421-739-797-865-821-1105-1105	
	Amino-acid Changes on *L.brevis gadC*			
N.No	**Nucleotide Changes**	**Codons**	**C. Amino-acids**	**R.No**
20. nt	A change to G	GAA/GGA	Glutamic acid/Glycine	7. aa
80. nt	C change to T	ACG/ATG	Threonine/Methionine	27. aa
278. nt	G change to T	GGT/GTT	Glycine/Valine	93. aa
421. nt	A change to G	ATC/GTT	Isoleucine/Valine	141. aa
512. nt	T change to C	ATC/ACC	Isoleucine/Threonine	171. aa
739. nt	G change to A	GTT/ATT	Valine/Isoleucine	247. aa
797. nt	A change to G	AAT/AGT	Asparagine/Serine	266. aa
865. nt	G change to A	GGC/AGC	Glycine/Serine	289. aa
871. nt	G change to A	GTC/ATT	Valine/Isoleucine	291. aa
995. nt	G change to A	CGC/CAC	Arginine/Histidine	332. aa
1028. nt	C change to T	ACC/ATC	Threonine/Isoleucine	343. aa
1099. nt	A change to G	AAC/GAC	Asparagine/Aspartic acid	367. aa
1105. nt	G change to T	GCG/TCG	Alanine/Serine	369. aa
1105. nt	G change to A	GCG/ACG	Alanine/Threonine	369. aa
1313. nt	T change to C	TTC/TCC	Phenylalanine/Serine	438. aa

71

Table 3.10: Biochemical features of amino acids mutated in *L. brevis gadC* gene.

R.No	Amino Acid	1-Letter Code	Class	Charge Features			Polarity		Interaction Modes	B	pI	H	Sing.	PROVEAN
				Positive	Negative	Neutral	Polar	Nonpolar						
7	Glutamic Acid	E	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.2	0.458		
	Glycine	G	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.770	YES	Neutral
27	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634	YES	Deleterious
	Methionine	M	Sulfuric	-	-	YES	-	YES	van der Waals	0	5.7	0.811		
93	Glycine	G	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.770	YES	Deleterious
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923		
141	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000	YES	Neutral
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923		
171	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000	NO	Deleterious
	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634		
247	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	YES	Neutral
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000		
266	Asparagine	N	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.4	0.448	YES	Neutral
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601		
289	Glycine	G	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.770	YES	Neutral
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601		
291	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	YES	Neutral
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000		
332	Arginine	R	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000		
	Histidine	H	Basic aromatic	YES	-	-	YES	-	Ionic, H-bonds, aromatic stacking, van der Waals	3	7.6	0.548	YES	Neutral

Table 3.10: (cont.).

	Amino Acid	1-Letter Code	Class	Charge Features			Polarity		Interaction Modes	B	pI	H	Sing.	PROVEAN
				Positive	Negative	Neutral	Polar	Nonpolar						
343	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634	YES	Neutral
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000		
367	Asparagine	N	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.4	0.448	NO	Neutral
	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417		
369	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806	YES	Neutral
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601		
369	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806	YES	Neutral
	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634		
438	Phenylalanine	F	Aromatic	-	-	YES	-	YES	Aromatic stacking, van der Waals	0	5.5	0.951	YES	Deleterious
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601		

3.4.3 Replacement Sites of *L. brevis gad1* Gene

The changes that lead to amino acid changes detected in the *L. brevis gad1* gene and their biochemical properties are shown in the tables below (Table 3.11 and Table 3.12). Mutations in the same nucleotide occurring on the *L. brevis gad1* gene were found in many isolation groups. Notably, the change in the 1413th nucleotide was observed in all isolation groups except the fermented vegetable group (Table 3.11). As a result of the PROVEAN analysis, it was determined that the change in only the 40th amino acid in the *L. brevis gad1* gene could be deleterious. However, the mutation found as deleterious is a singleton, as indicated in the table (Table 3.12).

Table 3.11: Replacement(Non-synonymous) sites in the *L. brevis gad1* gene and their corresponding amino acids. The nucleotides changed in the triple codon structure, and the new amino acids formed by the change are shown in red.

L.brevis gad1 1440 nt	Isolation Groups	Num. RP	Nucleotide Numbers of Mutations	
	Fermented Vegetable	3	26-719-1414	
	Sourdough	3	119-1016-1413	
	Fermented Dairy P.	3	698-1016-1413	
	Fermented Beverage	1	1413	
	Feces	6	112-701-719-1016-1219-1413	
Amino-acid Changes on *L.brevis gad1*				
N.No	**Nucleotide Changes**	**Codons**	**C. Amino-acids**	**R.No**
26. nt	A change to G	CAG/CGG	Glutamine/Arginine	9. aa
112. nt	G change to A	GAT/AAT	Aspartic acid/Asparagine	38. aa
119. nt	C change to T	CCC/CTC	Proline/Leucine	40. aa
701. nt	C change to T	ACG/ATG	Threonine/Methionine	234. aa
719. nt	C change to G	ACC/AGC	Threonine/Serine	240. aa
968. nt	C change to T	ACC/ATC	Threonine/Isoleucine	323. aa
1016. nt	G change to A	AGT/AAT	Serine/Asparagine	339. aa
1219. nt	G change to A	GAG/AAG	Glutamic acid/Lysine	407. aa
1413. nt	A change to C	CAA/CAC	Glutamine/Histidine	471. aa
1414. nt	G change to A	GAT/AAT	Aspartic acid/Asparagine	472. aa

Table 3.12: Biochemical features of amino acids mutated in *L. brevis gad1* gene.

R.No	Amino Acid	1-Letter Code	Class	Charge Features			Polarity		Interaction Modes	B	pI	H	Sing.	PROVEAN
				Positive	Negative	Neutral	Polar	Nonpolar						
9	Glutamine	Q	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.7	0.430		
	Arginine	R	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000	YES	Neutral
38	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417	YES	Neutral
	Asparagine	N	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.4	0.448		
40	Proline	P	Cyclic	-	-	YES	-	YES	van der Waals	0	6.3	0.678	YES	Deleterious
	Leucine	L	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.918		
234	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634	YES	Neutral
	Methionine	M	Sulfuric	-	-	YES	-	YES	van der Waals	0	5.7	0.811		
240	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634	NO	Neutral
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601		
323	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634	YES	Neutral
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000		
339	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601	NO	Neutral
	Asparagine	N	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.4	0.448		
407	Glutamic acid	E	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.2	0.458	YES	Neutral
	Lysine	K	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	3	9.7	0.263		
471	Glutamine	Q	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.7	0.430	NO	Neutral
	Histidine	H	Basic aromatic	YES	-	-	YES	-	Ionic, H-bonds, aromatic stacking, van der Waals	3	7.6	0.548		
472	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417	YES	Neutral
	Asparagine	N	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.4	0.448		

75

3.4.4 Replacement Sites of *L. brevis gad2* Gene

The *gad2* is the gene in which the most replacement changes were detected among the *L. brevis* GAD system genes. As shown in Table 3.13, almost equal numbers of mutations are seen in all isolation groups. Significantly, all isolation groups include changes in the *L. brevis gad2* gene at sites 50,68,458 and 1246.

Eight of the non-synonymous mutations in the *gad2* gene were detected in several strains. Also, non-singleton mutations in the *gad2* gene were classified as neutral modifications by PROVEAN scores. Moreover, amino acid changes of the *gad2* gene that were predicted as deleterious were found as singletons in the *L. brevis* population (Table 3.14).

Table 3.13: Replacement(Non-synonymous) sites in the *L. brevis gad2* gene and their corresponding amino acids. The nucleotides changed in the triple codon structure, and the new amino acids formed by the change are shown in red.

L.brevis gad2 1407 nt	Isolation Groups	Num. RP	Nucleotide Numbers of Mutations	
	Fermented Vegetable	8	50-68-458-468-679-685-984-1246	
	Sourdough	8	50-68-415-458-468-518-671-1246	
	Fermented Dairy P.	9	34-46-50-68-120-458-671-973-1246	
	Fermented Beverage	9	50-68-281-458-468-679-685-1058-1246	
	Feces	7	50-68-458-468-469-671-1246	
Amino-acid Changes on *L.brevis gad2*				
N.No	Nucleotide Changes	Codons	C. Amino-acids	R.No
34. nt	A change to G	ACA/GCA	Threonine/Alanine	12. aa
46. nt	C change to T	CTC/TTC	Leucine/Phenylalanine	16. aa
50. nt	C change to A	ACA/AAA	Threonine/Lysine	17. aa
68. nt	G change to C	AGC/ACC	Serine/Threonine	23. aa
120. nt	G change to T	GAG/GAT	Glutamic acid/Aspartic acid	40. aa
281. nt	G change to A	CGG/CAG	Arginine/Glutamine	94. aa
415. nt	T change to C	TTT/CTT	Phenylalanine/Leucine	139. aa
458. nt	C change to A	ACT/AAT	Threonine/Asparagine	153. aa
468. nt	A change to C	CAA/CAC	Glutamine/Histidine	156. aa
469. nt	C change to T	CCT/TCT	Proline/Serine	157. aa
518. nt	G change to C	TGT/TCT	Cysteine/Serine	173. aa
671. nt	C change to T	GCC/GTC	Alanine/Valine	224. aa
679. nt	G change to A	GAT/AAT	Aspartic acid/Asparagine	227. aa
685. nt	G change to A	GTT/ATT	Valine/Isoleucine	229. aa
973. nt	T change to G	TCC/GCC	Serine/Alanine	325. aa
984. nt	C change to G	ATC/ATG	Isoleucine/Methionine	328. aa
1058. nt	C change to T	GCC/GTC	Alanine/Valine	353. aa
1246. nt	G change to A	GCG/ACG	Alanine/Threonine	416. aa

Table 3.14: Biochemical features of amino acids mutated in *L. brevis gad2* gene.

R.No	Amino Acid 1-Letter Code		Class	Charge Features			Polarity		Interaction Modes	B	pI	H	Sing.	PROVEAN
				Positive	Negative	Neutral	Polar	Nonpolar						
12	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634		
	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806	YES	Neutral
16	Leucine	L	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.918		
	Phenylalanine	F	Aromatic	-	-	YES	-	YES	Aromatic stacking, van der Waals	0	5.5	0.951	YES	Neutral
17	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634		
	Lysine	K	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	3	9.7	0.263	NO	Neutral
23	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634		
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601	NO	Neutral
40	Glutamic acid	E	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.2	0.458		
	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417	YES	Neutral
94	Arginine	R	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000		
	Glutamine	Q	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.7	0.430	YES	Neutral
139	Phenylalanine	F	Aromatic	-	-	YES	-	YES	Aromatic stacking, van der Waals	0	5.5	0.951		
	Leucine	L	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.918	YES	Deleterious
153	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634		
	Asparagine	N	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.4	0.448	NO	Neutral
156	Glutamine	Q	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.7	0.430		
	Histidine	H	Basic aromatic	YES	-	-	YES	-	Ionic, H-bonds, aromatic stacking, van der Waals	3	7.6	0.548	NO	Neutral

Table 3.14: (cont.).

R.No	Amino Acid	1-Letter Code	Class	Charge Features			Polarity		Interaction Modes	B	pI	H	Sing.	PROVEAN
				Positive	Negative	Neutral	Polar	Nonpolar						
157	Proline	P	Cyclic	-	-	YES	-	YES	van der Waals	0	6.3	0.678		
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601	YES	Deleterious
173	Cysteine	C	Sulfuric	-	-	YES	YES	-	Covalent disulfide bonds, van der Waals	0	5.0	0.721		
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601	YES	Deleterious
224	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806		
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	NO	Neutral
227	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417		
	Asparagine	N	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.4	0.448	NO	Neutral
229	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923		
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000	NO	Neutral
325	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601		
	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806	YES	Neutral
328	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000		
	Methionine	M	Sulfuric	-	-	YES	-	YES	van der Waals	0	5.7	0.811	YES	Neutral
353	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806		
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	YES	Deleterious
416	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806		
	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634	NO	Neutral

3.4.5 Replacement Sites of *L. plantarum gadB* Gene

Table 3.15: Replacement(Non-synonymous) sites in the *L. plantarum gadB* gene and their corresponding amino acids. The nucleotides changed in the triple codon structure, and the new amino acids formed by the change are shown in red.

L.plantarum gadB 1410 nt	Isolation Groups	Num. RP	Nucleotide Numbers of Mutations
	Kimchi	20	340-341-418-419-500-540-544-572-580-583-595-658-674-704-712-1022-1108-1153-1337-1353
	Food	14	52-500-540-544-583-595-704-712-901-1019-1022-1023-1153-1353
	Raw Food	4	514-658-1153-1353
	Kefir	11	40-500-544-572-583-595-658-704-712-953-1353
	Cheese	23	53-98-340-341-418-419-425-500-540-544-580-583-595-674-704-712-764-793-953-1022-1108-1153-1353
	Beverage	6	540-683-850-871-951-1353
	Meat	11	500-544-583-595-704-712-850-951-1022-1108-1353
	Feces	4	1105-1108-1153-1353
	Other Milk-based P.	14	500-540-544-572-583-595-658-704-712-793-934-953-1022-1353

Amino-acid Changes on *L. plantarum gadB*				

N.No	Nucleotide Changes	Codons	C. Amino-acids	R.No
40. nt	G change to A	GAA/AAA	Glutamic acid/Lysine	14. aa
52. nt	C change to A	CCA/ACA	Proline/Threonine	18. aa
53. nt	C change to T	CCA/CTA	Proline/Leucine	18. aa
98. nt	G change to T	CGG/CTG	Arginine/Leucine	33. aa
340. nt	G change to A	GAT/AGT	Aspartic acid/Serine	114. aa
341. nt	A change to G			
418. nt	G change to A	GCC/AGT	Alanine/Serine	140. aa
419. nt	C change to G			
425. nt	G change to T	CGT/CTT	Arginine/Leucine	142. aa
500. nt	T change to C	GTT/GCT	Valine/Alanine	167. aa
514. nt	T change to G	TTT/GTT	Phenylalanine/Valine	172. aa
540. nt	G change to A	ATG/ATA	Methionine/Isoleucine	180. aa
544. nt	G change to A	GTG/ATG	Valine/Methionine	182. aa
572. nt	C change to T	GCC/GTC	Alanine/Valine	191. aa
580. nt	G change to A	GTT/ATT	Valine/Isoleucine	194. aa
583. nt	A change to G	AAC/GAC	Asparagine/Aspartic acid	195. aa
595. nt	G change to A	GAC/AAC	Aspartic acid/Asparagine	199. aa
658. nt	T change to G	TAT/GAT	Tyrosine/Aspartic acid	220. aa
674. nt	C change to G	GCA/GGA	Alanine/Glycine	225. aa
683. nt	A change to G	AAG/AGG	Lysine/Arginine	228. aa
704. nt	A change to G	CAT/CGT	Histidine/Arginine	235. aa
712. nt	C change to T	CCC/TCC	Proline/Serine	238. aa
764. nt	C change to T	ACC/ATC	Threonine/Isoleucine	255. aa
793. nt	G change to A	GAC/AAC	Aspartic acid/Asparagine	265. aa

Table 3.15: (cont.).

N.No	Nucleotide Changes	Codons	C. Amino-acids	R.No
850. nt	G change to A	GTT/ATT	Valine/Isoleucine	284. aa
871. nt	G change to A	GTC/ATC	Valine/Isoleucine	291. aa
901. nt	C change to T	CCA/TCA	Proline/Serine	301. aa
934. nt	G change to A	GGG/AGG	Glycine/Arginine	312. aa
951. nt	G change to A	ATG/ATA	Methionine/Isoleucine	317. aa
953. nt	C change to T	GCG/GTG	Alanine/Valine	318. aa
1019. nt	T change to A	ATG/AAG	Methionine/Lysine	340. aa
1022. nt	A change to G	GAC/GGC	Aspartic acid/Glycine	341. aa
1023. nt	C change to A	GAC/GAA	Aspartic acid/Glutamic acid	341. aa
1105. nt	A change to G	ATG/GTG	Methionine/Valine	369. aa
1108. nt	A change to G	ATC/GTC	Isoleucine/Valine	370. aa
1153. nt	C change to T	CCG/TCG	Proline/Serine	385. aa
1337. nt	T change to G	GTC/GGC	Valine/Glycine	446. aa
1353. nt	A change to C	CAA/CAC	Glutamine/Histidine	451. aa

The *L. plantarum gadB* gene has much more replacement mutations than the *L. brevis gad* genes. Table 3.15 shows that *L. plantarum gadB* gene non-synonymous changes are mostly distributed between the 300th and 1200th nucleotides. Furthermore, as seen in the table, amino acid changes caused by the change of two nucleotide changes were also found in the *gadB* gene. In Table 3.16, it is shown that most of the replacement mutations detected in the *L. plantarum gadB* gene are non-singleton changes. Of the 38 non-synonymous changes in the gadB gene, only ten are singleton changes. Besides, when the deleterious or neutral status of the mutations was evaluated, 13 mutations in the *L. plantarum gadB* gene were determined could be deleterious.

In contrast to the situation observed in *L. brevis gad* genes, the tables show that the majority of the deleterious mutations in the *L. plantarum gadB* gene were detected in more than one strain. However, in the *gadB* deleterious changes seen in many *L. plantarum* strains, it is noteworthy that the biochemical properties of the previous amino acid and the new amino acid are mainly similar.

Table 3.16: Biochemical features of amino acids mutated in *L. plantarum gadB* gene.

R.No	Amino Acid	1-Letter Code	Class	Charge Features			Polarity		Interaction Modes	B	pI	H	Sing.	PROVEAN
				Positive	Negative	Neutral	Polar	Nonpolar						
14	Glutamic acid	E	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.2	0.458		
	Lysine	K	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	3	9.7	0.263	YES	Neutral
18	Proline	P	Cyclic	-	-	YES	-	YES	van der Waals	0	6.3	0.678		
	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634	YES	Deleterious
18	Proline	P	Cyclic	-	-	YES	-	YES	van der Waals	0	6.3	0.678		
	Leucine	L	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.918	NO	Deleterious
33	Arginine	R	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000		
	Leucine	L	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.918	YES	Deleterious
114	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417		
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601	NO	Neutral
140	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806		
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601	NO	Neutral
142	Arginine	R	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000		
	Leucine	L	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.918	NO	Deleterious
167	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923		
	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806	NO	Deleterious
172	Phenylalanine	F	Aromatic	-	-	YES	-	YES	Aromatic stacking, van der Waals	0	5.5	0.951		
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	YES	Deleterious
180	Methionine	M	Sulfuric	-	-	YES	-	YES	van der Waals	0	5.7	0.811		
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1.000	NO	Neutral

Table 3.16: (cont.).

R.No	Amino Acid	1-Letter Code	Class	Charge Features			Polarity		Interaction Modes	B	pI	H	Sing.	PROVEAN
				Positive	Negative	Neutral	Polar	Nonpolar						
182	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	NO	Neutral
	Methionine	M	Sulfuric	-	-	YES	-	YES	van der Waals	0	5.7	0.811		
191	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806	NO	
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923		Neutral
194	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	NO	
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1.000		Neutral
195	Asparagine	N	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.4	0.448	NO	Neutral
	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417		
199	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417	NO	Neutral
	Asparagine	N	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.4	0.448		
220	Tyrosine	Y	Aromatic	-	-	YES	YES	-	Ionic, H-bonds, aromatic stacking, van der Waals	8?	5.6	?	NO	Deleterious
	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417		
225	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806	NO	Deleterious
	Glycine	G	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.770		
228	Lysine	K	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	3	9.7	0.263	NO	Neutral
	Arginine	R	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000		
235	Histidine	H	Basic aromatic	YES	-	-	YES	-	Ionic, H-bonds, aromatic stacking, van der Waals	3	7.6	0.548	NO	Neutral
	Arginine	R	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000		
238	Proline	P	Cyclic	-	-	YES	-	YES	van der Waals	0	6.3	0.678	NO	Neutral
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601		

Table 3.16: (cont.).

R.No	Amino Acid	1-Letter Code	Class	Positive	Negative	Neutral	Polar	Nonpolar	Interaction Modes	B	pI	H	Sing.	PROVEAN
255	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634	YES	Neutral
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000		
265	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417	NO	Deleterious
	Asparagine	N	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.4	0.448		
284	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	NO	Neutral
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000		
291	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	YES	Neutral
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000		
301	Proline	P	Cyclic	-	-	YES	-	YES	van der Waals	0	6.3	0.678	YES	Neutral
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601		
312	Glycine	G	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.770	YES	Deleterious
	Arginine	R	Basic	YES	-		YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000		
317	Methionine	M	Sulfuric	-	-	YES	-	YES	van der Waals	0	5.7	0.811	NO	Neutral
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000		
318	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806	NO	Deleterious
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923		
340	Methionine	M	Sulfuric	-	-	YES	-	YES	van der Waals	0	5.7	0.811	YES	Neutral
	Lysine	K	Basic	YES	-		YES	-	Ionic, H-bonds, van der Waals	3	9.7	0.263		
341	Aspartic acid	D	Acid	-	YES	-	YES	-	Ionic, H-bonds, van der Waals	4	3.0	0.417	NO	Deleterious
	Glycine	G	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.770		

Table 3.16: (cont.).

R.No	Amino Acid	1-Letter Code	Class	Charge Features			Polarity		Interaction Modes	B	pI	H	Sing.	PROVEAN
				Positive	Negative	Neutral	Polar	Nonpolar						
369	Methionine	M	Sulfuric	-	-	YES	-	YES	van der Waals	0	5.7	0.811		
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	YES	Neutral
370	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1.000		
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	NO	Neutral
385	Proline	P	Cyclic	-	-	YES	-	YES	van der Waals	0	6.3	0.678	NO	Neutral
	Serine	S	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.7	0.601		
446	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	NO	Deleterious
	Glycine	G	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.770		
451	Glutamine	Q	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.7	0.430		
	Histidine	H	Basic aromatic	YES	-	-	YES	-	Ionic, H-bonds, aromatic stacking, van der Waals	3	7.6	0.548	NO	Neutral

3.4.6 Replacement Sites of *L. plantarum yjeM* Gene

Table 3.17 lists the replacement changes detected in the *L. plantaum yjeM* gene. The absence of any changes in the N-termina l(1-378nt) part of the *yjeM* gene analyzed using 88 *L. plantarum* strain data indicates that this part is well conserved. Moreover, most replacement mutations detected for the *yjeM* gene have been observed in more than one strain. Furthermore, most of the replacement mutations discovered in the *yjeM* gene were found in more than one strain, and their impact on protein function was determined to be neutral by PROVEAN tests (Table 3.18).

Table 3.17: Replacement(Non-synonymous) sites in the *L. plantarum yjeM* gene and their corresponding amino acids. The nucleotides changed in the triple codon structure, and the new amino acids formed by the change are shown in red.

L.plantarum yjeM 1485 nt	Isolation Groups	Num. RP	Nucleotide Numbers of Mutations	
	Kimchi	7	379-400-409-898-1300-1387	
	Food	3	409-898-1300	
	Raw Food	1	1300	
	Kefir	4	400-592-898-1076	
	Cheese	7	898-923-1021-1238-1300-1385-1438	
	Beverage	2	1076-1300	
	Meat	3	400-898-1300	
	Feces	3	409-898-1387	
	Other Milk-based P.	4	409-592-898-1300	
Amino-acid Changes on *L .plantarum gadB*				
N.No	**Nucleotide Changes**	**Codons**	**C. Amino-acids**	**R.No**
379. nt	T change to C	TTT/CTT	Phenylalanine/Leucine	127. aa
400. nt	A change to G	ATG/GTG	Methionine/Valine	134. aa
409. nt	A change to C	ATT/CTT	Isoleucine/Leucine	137. aa
592. nt	A change to G	ATG/GTG	Methionine/Valine	198. aa
896. nt	C change to T	GCG/GTG	Alanine/Valine	299. aa
898. nt	C change to T	CTT/TTT	Leucine/Phenylalanine	300. aa
923. nt	A change to C	AAG/ACG	Lysine/Threonine	308. aa
1021. nt	C change to A	CTT/ATT	Leucine/Isoleucine	341. aa
1076. nt	A change to G	AAA/AGA	Lysine/Arginine	359. aa
1238. nt	C change to T	GCG/GTG	Alanine/Valine	413. aa
1300. nt	A change to G	ACT/GCT	Threonine/Alanine	434. aa
1385. nt	C change to A	ACT/AAT	Threonine/Asparagine	462. aa
1387. nt	G change to A	GGG/AGG	Glycine/Arginine	463. aa
1438. nt	G change to T	GTG/TTG	Valine/Leucine	480. aa

Table 3.18: Biochemical features of amino acids mutated in *L. plantarum yjeM* gene.

R.No	Amino Acid	1-Letter Code	Class	Charge Features			Polarity		Interaction Modes	B	pI	H	Sing.	PROVEAN
				Positive	Negative	Neutral	Polar	Nonpolar						
127	Phenylalanine	F	Aromatic	-	-	YES	-	YES	Aromatic stacking, van der Waals	0	5.5	0.951	YES	Neutral
	Leucine	L	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.918		
134	Methionine	M	Sulfuric	-	-	YES	-	YES	van der Waals	0	5.7	0.811	NO	Neutral
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923		
137	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000	NO	Neutral
	Leucine	L	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.918		
198	Methionine	M	Sulfuric	-	-	YES	-	YES	van der Waals	0	5.7	0.811	NO	Neutral
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923		
299	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806	YES	Deleterious
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923		
300	Leucine	L	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.918	NO	Neutral
	Phenylalanine	F	Aromatic	-	-	YES	-	YES	Aromatic stacking, van der Waals	0	5.5	0.951		
308	Lysine	K	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	3	9.7	0.263	NO	Neutral
	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634		
341	Leucine	L	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.918	YES	Neutral
	Isoleucine	I	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	1,000		
359	Lysine	K	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	3	9.7	0.263	NO	Neutral
	Arginine	R	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000		

(cont. on next page)

86

Table 3.18: (cont.).

R.No	Amino Acid	1-Letter Code	Class	Charge Features			Polarity		Interaction Modes	B	pI	H	Sing.	PROVEAN
				Positive	Negative	Neutral	Polar	Nonpolar						
413	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806		
	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923	YES	Deleterious
434	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634		
	Alanine	A	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.806	NO	Neutral
462	Threonine	T	Hydroxylic	-	-	YES	YES	-	H-bonds, van der Waals	3	5.6	0.634		
	Asparagine	N	Amide	-	-	YES	YES	-	H-bonds, van der Waals	5	5.4	0.448	NO	Neutral
463	Glycine	G	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.770		
	Arginine	R	Basic	YES	-	-	YES	-	Ionic, H-bonds, van der Waals	7	10.8	0.000	NO	Deleterious
480	Valine	V	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.923		
	Leucine	L	Aliphatic	-	-	YES	-	YES	van der Waals	0	6.0	0.918	NO	Neutral

3.5 The Influence of Replacement Changes on the Charge Distribution of GAD Pathway Proteins

This section has studied the effects of replacement mutations detected in *L. brevis* and *L .plantarum* GAD system genes on pathway proteins. First, strains with replacement changes were determined for each GAD system gene. Then, the pI and charge calculations were made at varying pH using protein sequences of strains containing these replacement changes. Each strain belonging to *L. brevis* and *L. plantarum* species contains different replacement polymorphisms in their GAD system genes.Therefore, the amino acid changes in the genes of the strains also differ. While some strains have many amino acid changes, some strains have replacement polymorphisms that cause only a few amino acid changes. If there are strains with the same replacement changes, only one of them has been selected and used in the analysis.

When amino acid sequences of *L. brevis* GAD pathway proteins that do not contain replacement mutations were analyzed, their protein pI values were calculated as 5,891 for GadR, 8,996 for GadC, 4,905 for Gad1, and 5,363 for Gad2. Also, *L. plantarum* protein sequences without replacement were analyzed, and the isoelectric points of GadB and YjeM proteins were found as 5,577 and 9,519, respectively. The total pI value of *L. brevis* Gad2 protein was found to be similar to that of *L. plantarum* GadB protein. However, it differed significantly from the pI of the *L. brevis* Gad1 protein.

 All combinations of replacement alterations found in *L. brevis* and *L. plantarum* strains and their calculated values are reported in the tables created for pI and charge calculations (Table 3.19 and Table 3.20).

Protein sequences from strains with replacement changes were examined, and shown that mutations resulting in amino acid substitution did not noticeably impact the pI and charge values of the *L. brevis* and *L. plantarum* GAD pathway proteins. Moreover, it can be seen in the pI tables that mutations in a strain balanced each other and were not significantly affect the overall pI value of the protein.

Table 3.19: The pI and charge calculations of *L. brevis* GAD pathway proteins.

L.brevis **GadR Protein 196 aa**			Neutral	Estimated Charge Over pH Range					
R.No on the protein	**Replacement Changes**	**Estimated pI**	**pH 7.00**	**pH 4.00**	**pH 4.50**	**pH 5.00**	**pH 5.50**	**pH 6.00**	**pH 6.50**
No Rep.	0	5.891	-4.794	+19.856	+12.581	+6.554	+2.546	-0.667	-3.232
55-70-81	3	5,894	-4.802	+20.066	+12.791	+6.673	+2.593	-0.652	-3.23
93	1	5.891	-4.794	+19.856	+12.581	+6.554	+2.546	-0.667	-3.232
191	1	5.891	-4.794	+19.856	+12.581	+6.554	+2.546	-0.667	-3.232
L.brevis **GadC Protein 501 aa**									
No Rep.	0	8.996	+6.179	+30.118	+23.11	+17.334	+13.463	+10.303	+7.752
7-141-247-266-289-291-369	7	9.097	+7.176	+30.38	+23.639	+18.114	+14.381	+11.276	+8.743
27-438	2	8.996	+6.179	+30.118	+23.11	+17.334	+13.463	+10.303	+7.752
93	1	8.996	+6.179	+30.118	+23.11	+17.334	+13.463	+10.303	+7.752
171	1	8.996	+6.179	+30.118	+23.11	+17.334	+13.463	+10.303	+7.752
332	1	8.877	+5.266	+30.108	+23.078	+17.239	+13.214	+9.792	+6.984
343-367	2	8.876	+5.18	+29.647	+22.372	+16.435	+12.497	+9.314	+6.755
367	1	8.876	+5.18	+29.647	+22.372	+16.435	+12.497	+9.314	+6.755
369	1	8.996	+6.179	+30.118	+23.11	+17.334	+13.463	+10.303	+7.752
L.brevis **Gad1 Protein 479 aa**									
No.Rep	0	4,905	-21.359	+26.51	+10.32	-1.877	-9.283	-14.765	-18.939
9	1	4,954	-20.359	+27.51	+11.32	-0.877	-8.283	-13.765	-17.939
38-234-407	3	5.045	-18.364	+28.243	+12.586	+0.802	-6.4	-11.804	-15.951
40	1	4,905	-21.359	+26.51	+10.32	-1.877	-9.283	-14.765	-18.939
240	1	4,905	-21.359	+26.51	+10.32	-1.877	-9.283	-14.765	-18.939
323	1	4,905	-21.359	+26.51	+10.32	-1.877	-9.283	-14.765	-18.939

Table 3.19: (cont.).

L.brevis Gad1 Protein 479 aa (cont.)

R.No on the protein	Replacement Changes	Estimated pI	Neutral pH 7.00	Estimated Charge Over pH Range pH 4.00	pH 4.50	pH 5.00	pH 5.50	pH 6.00	pH 6.50
339-471	2	4.950	-21.272	+27.499	+11.287	-0.972	-8.532	-14.277	-18.707
471	1	4.950	-21.272	+27.499	+11.287	-0.972	-8.532	-14.277	-18.707
472	1	4.949	-20.36	+26.981	+11.058	-0.978	-8.318	-13.776	-17.942
L.brevis Gad2 Protein 468 aa									
No.Rep	0	5.363	-18.832	+39.085	+21.164	+6.992	-2.238	-9.655	-15.493
12-16-23-40-325	5	5.359	-18.834	+38.876	+20.955	+6.873	-2.285	-9.671	-15.499
17-416	2	5.423	-17.833	+40.085	+22.164	+7.992	-1.238	-8.655	-14.494
17-328-416	3	5.423	-17.833	+40.085	+22.164	+7.992	-1.238	-8.655	-14.494
17-224-416	3	5.423	-17.833	+40.085	+22.164	+7.992	-1.238	-8.655	-14.494
17-153-224	3	5.423	-17.833	+40.085	+22.164	+7.992	-1.238	-8.655	-14.494
23-153-156	3	5.410	-18.745	+40.075	+22.132	+7.898	-1.487	-9.167	-15.261
23-94-153-156	4	5.352	-19.745	+39.075	+21.132	+6.898	-2.487	-10.167	-16.261
23	1	5.363	-18.832	+39.085	+21.164	+6.992	-2.238	-9.655	-15.493
23-153-156-157	4	5.410	-18.745	+40.075	+22.132	+7.898	-1.487	-9.167	-15.261
23-153-156	3	5.410	-18.745	+40.075	+22.132	+7.898	-1.487	-9.167	-15.261
139-173-416	3	5.363	-18.832	+39.085	+21.164	+6.992	-2.238	-9.655	-15.493
224-416	2	5.363	-18.832	+39.085	+21.164	+6.992	-2.238	-9.655	-15.493
227-229	2	5.421	-17.834	+39.557	+21.902	+7.891	-1.272	-8.666	-14.497
227-229-353	3	5.421	-17.834	+39.557	+21.902	+7.891	-1.272	-8.666	-14.497
416	1	5.363	-18.832	+39.085	+21.164	+6.992	-2.238	-9.655	-15.493

Table 3.20: The pI and charge calculations of *L. plantarum* GAD pathway proteins.

L.plantarum GadB Protein 469 aa			Neutral	Estimated Charge Over pH Range					
R.No on the protein	Replacement Changes	Estimated pI	pH 7.00	pH 4.00	pH 4.50	pH 5.00	pH 5.50	pH 6.00	pH 6.50
No.Rep	0	**5.577**	**-18.399**	+41.39	+24.694	+11.014	+1.347	-7.268	-14.35
180-228	2	5.577	-18.398	+41.39	+24.694	+11.014	+1.347	-7.268	-14.35
291	1	5.577	-18.398	+41.39	+24.694	+11.014	+1.347	-7.268	-14.35
142	1	5.520	-19.399	+40.39	+23.694	+10.014	+0.347	-8.268	-15.35
18-167-182-195-199-235-238	7	5.593	-17.486	+41.4	+24.726	+11.109	+1.596	-6.756	-13.582
33-167-182-195-199-235-238-265-318	9	5.592	-17.487	+40.871	+24.464	+11.008	+1.562	-6.767	-13.586
255-385	2	5.577	-18.399	+41.39	+24.694	+11.014	+1.347	-7.268	-14.35
180	1	5.577	-18.399	+41.39	+24.694	+11.014	+1.347	-7.268	-14.35
369	1	5.577	-18.399	+41.39	+24.694	+11.014	+1.347	-7.268	-14.35
370-451	2	5.617	-18.312	+42.379	+25.662	+11.919	+2.098	-6.779	-14.118
180-301-340	3	5.634	-17.4	+42.39	+25.694	+12.014	+2.347	-6.268	-13.351
341-385-451	3	5.619	-18.31	+42.589	+25.872	+12.038	+2.146	-6.763	-14.113
341-451	2	5.671	-17.313	+42.851	+26.4	+12.818	+3.064	-5.79	-13.122
18-451	2	5.617	-18.312	+42.379	+25.662	+11.919	+2.098	-6.779	-14.118
167-182-195-199-235-238	6	5.593	-17.486	+41.4	+24.726	+11.109	+1.596	-6.756	-13.582
14-318-451	3	5.726	-16.316	+43.641	+27.191	+13.699	+4.016	-4.807	-12.128
167-182-195-199-235-238-318	7	5.593	-17.486	+41.4	+24.726	+11.109	+1.596	-6.756	-13.582
191-451	2	5.617	-18.312	+42.379	+25.662	+11.919	+2.098	-6.779	-14.118
220-451	2	5.563	-19.31	+41.908	+24.924	+11.02	+1.133	-7.768	-15.115
451	1	5.617	-18.312	+42.379	+25.662	+11.919	+2.098	-6.779	-14.118

Table 3.20: (cont.).

R.No on the protein	Replacement Changes	Estimated pI	Neutral	Estimated Charge Over pH Range					
			pH 7.00	pH 4.00	pH 4.50	pH 5.00	pH 5.50	pH 6.00	pH 6.50
L.plantarum GadB Protein 469 aa (cont.)									
140-194-195-225-235	5	5.537	-18.485	+40.929	+23.988	+10.21	+0.63	-7.745	-14.579
385-446-451	3	5.617	-18.312	+42.379	+25.662	+11.919	+2.098	-6.779	-14.118
114-140-194-195-225	5	5.577	-18.399	+41.39	+24.694	+11.014	+1.347	-7.268	-14.35
284-317-451	3	5.617	-18.312	+42.379	+25.662	+11.919	+2.098	-6.779	-14.118
220-341-451	3	5.671	-17.313	+42.851	+26.4	+12.818	+3.064	-5.79	-13.122
180-312	2	5.634	-17.4	+42.39	+25.694	+12.014	+2.347	-6.268	-13.351
172-220-451	3	5.563	-19.31	+41.908	+24.924	+11.02	+1.133	-7.768	-15.115
385	1	5.577	-18.399	+41.39	+24.694	+11.014	+1.347	-7.268	-14.35
L.plantarum YjeM Protein 494 aa									
No.Rep	0	9.519	+13.81	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772
127	1	9.519	+13.81	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772
134	1	9.519	+13.81	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772
137	1	9.519	+13.81	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772
137-434	2	9.519	+13.81	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772
198	1	9.519	+13.81	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772
299	**1**	9.519	+13.81	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772
300	1	9.519	+13.81	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772
300-463	2	9.568	+14.81	+29.244	+24.991	+21.348	+18.973	+17.189	+15.772
308-462	2	9.479	+12.811	+27.244	+22.991	+19.348	+16.973	+15.189	+13.772
308-341-462	3	9.479	+12.811	+27.244	+22.991	+19.348	+16.973	+15.189	+13.772
359	1	9.532	+13.811	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772

Table 3.20: (cont.).

L.plantarum YjeM Protein 494 aa (Cont.)

R.No on the protein	Replacement Changes	Estimated pI	Neutral	Estimated Charge Over pH Range						
			pH 7.00	pH 4.00	pH 4.50	pH 5.00	pH 5.50	pH 6.00	pH 6.50
413	1	9.519	+13.81	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772
434	1	9.519	+13.81	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772
480	1	9.519	+13.81	+28.244	+23.991	+20.348	+17.973	+16.189	+14.772

3.6 Codon Usage Bias of GAD Pathway Proteins

The CBI and CAI values of the *L. brevis* and *L. plantarum* GAD system genes are shown in Table 3.21. The calculated CBI values for the GAD pathway genes of the *L. brevis* and *L. plantarum* species are values greater than zero. Accordingly, it can be said that preferred codons are used instead of random codons in the synthesis of GAD pathway proteins. In addition, CBI values can detect a foreign gene in the genome. Interestingly, the CBI values calculated for the two gad genes of *L. brevis* are different from each other.

The expression levels of the *L. brevis* and *L .plantarum* GAD system genes were tested with CAI calculations. The table below shows that the gene with the highest expression level among the GAD pathway genes is the *L. brevis gad1* gene (Table 3.21). In the *L. brevis* GAD operon, the expression levels of genes appear to be ordered as *gad1*, followed by *gadR* and *gadC*. Also, the expression level of the *L. brevis gad2* gene located outside the operon is lower than other genes. Unlike *L. brevis*, the calculated expression levels for *gadB* and *yjeM* are the same in *L. plantarum*.

Table 3.21: CBI and CAI calculations of *L.brevis*/*L.plantarum* GAD system proteins

Organisms	Protein	CBI	Scaled Chi-square, SChi2	CAI
L. brevis				
	Gad1	0.39	0.36	0.83
	Gad2	0.30	0.27	0.77
	GadC	0.33	0.28	0.79
	GadR	0.45	0.39	0.80
L.plantarum				
	GadB	0.29	0.26	0.79
	YjeM	0.33	0.28	0.79

3.7 Identification of Positively Selected Sites on GAD Pathway Proteins

L. brevis and *L. plantarum* GAD system genes are protein-coding genes, and site-model analysis was performed using models with different ω values in order to detect positively selected residues on these genes. As a result of site model analysis, statistically significant positively selected sites were detected only in the *gad2* and *gadB* genes among the *L. brevis* and *L. plantarum* GAD system genes. However, no significant difference was observed when comparing the maximum likelihood scores of the tested models.

Table 3.22 : PAML site model analysis of *L. brevis gad2* gene among different strains from various environments.

Model	Np	Estimated Parameters	InL	Model Comparisons	2ΔInL	Positively Selected Sites
M0 (one ratio)	21	ω = 0.10357	-2552.79			No sites
M1 (neutral)	22	p0 = 0.89389, p1 = 0.10611 ω0 = 0.00000, ω1 = 1.00000	-2546.77			
M2 (selection)	24	p0 = 0.89611, p1 = 0.00000 p2 = 0.10389, ω0 = 0.00000 ω1 =1.00000 , ω2 = 1.02924	-2546.77	M2 vs. M1	0	12T**,16L**, 17T**,23S**, 40E**,94R**, 139F**,153T**, 156Q**,157P**, 173C**,224A**, 227N**,229I**, 325S**,328I**, 353V**,416A**
M7 (beta)	22	p = 0.00848,q = 0.06238	-2546.79			
M8 (beta & ω)	24	p0 = 0.89611,p = 0.00500 q = 1.84034 ,p1 = 0.10389 ω = 1.02923	-2546.77	M8 vs. M7	0.04	12T**,16L**, 17T**,23S**, 40E**,94R**, 139F**,153T**, 156Q**,157P**, 173C**,224A**, 227N**,229I**, 325S**,328I**, 353V**,416A**
(*: P>95%; **: P>99%)						

95

Table 3.23 : PAML site model analysis of *L. plantarum gadB* gene among different strains from various environments.

Model	Np	Estimated Parameters	InL	Model Comparisons	2ΔInL	Positively Selected Sites
M0 (one ratio)	49	$\omega = 0.18776$	-2770.93			No sites
M1 (neutral)	50	p0 = 0.82224, p1 = 0.17776 ω0 = 0.01004, ω1 = 1.00000	-2763.32			
M2 (selection)	52	p0 = 0.91063, p1 = 0.00000 p2 = 0.08937, ω0 = 0.05941 ω1 = 1.0000, ω2 = 1.52501	-2763.24	M2 vs. M1	0.16	14E,**18P***,33R, **114D*,140A*,** 142R,167V, 172F,180M, 182V,191A, 194V,**195N***, 199D,220Y, 225A,**235H*,** 238P,265D, 284V,291V, 301P,312G, 318A,**341D*,** 370I,385P, 446V,**451Q****
M7 (beta)	50	p = 0.01345, q = 0.06516	-2763.35			
M8 (beta & ω)	52	p0 = 0.91135, p = 6.37004 q = 99.00000, p1 = 0.08865 ω = 1.52936	-2763.24	M8 vs. M7	0.22	14E,**18P***,33R, **114D*,140A*,** 142R,167V, 172F,180M, 182V,191A, 194V,**195N***, 199D,220Y, 225A,**235H*,** 238P,265D, 284V,291V, 301P,312G, 318A,**341D*,** 370I,385P, 446V,**451Q****
(*: P>95%; **: P>99%)						

3.8 Structure Predictions of GAD Pathway Genes

In this section, the results of the prediction servers specified in Chapter 2 are presented by comparing them. Since the variety of information provided by each server is different, tables containing the results were created. Moreover, the structural features of the replacement sites that change the amino acid sequences of the GAD system proteins were focused on in the tables.

3.8.1 Structural Features of *L. brevis and L. plantarum* GAD Enzymes

The PDB code 5GP4 structure was the most comparable for the *L. brevis* and *L. plantarum gad* genes in the investigations performed on the prediction servers. The structure of 5GP4 was utilized as a template by the prediction algorithms. Furthermore, the chosen template's similarity to *L. brevis* Gad1, *L. brevis* Gad2, and *L. plantarum* GadB proteins were determined as 52%, 99%, and 84%, respectively. It has been observed in structure prediction studies that the *L. brevis* GAD enzyme, whose structure is known in the literature, is the *L. brevis* Gad2 protein. Secondary structures of *L. brevis* and *L. plantarum gad* genes predicted by the I-TASSER server are shown in Figure 3.7. As can be seen from the figure, the secondary structures of the three indicated gad genes are pretty similar. Although the secondary structures of the three enzymes are quite similar to each other, it is noteworthy that *L. brevis* Gad2 and *L. plantarum* GadB proteins are more similar to each other.

Glutamate decarboxylase enzyme is a PLP-dependent enzyme, and there are essential residues in the enzyme to which both PLP and substrate bind. Moreover, L-glutamate, which is the substrate of the glutamate decarboxylase enzyme, binds very closely to PLP in the active site of the enzyme, and many amino acids in the active site form bonds with both molecules. Critical residues on the enzyme structure are presented in Chapter 1. Also, the alignment of the *L. brevis* and *L. plantarum gad* genes in Section 3.3 was used the find which sites correspond to the essential amino acids(Table 3.24). Afterward, both important sites and sites with replacement changes were visualized and examined on the predicted 3D structures of *L. brevis* and *L. plantarum gad* genes (Figure 3.8).
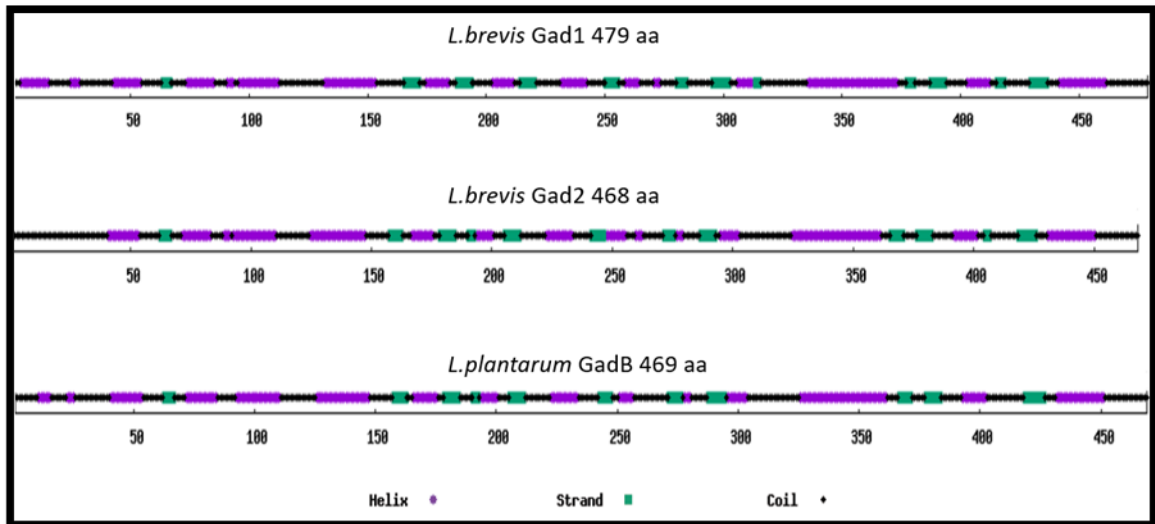
Figure 3.7: Schematic representation of secondary structures of *L. brevis* and *L. plantarum* GAD enzyme genes.

Table 3.24: Important and PLP/L-Glu binding sites on the *L. brevis* and *L. plantarum* GAD enzymes.

L.brevis Gad1	*L.brevis* Gad2	*L.plantarum* GadB	Bond Information
T66	T64	T64	Hyrogen bond with L-Glu
F67	F65	F65	Conserved and critical for catalytic activity
C68	C66	C66	Hydrogen bond with L-Glu
N87	N85	N85	Conserved and critical for catalytic activity
S132	S126	S126	Interaction with PLP
S133	S127	S127	Hyrogen bond with PLP
Q175	Q166	Q166	Hydrogen bond with L-Glu
V177	C168	C168	Hydrophobic interaction with pyridine ring of PLP
I220	I211	I211	Hydrophobic interaction with pyridine ring of PLP
T224	T215	T215	Necessary for the enzyme activity
A257	A248	A249	Hydrophobic interaction with pyridine ring of PLP
S285	S276	S277	Hydrogen bond with PLP
H287	H278	H279	Salt bridge with L-Glu, Hydrogen bond with PLP
K288	K279	K280	Salt bridge with L-Glu, PLP is covalently attached to the catalytic K279(Gad2)
F331	F320	F321	Conserved and critical for catalytic activity
S332	S321	S322	Hydrogen bond with L-Glu
R433	R422	R423	Salt bridge with L-Glu
Flexible Loop			
Y319	Y308	Y309	It covers the active site and provides a catalytic environment for GABA production.
L320	L309	L310	
G321	G310	G311	
K322	G311	G312	
T323	E312	E313	
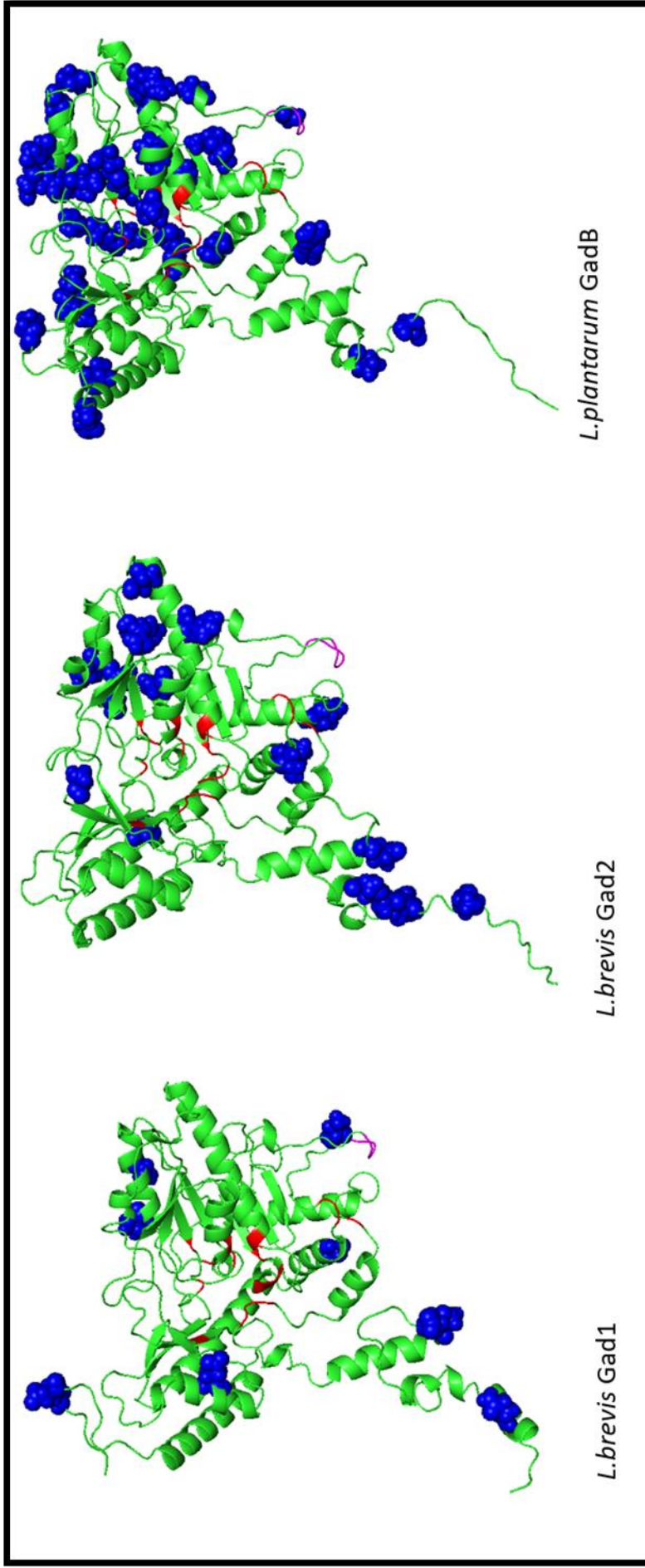
Figure 3.8: Predicted 3D structures of *L. brevis* and *L. plantarum* GAD enzymes. The 3D structures used in the image are structures predicted with the AlphaFold algorithm. Atoms shown as blue spheres represent replacement sites, while critical binding amino acids are colored in red, and pink represents the flexible loop. PyMol was used for the coloring of structures.

99

3.8.2 Structural Features of *L. brevis* and *L. plantarum* Glutamate-GABA Antiporters

The structure found as a template for the *L. brevis gadC* and *L. plantarum yjeM* genes by structure prediction analysis was the *E. coli* glutamate/GABA antiporter protein held on the PDB with the 4DJI code. The prediction studies found that protein secondary structures of *L. brevis gadC* and *L .plantarum yjeM* genes were formed by long α-helixes and short coils between α-helixes (Figure 3.9). Moreover, 12 TM-helixes for the 3D structure of both transporter genes were predicted by the Phyre2 server (Figure 3.10).

The amino acid sequences of the *L. brevis gadC* gene and the *L. plantarum yjeM* gene are only around 18% identical. Although the amino acid sequences of the *L. brevis gadC* and *L. plantarum yjeM* genes differ, their protein structures are very similar.
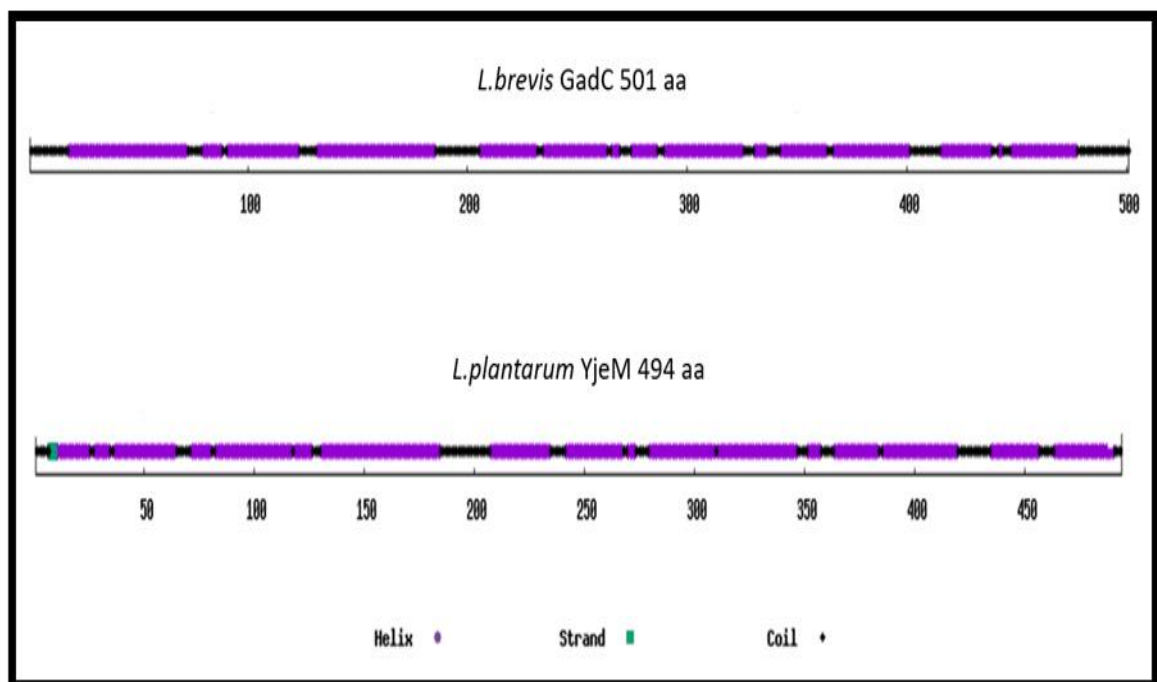
Figure 3.9: Schematic representation of *L. brevis* GadC and *L. plantarum* YjeM proteins' secondary structures.
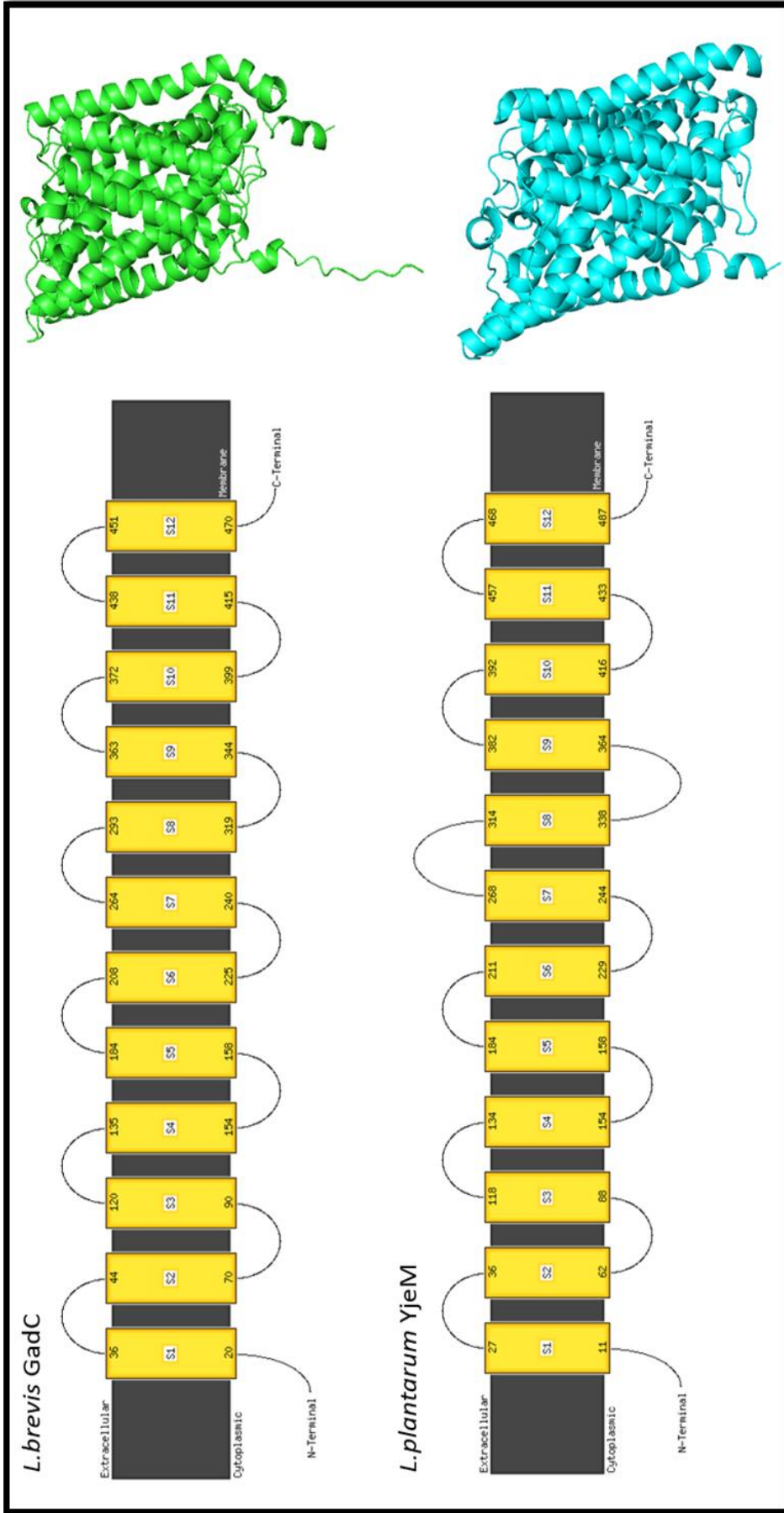
Figure 3.10: 3D structures and TM-helix (predicted with Phyre2) positions of *L. brevis* GadC and *L. plantarum* YjeM proteins. The green color represents GadC, and the cyan color represents YjeM. Structures were predicted with the AlphaFold server.

3.8.3 Structural Features of L.brevis GadR Protein

The GAD system regulator protein's specified three-dimensional structure is not found in the PDB database. Furthermore, the *gadR* gene is not present in all LABs. The similarities between the genes of LABs that include the *gadR* gene are quite low. The prediction servers used the regulatory proteins with the most identical protein sequences to *L. brevis* GadR in the PDB as templates to predict the three-dimensional protein structure of the *L. brevis gadR* gene. As a result of structure prediction analysis, it was found that the *L. brevis* GadR is a small protein consisting of α-helix structures. Furthermore, according to Phyre2 server analysis, the *L. brevis* GadR protein could contain a TM-helix structure between amino acids 136 and 151.
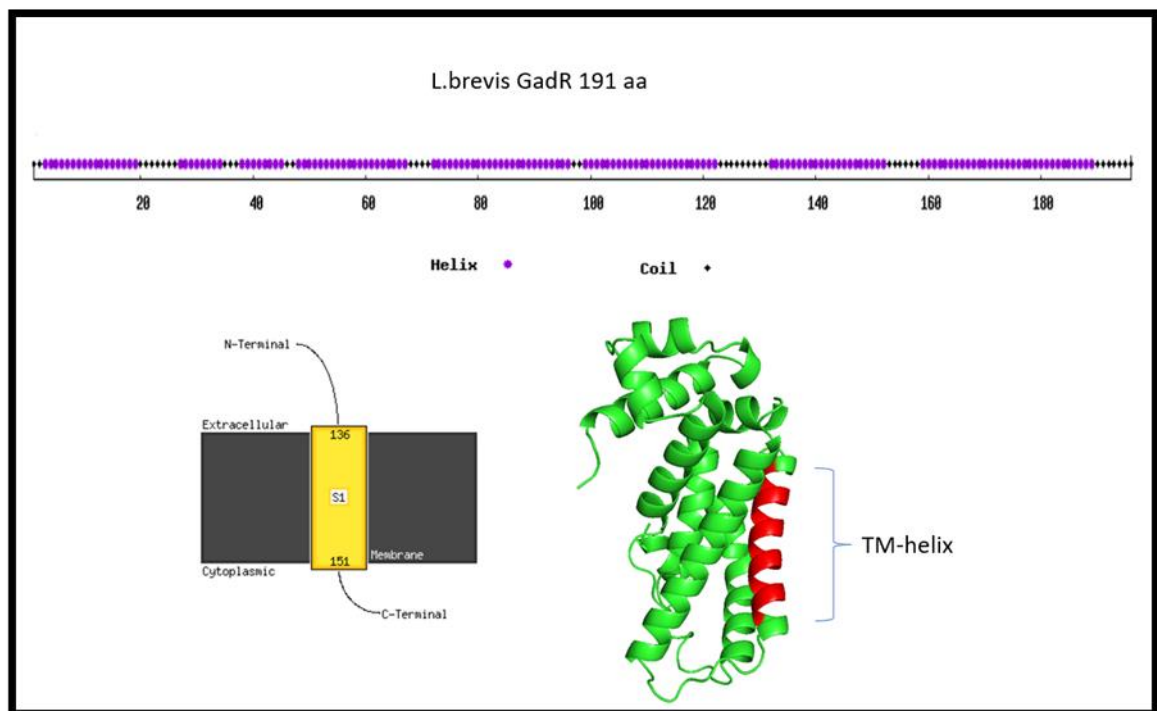
Figure 3.11: Secondary and 3D structure of *L. brevis* GadR protein predicted by I-TASSER and AlphaFold servers. The TM-helix structure predicted by the Phyre2 server is colored red. Coloring was done with PyMol.

3.9 Sliding Windows Analysis of Replacement Sites in GAD Pathway Genes

Sliding window analyses on the GAD pathway genes provide information on the non-synonymous π and TD values of the replacement polymorphisms observed in the genes of *L. brevis* and *L. plantarum* GAD system. Additionally, the study aimed to understand where replacement mutations are abundant in the protein structure by adding secondary structure predictions of GAD pathway genes to the π and TD values images in sliding window analysis.

The figures were prepared to compare the genes with each other. First, the *L. brevis gad1* and *gad2* genes were compared (Figure 3.12). The figure illustrates the distribution of replacement mutations found in the *L. brevis gad1* and *gad2* genes. While approximately 600 nucleotides of the *L. brevis gad1* gene were conserved, several mutations in the same region of the *L. brevis gad2* gene were found. Another contrast is that the non-synonymous TD values of the identified replacement mutations in the *gad1* gene are negative. In conversely to the *gad1* gene, many changes in the *gad2* gene have positive non-synonymous TD values. In previous analyses, the *L. brevis gad2* gene was very similar to the *L. plantarum gadB* gene in amino acid sequence. Therefore, the distribution of replacement mutations on *L .brevis gad1* and *L. plantarum gadB* genes was compared in the second figure (Figure 3.13). The number of changes found in the *L. plantarum gadB* gene is relatively high compared to the *L. brevis gad2* gene, and it is seen that these changes are concentrated in the middle PLP-binding domain of the *gadB* gene.

Additionally, the GAD system transporter gene *L. brevis gadC* was compared to the *yjeM* gene, which is assumed to be involved in the GAD system in *L. plantarum* (Figure 3.14). Replacement mutations in the *L. brevis gadC* gene were spread throughout the gene. Nevertheless, the first part of the *L. plantarum yjeM* gene is relatively conserved, and the changes were found to be predominantly distributed in the C-terminal regions of the gene.
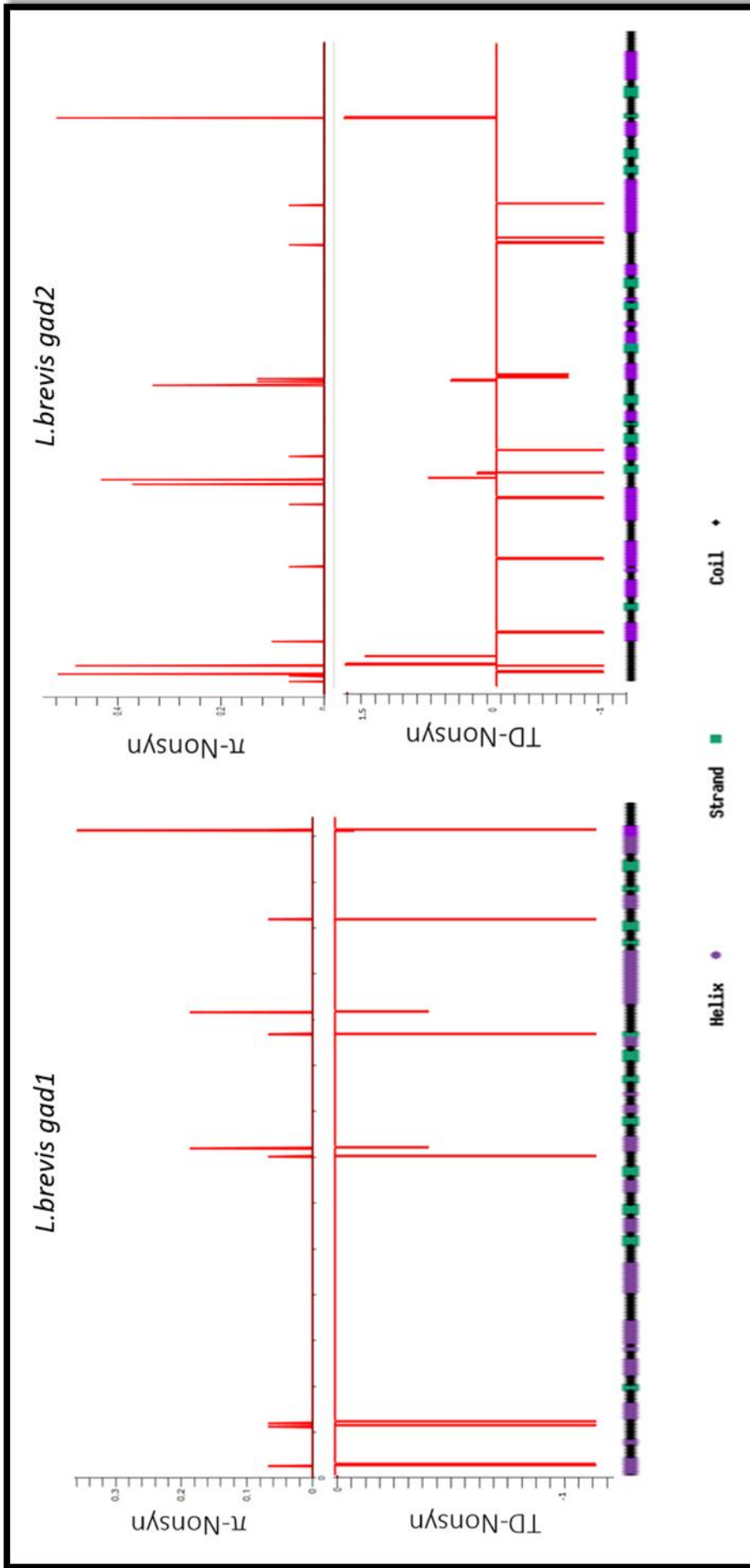
Figure 3.12: Sliding window analysis of *L. brevis gad* genes non-synonymous nucleotide diversity and neutrality statistics. Windows were arranged 1 kb in width with a step size of 1 kb.

Figure 3.13: Sliding window analysis of *L. brevis gad1* and *L. plantarum gadB* genes non-synonymous nucleotide diversity and neutrality statistics. The windows are prepared 1 kb in width with a step size of 1 kb.
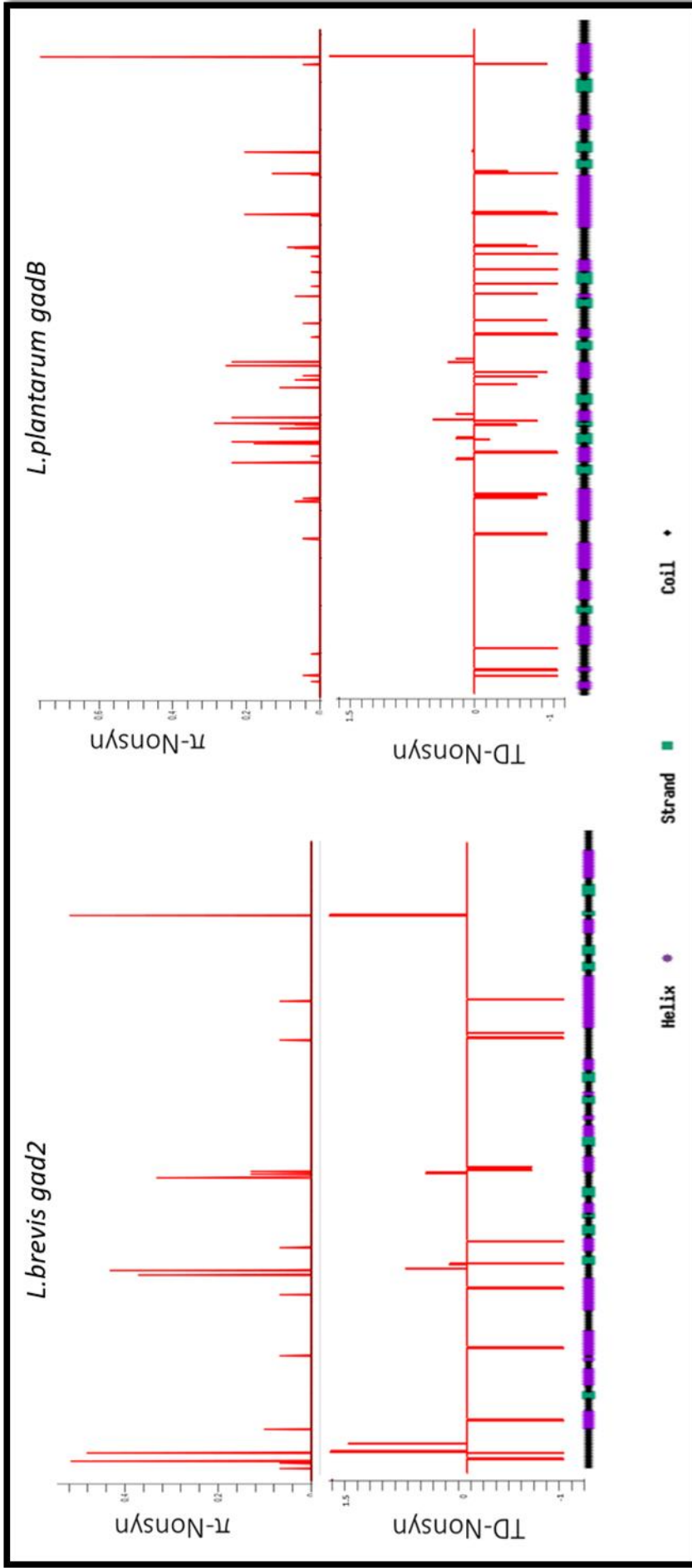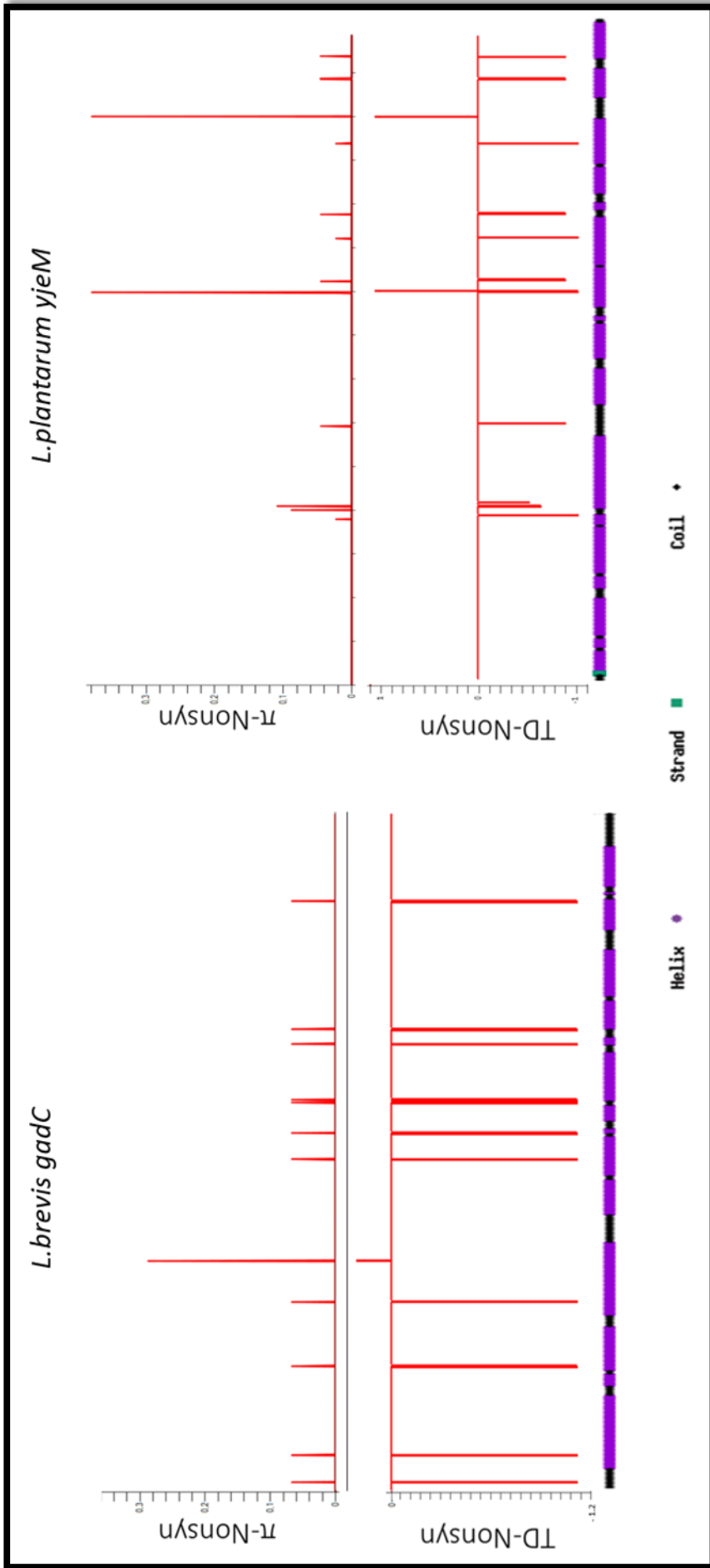
Figure 3.14: Sliding window analysis of *L. brevis gadC* and *L. plantarum yjeM* genes non-synonymous nucleotide diversity and neutrality statistics. The step size and width of the windows were 1kb.

106

3.10 Interspecies Comparison of GAD System Genes Among LAB Species

In interspecies studies, *gad* and transporter *gadC*(or *yjeM*) gene data from 32 bacterial species, including *L. brevis* and *L. plantarum*, were used. Interspecific MK and DoS analyses were performed for *L. brevis* and *L. plantarum* genes with all other species and listed in tables. Comparing *L. brevis gad1* and *gad2* gene sequences with other species revealed statistically significant MK values indicating positive selection. Comparisons of *L. brevis gad1* and *gad2* genes with some of the other species, on the other hand, supported the neutral selection theory based on NI and DoS values. It was discovered that the species *L. spicheri*, *L. angrenensis*, *L. cerevisiae*, *L. zymae*, and *L. sakei* yielded results consistent with the neutral selection hypothesis when compared to the *L. brevis gad1* gene. Significantly, the gad gene sequences of *L. zymae* and *L. sakei* species are quite similar to the *L. brevis gad1* gene (Table 3.25).

When gad genes are analyzed with the *L. brevis gad2* gene, the species that give MK and DoS values indicating neutral selection are *L. herbarum*, *L. paraplantarum*, *L. argentoratensis*, *L. plantarum*, and *C.futsaii* (Table 3.26). Interestingly, although the *gad1* and *gad2* genes of *L.brevis* are homologous genes expressing the same enzyme, the MK and DoS values obtained from comparing these genes with each other indicate positive selection. The tables show that fixed non-synonymous changes between *gad1* and *gad2* genes are more than synonymous changes.

A different scenario than *L. brevis* was observed in the results obtained comparing the *L. plantarum gadB* gene with other bacterial species. The species whose gad genes are very similar to the *L. plantarum gadB* gene are *L. argentoratensis* and *C. futsaii*. It was also found in species with NI values indicating negative selection among the compared species. The NI value obtained is quite large and statistically significant when only the *L.paraplantarum gad* gene and the *L.plantarum gadB* gene are compared among the *L. brevis* (*gad2*), *L. herbarum*, and *L. paraplantarum* species with NI values greater than one (Table 3.27). Fixed synonymous changes between the *L.plantarum gadB* and *L. paraplantarum gad* genes are almost nine times big as non-synonymous changes. Moreover, DoS calculations support NI values indicating negative selection.

Together with the *gad1* gene, the *L. brevis gadC* gene forms an operon structure. The *gadC* gene may not be present in all LABs that have the gad gene. The genomes of

the 32 bacterial species included in the analyses were searched, and data on the *gadC* gene were gathered. The *yjeM* gene data were collected from species that did not have the *gadC* gene. When the *L. brevis gadC* gene is compared with the transporter genes of other bacterial species, it is seen that four species give results close to the neutral selection hypothesis. Comparing the *L. angrenensis* and *L. cerevisiae* transporter genes with the *L. brevis gadC* gene provides the MK with DoS values that are not statistically significant and support neutral selection. Besides, the *L. zymae* and *L. sakei* transporter genes are identical to the *L. brevis gadC* gene. No fixed synonymous or non-synonymous changes were detected between these species (Table 3.28).

The *L .plantarum yjeM* gene was compared among 32 bacterial species that only have the *yjeM* gene instead of the *gadC* gene. As a result of the analysis, NI and DoS values obtained from the comparison of *L. paraplantarum* transporter gene and *L. plantarum yjeM*, as for *L. plantarum gadB* gene, are statistically significant values indicating negative selection. In addition, the *L. argentoratensis yjeM* gene is identical to the *L. plantarum yjeM* gene. There is no fixed synonymous or non-synonymous polymorphism between the *yjeM* genes of these two species (Table 3.29).

According to the results of MK analysis, species close to *L. brevis* and *L. plantarum* GAD pathway genes were chosen as an outgroup, and Fu-Li's tests were performed with similar species for each GAD pathway gene (Table 3.30 and 3.31).

L. brevis gad1 gene was analysed with closely related species such as *L. sakei*, *L. zymae*, *L. crevisiae*, *L. angrenensis*, and *L. spicheri*. Similarly, the *L. brevis gadC* gene was analysed with the same species. However, the *L. brevis gad2* gene was analysed with other species like *L. herbarum*, *L. paraplantarum*, *L. argentoratensis*, *L. platarum*, and *C. futsaii*. The *L. plantarum gadB* gene was tested with the closely determined *L. brevis gad2*, *L. herbarum*, *L. parplantarum*, *L. argentoratensis*, and *C. futsaii gad* genes. In MK analyses, the similarity of *L. brevis gad2* and *L. palantarum gadB* genes to each other was also observed with the similarity of the closely related species. Besides these, the *L. plantarym yjeM* gene was also tested with *L. paraplantarum* and *L. argentoratensis yjeM* genes.

Table 3.25: Interspecific comparisons of *L. brevis gad1* gene with *gad* genes of different bacterial species.

	Fixed	Polymorphic	P	NI	Alpha	DoS
L. oris						
Synonymous	244	40	0,000***	0.12	0.88	0.44
Nonsynonymous	327	6				
L. antri						
Synonymous	255	40	0,000***	0.12	0.88	0.43
Nonsynonymous	323	6				
L. reuteri						
Synonymous	253	40	0,000***	0.12	0.88	0.43
Nonsynonymous	324	6				
L. fermentum						
Synonymous	270	40	0,000***	0.13	0.87	0.42
Nonsynonymous	324	6				
L. gastricus						
Synonymous	249	40	0,000***	0.11	0.89	0.44
Nonsynonymous	330	6				
P. suebicus						
Synonymous	257	40	0,000***	0.12	0.88	0.42
Nonsynonymous	320	6				
L. lactis						
Synonymous	243	40	0,000***	0.11	0.89	0.44
Nonsynonymous	322	6				
E. faecium						
Synonymous	257	40	0,000***	0.12	0.88	0.43
Nonsynonymous	329	6				
L.brevis gad2						
Synonymous	275	40	0,000***	0.13	0.87	0.41
Nonsynonymous	322	6				
L. herbarum						
Synonymous	260	40	0,000***	0.13	0.87	0.41
Nonsynonymous	310	6				
L. paraplantarum						
Synonymous	270	40	0,000***	0.13	0.87	0.41
Nonsynonymous	323	6				
L. argentoratensis						
Synonymous	266	40	0,000***	0.12	0.88	0.42
Nonsynonymous	321	6				
L.plantarum						
Synonymous	272	40	0,000***	0.13	0.87	0.41
Nonsynonymous	317	6				
C. futsaii						
Synonymous	272	40	0,000***	0.13	0.87	0.41
Nonsynonymous	318	6				
C. nuruki						
Synonymous	242	40	0,000***	0.11	0.89	0.44
Nonsynonymous	324	6				
S. paracollinoides						
Synonymous	258	40	0,000***	0.12	0.88	0.44
Nonsynonymous	337	6				
F. rossiae						
Synonymous	241	40	0,001**	0.27	0.73	0.23
Nonsynonymous	135	6				
L. buchneri						
Synonymous	260	40	0,002**	0.27	0.73	0.23
Nonsynonymous	147	6				

Table 3.25: (cont.).

	Fixed	Polymorphic	P	NI	Alpha	DoS
L. paracasei						
Synonymous	259	40	0,001***	0.26	0.74	0.23
Nonsynonymous	149	6				
L. parakefiri						
Synonymous	250	40	0,003**	0.29	0.71	0.21
Nonsynonymous	131	6				
L. curvatus						
Synonymous	229	40	0,000***	0.15	0.85	0.31
Nonsynonymous	154	4				
L. rennini						
Synonymous	222	40	0,0003***	0.23	0.77	0.26
Nonsynonymous	144	6				
L. coleohominis						
Synonymous	219	38	0,000***	0.09	0.91	0.40
Nonsynonymous	194	3				
L. senmaizukei						
Synonymous	267	41	0,015*	0.35	0.65	0.17
Nonsynonymous	111	6				
L. tangyuanensis						
Synonymous	263	40	0,022*	0.36	0.64	0.16
Nonsynonymous	109	6				
L. spicheri						
Synonymous	223	40	0,412	0.64	0.36	0.06
Nonsynonymous	52	6				
L. angrenensis						
Synonymous	232	41	0,320	0.62	0.38	0.06
Nonsynonymous	55	6				
L. cerevisiae						
Synonymous	230	41	0,177	0.52	0.48	0.09
Nonsynonymous	65	6	(not sig.)			
L. zymae						
Synonymous	0	41	-	-	-	-
Nonsynonymous	0	6				
L. sakei						
Synonymous	0	41	0,145	0.00	1.00	0.87
Nonsynonymous	1	6	(not sig.)			
S. thermophilus						
Synonymous	269	41	0,000***	0.10	0.90	0.47
Nonsynonymous	397	6				
E.coli-alpha						
Synonymous	269	39	0,000***	0.09	0.91	0.49
Nonsynonymous	439	6				
E.coli-beta						
Synonymous	271	40	0,000***	0.08	0.92	0.49
Nonsynonymous	440	5				

Table 3.26: Interspecific comparisons of *L. brevis gad2* gene with *gad* genes of different bacterial species.

	Fixed	Polymorphic	P	NI	Alpha	DoS
L. oris						
Synonymous	262	55	0,0003***	0.34	0.66	0.23
Nonsynonymous	196	14				
L. antri						
Synonymous	269	55	0,001***	0.36	0.64	0.21
Nonsynonymous	189	14				
L. reuteri						
Synonymous	278	55	0,0002***	0.34	0.66	0.23
Nonsynonymous	210	14				
L. fermentum						
Synonymous	262	55	0,001***	0.35	0.65	0.22
Nonsynonymous	190	14				
L. gastricus						
Synonymous	252	55	0,001***	0.35	0.65	0.22
Nonsynonymous	182	14				
P. suebicus						
Synonymous	261	55	0,001***	0.37	0.63	0.21
Nonsynonymous	180	14				
L. lactis						
Synonymous	261	55	0,001**	0.38	0.62	0.20
Nonsynonymous	177	14				
E. faecium						
Synonymous	255	55	0,001**	0.38	0.62	0.20
Nonsynonymous	171	14				
L. herbarum						
Synonymous	262	55	0,236	0.67	0.33	0.07
Nonsynonymous	99	14	(not sig.)			
L. paraplantarum						
Synonymous	251	55	0,186	0.65	0.35	0.08
Nonsynonymous	98	14	(not sig.)			
L. argentoratensis						
Synonymous	262	55	0,185	0.64	0.36	0.08
Nonsynonymous	104	14	(not sig.)			
L.plantarum						
Synonymous	263	55	0,235	0.66	0.34	0.07
Nonsynonymous	101	14	(not sig.)			
C. futsaii						
Synonymous	265	55	0,235	0.66	0.34	0.08
Nonsynonymous	102	14	(not sig.)			
C. nuruki						
Synonymous	274	55	0,0003***	0.35	0.65	0.22
Nonsynonymous	201	14				
S. paracollinoides						
Synonymous	240	55	0,001***	0.36	0.64	0.21
Nonsynonymous	170	14				
F. rossiae						
Synonymous	270	55	0,000***	0.22	0.78	0.31
Nonsynonymous	290	13				
L. buchneri						
Synonymous	273	55	0,000***	0.21	0.79	0.34
Nonsynonymous	329	14				
L. paracasei						
Synonymous	273	55	0,000***	0.21	0.79	0.34
Nonsynonymous	329	14				

Table 3.26: (cont.).

	Fixed	Polymorphic	P	NI	Alpha	DoS
L. parakefiri						
Synonymous	284	55	0,000***	0.23	0.77	0.32
Nonsynonymous	314	14				
L. curvatus						
Synonymous	257	53	0,000***	0.23	0.77	0.32
Nonsynonymous	291	14				
L. rennini						
Synonymous	266	54	0,000***	0.22	0.78	0.33
Nonsynonymous	313	14				
L. coleohominis						
Synonymous	239	51	0,000***	0.16	0.84	0.38
Nonsynonymous	285	10				
L. senmaizukei						
Synonymous	273	55	0,000***	0.22	0.78	0.33
Nonsynonymous	316	14				
L. tangyuanensis						
Synonymous	280	54	0,000***	0.24	0.74	0.32
Nonsynonymous	308	14				
L. spicheri						
Synonymous	275	54	0,000***	0.23	0.77	0.33
Nonsynonymous	315	14				
L. angrenensis						
Synonymous	280	55	0,000***	0.22	0.78	0.33
Nonsynonymous	320	14				
L. cerevisiae						
Synonymous	279	55	0,000***	0.22	0.78	0.34
Nonsynonymous	330	14				
L. brevis gad1						
Synonymous	264	55	0,000***	0.20	0.80	0.35
Nonsynonymous	312	13				
L. zymae						
Synonymous	262	55	0,000***	0.21	0.79	0.35
Nonsynonymous	318	14				
L. sakei						
Synonymous	261	55	0,000***	0.21	0.79	0.35
Nonsynonymous	319	14				
S. thermophilus						
Synonymous	252	55	0,000***	0.18	0.82	0.38
Nonsynonymous	353	14				
E. coli-alpha						
Synonymous	249	54	0,000***	0.16	0.84	0.40
Nonsynonymous	376	13				
E. coli-beta						
Synonymous	251	55	0,000***	0.15	0.85	0.41
Nonsynonymous	385	13				

Table 3.27: Interspecific comparisons of *L. plantarum gadB* gene with *gad* genes of different bacterial species.

	Fixed	Polymorphic	P	NI	Alpha	DoS
L. oris						
Synonymous	257	82	0,010**	0.57	0.43	0.13
Nonsynonymous	209	38				
L. antri						
Synonymous	251	82	0,010**	0.57	0.43	0.13
Nonsynonymous	204	38				
L. reuteri						
Synonymous	271	82	0,009**	0.56	0.44	0.13
Nonsynonymous	217	37				
L. fermentum						
Synonymous	253	82	0,021*	0.60	0.40	0.12
Nonsynonymous	191	37				
L. gastricus						
Synonymous	240	82	0,007**	0.55	0.45	0.14
Nonsynonymous	196	37				
P. suebicus						
Synonymous	236	82	0,011*	0.57	0.43	0.13
Nonsynonymous	186	37				
L. lactis						
Synonymous	252	82	0,028*	0.62	0.38	0.11
Nonsynonymous	189	38				
E. faecium						
Synonymous	243	82	0,045*	0.64	0.36	0.10
Nonsynonymous	175	38				
L.brevis gad2						
Synonymous	253	82	0,644 (not sig.)	1.14	-0.14	-0.03
Nonsynonymous	103	38				
L. herbarum						
Synonymous	215	82	0,181 (not sig.)	1.39	-0.39	-0.07
Nonsynonymous	72	38				
L. paraplantarum						
Synonymous	191	82	0,000***	4.02	-3.02	-0.21
Nonsynonymous	22	38				
L. argentoratensis						
Synonymous	1	82	0,105 (not sig.)	0.15	0.85	0.43
Nonsynonymous	3	38				
C. futsaii						
Synonymous	2	82	1,000 (not sig.)	0.93	0.07	0.02
Nonsynonymous	1	38				
C. nuruki						
Synonymous	253	82	0,009**	0.57	0.43	0.13
Nonsynonymous	202	37				
S. paracollinoides						
Synonymous	243	82	0,055 (not sig.)	0.64	0.36	0.10
Nonsynonymous	171	37				
F. rossiae						
Synonymous	252	82	0,000***	0.38	0.62	0.23
Nonsynonymous	296	37				
L. buchneri						
Synonymous	274	82	0,000***	0.39	0.61	0.23
Nonsynonymous	318	37				
L. paracasei						
Synonymous	274	82	0,000***	0.39	0.61	0.23
Nonsynonymous	319	37				

Table 3.27: (cont.).

	Fixed	Polymorphic	P	NI	Alpha	DoS
L. parakefiri						
Synonymous	267	82	0,000***	0.39	0.61	0.23
Nonsynonymous	309	37				
L. rennini						
Synonymous	244	82	0,000***	0.34	0.66	0.26
Nonsynonymous	326	37				
L. coleohominis						
Synonymous	235	76	0,000***	0.38	0.62	0.23
Nonsynonymous	278	34				
L. senmaizukei						
Synonymous	259	81	0,000***	0.38	0.62	0.23
Nonsynonymous	308	37				
L. tangyuanensis						
Synonymous	257	82	0,000***	0.39	0.61	0.23
Nonsynonymous	298	37				
L. spicheri						
Synonymous	259	82	0,000***	0.39	0.61	0.23
Nonsynonymous	303	37				
L. angrenensis						
Synonymous	243	82	0,000***	0.36	0.64	0.25
Nonsynonymous	309	37				
L. cerevisiae						
Synonymous	250	81	0,000***	0.36	0.64	0.24
Nonsynonymous	315	37				
L. brevis-gad1						
Synonymous	253	82	0,000***	0.36	0.64	0.24
Nonsynonymous	301	35				
L. zymae						
Synonymous	255	82	0,000***	0.37	0.63	0.24
Nonsynonymous	309	37				
L. sakei						
Synonymous	255	82	0,000***	0.37	0.63	0.24
Nonsynonymous	310	37				
S. thermophilus						
Synonymous	250	82	0,000***	0.33	0.67	0.27
Nonsynonymous	342	37				
E. coli-alpha						
Synonymous	244	81	0,000***	0.31	0.69	0.28
Nonsynonymous	366	38				
E. coli-beta						
Synonymous	250	81	0,000***	0.32	0.68	0.28
Nonsynonymous	369	38				

Table 3.28: Interspecific comparisons of *L. brevis gadC* gene with *gadC* or *yjeM* genes of different bacterial species.

	Fixed	Polymorphic	P	NI	Alpha	DoS
L. zymae						
Synonymous	0	52	-	-	-	-
Nonsynonymous	0	7				
L. sakei						
Synonymous	0	52	-	-	-	-
Nonsynonymous	0	7				
L. angrenensis						
Synonymous	240	52	0,198	0.56	0.44	0.08
Nonsynonymous	58	7	(not sig.)			
L. cerevisiae						
Synonymous	263	52	0,151	0.53	0.47	0.08
Nonsynonymous	67	7	(not sig.)			
L. spicheri						
Synonymous	265	52	0,043*	0.44	0.56	0.12
Nonsynonymous	82	7				
L. senmaizukei						
Synonymous	283	52	0,002**	0.31	0.69	0.19
Nonsynonymous	124	7				
L. tangyuanensis						
Synonymous	280	52	0,002**	0.30	0.70	0.20
Nonsynonymous	125	7				
F. rossiae						
Synonymous	284	52	0,001***	0.27	0.73	0.21
Nonsynonymous	142	7				
L. parakefiri						
Synonymous	250	50	0,000***	0.19	0.81	0.28
Nonsynonymous	159	6				
L. buchneri						
Synonymous	276	52	0,000***	0.23	0.77	0.25
Nonsynonymous	165	7				
L. rennini						
Synonymous	255	52	0,000***	0.20	0.80	0.28
Nonsynonymous	169	7				
L. curvatus						
Synonymous	259	49	0,000***	0.20	0.80	0.23
Nonsynonymous	188	7				
L. coleohominis						
Synonymous	265	50	0,000***	0.17	0.83	0.33
Nonsynonymous	221	7				
E. coli						
Synonymous	333	52	0,000***	0.09	0.91	0.46
Nonsynonymous	429	6				
L. lactis						
Synonymous	312	52	0,000***	0.09	0.91	0.46
Nonsynonymous	399	6				
E. faecium						
Synonymous	280	51	0,000***	0.08	0.92	0.49
Nonsynonymous	404	6				
L. reuteri						
Synonymous	312	52	0,000***	0.08	0.92	0.47
Nonsynonymous	427	6				
L. oris						
Synonymous	296	52	0,000***	0.09	0.91	0.47
Nonsynonymous	400	6				

Table 3.28: (cont.).

	Fixed	Polymorphic	P	NI	Alpha	DoS
L. antri						
Synonymous	290	52	0,000***	0.08	0.92	0.47
Nonsynonymous	397	6				
S. thermophilus						
Synonymous	280	52	0,000***	0.08	0.92	0.52
Nonsynonymous	497	7				
C. nuruki						
Synonymous	284	45	0,000***	0.08	0.92	0.52
Nonsynonymous	538	7				
C. futsaii						
Synonymous	274	44	0,000***	0.08	0.92	0.52
Nonsynonymous	530	7				
L. plantarum						
Synonymous	292	47	0,000***	0.09	0.91	0.50
Nonsynonymous	506	7				
L. paraplantarum						
Synonymous	279	45	0,000***	0.09	0.91	0.51
Nonsynonymous	507	7				
L. argentoratensis						
Synonymous	295	47	0,000***	0.09	0.91	0.51
Nonsynonymous	513	7				
S. paracollinoides						
Synonymous	284	46	0,000***	0.07	0.93	0.53
Nonsynonymous	514	6				
L. fermentum						
Synonymous	284	46	0,000***	0.08	0.92	0.51
Nonsynonymous	514	7				
L. gastricus						
Synonymous	260	45	0,000***	0.08	0.92	0.53
Nonsynonymous	517	7				

Table 3.29: Interspecific comparisons of *L. plantarum yjeM* gene with *yjeM* genes of different bacterial species.

	Fixed	Polymorphic	P	NI	Alpha	DoS
C. nuruki						
Synonymous	279	51	0,000***	0.22	0.78	0.34
Nonsynonymous	330	13				
C. futsaii						
Synonymous	273	51	0,000***	0.20	0.80	0.36
Nonsynonymous	347	13				
L. paraplantarum						
Synonymous	125	51	0,001**	4.9	-3.9	-0.16
Nonsynonymous	7	14				
L. argentoratensis						
Synonymous	0	51	-	-	-	-
Nonsynonymous	0	14				
S. paracollinoides						
Synonymous	268	51	0,000***	0.20	0.80	0.36
Nonsynonymous	362	14				
L. fermentum						
Synonymous	287	50	0,000***	0.20	0.80	0.37
Nonsynonymous	385	13				
L. gastricus						
Synonymous	293	51	0,000***	0.18	0.82	0.38
Nonsynonymous	418	13				

Table 3.30: *L. brevis* GAD pathway (*gad1-gadC* and *gad2*) genes Fu -Li's Tests with outgroups.

	Total Sites	S	Eta	Eta (e)	Eta (s)	FL-D	FL-F	FW-Hn	ZE	AY
Levilactobacillus brevis gad1 gene with;										
L.sakei										
	1437	47	47	11	11	0.40	0.42	0.92	-0.74	0.06
L.zymae										
	1437	47	47	11	11	0.40	0.42	0.92	-0.74	0.06
L.cerevisiae										
	1437	38	38	8	11	0.55	0.48	-0.94	2.14	-0.22
L.angrenensis										
	1437	37	37	8	10	0.51	0.45	-0.88	2.24	-0.22
L.spicheri										
	1407	36	36	5	9	1.03	0.93	-1.43	2.84	-0.39
Levilactobacillus brevis gadC gene with;										
L.sakei										
	1503	59	59	18	18	-0.10	-0.27	1.09	-1.58	-0.59
L zymae										
	1503	59	59	18	18	-0.10	-0.27	1.09	-1.58	-0.59
L.cerevisiae										
	1503	48	48	10	14	0.57	0.30	-2.52	3.40	-1.08
L.angrenensis										
	1503	46	46	9	13	0.66	0.39	-1.95	3.16	-1.13
L.spicheri										
	1503	48	48	10	14	0.58	0.31	-2.46	3.37	-1.08
Levilactobacillus brevis gad2 gene with;										
L.herbarum										
	1398	50	51	13	24	0.25	-0.01	-3.38	4.54	-1.00
L. paraplantarum										
	1398	48	49	12	24	0.32	-0.03	-3.34	4.40	-1.05
L.argentoratensis										
	1398	47	48	10	22	0.57	0.29	-3.13	4.63	-1.14
L.plantarum										
	1398	47	48	10	22	0.57	0.29	-3.22	4.70	-1.14
C.fursaii										
	1398	47	48	10	22	0.57	0.29	-3.22	4.70	-1.14

Table 3.31: *L. plantarum* GAD pathway (*gadB* and *yjeM*) genes Fu -Li's Tests with outgroups.

	Total Sites	S	Eta	Eta (e)	Eta (s)	FL-D	FL-F	FW-Hn	ZE	AY
Lactiplantibacillus plantarum gadB gene with;										
L. brevis-gad2										
	1398	88	88	18	26	-0.08	-0.91	-4.44	5.06	-2.03
L. herbarum										
	1401	101	101	17	33	-0.39	-0.64	-4.99	4.38	-2.04
L.paraplantarum										
	1401	103	103	23	32	-0.33	-1.10	-3.98	3.52	-1.96
L. argentoratensis										
	1401	120	120	33	39	-1.04	-1.62	-2.62	0.84	-1.79
C. futsaii										
	1401	120	120	39	39	-1.72	-2.07#	0.60	-2.09	-1.67
Lactiplantibacillus plantarum YjeM gene with;										
L. paraplantarum										
	1482	64	64	21	30	-1.60	-1.87	-3.19	1.6	-1.23
L. argentoratensis										
	1482	65	65	28	30	-2.87*	-2.70*	-0.10	-1.13	-0.83

3.11 Phylogenetic Gene and Species Trees of *L. brevis and L. plantarum* GAD Pathway Genes with Other Bacterial Species

In order to determine the evolutionary phylogenetic relationships of 32 bacterial species used in interspecies analyses, a species tree was constructed with 16S rRNA data of these species (Figure 3.15). In addition, the presence and organization of the GAD system genes in each bacterial species were determined by checking their genomes over NCBI. The bacterial species examined differ in having the GAD system genes. In some species, the gad gene forms an operon structure with *gadC* and *gadR*, while in some species, the *gltX* gene is included in this operon structure to create a gene cluster. Also, some species contain only the gad gene in their genome without the operon structure (Figure 3.16).

From the GAD system genes, gene trees were constructed with the *gad* gene and the transporter *gadC* gene nucleotide sequences. Data for the *yjeM* gene were gathered from organisms that lacked the *gadC* gene. The three species were not included in the transporter gene tree since there is no information on them in the NCBI. Rooted and unrooted gene trees were generated to observe the species' phylogenetic distances better.

When the species and gene trees of 32 bacterial species are reviewed, it is observed that the trees have different topologies. Also, the distribution of GAD operon structure on the LAB species tree (*E. coli* used as an outgroup) suggested loss of operon structure but maintenance of GAD system activity via homologous genes more than once in different LAB clades. It is noteworthy that some species that are phylogenetically distant from each other in the species tree have a closer evolutionary relationship in terms of GAD system genes. Moreover, a specific pattern is seen in the gene trees of the GAD pathway.

Figure 3.17 shows the gene tree constructed from the *gad* genes of 31 LAB species and outgroup *E .coli*. As it is clearly seen in the unrooted gene tree with orange coloring, the species that contain the *gltX* gene next to the GAD operon are a separate group from the other species in the phylogenetic tree. The species in the branches marked with purple are the species that have only the *gad* gene in their genome, which is not in the operon form, and they are grouped closely with each other. The species in the branches marked with blue contain the *gadC* and *gad* gene in the GAD operon.No *gltX* gene was found next to the GAD operon in the blue group species' genomes. In the phylogenetic tree, the grouping of the *gad* genes of the species in the blue group together with the *gad* genes of the species that do not have the GAD operon structure in their genomes may indicate the loss of the operon structure in the evolutionary process.

Moreover, the functionally conserved two gad copies in the *L.brevis* genome clustered with two distinct *gad* clades indicating horizontal gene transfer from a phylogenetically distant LAB species rather than gene duplication for the origin of the second *gad* gene. Similar to *L .brevis*, *E. coli* also has two *gad* genes that are functionally homologous to each other. However, it is understood that *E.coli gad* genes are formed due to duplication as they are found in the same branch of the gene tree.

Similar results to the *gad* gene tree were obtained from the gene tree created for the GAD system transporter gene. While the *gadC* genes of the species that have the *gltX* gene next to the GAD operon structure are gathered together in the gene tree, the *gadC* genes of the species in the blue group are again grouped separately from the orange group (Figure 3.18). As expected, *yjeM* genes collected from species that do not have the *gadC* gene in their genomes (purple group) are grouped in the gene tree away from species that have the *gadC* gene in their genomes.
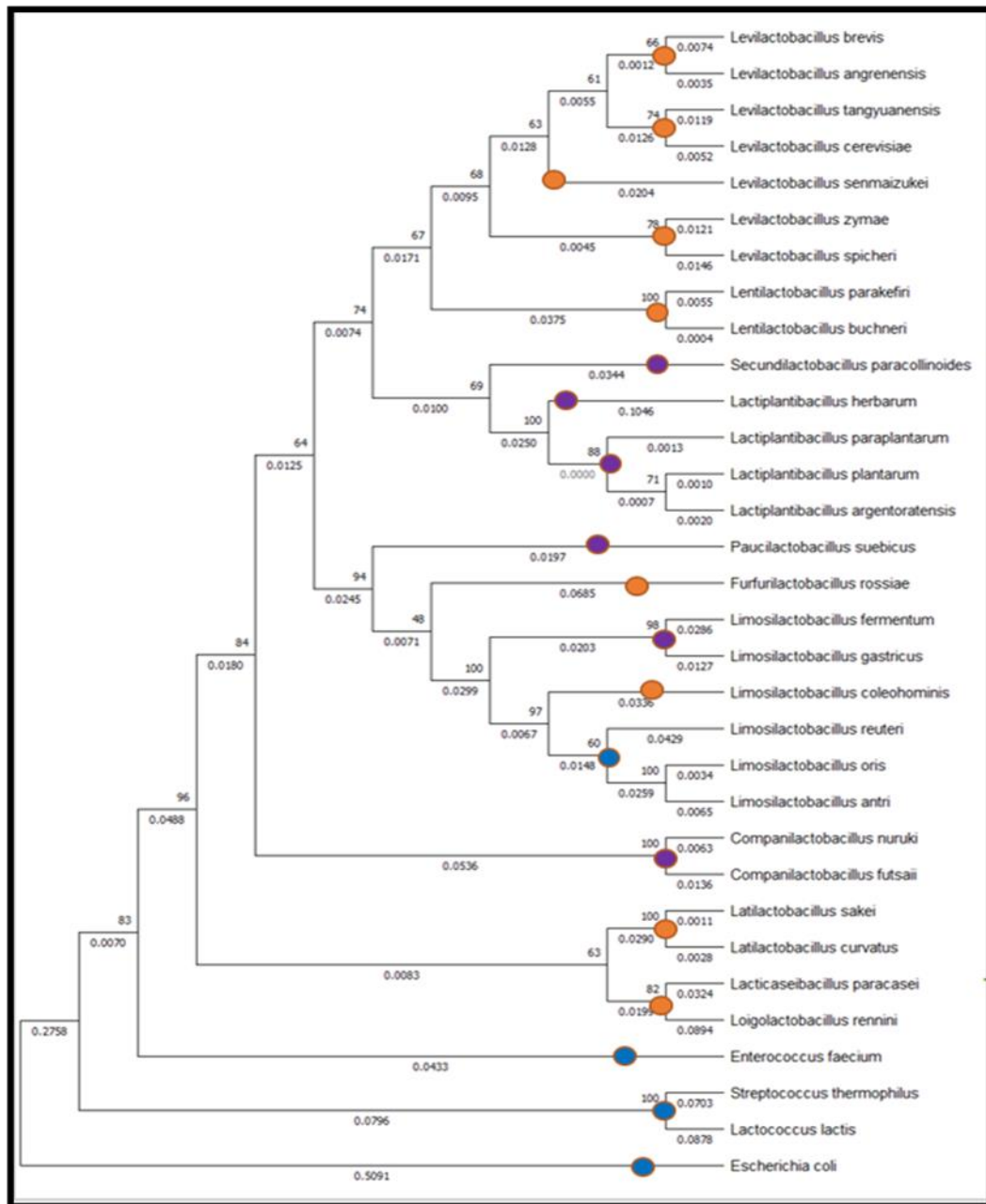
Figure 3.15: Phylogenetic species tree of 32 bacterial species containing GAD system. The tree was created with MEGA-X using the Maximum Likelihood method and the Kimura 2-parameter model[193]. One thousand replicates were used to generate the bootstrap consensus tree. To describe evolutionary rate variations across sites, a discrete Gamma(G) distribution was applied. The complete delete option was chosen (all gaps and missing data eliminated). The colorings in the branches indicate the presence of the GAD system genes of the species. Purple means that they contain only the *gad* gene, blue indicates the *gadC* and *gad* gene in the operon, and orange indicates that they have the *gltX* gene next to the *gad* gene in the operon structure.

121

Organisim	gadR	gadC	gad1	gltx	I	gad2
Levilactobacillus brevis	CP015398(363239..363829)	CP015398 (364023..365528)	CP015398 (365614..367023)	CP015398 (367102..368595)		CP015398 (2255135..2256541)
Levilactobacillus angrenensis	NZ_RHOB01000001 (202138..202728)	NZ_RHOB01000001 (202939..204444)	NZ_RHOB01000001 (204504..205949)	NZ_RHOB01000001 (206015..207526)		#
Levilactobacillus tangyuanensis	NZ_RHOA01000007 complement(6320..6910)	NZ_RHOA01000007 complement(4619..6118)	NZ_RHOA01000007 complement(3115..4521)	NZ_RHOA01000007 complement(1526..3040)		#
Levilactobacillus cerevisiae	NZ_RHNN01000018 complement(20568..21158)	NZ_RHNN01000018 complement(18850..20355)	NZ_RHNN01000018 complement(17346..18791)	NZ_RHNN01000018 complement(15773..17284)		#
Levilactobacillus senmaizukei	AYZH01000037 (275..865)	AYZH01000037 (1067..2566)	AYZH01000037 (2634..4070)	AYZH01000037 (4280..5659)		#
Levilactobacillus zymae	*	KF690144	KF690143	*		#
Levilactobacillus spicheri	JZCR01000015 complement(8136..8726)	JZCR01000015 complement(6431..7936)	JZCR01000015 complement(4919..6346)	JZCR01000015 complement(3349..4860)		#
Lentilactobacillus parakefiri	*	BDGB01000146 (282..1628)	BDGB01000146 (1652..3091)	BDGB01000146 (3136..4671)		#
Lentilactobacillus buchneri	CP002652 (101539..102132)	CP002652 (102412..103932)	CP002652 (104011..105459)	CP002652 (105492..106997)		#
Secundilactobacillus paracollinoides	#	#	#	#		CP014915 complement(3407511..3408911)
Lactiplantibacillus herbarum	#	#	#	#		NZ_LFEE01000043 (18702..20105)
Lactiplantibacillus paraplantarum	#	#	#	#		CP032744 (3003970..3005379)
Lactiplantibacillus plantarum	#	#	#	#		CP022294 (2995949..2997358)
Lactiplantibacillus argentoratensis	#	#	#	#		CP032751 (2955520..2956929)
Pauciclactobacillus suebicus	#	#	#	*		AZGF01000003 complement(6677..8077)
Furfurilactobacillus rossiae	AZFF01000004 (99515..100093)	AZFF01000004 (100177..101781)	AZFF01000004 (101836..103251)	AZFF01000004 (103381..104835)		#
Limosilactobacillus fermentum	#	#	KM982734	#		AB856984
Limosilactobacillus gastricus	#	#	AB295641	#		NZ_AICN01000052 complement(6599..8002)
Limosilactobacillus coleohominis	**	AZEW01000121 (825..2294)	AZEW01000005 (13..1290)	AZEW01000005 (1435..2943)		#
Limosilactobacillus reuteri	**	JOKX02000003 complement(232782..234314)	JOKX02000003 complement(231353..232759)	**		#
Limosilactobacillus oris	**	AZGE01000002 complement(17703..19268)	AZGE01000002 complement(16282..17688)	**		#
Limosilactobacillus antri	**	ACLL01000020 complement(35442..37007)	ACLL01000020 complement(34021..35427)	**		#
Companilactobacillus nuruki	#	#	#	**		NIPR01000025 (7568..8968)
Companilactobacillus futsaii	*	#	#	*		AB986192
Latilactobacillus sakei	*	KP310071	#	*		#
Latilactobacillus curvatus	**	CP022475 complement(1766348..1767805)	CP022475 complement(1764918..1766273)	CP022475 complement(1763295..1764809)		#
Lacticaseibacillus paracasei		*	#	**		#
Loigolactobacillus rennini	NZ_AYYI01000075 complement(17950..18552)	NZ_AYYI01000075 complement(16191..17708)	NZ_AYYI01000075 complement(14685..16121)	NZ_AYYI01000075 complement(13168..14664)		#
Enterococcus faecium	**	LKPH01000077 complement(3248..4759)	LKPH01000077 complement(1835..3235)	**		#
Streptococcus thermophilus	**	CP061019 (390806..392239)	CP061019 (389390..390769)	**		#
Lactococcus lactis	CP042408 complement(799453..800283)	CP042408 complement(797798..799309)	CP042408 complement(796377..797777)	**		#
Escherichia coli	**	CACSHO01000111 complement(39744..41279)	CACSHO01000111 complement(41435..42835)	**		CACSHO01000110 complement(42246..43646)

Figure 3.16: Presence and organization of GAD pathway genes in 32 different bacterial genomes. The locations of the genes on the genome and the accession numbers for each species' genes are shown. The *gad1* represents the *gad* gene found in the operon structure. The gene named *gad2* represents the *gad* gene of species that contain only the *gad* gene in their genomes. * Indicates lack of data on NCBI. ** Used for genes not included in the operon structure. The GAD operon consists only of the *gadC* and *gad1* genes in some species. # means that there is no gene data in the genome of the species.
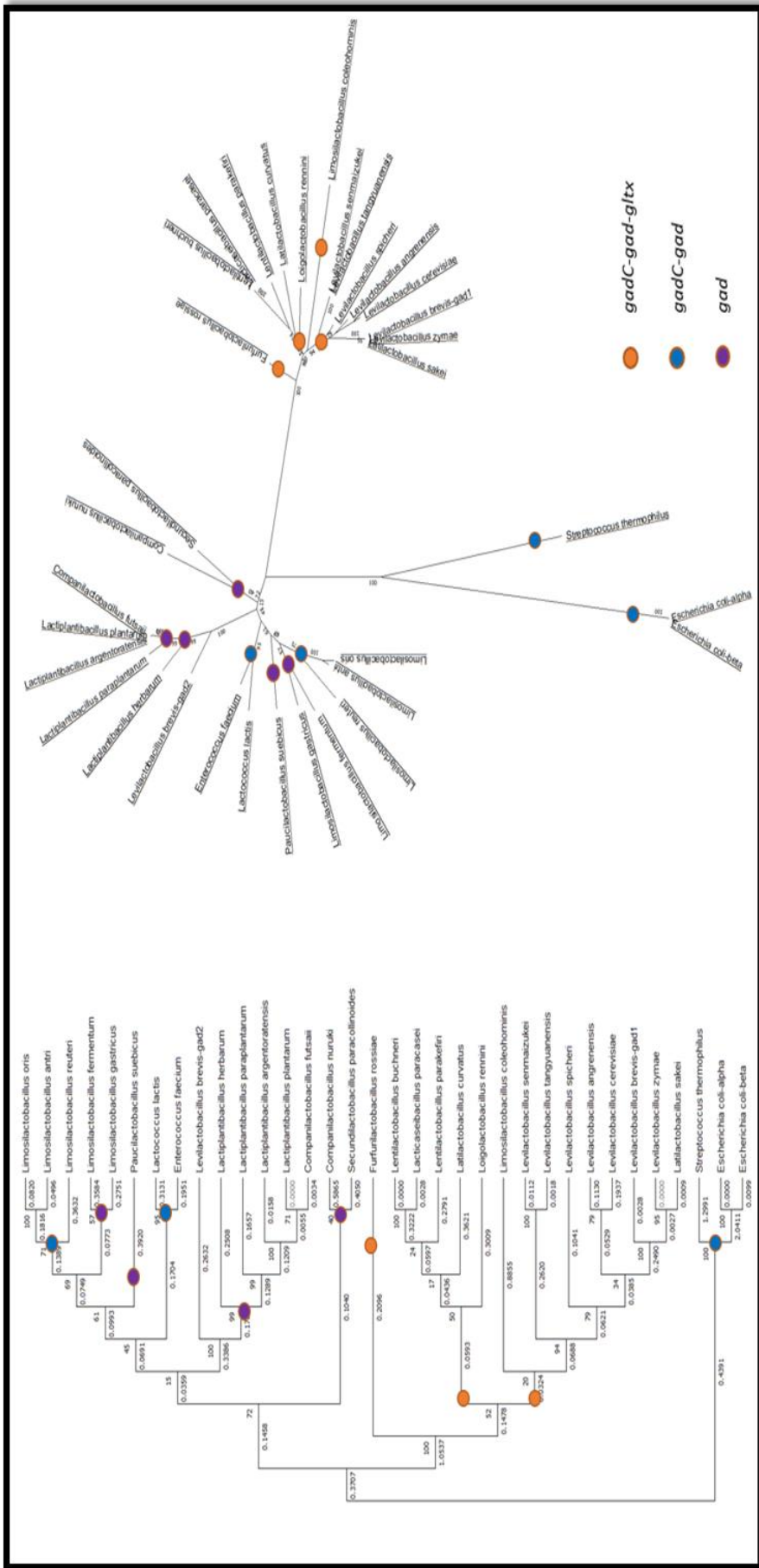
Figure 3.17: Rooted and unrooted gene trees of 32 bacterial species' gad gene. Trees were produced with the Maximum Likelihood method and General Time Reversible(GTR) model[194] using the 1000 bootstrap replicates. G-distribution and complete deletion options were selected on the MEGA-X. Colorings indicate the organization of GAD pathway genes.
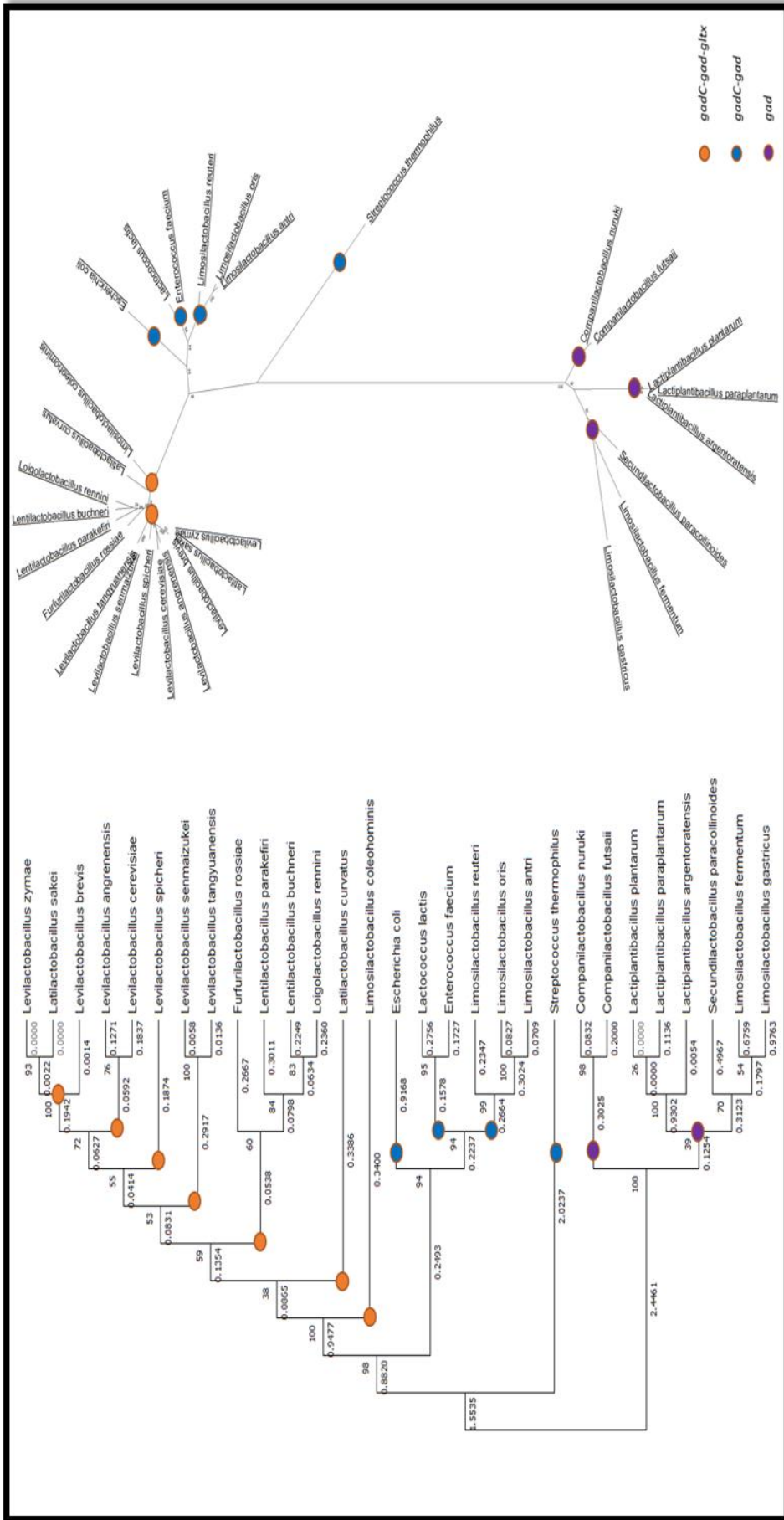
Figure 3.18: Rooted and unrooted Maximum Likelihood gene trees of 29 bacterial species transporter gene (*gadC* or *yjeM*). With the 1000 bootstrap replicates, the GTR+G model[194] and complete deletion were applied on MEGA-X. Coloring represents gene organizations.

3.12 Comparative Interpretation of Molecular Evolution Results with Experimental Results

LAB species are commonly utilized in the fermentation of foods. To defend themselves from the acidified food environment, LABs have several metabolic pathways, and they make various compounds by using the nutrients in the environment through these metabolic pathways[42, 43]. The GAD system is one of the acid-resistant metabolic pathways that synthesize the GABA molecule, essential for human health (Figure 1.9). In recent years, species that produce GABA and their ability to synthesize GABA molecules in fermented foods have been widely studied. Among the LAB species reported in the literature, *L .brevis* is the one that can be found in various fermented food environments and has a high GABA production potential[125]. *L. brevis* contains all of the genes belonging to the GAD system in its genome as an operon. Moreover, in addition to the GAD operon (*gadR-gadC-gad1*), *L. brevis* contains a second *gad* gene (*gad2*) synthesizing the GAD enzyme[128]. *L. plantarum* is another LAB species discovered in diverse food environments and known to produce GABA. In contrast to *L. brevis*, no GAD operon structure was reported in the *L. plantarum* genome. Only the *gadB* gene, which encodes the GAD enzyme, has been identified in *L. plantarum* strains[127]. There are many experimental studies about *L. brevis* and *L. plantarum* GAD system proteins. These studies are mainly based on isolating strains with the potential to produce GABA and the determining pH ranges in which GAD enzymes are most active. However, there is no *in silico* study examining the population genetics of intraspecific and interspecific relationships of GAD system genes by using diverse strain data.

One of the studies discovered that the GAD operon (*gadC+gad*) in *L. brevis* is acid-sensitive[125]. They came to the conclusion that the *L. brevis gadC* and *gad* gene within the operon are both required for acid resistance in bacteria. However, they said that the *L. brevis gad* gene outside the operon has only a minimal role in the acid resistance according to the experiments done with the GAD system mutants[125]. Based on the experimental results, we expected that the *gadC* and *gad1* genes, which form the *L. brevis* GAD operon structure, would be more conserved than the *L.brevis gad2* gene. The thesis's polymorphism analysis revealed that the *L. brevis gad2* ($\pi = 31.9 \times 10^{-4}$) gene has more nucleotide diversity than the *L. brevis gadC* ($\pi = 9 \times 10^{-4}$) and *gad1* ($\pi = 9.7 \times 10^{-4}$) genes. Moreover, the CBI values of *L. brevis gadC* (0.33) and *gad1* (0.39) genes were higher

than the *L. brevis gad2* (0.30) gene. This result indicates that more preferred codons are used in *L .brevis gadC* and *gad1* genes compared to the *L.brevis gad2* gene. Besides, when the expression levels(CAI) of the *L. brevis* GAD system genes were compared, it was found that the estimated expression levels of the *L. brevis gadC* (0.79) and *gad1* (0.83) genes were higher than the *L. brevis gad2* (0.77) gene (Table 3.21). The results can give a similar conclusion with the state that the genes of *L. brevis gad1* and *gadC* are essential for acid resistance because more conservation and expression levels indicate more activity in the cell.

According to the results of enzyme activity tests conducted in many studies, it is known that *L. brevis* Gad2 and *L. plantarum* GadB proteins exhibit similar properties[127]. In addition, in many features such as protein sequence similarity, enzyme activity, and thermostability, the *L. brevis gad2* gene is closer to the *L. plantarum gadB* gene than the *L. brevis gad1* gene. A similar trend was observed in the phylogenetic studies conducted in the thesis. While the amino acid sequence similarity of the *L. brevis gad2* gene with the *L. plantarum gadB* gene is 82 percent, it is around 52 percent with the *L. brevis gad1* gene. Furthermore, as seen in the GAD enzyme gene tree, the *L. brevis gad2* gene is located in a different clade close to the *L. plantarum gadB* gene, away from the *L.brevis gad1* gene (Figure 3.17). These results reveal the approach that the *L. brevis gad2* gene may have been acquired by gene transfer from a different LAB species.

In the literature, studies conducted with *L. plantarum* GadB and *L. brevis gad* gene (outside the GAD operon) product have found that the C-terminal regions of these proteins are pH-dependent and adjust the catalytic activity of the enzyme according to the pH change[136]. Moreover, it is stated in the articles that the C-terminal part forms the C-plug structure and curls towards the active site of the GAD enzyme[195]. In the thesis's analyses, there was no RP that would cause any amino acid change in the regions corresponding to the protein C-terminal part of the *L. brevis gad2* and *L. plantarum gadB* genes, supporting the experimental results. Indicating that the C-terminal part of the *L. brevis gad2* and *L. plantarum gadB* is essential for the enzyme activity; therefore, the C-terminal region is highly conserved. Moreover, it was observed that the C-terminal part of the predicted three-dimensional structures of *L. brevis* Gad2 and *L. plantarum* GadB proteins curved towards the active site. The predicted three-dimensional structure of the *L. brevis* Gad1 protein, on the other hand, varies from the other two GAD proteins according to its C-terminus. The C-terminal site of the predicted *L. brevis* Gad1 protein

was not found to fold towards the active site as in the *L. brevis* Gad2 protein. In addition, in *L .brevis gad1* gene polymorphism analyses, RPs were found in the region that will form the C-terminal of the protein.

When the N-terminal regions of the three GAD enzymes were examined, it was discovered that the nucleotides that would form the N-terminal site of the *L. brevis* Gad1 protein were highly conserved since no RPs found in this region. These results support previous reports that suggest a role for N-terminal residues in appropriate quaternary structure formation.

As discovered in previous studies, the states of *L. brevis* and *L. plantarum* GAD enzymes in the bacterial cell differ. Hiraga and his colleagues[134] indicated that the *L. brevis* Gad1 enzyme is active in the tetramer state, unlike the other two GAD enzymes. Additionally, it is stated in the article that the N-terminus of the enzyme is responsible for the formation of the tetramer state. A similar quaternary structure is not reported for *L. plantarum* GadB and *L.brevis* Gad2 enzymes. The states in which *L. plantarum* GadB and *L. brevis* Gad2 enzymes are active in the bacterial cell are in dimer[136] and monomer[135] forms, respectively. Sliding window analyses also show that there were many RP changes in the N-terminus of *L. brevis* Gad2 and *L. plantarum* GadB proteins (Figure 3.13).

The antiporter gene *gadC* is not found in ant of the LAB species studies so far, and the three-dimensional structure of GadC protein isolated from LAB is not yet solved. However, information on the structure of the *E. coli* GadC protein, which is not a LAB species, is available in the literature. According to Ma et al.[140], *E. coli* GadC protein has a C-plug structure that changes conformationally according to pH. It was thought that the C-terminus part of the *L. brevis* GadC antiporter protein examined in the thesis might have a similar function as the *E. coli* GadC C-plug. *L. brevis gadC* gene polymorphism analyses supported the above hypothesis by revealing that the gene region forming the C-terminal section was notably conserved in all *L. brevis* strains (Figure 3.14).

LABs used in the production of fermented products are first added in small amounts to food media, regardless of industrial or traditional production. LAB species added to a nutrient-rich environment dominate the fermented food niche by inhibiting the proliferation of other microorganisms with the lactic acid they secrete[40]. In this way, they can increase their numbers rapidly in the nutrient-rich fermented medium. As noted in the literature, rare alleles are observed, indicating negative[196] TD and Fu/Li(D* and F*) values in the genes of bacterial populations that overgrow in nutrient-rich environments

after a bottleneck or selective sweeping[197]. The TD and Fu/Li analyses conducted in the thesis were negative for *L. brevis* and *L. plantarum* GAD system genes, in line with the approach that negative neutrality statistics obtained from sudden overgrowth bacterial populations in a nutrient-rich environment.

Many reports[198] suggest that gene trees often do not reflect the true interspecies relationships since a gene tree is a phylogenetic network that shows the evolution of a gene not only through nucleotide changes but also through various processes such as duplication, gene extinction, or horizontal gene transfer[199]. The species tree and GAD system gene trees created for LAB species exhibited different topologies from each other. In this thesis, the *gad* (Figure 3.17) and *gadC* (Figure 3.18) gene trees showed a topology that supports the hypothesis that species have the same GAD operon organization groups in the same clade in the gene tree unlike the 16S rRNA species tree (Figure 3.15). Species that contain the *gltX* gene in addition to the GAD operon were grouped in different clades of gene trees from those that do not contain the *gltX* gene in the GAD gene cluster. It has also been demonstrated that the *L. brevis gad1* and *gad2* genes are located in distant phylogenetic branches. The results suggest the process of horizontal gene transfer might have a role in *L .brevis* gaining the second gad gene.

CHAPTER 4

CONCLUSION

The GAD system is an efficient mechanism that produces the GABA molecule, a bioactive compound. With the discovery of the therapeutic benefits of the GABA molecule on human health, the demand for it and interest in GABA-containing functional foods is expanding. Since the production of GABA in functional fermented foods depends on the GAD activity of the LAB species, it is crucial to identify the LAB species that contain the GAD metabolic pathway and understand the biochemical properties of enzymes that perform GABA biosynthesis. Examination of the working principle of the GAD system of LAB species and the collection of biochemical and genomic data on the system proteins will facilitate the selection of suitable LAB strains in the production of GABA-enriched foods. Although there are many studies on increasing GABA production in fermented foods, there is still a lack of information in the literature about LAB species that produce GABA, regulation of the GAD system, structures, and similarities of system proteins. The majority of studies are focused on increasing the expression of GAD system genes, adding products containing the GAD enzyme's substrate L-Glu into fermented food media, and improving the activity of the GAD enzyme through protein engineering and gene mutation studies. Nowadays, genome information of many GAD pathway active LAB species is readily accessible from several databases, allowing cost-effective and time-efficient bioinformatics analyses.

In this thesis, GAD system genes of two LAB species (*L. brevis* and *L. plantarum*) with high GABA production potential were investigated by in silico analyses. Many evolutionary population genetic analyses have been performed using the nucleotide and amino acid sequences of the GAD system genes belonging to the *L. brevis* and *L .plantarum* strains' genomes. Intraspecies polymorphism analyses revealed that the genes with the highest nucleotide diversity were *L. brevis gad2* and *L. plantarum gadB*, which did not form an operon structure. When all GAD pathway genes were evaluated together, it was found that the nucleotide diversity of *L. brevis* and *L. plantarum* strains isolated

from dairy products was higher than in other food environments. Moreover, although the isolation environments of *L. brevis* and *L. plantarum* strains are quite diverse, intraspecies neutrality statistics tests indicate that the GAD pathway genes are highly conserved. Neutrality tests showed that changes in the GAD pathway genes are rare alleles, and mutations are mostly found in the external branches of the phylogenetic gene trees. On the other hand, It was observed many times in the thesis that the two gad genes of the *L. brevis* are quite different in terms of nucleotide and proteins sequence similarity, structural, and activity prediction analysis results. Considering much higher amino acid sequence similarity between *L. brevis gad2* and *L. plantarum gadB* compared to the *L. brevis gad1* gene, a horizontal gene transfer event is concluded to be more plausible compared to a gene duplication event in the *L. brevis* genome. Interspecies evolutionary and phylogenetic analyzes with many LAB species also revealed the differences between the two *gad* genes of *L. brevis*. In addition, species that may have had gene transfer between each other in terms of GAD system genes were determined by interspecies analyses and phylogenetic trees. These results might facilitate the selection of LAB species with similar GAD system proteins to produce fermented foods in future studies.

With the new machine learning algorithms that are constantly developing, the protein's three-dimensional structure can be predicted from the protein's amino acid sequence with very close to experimental structures. In the thesis, we also used several protein structure prediction tools to gather more information about the GAD pathway proteins' secondary structures of *L. brevis* and *L. plantarum* and also to understand how GAD pathway proteins fold in the bacterial cell. We examined how the mutations determined from intraspecies polymorphism analyses are distributed, especially around the active site of the GAD enzymes. Since there is very little information about the structure of LAB species' GAD system proteins in the literature, all the information about the structures of GAD pathway proteins becomes important. Protein structure prediction analyses have built similar structures for *L. brevis* and *L. plantarum* GAD enzymes. However, the similarity of *L. brevis* Gad2 protein to *L. plantarum* GadB protein more than the *L. brevis* Gad1 protein is also remarkable in the prediction results. Moreover, the predicted structure of the YjeM protein, which is thought to be part of the GAD system in the *L. plantarum* genome, was found to be similar to the predicted structure of *L. brevis* GadC transporter protein. However, It is necessary to conduct experimental mutant studies with the *L. plantarum yjeM* gene to understand whether the *yjeM* gene is needed

or not for the transportation of L-Glu and GABA molecules without jumping to a direct conclusion.

As explained in detail in Chapter 3, the data obtained from in silico analyses support the experimental results of *L. brevis* and *L. plantarum* GAD enzymes. As stated in many studies, the C and N termini of *L. brevis* and *L. plantarum* GAD enzymes have an important role in adjusting the pH at which the enzyme is active and forming a quaternary enzyme structure. The analyses performed with diverse strains of *L. brevis* and *L. plantarum* genomic data revealed that the experimentally emphasized regions are well conserved in the bacterial genomes. Furthermore, we showed that the structure predictions of the GAD enzyme and L-Glu/GABA antiporter give supporting information that agreed with their experimentally discovered structures.

This study has the potential to provide detailed information to the literature as it is the first study to examine the *L. brevis* and *L. plantarum* acid-resistant GAD pathway genes with molecular and evolutionary analyses. Since all the genomic data of the GAD system are collected through databases, the situation that limits the study may be the sample size of the gene data of LAB strains taken from some isolation environments. In future studies, carrying out *in silico* analyses together with experimental studies, for instance, sequencing the genomes of LAB strains collected from a specific environment, and examining their genomes with molecular evolutionary tests in terms of acid-resistant systems, will provide much more detailed and precise information. In addition to the lack of information about LAB species' GAD system, there is also a significant information gap in the literature about other acid-resistant systems found in LABs. Studies with these systems will be beneficial in the future, as other acid-regulating metabolic pathways in LABs may affect the content and quality of the food environments in which LABs are found.

REFERENCES

1. Von Wright, A.; Axelsson, L., Lactic Acid Bacteria: An Introduction. In *Lactic Acid Bacteria*, CRC Press: 2019; pp 1-16.

2. Sheeladevi, A.; Ramanathan, N., Lactic Acid Production Using Lactic Acid Bacteria Under Optimized Conditions. *Int. J. Pharm. Biol. Arch* **2011,** *2* (6).

3. Mokoena, M. P., Lactic Acid Bacteria And Their Bacteriocins: Classification, Biosynthesis And Applications Against Uropathogens: A Mini-Review. *Molecules* **2017,** *22* (8), 1255.

4. Salminen, S.; Bouley, C.; Boutron, M.-C.; Cummings, J.; Franck, A.; Gibson, G.; Isolauri, E.; Moreau, M.-C.; Roberfroid, M.; Rowland, I., Functional Food Science And Gastrointestinal Physiology And Function. *British Journal Of Nutrition* **1998,** *80* (S1), S147-S171.

5. Prückler, M.; Lorenz, C.; Endo, A.; Kraler, M.; Dürrschmid, K.; Hendriks, K.; da Silva, F. S.; Auterith, E.; Kneifel, W.; Michlmayr, H., Comparison Of Homo- And Heterofermentative Lactic Acid Bacteria For Implementation Of Fermented Wheat Bran In Bread. *Food Microbiology* **2015,** *49*, 211-219.

6. Ayivi, R. D.; Gyawali, R.; Krastanov, A.; Aljaloud, S. O.; Worku, M.; Tahergorabi, R.; Silva, R. C. d.; Ibrahim, S. A., Lactic Acid Bacteria: Food Safety And Human Health Applications. *Dairy* **2020,** *1* (3), 202-232.

7. Higuchi, M.; Yamamoto, Y.; Kamio, Y., Molecular Biology Of Oxygen Tolerance In Lactic Acid Bacteria: Functions Of NADH Oxidases And Dpr In Oxidative Stress. *Journal Of Bioscience And Bioengineering* **2000,** *90* (5), 484-493.

8. Kaban, G.; Kaya, M., Identification Of Lactic Acid Bacteria And Gram-Positive Catalase-Positive Cocci Isolated From Naturally Fermented Sausage (Sucuk). *Journal Of Food Science* **2008,** *73* (8), M385-M388.

9. Quinto, E. J.; Jiménez, P.; Caro, I.; Tejero, J.; Mateo, J.; Girbés, T., Probiotic Lactic Acid Bacteria: A Review. *Food And Nutrition Sciences* **2014,** *5* (18), 1765.

10. Liu, W.; Bao, Q.; Qing, M.; Chen, X.; Sun, T.; Li, M.; Zhang, J.; Yu, J.; Bilige, M.; Sun, T., Isolation And Identification Of Lactic Acid Bacteria From Tarag In Eastern Inner Mongolia Of China By 16S rRNA Sequences And DGGE Analysis. *Microbiological Research* **2012,** *167* (2), 110-115.

11. Sharma, A.; Kaur, J.; Lee, S.; Park, Y.-S., Tracking Of Intentionally Inoculated Lactic Acid Bacteria Strains In Yogurt And Probiotic Powder. *Microorganisms* **2019,** *8* (1), 5.

12. Mohania, D.; Nagpal, R.; Kumar, M.; Bhardwaj, A.; Yadav, M.; Jain, S.; Marotta, F.; Singh, V.; Parkash, O.; Yadav, H., Molecular Approaches For Identification And Characterization Of Lactic Acid Bacteria. *Journal Of Digestive Diseases* **2008,** *9* (4), 190-198.

13. George, F.; Daniel, C.; Thomas, M.; Singer, E.; Guilbaud, A.; Tessier, F. J.; Revol-Junelles, A.-M.; Borges, F.; Foligné, B., Occurrence And Dynamism Of Lactic Acid Bacteria In Distinct Ecological Niches: A Multifaceted Functional Health Perspective. *Frontiers In Microbiology* **2018,** *9*, 2899.

14. Stiles, M. E.; Holzapfel, W. H., Lactic Acid Bacteria Of Foods And Their Current Taxonomy. *International Journal Of Food Microbiology* **1997,** *36* (1), 1-29.

15. Vaughan, E. E.; Heilig, H. G.; Ben-Amor, K.; De Vos, W. M., Diversity, Vitality And Activities Of Intestinal Lactic Acid Bacteria And Bifidobacteria Assessed By Molecular Approaches. *FEMS Microbiology Reviews* **2005,** *29* (3), 477-490.

16. Salvetti, E.; Torriani, S.; Felis, G. E., The Genus Lactobacillus: A Taxonomic Update. *Probiotics And Antimicrobial Proteins* **2012,** *4* (4), 217-226.

17. Claesson, M. J.; Van Sinderen, D.; O'Toole, P. W., The Genus Lactobacillus—A Genomic Basis For Understanding Its Diversity. *FEMS Microbiology Letters* **2007,** *269* (1), 22-28.

18. Felis, G. E.; Dellaglio, F., Taxonomy Of Lactobacilli And Bifidobacteria. *Current Issues In Intestinal Microbiology* **2007,** *8* (2), 44.

19. Holzapfel, W. H.; Wood, B. J., *Lactic Acid Bacteria: Biodiversity And Taxonomy*. John Wiley & Sons: 2014.

20. Salvetti, E.; Harris, H. M.; Felis, G. E.; O'Toole, P. W., Comparative Genomics Of The Genus Lactobacillus Reveals Robust Phylogroups That Provide The Basis

For Reclassification. *Applied And Environmental Microbiology* **2018,** *84* (17), e00993-18.

21. Inglin, R. C.; Meile, L.; Stevens, M. J., Clustering Of Pan-And Core-Genome Of Lactobacillus Provides Novel Evolutionary Insights For Differentiation. *BMC Genomics* **2018,** *19* (1), 1-15.

22. Zheng, J.; Wittouck, S.; Salvetti, E.; Franz, C. M.; Harris, H. M.; Mattarelli, P.; O'toole, P. W.; Pot, B.; Vandamme, P.; Walter, J., A Taxonomic Note On The Genus Lactobacillus: Description Of 23 Novel Genera, Emended Description Of The Genus Lactobacillus Beijerinck 1901, And Union Of Lactobacillaceae And Leuconostocaceae. *International Journal Of Systematic And Evolutionary Microbiology* **2020,** *70* (4), 2782-2858.

23. McAuliffe, O., Symposium Review: *Lactococcus lactis* From Nondairy Sources: Their Genetic And Metabolic Diversity And Potential Applications In Cheese. *Journal Of Dairy Science* **2018,** *101* (4), 3597-3610.

24. Chen, Y. S.; Yanagida, F.; Shinohara, T., Isolation And Identification Of Lactic Acid Bacteria From Soil Using An Enrichment Procedure. *Letters In Applied Microbiology* **2005,** *40* (3), 195-200.

25. Fhoula, I.; Najjari, A.; Turki, Y.; Jaballah, S.; Boudabous, A.; Ouzari, H., Diversity And Antimicrobial Properties Of Lactic Acid Bacteria Isolated From Rhizosphere Of Olive Trees And Desert Truffles Of Tunisia. *BioMed Research International* **2013,** *2013*.

26. Sirichoat, A.; Flórez, A. B.; Vázquez, L.; Buppasiri, P.; Panya, M.; Lulitanond, V.; Mayo, B., Antibiotic Susceptibility Profiles Of Lactic Acid Bacteria From The Human Vagina And Genetic Basis Of Acquired Resistances. *International Journal Of Molecular Sciences* **2020,** *21* (7), 2594.

27. Silvester, M.; Dicks, L., Identification Of Lactic Acid Bacteria Isolated From Human Vaginal Secretions. *Antonie Van Leeuwenhoek* **2003,** *83* (2), 117-123.

28. Borges, S.; Silva, J.; Teixeira, P., The Role Of Lactobacilli And Probiotics In Maintaining Vaginal Health. *Archives Of Gynecology And Obstetrics* **2014,** *289* (3), 479-489.

29. O'Hanlon, D. E.; Moench, T. R.; Cone, R. A., In Vaginal Fluid, Bacteria Associated With Bacterial Vaginosis Can Be Suppressed With Lactic Acid But Not Hydrogen Peroxide. *BMC Infectious Diseases* **2011,** *11* (1), 1-8.

30. Jin, L.; Tao, L.; Pavlova, S.; So, J. S.; Kiwanuka, N.; Namukwaya, Z.; Saberbein, B.; Wawer, M., Species Diversity And Relative Abundance Of Vaginal Lactic Acid Bacteria From Women In Uganda And Korea. *Journal Of Applied Microbiology* **2007,** *102* (4), 1107-1115.

31. Ling, Z.; Kong, J.; Liu, F.; Zhu, H.; Chen, X.; Wang, Y.; Li, L.; Nelson, K. E.; Xia, Y.; Xiang, C., Molecular Analysis Of The Diversity Of Vaginal Microbiota Associated With Bacterial Vaginosis. *BMC Genomics* **2010,** *11* (1), 1-16.

32. Conway, P.; Gorbach, S.; Goldin, B., Survival Of Lactic Acid Bacteria In The Human Stomach And Adhesion To Intestinal Cells. *Journal Of Dairy Science* **1987,** *70* (1), 1-12.

33. Hove, H.; Nørgaard, H.; Brøbech Mortensen, P., Lactic Acid Bacteria And The Human Gastrointestinal Tract. *European Journal Of Clinical Nutrition* **1999,** *53* (5), 339-350.

34. Al Jassim, R. A.; Scott, P. T.; Trebbin, A. L.; Trott, D.; Pollitt, C. C., The Genetic Diversity Of Lactic Acid Producing Bacteria In The Equine Gastrointestinal Tract. *FEMS Microbiology Letters* **2005,** *248* (1), 75-81.

35. Gollwitzer, E. S.; Saglani, S.; Trompette, A.; Yadava, K.; Sherburn, R.; McCoy, K. D.; Nicod, L. P.; Lloyd, C. M.; Marsland, B. J., Lung Microbiota Promotes Tolerance To Allergens In Neonates Via PD-L1. *Nature Medicine* **2014,** *20* (6), 642-647.

36. Turturice, B. A.; McGee, H. S.; Oliver, B.; Baraket, M.; Nguyen, B. T.; Ascoli, C.; Ranjan, R.; Rani, A.; Perkins, D. L.; Finn, P. W., Atopic Asthmatic Immune Phenotypes Associated With Airway Microbiota And Airway Obstruction. *PLoS One* **2017,** *12* (10), e0184566.

37. Remot, A.; Descamps, D.; Noordine, M.-L.; Boukadiri, A.; Mathieu, E.; Robert, V.; Riffault, S.; Lambrecht, B.; Langella, P.; Hammad, H., Bacteria Isolated From Lung Modulate Asthma Susceptibility In Mice. *The ISME Journal* **2017,** *11* (5), 1061-1074.

38. Harata, G.; He, F.; Hiruta, N.; Kawase, M.; Kubota, A.; Hiramatsu, M.; Yausi, H., Intranasal Administration Of *Lactobacillus rhamnosus GG* Protects Mice From H1N1 Influenza Virus Infection By Regulating Respiratory Immune Responses. *Letters In Applied Microbiology* **2010,** *50* (6), 597-602.

39. Nuraida, L., A Review: Health Promoting Lactic Acid Bacteria In Traditional Indonesian Fermented Foods. *Food Science And Human Wellness* **2015,** *4* (2), 47-55.

40. Lamont, J. R.; Wilkins, O.; Bywater-Ekegärd, M.; Smith, D. L., From Yogurt To Yield: Potential Applications Of Lactic Acid Bacteria In Plant Production. *Soil Biology And Biochemistry* **2017,** *111*, 1-9.

41. Molina, V.; Médici, M.; de Valdez, G. F.; Taranto, M. P., Soybean-Based Functional Food With Vitamin B12-Producing Lactic Acid Bacteria. *Journal Of Functional Foods* **2012,** *4* (4), 831-836.

42. Yu, J.; Wang, W.; Menghe, B.; Jiri, M.; Wang, H.; Liu, W.; Bao, Q.; Lu, Q.; Zhang, J.; Wang, F., Diversity Of Lactic Acid Bacteria Associated With Traditional Fermented Dairy Products In Mongolia. *Journal Of Dairy Science* **2011,** *94* (7), 3229-3241.

43. Di Cagno, R.; Coda, R.; De Angelis, M.; Gobbetti, M., Exploitation Of Vegetables And Fruits Through Lactic Acid Fermentation. *Food Microbiology* **2013,** *33* (1), 1-10.

44. Steinkraus, K. H., Lactic Acid Fermentation In The Production Of Foods From Vegetables, Cereals And Legumes. *Antonie Van Leeuwenhoek* **1983,** *49* (3), 337-348.

45. McKay, L. L.; Baldwin, K. A., Applications For Biotechnology: Present And Future Improvements In Lactic Acid Bacteria. *FEMS Microbiology Reviews* **1990,** *7* (1-2), 3-14.

46. De Vuyst, L.; Leroy, F., Bacteriocins From Lactic Acid Bacteria: Production, Purification, And Food Applications. *Microbial Physiology* **2007,** *13* (4), 194-199.

47. O'sullivan, L.; Ross, R.; Hill, C., Potential Of Bacteriocin-Producing Lactic Acid Bacteria For Improvements In Food Safety And Quality. *Biochimie* **2002,** *84* (5-6), 593-604.

48. Samelis, J.; Maurogenakis, F.; Metaxopoulos, J., Characterisation Of Lactic Acid Bacteria Isolated From Naturally Fermented Greek Dry Salami. *International Journal Of Food Microbiology* **1994,** *23* (2), 179-196.

49.   Temmerman, R.; Huys, G.; Swings, J., Identification Of Lactic Acid Bacteria: Culture-Dependent And Culture-Independent Methods. *Trends In Food Science & Technology* **2004,** *15* (7-8), 348-359.

50.   Kim, B.; Mun, E.-G.; Kim, D.; Kim, Y.; Park, Y.; Lee, H.-J.; Cha, Y.-S., A Survey Of Research Papers On The Health Benefits Of Kimchi And Kimchi Lactic Acid Bacteria. *Journal Of Nutrition And Health* **2018,** *51* (1), 1-13.

51.   Tokatlı, M.; Gülgör, G.; Bağder Elmacı, S.; Arslankoz İşleyen, N.; Özçelik, F., In Vitro Properties Of Potential Probiotic Indigenous Lactic Acid Bacteria Originating From Traditional Pickles. *BioMed Research International* **2015,** *2015*.

52.   Wilkinson, M. G.; LaPointe, G., Invited Review: Starter Lactic Acid Bacteria Survival In Cheese: New Perspectives On Cheese Microbiology. *Journal Of Dairy Science* **2020,** *103* (12), 10963-10985.

53.   Mainville, I.; Robert, N.; Lee, B.; Farnworth, E. R., Polyphasic Characterization Of The Lactic Acid Bacteria In Kefir. *Systematic And Applied Microbiology* **2006,** *29* (1), 59-68.

54.   Muyanja, C.; Narvhus, J. A.; Treimo, J.; Langsrud, T., Isolation, Characterisation And Identification Of Lactic Acid Bacteria From Bushera: A Ugandan Traditional Fermented Beverage. *International Journal Of Food Microbiology* **2003,** *80* (3), 201-210.

55.   Kavitake, D.; Kandasamy, S.; Devi, P. B.; Shetty, P. H., Recent Developments On Encapsulation Of Lactic Acid Bacteria As Potential Starter Culture In Fermented Foods–A Review. *Food Bioscience* **2018,** *21*, 34-44.

56.   Ouoba, L.; Nyanga-Koumou, C.; Parkouda, C.; Sawadogo, H.; Kobawila, S.; Keleke, S.; Diawara, B.; Louembe, D.; Sutherland, J., Genotypic Diversity Of Lactic Acid Bacteria Isolated From African Traditional Alkaline-Fermented Foods. *Journal Of Applied Microbiology* **2010,** *108* (6), 2019-2029.

57.   van Hylckama Vlieg, J.; Hugenholtz, J., Mining Natural Diversity Of Lactic Acid Bacteria For Flavour And Health Benefits. *International Dairy Journal* **2007,** *17* (11), 1290-1297.

58.   Phan, Y. T. N.; Tang, M. T.; Tran, T. T. M.; Nguyen, T. H.; Tsuruta, T.; Nishino, N., Diversity Of Lactic Acid Bacteria In Vegetable-Based And Meat-Based Fermented Foods Produced In The Central Region Of Vietnam. *AIMS Microbiology* **2017,** *3* (1), 61.

59. Plavec, T. V.; Berlec, A., Safety Aspects Of Genetically Modified Lactic Acid Bacteria. *Microorganisms* **2020,** *8* (2), 297.

60. Azaïs-Braesco, V.; Bresson, J.; Guarner, F.; Corthier, G., Not All Lactic Acid Bacteria Are Probiotics,… But Some Are. *British Journal of Nutrition* **2010,** *103* (7), 1079-1081.

61. Sharma, R.; Sanodiya, B. S.; Bagrodia, D.; Pandey, M.; Sharma, A.; Bisen, P. S., Efficacy And Potential Of Lactic Acid Bacteria Modulating Human Health. *International Journal Of Pharma And Bio Sciences* **2012,** *3* (4), 935-948.

62. Jeong, Y.; Kim, H.; Lee, J. Y.; Won, G.; Choi, S.-I.; Kim, G.-H.; Kang, C.-H., The Antioxidant, Anti-Diabetic, And Anti-Adipogenesis Potential And Probiotic Properties Of Lactic Acid Bacteria Isolated From Human And Fermented Foods. *Fermentation* **2021,** *7* (3), 123.

63. Zapaśnik, A.; Sokołowska, B.; Bryła, M., Role of Lactic Acid Bacteria in Food Preservation and Safety. *Foods* **2022,** *11* (9), 1283.

64. Seo, K.-H.; Kim, D.-H.; Yokoyama, W. H.; Kim, H., Synbiotic Effect Of Whole Grape Seed Flour And Newly Isolated Kefir Lactic Acid Bacteria On Intestinal Microbiota Of Diet-Induced Obese Mice. *Journal Of Agricultural And Food Chemistry* **2020,** *68* (46), 13131-13137.

65. Wu, J.; Zhang, Y.; Ye, L.; Wang, C., The Anti-Cancer Effects And Mechanisms Of Lactic Acid Bacteria Exopolysaccharides In Vitro: A Review. *Carbohydrate Polymers* **2021,** *253*, 117308.

66. Thirabunyanon, M.; Boonprasom, P.; Niamsup, P., Probiotic Potential Of Lactic Acid Bacteria Isolated From Fermented Dairy Milks On Antiproliferation Of Colon Cancer Cells. *Biotechnology Letters* **2009,** *31* (4), 571-576.

67. Bassyouni, R. H.; Abdel-all, W. S.; Abdel-all, M. G. F. S.; Kamel, Z., Characterization Of Lactic Acid Bacteria Isolated From Dairy Products In Egypt As A Probiotic. **2012**.

68. Heyman, M., Effect Of Lactic Acid Bacteria On Diarrheal Diseases. *Journal Of The American College Of Nutrition* **2000,** *19* (sup2), 137S-146S.

69. Maldonado, N. C.; Chiaraviglio, J.; Bru, E.; De Chazal, L.; Santos, V.; Nader-Macías, M., Effect Of Milk Fermented With Lactic Acid Bacteria On Diarrheal Incidence, Growth Performance And Microbiological And Blood Profiles Of

Newborn Dairy Calves. *Probiotics And Antimicrobial Proteins* **2018,** *10* (4), 668-676.

70. Mathur, H.; Beresford, T. P.; Cotter, P. D., Health Benefits Of Lactic Acid Bacteria (LAB) Fermentates. *Nutrients* **2020,** *12* (6), 1679.

71. Song, Y.-O., The Functional Properties Of Kimchi For The Health Benefits. *Food Industry And Nutrition* **2004,** *9* (3), 27-33.

72. Tsai, T.-Y.; Chu, L.-H.; Lee, C.-L.; Pan, T.-M., Atherosclerosis-Preventing Activity Of Lactic Acid Bacteria-Fermented Milk− Soymilk Supplemented With Momordica Charantia. *Journal Of Agricultural And Food Chemistry* **2009,** *57* (5), 2065-2071.

73. Guan, X.; Xu, Q.; Zheng, Y.; Qian, L.; Lin, B., Screening And Characterization Of Lactic Acid Bacterial Strains That Produce Fermented Milk And Reduce Cholesterol Levels. *Brazilian Journal Of Microbiology* **2017,** *48*, 730-739.

74. Cabello-Olmo, M.; Oneca, M.; Torre, P.; Sainz, N.; Moreno-Aliaga, M. J.; Guruceaga, E.; Díaz, J. V.; Encio, I. J.; Barajas, M.; Araña, M., A Fermented Food Product Containing Lactic Acid Bacteria Protects Zdf Rats From The Development Of Type 2 Diabetes. *Nutrients* **2019,** *11* (10), 2530.

75. Cross, M.; Stevenson, L.; Gill, H., Anti-Allergy Properties Of Fermented Foods: An Important Immunoregulatory Mechanism Of Lactic Acid Bacteria? *International Immunopharmacology* **2001,** *1* (5), 891-901.

76. Kwak, S.-H.; Cho, Y.-M.; Noh, G.-M.; Om, A.-S., Cancer Preventive Potential Of Kimchi Lactic Acid Bacteria (*Weissella cibaria, Lactobacillus clantarum*). *Journal Of Cancer Prevention* **2014,** *19* (4), 253.

77. Arena, M. P.; Capozzi, V.; Russo, P.; Drider, D.; Spano, G.; Fiocco, D., Immunobiosis And Probiosis: Antimicrobial Activity Of Lactic Acid Bacteria With A Focus On Their Antiviral And Antifungal Properties. *Applied microbiology and biotechnology* **2018,** *102* (23), 9949-9958.

78. González, L.; Sandoval, H.; Sacristán, N.; Castro, J.; Fresno, J.; Tornadijo, M., Identification Of Lactic Acid Bacteria Isolated From Genestoso Cheese Throughout Ripening And Study Of Their Antimicrobial Activity. *Food Control* **2007,** *18* (6), 716-722.

79. Awad, S.; Ahmed, N.; El Soda, M., Evaluation Of Isolated Starter Lactic Acid Bacteria In Ras Cheese Ripening And Flavour Development. *Food Chemistry* **2007,** *104* (3), 1192-1199.

80. Özogul, F.; Hamed, I., The Importance Of Lactic Acid Bacteria For The Prevention Of Bacterial Growth And Their Biogenic Amines Formation: A Review. *Critical Reviews In Food Science And Nutrition* **2018,** *58* (10), 1660-1670.

81. Spano, G.; Russo, P.; Lonvaud-Funel, A.; Lucas, P.; Alexandre, H.; Grandvalet, C.; Coton, E.; Coton, M.; Barnavon, L.; Bach, B., Biogenic Amines In Fermented Foods. *European Journal Of Clinical Nutrition* **2010,** *64* (3), S95-S100.

82. Elsanhoty, R. M.; Ramadan, M. F., Genetic Screening Of Biogenic Amines Production Capacity From Some Lactic Acid Bacteria Strains. *Food Control* **2016,** *68*, 220-228.

83. Deepika Priyadarshani, W. M.; Rakshit, S. K., Screening Selected Strains Of Probiotic Lactic Acid Bacteria For Their Ability To Produce Biogenic Amines (Histamine And Tyramine). *International Journal Of Food Science & Technology* **2011,** *46* (10), 2062-2069.

84. Barbieri, F.; Montanari, C.; Gardini, F.; Tabanelli, G., Biogenic Amine Production By Lactic Acid Bacteria: A Review. *Foods* **2019,** *8* (1), 17.

85. Lucas, P. M.; Blancato, V. S.; Claisse, O.; Magni, C.; Lolkema, J. S.; Lonvaud-Funel, A., Agmatine Deiminase Pathway Genes In Lactobacillus Brevis Are Linked To The Tyrosine Decarboxylation Operon In A Putative Acid Resistance Locus. *Microbiology* **2007,** *153* (7), 2221-2230.

86. Kajfasz, J. K.; Quivey, R. G., Responses Of Lactic Acid Bacteria To Acid Stress. In *Stress Responses Of Lactic Acid Bacteria*, Springer: 2011; pp 23-53.

87. Serrazanetti, D. I.; Guerzoni, M. E.; Corsetti, A.; Vogel, R., Metabolic Impact And Potential Exploitation Of The Stress Reactions In Lactobacilli. *Food Microbiology* **2009,** *26* (7), 700-711.

88. Wu, C.; Huang, J.; Zhou, R., Progress In Engineering Acid Stress Resistance Of Lactic Acid Bacteria. *Applied Microbiology And Biotechnology* **2014,** *98* (3), 1055-1063.

89. Wu, C.; Zhang, J.; Wang, M.; Du, G.; Chen, J., *Lactobacillus casei* Combats Acid Stress By Maintaining Cell Membrane Functionality. *Journal Of Industrial Microbiology And Biotechnology* **2012,** *39* (7), 1031-1039.

90. Yokota, A.; Amachi, S.; Ishii, S.; Tomita, F., Acid Sensitivity Of A Mutant Of *Lactococcus lactis subsp. lactis C2* with reduced Membrane-Bound ATPase Activity. *Bioscience, Biotechnology, And Biochemistry* **1995,** *59* (10), 2004-2007.

91. Wu, C.; Zhang, J.; Chen, W.; Wang, M.; Du, G.; Chen, J., A Combined Physiological And Proteomic Approach To Reveal Lactic-Acid-Induced Alterations In *Lactobacillus casei Zhang* and its mutant With Enhanced Lactic Acid Tolerance. *Applied Microbiology And Biotechnology* **2012,** *93* (2), 707-722.

92. Wang, C.; Cui, Y.; Qu, X., Mechanisms And Improvement Of Acid Resistance In Lactic Acid Bacteria. *Archives Of Microbiology* **2018,** *200* (2), 195-201.

93. Costantini, A.; Pietroniro, R.; Doria, F.; Pessione, E.; Garcia-Moruno, E., Putrescine Production From Different Amino Acid Precursors By Lactic Acid Bacteria From Wine And Cider. *International Journal Of Food Microbiology* **2013,** *165* (1), 11-17.

94. De Angelis, M.; Mariotti, L.; Rossi, J.; Servili, M.; Fox, P. F.; Rollán, G.; Gobbetti, M., Arginine Catabolism By Sourdough Lactic Acid Bacteria: Purification And Characterization Of The Arginine Deiminase Pathway Enzymes From Lactobacillus Sanfranciscensis CB1. *Applied And Environmental Microbiology* **2002,** *68* (12), 6193-6201.

95. de Nadra, M. C. M., Nitrogen Metabolism In Lactic Acid Bacteria From Fruits. *Communicating Current Research And Educational Topics And Trends In Applied Microbiology* **2007**.

96. Griswold, A. R.; Chen, Y.-Y. M.; Burne, R. A., Analysis Of An Agmatine Deiminase Gene Cluster In *Streptococcus mutans* UA159. *Journal Of Bacteriology* **2004,** *186* (6), 1902-1904.

97. Alberto, M. R.; Manca de Nadra, M. C.; Arena, M. E., Influence Of Phenolic Compounds On The Growth And Arginine Deiminase System In A Wine Lactic Acid Bacterium. *Brazilian Journal of Microbiology* **2012,** *43*, 167-176.

98. Fernández, M.; Zúñiga, M., Amino Acid Catabolic Pathways Of Lactic Acid Bacteria. *Critical Reviews In Microbiology* **2006,** *32* (3), 155-183.

99. Moreno-Arribas, V.; Torlois, S.; Joyeux, A.; Bertrand, A.; Lonvaud-Funel, A., Isolation, Properties And Behaviour Of Tyramine-Producing Lactic Acid Bacteria From Wine. *Journal Of Applied Microbiology* **2000,** *88* (4), 584-593.

100. Lucas, P.; Landete, J.; Coton, M.; Coton, E.; Lonvaud-Funel, A., The Tyrosine Decarboxylase Operon Of *Lactobacillus brevis IOEB 9809*: Characterization And Conservation In Tyramine-Producing Bacteria. *FEMS Microbiology Letters* **2003,** *229* (1), 65-71.

101. Martín, M. C.; Fernández, M.; Linares, D. M.; Alvarez, M. A., Sequencing, Characterization And Transcriptional Analysis Of The Histidine Decarboxylase Operon Of *Lactobacillus buchneri*. *Microbiology* **2005,** *151* (4), 1219-1228.

102. Marcobal, A.; de las RIVAS, B.; Moreno-Arribas, M. V.; Munoz, R., Multiplex PCR Method For The Simultaneous Detection Of Histamine-, Tyramine-, And Putrescine-Producing Lactic Acid Bacteria In Foods. *Journal Of Food Protection* **2005,** *68* (4), 874-878.

103. Romano, A.; Trip, H.; Lonvaud-Funel, A.; Lolkema, J. S.; Lucas, P. M., Evidence Of Two Functionally Distinct Ornithine Decarboxylation Systems In Lactic Acid Bacteria. *Applied And Environmental Microbiology* **2012,** *78* (6), 1953-1961.

104. Romano, A.; Ladero, V.; Alvarez, M. A.; Lucas, P. M., Putrescine Production Via The Ornithine Decarboxylation Pathway Improves The Acid Stress Survival Of *Lactobacillus brevis* And Is Part Of A Horizontally Transferred Acid Resistance Locus. *International Journal Of Food Microbiology* **2014,** *175*, 14-19.

105. Hazards, E. P. o. B., Scientific Opinion On Risk Based Control Of Biogenic Amine Formation In Fermented Foods. *Efsa Journal* **2011,** *9* (10), 2393.

106. Shalaby, A. R., Significance Of Biogenic Amines To Food Safety And Human Health. *Food Research International* **1996,** *29* (7), 675-690.

107. Colombo, F. M.; Cattaneo, P.; Confalonieri, E.; Bernardi, C., Histamine Food Poisonings: A Systematic Review And Meta-Analysis. *Critical Reviews In Food Science And Nutrition* **2018,** *58* (7), 1131-1151.

108. De Biase, D.; Pennacchietti, E., Glutamate Decarboxylase-Dependent Acid Resistance In Orally Acquired Bacteria: Function, Distribution And Biomedical Implications Of The *gadBC* Operon. *Molecular Microbiology* **2012,** *86* (4), 770-786.

109. Feehily, C.; Karatzas, K., Role Of Glutamate Metabolism In Bacterial Responses Towards Acid And Other Stresses. *Journal Of Applied Microbiology* **2013**, *114* (1), 11-24.

110. Magasanik, B., Genetic Control On Nitrogen Assimilation In Bacteria. **1982**.

111. Yang, H.; He, M.; Wu, C., Cross Protection Of Lactic Acid Bacteria During Environmental Stresses: Stress Responses And Underlying Mechanisms. *Lwt* **2021**, *144*, 111203.

112. Cotter, P. D.; Hill, C., Surviving The Acid Test: Responses Of Gram-Positive Bacteria To Low Ph. *Microbiology And Molecular Biology Reviews* **2003,** *67* (3), 429-453.

113. Karatzas, K.-A. G.; Brennan, O.; Heavin, S.; Morrissey, J.; O'Byrne, C. P., Intracellular Accumulation Of High Levels Of γ-Aminobutyrate By *Listeria monocytogenes 10403s* In Response To Low pH: Uncoupling Of γ-Aminobutyrate Synthesis From Efflux In A Chemically Defined Medium. *Applied And Environmental Microbiology* **2010,** *76* (11), 3529-3537.

114. Cotter, P. D.; Gahan, C. G.; Hill, C., A Glutamate Decarboxylase System Protects Listeria Monocytogenes In Gastric Fluid. *Molecular Microbiology* **2001,** *40* (2), 465-475.

115. Cotter, P. D.; Ryan, S.; Gahan, C. G.; Hill, C., Presence Of GadD1 Glutamate Decarboxylase In Selected Listeria Monocytogenes Strains Is Associated With An Ability To Grow At Low Ph. *Applied And Environmental Microbiology* **2005,** *71* (6), 2832-2839.

116. Smith, D.; Kassam, T.; Singh, B.; Elliott, J., *Escherichia coli* has Two Homologous Glutamate Decarboxylase Genes That Map To Distinct Loci. *Journal Of Bacteriology* **1992,** *174* (18), 5820-5826.

117. Ma, Z.; Gong, S.; Richard, H.; Tucker, D. L.; Conway, T.; Foster, J. W., GadE (YhiE) Activates Glutamate Decarboxylase-Dependent Acid Resistance In *Escherichia coli K-12*. *Molecular Microbiology* **2003,** *49* (5), 1309-1320.

118. Yunes, R.; Poluektova, E.; Dyachkova, M.; Klimina, K.; Kovtun, A.; Averina, O.; Orlova, V.; Danilenko, V., GABA Production And Structure Of *gadB/gadC* Genes In Lactobacillus And Bifidobacterium Strains From Human Microbiota. *Anaerobe* **2016,** *42*, 197-204.

119. Redruello, B.; Saidi, Y.; Sampedro, L.; Ladero, V.; Del Rio, B.; Alvarez, M. A., GABA-Producing *Lactococcus lactis* Strains Isolated From Camel's Milk As Starters For The Production Of GABA-Enriched Cheese. *Foods* **2021,** *10* (3), 633.

120. Yang, S.-Y.; Lü, F.-X.; Lu, Z.-X.; Bie, X.-M.; Jiao, Y.; Sun, L.-J.; Yu, B., Production Of γ-Aminobutyric Acid By *Streptococcus salivarius subsp. thermophilus Y2* Under Submerged Fermentation. *Amino Acids* **2008,** *34* (3), 473-478.

121. Wang, Q.; Liu, X.; Fu, J.; Wang, S.; Chen, Y.; Chang, K.; Li, H., Substrate Sustained Release-Based High Efficacy Biosynthesis Of GABA By *Lactobacillus brevis NCL912. Microbial Cell Factories* **2018,** *17* (1), 1-8.

122. Di Cagno, R.; Mazzacane, F.; Rizzello, C. G.; De Angelis, M.; Giuliani, G.; Meloni, M.; De Servi, B.; Gobbetti, M., Synthesis Of Γ-Aminobutyric Acid (GABA) By Lactobacillus Plantarum DSM19463: Functional Grape Must Beverage And Dermatological Applications. *Applied Microbiology And Biotechnology* **2010,** *86* (2), 731-741.

123. Park, K.-B.; Oh, S.-H., Isolation And Characterization Of Lactobacillus Buchneri Strains With High ${\Gamma}$-Aminobutyric Acid Producing Capacity From Naturally Aged Cheese. *Food Science And Biotechnology* **2006,** *15* (1), 86-90.

124. Komatsuzaki, N.; Nakamura, T.; Kimura, T.; Shima, J., Characterization Of Glutamate Decarboxylase From A High γ-aminobutyric Acid (GABA)-Producer, *Lactobacillus paracasei. Bioscience, Biotechnology, And Biochemistry* **2008,** *72* (2), 278-285.

125. Lyu, C.; Zhao, W.; Peng, C.; Hu, S.; Fang, H.; Hua, Y.; Yao, S.; Huang, J.; Mei, L., Exploring The Contributions Of Two Glutamate Decarboxylase Isozymes In *Lactobacillus brevis* To Acid Resistance And γ-Aminobutyric Acid Production. *Microbial Cell Factories* **2018,** *17* (1), 1-14.

126. Cui, Y.; Miao, K.; Niyaphorn, S.; Qu, X., Production Of Gamma-Aminobutyric Acid From Lactic Acid Bacteria: A Systematic Review. *International Journal Of Molecular Sciences* **2020,** *21* (3), 995.

127. Wu, Q.; Tun, H. M.; Law, Y.-S.; Khafipour, E.; Shah, N. P., Common Distribution Of Gad Operon In *Lactobacillus brevis* And Its GadA Contributes To Efficient GABA Synthesis Toward Cytosolic Near-Neutral pH. *Frontiers In Microbiology* **2017,** *8*, 206.

128. Wu, Q.; Shah, N. P., High γ-Aminobutyric Acid Production From Lactic Acid Bacteria: Emphasis On *Lactobacillus brevis* As A Functional Dairy Starter. *Critical Reviews In Food Science And Nutrition* **2017,** *57* (17), 3661-3672.

129. Lu, P.; Ma, D.; Chen, Y.; Guo, Y.; Chen, G.-Q.; Deng, H.; Shi, Y., L-Glutamine Provides Acid Resistance For *Escherichia coli* Through Enzymatic Release Of Ammonia. *Cell Research* **2013,** *23* (5), 635-644.

130. Lim, H. S.; Seo, D.-H.; Cha, I.-T.; Lee, H.; Nam, Y.-D.; Seo, M.-J., Expression And Characterization Of Glutamate Decarboxylase From *Lactobacillus brevis HYE1* Isolated From Kimchi. *World Journal Of Microbiology And Biotechnology* **2018,** *34* (3), 1-10.

131. Sandmeier, E.; Hale, T. I.; Christen, P., Multiple Evolutionary Origin Of Pyridoxal-5′-Phosphate-Dependent Amino Acid Decarboxylases. *European Journal Of Biochemistry* **1994,** *221* (3), 997-1002.

132. Yogeswara, I. B. A.; Maneerat, S.; Haltrich, D., Glutamate Decarboxylase From Lactic Acid Bacteria—A Key Enzyme In GABA Synthesis. *Microorganisms* **2020,** *8* (12), 1923.

133. Huang, J.; Fang, H.; Gai, Z.-C.; Mei, J.-Q.; Li, J.-N.; Hu, S.; Lv, C.-J.; Zhao, W.-R.; Mei, L.-H., *Lactobacillus brevis CGMCC 1306* Glutamate Decarboxylase: Crystal Structure And Functional Analysis. *Biochemical And Biophysical Research Communications* **2018,** *503* (3), 1703-1709.

134. Hiraga, K.; Ueno, Y.; Oda, K., Glutamate Decarboxylase From Lactobacillus Brevis: Activation By Ammonium Sulfate. *Bioscience, Biotechnology, And Biochemistry* **2008**, 0803250825-0803250825.

135. Shi, F.; Xie, Y.; Jiang, J.; Wang, N.; Li, Y.; Wang, X., *Directed evolution and mutagenesis of glutamate decarboxylase from Lactobacillus brevis Lb85 t*o *broaden the range of its activity toward a near-neutral* pH. *Enzyme And Microbial Technology* **2014,** *61*, 35-43.

136. Shin, S.-M.; Kim, H.; Joo, Y.; Lee, S.-J.; Lee, Y.-J.; Lee, S. J.; Lee, D.-W., Characterization Of Glutamate Decarboxylase From *Lactobacillus plantarum* and its C-Terminal Function For The pH Dependence Of Activity. *Journal Of Agricultural And Food Chemistry* **2014,** *62* (50), 12186-12193.

137. Lyu, C.-j.; Zhao, W.-r.; Hu, S.; Huang, J.; Lu, T.; Jin, Z.-h.; Mei, L.-h.; Yao, S.-j., Physiology-Oriented Engineering Strategy To Improve Gamma-Aminobutyrate Production In

Lactobacillus brevis Journal Of Agricultural And Food Chemistry **2017,** *65* (4), 858-866.

138. Gänzle, M. G., Lactic Metabolism Revisited: Metabolism Of Lactic Acid Bacteria In Food Fermentations And Food Spoilage. *Current Opinion In Food Science* **2015,** *2*, 106-117.

139. Li, H.; Li, W.; Liu, X.; Cao, Y., *gadA* Gene Locus In *Lactobacillus brevis NCL912* and its Expression During Fed-Batch Fermentation. *FEMS Microbiology Letters* **2013,** *349* (2), 108-116.

140. Ma, D.; Lu, P.; Yan, C.; Fan, C.; Yin, P.; Wang, J.; Shi, Y., Structure And Mechanism Of A Glutamate–GABA Antiporter. *Nature* **2012,** *483* (7391), 632-636.

141. Hommais, F.; Krin, E.; Coppee, J.-Y.; Lacroix, C.; Yeramian, E.; Danchin, A.; Bertin, P., GadE (YhiE): A Novel Activator Involved In The Response To Acid Environment In *Escherichia coli*. *Microbiology* **2004,** *150* (1), 61-72.

142. Gong, L.; Ren, C.; Xu, Y., Deciphering The Crucial Roles Of Transcriptional Regulator *gadR* On Gamma-Aminobutyric Acid Production And Acid Resistance In *Lactobacillus brevis*. *Microbial Cell Factories* **2019,** *18* (1), 1-12.

143. Lyu, C.-J.; Fei, J.-Y.; Yan, J.-P.; Xu, Q.-C.; Mei, J.-Q.; Yue, H.-Y.; Chen, L.-L.; Mei, L.-H.; Huang, J.; Yao, S.-J., Improvement Of γ-Aminobutyrate Biosynthesis By Genetically Engineered *Lactococcus lactis*. *Biochemical Engineering Journal* **2020,** *157*, 107525.

144. Sarasa, S. B.; Mahendran, R.; Muthusamy, G.; Thankappan, B.; Selta, D. R. F.; Angayarkanni, J., A Brief Review On The Non-Protein Amino Acid, Gamma-Amino Butyric Acid (GABA): Its Production And Role In Microbes. *Current Microbiology* **2020,** *77* (4), 534-544.

145. Lancien, M.; Roberts, M. R., Regulation Of *Arabidopsis thaliana 14-3-3* gene Expression By γ-Aminobutyric Acid. *Plant, Cell & Environment* **2006,** *29* (7), 1430-1436.

146. Decavel, C.; Van Den Pol, A. N., GABA: A Dominant Neurotransmitter In The Hypothalamus. *Journal Of Comparative Neurology* **1990,** *302* (4), 1019-1037.

147. Chuang, C.-Y.; Shi, Y.-C.; You, H.-P.; Lo, Y.-H.; Pan, T.-M., Antidepressant Effect Of GABA-Rich Monascus-Fermented Product On Forced Swimming Rat Model. *Journal Of Agricultural And Food Chemistry* **2011,** *59* (7), 3027-3034.

148. Inoue, K.; Shirai, T.; Ochiai, H.; Kasao, M.; Hayakawa, K.; Kimura, M.; Sansawa, H., Blood-Pressure-Lowering Effect Of A Novel Fermented Milk Containing Γ-Aminobutyric Acid (GABA) In Mild Hypertensives. *European Journal Of Clinical Nutrition* **2003,** *57* (3), 490-495.

149. Marques, T. M.; Patterson, E.; Wall, R.; O'Sullivan, O.; Fitzgerald, G. F.; Cotter, P. D.; Dinan, T. G.; Cryan, J. F.; Ross, R. P.; Stanton, C., Influence Of GABA And GABA-Producing *Lactobacillus brevis DPC 6108* On The Development Of Diabetes In A Streptozotocin Rat Model. *Beneficial Microbes* **2016,** *7* (3), 409-420.

150. Diana, M.; Quílez, J.; Rafecas, M., Gamma-Aminobutyric Acid As A Bioactive Compound In Foods: A Review. *Journal Of Functional Foods* **2014,** *10*, 407-420.

151. Gibney, G.; Baxevanis, A. D., Searching NCBI Databases Using Entrez. *Current Protocols In Bioinformatics* **2011,** *34* (1), 1.3. 1-1.3. 25.

152. Patel, J. B., 16S rRNA Gene Sequencing For Bacterial Pathogen Identification In The Clinical Laboratory. *Molecular Diagnosis* **2001,** *6* (4), 313-321.

153. Okonechnikov, K.; Golosova, O.; Fursov, M.; Team, U., Unipro UGENE: A Unified Bioinformatics Toolkit. *Bioinformatics* **2012,** *28* (8), 1166-1167.

154. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K., MEGA X: Molecular Evolutionary Genetics Analysis Across Computing Platforms. *Molecular Biology And Evolution* **2018,** *35* (6), 1547.

155. Edgar, R. C., MUSCLE: Multiple Sequence Alignment With High Accuracy And High Throughput. *Nucleic Acids Research* **2004,** *32* (5), 1792-1797.

156. Méric, G.; Yahara, K.; Mageiros, L.; Pascoe, B.; Maiden, M. C.; Jolley, K. A.; Sheppard, S. K., A Reference Pan-Genome Approach To Comparative Bacterial Genomics: Identification Of Novel Epidemiological Markers In Pathogenic Campylobacter. *PloS One* **2014,** *9* (3), e92798.

157. Anantharaman, K.; Hausmann, B.; Jungbluth, S. P.; Kantor, R. S.; Lavy, A.; Warren, L. A.; Rappé, M. S.; Pester, M.; Loy, A.; Thomas, B. C., Expanded

Diversity Of Microbial Groups That Shape The Dissimilatory Sulfur Cycle. *The ISME journal* **2018,** *12* (7), 1715-1728.

158. White, J. R.; Navlakha, S.; Nagarajan, N.; Ghodsi, M.-R.; Kingsford, C.; Pop, M., Alignment And Clustering Of Phylogenetic Markers-Implications For Microbial Diversity Studies. *BMC Bioinformatics* **2010,** *11* (1), 1-10.

159. Johnson, J. S.; Spakowicz, D. J.; Hong, B.-Y.; Petersen, L. M.; Demkowicz, P.; Chen, L.; Leopold, S. R.; Hanson, B. M.; Agresta, H. O.; Gerstein, M., Evaluation Of 16S rRNA Gene Sequencing For Species And Strain-Level Microbiome Analysis. *Nature Communications* **2019,** *10* (1), 1-11.

160. Waterhouse, A.; Procter, J.; Martin, D. A.; Barton, G. J., Jalview: Visualization And Analysis Of Molecular Sequences, Alignments, And Structures. *BMC Bioinformatics* **2005,** *6* (3), 1-1.

161. Felsenstein, J., Evolutionary Trees From DNA Sequences: A Maximum Likelihood Approach. *Journal Of Molecular Evolution* **1981,** *17* (6), 368-376.

162. Holmes, S., Bootstrapping Phylogenetic Trees: Theory And Methods. *Statistical Science* **2003,** *18* (2), 241-255.

163. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J. C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S. E.; Sánchez-Gracia, A., DnaSP 6: DNA Sequence Polymorphism Analysis Of Large Data Sets. *Molecular Biology And Evolution* **2017,** *34* (12), 3299-3302.

164. Tajima, F., Statistical Method For Testing The Neutral Mutation Hypothesis By DNA Polymorphism. *Genetics* **1989,** *123* (3), 585-595.

165. Simonsen, K. L.; Churchill, G. A.; Aquadro, C. F., Properties Of Statistical Tests Of Neutrality For DNA Polymorphism Data. *Genetics* **1995,** *141* (1), 413-429.

166. Wang, Z.; Sew, P.-H.; Ambrose, H.; Ryan, S.; Chong, S. S.; Lee, E. J.; Lee, C. G., Nucleotide Sequence Analyses Of the MRP 1 Gene In Four Populations Suggest Negative Selection On Its Coding Region. *BMC Genomics* **2006,** *7* (1), 1-18.

167. Moeller, A. H., Metagenomic Targets Of Balancing Selection In The Human Gut. *Authorea Preprints* **2021**.

168. Fu, Y.-X.; Li, W.-H., Statistical Tests Of Neutrality Of Mutations. *Genetics* **1993,** *133* (3), 693-709.

169. Proutski, V.; Holmes, E., SWAN: Sliding Window Analysis Of Nucleotide Sequence Variability. *Bioinformatics (Oxford, England)* **1998,** *14* (5), 467-468.

170. Rozas, J., DNA Sequence Polymorphism Analysis Using DnaSP. In *Bioinformatics For DNA Sequence Analysis*, Springer: 2009; pp 337-350.

171. Bivand, R., Implementing Spatial Data Analysis Software Tools In R. *Geographical Analysis* **2006,** *38* (1), 23-40.

172. McDonald, J. H.; Kreitman, M., Adaptive Protein Evolution At The Adh Locus In Drosophila. *Nature* **1991,** *351* (6328), 652-654.

173. Charlesworth, J.; Eyre-Walker, A., The Mcdonald–Kreitman Test And Slightly Deleterious Mutations. *Molecular Biology And Evolution* **2008,** *25* (6), 1007-1015.

174. Smith, N. G.; Eyre-Walker, A., Adaptive Protein Evolution In Drosophila. *Nature* **2002,** *415* (6875), 1022-1024.

175. Stoletzki, N.; Eyre-Walker, A., Estimation Of The Neutrality Index. *Molecular Biology And Evolution* **2011,** *28* (1), 63-70.

176. Schabauer, H.; Valle, M.; Pacher, C.; Stockinger, H.; Stamatakis, A.; Robinson-Rechavi, M.; Yang, Z.; Salamin, N. In *SlimCodeML: An Optimized Version Of Codeml For The Branch-Site Model*, 2012 IEEE 26th International Parallel And Distributed Processing Symposium Workshops & PhD Forum, IEEE: 2012; pp 706-714.

177. Gao, F.; Chen, C.; Arab, D. A.; Du, Z.; He, Y.; Ho, S. Y., EasyCodeML: A Visual Tool For Analysis Of Selection Using CodeML. *Ecology And Evolution* **2019,** *9* (7), 3891-3898.

178. Yang, Z.; Swanson, W. J., Codon-Substitution Models To Detect Adaptive Evolution That Account For Heterogeneous Selective Pressures Among Site Classes. *Molecular Biology And Evolution* **2002,** *19* (1), 49-57.

179. Yang, Z., PAML 4: Phylogenetic Analysis By Maximum Likelihood. *Molecular Biology And Evolution* **2007,** *24* (8), 1586-1591.

180. Xu, B.; Yang, Z., PAMLX: A Graphical User Interface For PAML. *Molecular Biology And Evolution* **2013,** *30* (12), 2723-2724.

181. Vihinen, M., Functional Effects Of Protein Variants. *Biochimie* **2021,** *180*, 104-120.

182. Xia, X., Protein Isoelectric Point. *Bioinformatics And The Cell: Modern Computational Approaches In Genomics, Proteomics And Transcriptomics* **2007**, 207-219.

183. Choi, Y.; Chan, A. P., PROVEAN Web Server: A Tool To Predict The Functional Effect Of Amino Acid Substitutions And Indels. *Bioinformatics* **2015,** *31* (16), 2745-2747.

184. Bahiri-Elitzur, S.; Tuller, T., Codon-Based Indices For Modeling Gene Expression And Transcript Evolution. *Computational and Structural Biotechnology Journal* **2021,** *19*, 2646-2663.

185. Fox, J. M.; Erill, I., Relative Codon Adaptation: A Generic Codon Bias Index For Prediction Of Gene Expression. *DNA Research* **2010,** *17* (3), 185-196.

186. DeLano, W. L., Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr* **2002,** *40* (1), 82-92.

187. Team, R. C., R: A Language And Environment For Statistical Computing. **2013**.

188. Wickham, H.; Chang, W.; Wickham, M. H., Package 'ggplot2'. *Create Elegant Data Visualisations Using The Grammar Of Graphics. Version* **2016,** *2* (1), 1-189.

189. Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C., SWISS-MODEL: An Automated Protein Homology-Modeling Server. *Nucleic Acids Research* **2003,** *31* (13), 3381-3385.

190. Zhang, Y., I-TASSER Server For Protein 3D Structure Prediction. *BMC Bioinformatics* **2008,** *9* (1), 1-8.

191. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J., The Phyre2 Web Portal For Protein Modeling, Prediction And Analysis. *Nature Protocols* **2015,** *10* (6), 845-858.

192. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A., Highly Accurate Protein Structure Prediction With AlphaFold. *Nature* **2021,** *596* (7873), 583-589.

193. Kimura, M., A Simple Method For Estimating Evolutionary Rates Of Base Substitutions Through Comparative Studies Of Nucleotide Sequences. *Journal Of Molecular Evolution* **1980,** *16* (2), 111-120.

194. Nei, M.; Kumar, S., *Molecular Evolution And Phylogenetics*. Oxford University Press, USA: 2000.

195. Lyu, C.; Yao, L.; Zhu, Q.; Mei, J.; Cao, Y.; Hu, S.; Zhao, W.; Huang, J.; Mei, L.; Yao, S., Reconstruction Of The Glutamate Decarboxylase System In *Lactococcus lactis* For Biosynthesis Of Food-Grade γ-Aminobutyric Acid. *Applied Microbiology And Biotechnology* **2021,** *105* (10), 4127-4140.

196. Hughes, A. L., Evidence For Abundant Slightly Deleterious Polymorphisms In Bacterial Populations. *Genetics* **2005,** *169* (2), 533-538.

197. Sharma, M.; Fomda, B. A.; Mazta, S.; Sehgal, R.; Bagicha Singh, B.; Malla, N., Genetic Diversity And Population Genetic Structure Analysis Of *Echinococcus granulosus* Sensu Stricto Complex Based On Mitochondrial DNA Signature. *PLoS One* **2013,** *8* (12), e82904.

198. Szöllősi, G. J.; Tannier, E.; Daubin, V.; Boussau, B., The Inference Of Gene Trees With Species Trees. *Systematic Biology* **2015,** *64* (1), e42-e62.

199. Boussau, B.; Scornavacca, C., Reconciling Gene Trees With Species Trees. *Phylogenetics In The Genomic Era* **2020**, 3.2: 1--3.2: 23.