

Comparative Genome Analysis of *Lactiplantibacillus paraplantarum*

Ruveyda BENK¹, Fatih ORTAKCI^{2*}

¹Department of Biomedical Engineering, Faculty of Engineering, Erciyes University, Kayseri, TR

²Department of Food Engineering, Faculty of Chemical and Metallurgical Engineering,

Istanbul Technical University, Istanbul, TR

(ORCID: [0000-0002-5831-9431](https://orcid.org/0000-0002-5831-9431)) (ORCID: [0000-0003-1319-0854](https://orcid.org/0000-0003-1319-0854))



Keywords: *Lactiplantibacillus paraplantarum*, comparative genomics, bacteriocin, prophage, carbohydrate, adaptation.

Abstract

Lactiplantibacillus paraplantarum is a lactic acid bacteria species that is associated with food microbiomes and has been found to be either detrimental or beneficial to specific food processes. In this study, an *in-silico* genomic approach was applied using JGI's IMG/MER and PATRIC to compare the genomes of the *L. paraplantarum* DSM10667, L-ZS9, and AS-7 strains to uncover metabolic differences and lifestyle adaptations between these isolates, and better utilizing these species in food bioprocesses. Bacteriocin and prophage screenings were performed using Bagel4 and PHASTER software, respectively. BRIG was used to identify alignments of strains with each other for visual inspection of each genome. KEGG was used to predict putative carbohydrate, pyruvate, and amino acid metabolisms. Genome sizes of DSM10667, L-ZS9, and AS-7 were 3.36, 3.14, and 3.01 Mbp, respectively. Unique genes were found to predict the evolutionary adaptation of strains against their corresponding microniche. For example, the gene encoding arginase was only found in sausage isolate L-ZS9, while the dextran-sucrase-encoding gene was unique to beer contaminant DSM10667. Three strains were predicted to carry the *plnA*EFJ operon for plantaricin biosynthesis, and the AS-7 genome contains leucocin K. Although DSM 10667 harbors four intact prophages, both L-ZS9 and AS-7 carried one prophage region, still showing the plasticity of the genome. Genome analysis predicted that isolation sources might potentially affect the metabolic capabilities of strains as part of the adaptation of the strains to their habitats.

1. Introduction

The genus *Lactobacilli* is a primary comprehensive group of lactic acid bacteria that was isolated from various ecological niches [1]. Before the re-taxonomic structuring of the *Lactobacillus* genus into 23 new genera, *Lactiplantibacillus* species were considered part of the *L. plantarum* group [2]. The *Lactiplantibacillus* genus is composed of 17 species: *Lactiplantibacillus argenteratensis*, *Lactiplantibacillus songbeiensis*, *Lactiplantibacillus dongliensis*, *Lactiplantibacillus daowaiensis*, *Lactiplantibacillus nangangensis*, *Lactiplantibacillus daoliensis*, *Lactiplantibacillus pingfangensis*,

Lactiplantibacillus garii, *Lactiplantibacillus modestisalitolans*, *Lactiplantibacillus plajomi*, *Lactiplantibacillus mudanjiangensis*, *Lactiplantibacillus xiangfangensis*, *Lactiplantibacillus herbarum*, *Lactiplantibacillus fabifermentans*, *Lactiplantibacillus pentosus*, *Lactiplantibacillus plantarum*, and *Lactiplantibacillus paraplantarum*.

L. paraplantarum strains have been isolated from a variety of different fermented foods, such as sourdough [3], Awa-bancha fermented tea [1], cheese [4], Rice Bran Pickles [5], kimchi [6], Tulum cheese [7], breast milk [8], gundruk (i.e., a fermented leafy

*Corresponding author: ortakci@itu.edu.tr

Received: 08.10.2022, Accepted: 09.11.2023

vegetable) [9], sliced ham [10], fermented sausage [11], fruit and vegetables [12], and beer [13]. Given that *L. paraplantarum* strains have been isolated from a variety of different food matrices, only four different strains of *Lactiplantibacillus paraplantarum* complete whole genome sequences are available in the Joint Genome Institute public database as of September 5, 2022, which limits our understanding of niche-specific adaptation strategies that those organisms developed. Those *L. paraplantarum* strains are: i) DSM 10667, isolated from beer as a contaminant; ii) L-ZS9, isolated from fermented sausage, and iii) AS-7, isolated from fruit and vegetables [14].

To avoid confusion between *L. plantarum* and *L. paraplantarum* species, Curk [19] reported that the G+C content of the *L. paraplantarum* DNA is around 44-45% while *L. plantarum* has a G+C content of 41% [19]. *Lactiplantibacillus paraplantarum* is a rod-shaped bacterium that exists in short chains or pairs. *L. paraplantarum* is phenotypically remarkably close to *L. plantarum*, which was initially characterized at the Pasteur Institute by French scientist Curk [19]. *L. paraplantarum* is a facultatively anaerobic, gram-positive, and catalase-negative species that forms creamy-colored, dome-shaped colonies on the MRS agar between pH 5 and pH 7 at 30-37°C. Besides, *L. paraplantarum* strains have been shown to have potential for use in food protection and preservation [15].

Looking into the literature, comparative genome analysis within several different *lactobacillus* species at strain level has been conducted to better elucidate their lifestyle adaptation to the microenvironments they inhabit, and strain level differences have been put forth [16–18]. However, to our knowledge, no comparative genomic studies were performed to understand strain level differences within *L. paraplantarum* strains. Thus, the present study fills an important gap in the literature by performing comparative genome analysis of three different strains of *L. paraplantarum* that were isolated from sausage, beer, or fruits and vegetables. An *in-silico* approach using bioinformatic tools was utilized to predict metabolic and functional characteristics of *L. paraplantarum* DSM 10667, L-ZS9, and AS-7 for unravelling lifestyle adaptations and strain-level genomic differences [19].

2. Material and Method

We performed a comparative genome analysis of three *L. paraplantarum* strains that were isolated from diverse ecosystems such as beer, sausage, fruit, and

vegetables. The genome sequence data for *L. paraplantarum* DSM 10667, L-ZS9, and AS-7 were downloaded from the Joint Genome Institute as shown in Table 1 with IMG genome IDs [14] and IMG/MER [20]. The average Nucleotide Identity of *L. paraplantarum* strains was calculated using the JGI IMG/MER Pairwise ANI tool [20]. The unique genes that are specific to each strain being evaluated were extracted using homologous and non-homologous functions of IMG/MER under the phylogenetic profilers' platform. Genome alignments were conducted on PATRIC services using the MAUVE genome alignment tool.

Visualization of genome alignments was conducted using the BLAST Ring Image Generator (BRIG) [21], including a ring for each genome and a ring for GC content and GC skew. A BLAST type of BLASTn [22] was used with the following options: a lower identity threshold of 70% and an upper identity threshold of 90% with a ring size of 30. Genomic features with low BLAST identity were inspected through visual genomic evaluation. The similarity of the three strains that were compared against each other and other closely related genomes was performed using the orthoANI algorithm in IMG/MER. A phylogenetic tree was built using the Type Strain Genome Server [36] for taxonomic analysis based on highly conserved 16s rRNA sequences and complete whole genomes.

Prediction of metabolic pathways of *L. paraplantarum* DSM10667, LZS-9, and AS-7 was conducted using the PATRIC KEGG database [23]. Putative glycolysis, pentose phosphate, galactose, fructose, rhamnose, maltose, pyruvate, nitrogen, histidine, and arginine pathways were elucidated and further compared across all three strains. Genome analysis for the presence of putative bacteriocin biosynthesis-responsible gene cluster(s) was performed using the web tool BAGEL 4 [24]. Each putative bacteriocin-encoding protein sequence was further confirmed using BLASTp against a non-redundant protein sequence database [25]. The prophage regions on each genome were identified and annotated using the PHASTER-Phage Search Tool Enhanced Release [26].

To better understand the evolutionary relationships across *L. paraplantarum* DSM10667, LZS-9, and AS-7, the MAUVE alignment tool, and IMG/MER's Dot Plot were performed. The latter employs the Mummer alignment tool to generate dot plot diagrams between two genomes [27] [28].

The presence of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) regions and Cas genes was screened using

CRISPRviz and CRISPR Cas Finder [29]. Carbohydrate active enzymes (CAZyme)-related genes were identified using the CAZy database (v10) on the dbCAN server [30] by HMMER version 3.1b2 according to the suggested protocol of dbCAN. Results of the CAZyme analysis were classified based on a suggested threshold minimum of 0.35 coverage and a cut-off E-value of 1e-15.

Table 1. JGI /IMG-MER Genome ID accession numbers and isolation sources of three *L. paraplantarum* strains

Strains	IMG Genome ID	Isolation Source
<i>Lactiplantibacillus paraplantarum</i> DSM10667	2896850075	Fruit and vegetables
<i>Lactiplantibacillus paraplantarum</i> LZS-9	2609460312	Fermented Sausage
<i>Lactiplantibacillus paraplantarum</i> AS-7	2922794270	Beer

3. Results and Discussion

3.1. Genomic Analysis

Genetic distances among *L. paraplantarum* strains

were determined via the Type Strain Genome server at <https://tygs.dsmz.de/> with default settings [31]. Strains were compared based on whole genome sequences (Figure 1). Among the 10 *L. paraplantarum* strains being analyzed, only DSM 10667, L-ZS9, and AS-7 possess complete genome sequence data that is publicly available in the Joint Genome Institute IMG/Genome Database [14]. According to Figure 1, *L. paraplantarum* DSM 10667 and AS-7 were closely related to each other and were isolated from beer and plant material. However, *L. paraplantarum* L-ZS9 was closely related to the *L. paraplantarum* KMB-599 strain, and the former was being isolated from sausage, though the isolation source for the *L. paraplantarum* KMB-599 strain is not described [32].

Table 2. Genomic characteristics of *Lactiplantibacillus paraplantarum* DSM 10667, L-ZS9 and AS-7.

Feature	DSM		
	10667	L-ZS9	AS-7
Size (bp)	3.36853	3.13973	3.00744
GC content (%)	43.89	44.0	44.1
Genes (total)	3399	2974	2905
Protein coding sequences	3063	2812	2689
tRNA	68	66	59
rRNA	16	6	None
Crispr Count	1	4	1
Pseudogenes	64	78	62
Plasmid	2	None	None

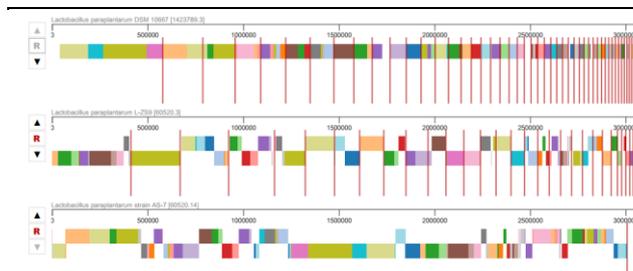


Figure 1. Whole genome comparisons. Mauve alignment of all complete *L. paraplantarum* genomes

3.1.1 Phylogenetic analysis

The phylogenetic tree was structured in Patric [23] to reveal the evolutionary relationships between ten *L. paraplantarum* strains. It was shown that *L. paraplantarum* AS-7 and DSM 10667 are closely related strains, while *L. paraplantarum* L-ZS9 is positioned far from these two strains. We speculate that since *L. paraplantarum* L-ZS9 has a quite different isolation source than AS-7 and DSM 10667 strains, the phylogenetic distance might be a result of this fact. Indeed, AS-7 and DSM 10667 do not have the exact same isolation source, but their environmental conditions are perhaps more closely related (i.e., plant-based) than the L-ZS9 (i.e., meat-based).

The genomic features of *L. paraplantarum* DSM10667, L-ZS9, and AS-7 are given in Table 2. The sizes of genomes are as follows: DSM 10667>L-ZS9>AS-7, which is in alignment with the total genes and protein-coding sequences of each strain. It was interesting to note that although L-ZS9 and AS-7 did

not carry any plasmids in their cytoplasm, DSM10667 contains 2 plasmids. The L-ZS9 strain carried four CRISPR loci, whereas DSM 10667 and AS-7 carried one CRISPR loci.

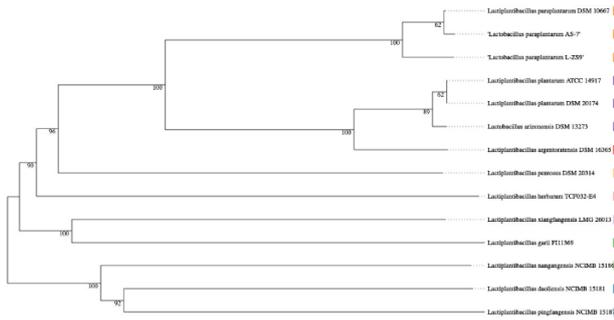


Figure 2. Phylogenetic distance tree of *L. paraplantarum* strains based on whole genome sequence alignment

3.1.2. Comparative Genome Analysis

Functional comparisons of DSM 10667, L-ZS9, and AS-7 based on whole genomes were summarized in Supplementary Table 1, Table 2, and Table 3. It was revealed that DSM 10667 harbors 438 unique genes that do not exist in L-ZS9 and AS-7. Although some of those genes were being annotated as hypothetical proteins, the DSM10667 excelled with 4 prophages and 3 bacteriocin-encoding genes. On the other hand, both L-ZS9 and AS-7 carried a single intact prophage in their genomes.

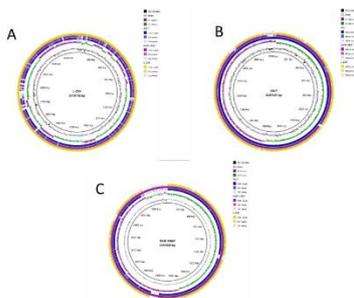


Figure 3. Genome wide BLAST comparison of all strains against reference strain: A)DSM10667 and AS-7 against L-ZS9 B)DSM 10667 and L-ZS9 against AS-7 C)L-ZS9 and AS-7 against DSM 10667 references

3.1.3 Unique genes

L. paraplantarum AS-7 contains 35 strain-specific genes (i.e., unique genes) composed of 13 hypothetical protein-encoding genes, which are not found in either *L. paraplantarum* DSM 10667 or L-ZS9. Likewise, L-ZS9 harbors 248 unique genes, including 89 hypothetical protein-encoding genes.

Across all three genomes, *L. paraplantarum* DSM 10667 harbors the highest number of unique genes (438), more than half of which contain hypothetical protein-encoding genes (Tables S1, S2, and S3).

3.1.4. Bacteriocins

The Bagel4 tool predicted that *L. paraplantarum* DSM 10667 encodes plantaricin A, E, F, J, and leucocin K [24]. On the other hand, both L-ZS9 and AS-7 contain class I bacteriocins of plantaricin A, E, F, J and K (Figure 4A, 4B, 4C). Evidence of putative bacteriocin biosynthesis-associated genes might imply that the potential antimicrobial activity of *L. paraplantarum* against competitive microorganisms (i.e., including pathogens) coexists in a similar environmental ecosystem that *L. paraplantarum* strains inhabit [33].

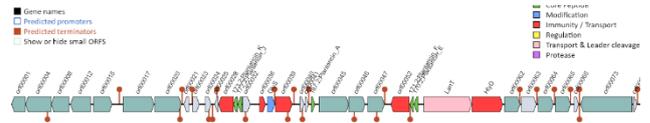


Figure 4. The predicted gene cluster of L-ZS9 responsible for the biosynthesis of Plantaricins by using the BAGEL4 webserver

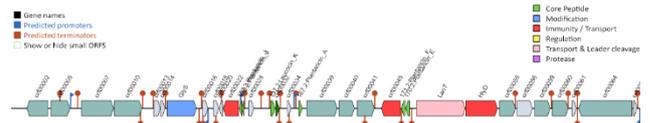


Figure 5. The predicted gene cluster of AS-7 responsible for the biosynthesis of Plantaricins by using the BAGEL4 webserver

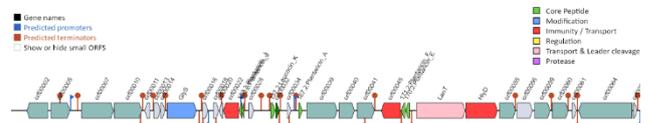


Figure 6. The predicted gene cluster of DSM 10667 responsible for the biosynthesis of Plantaricins and Leucocin K by using the BAGEL4 webserver

3.1.5. Prophages

In *L. paraplantarum* DSM10667, six prophage regions were predicted. Four of those were complete, one of which was incomplete and questionable (Figure 2). The sizes of the regions were 48.3 Kb, 24.9 Kb, 41.9 Kb, 59.2 Kb, 15 Kb, and 8.1 Kb. One of

these phages could be Lactob_Sha1, which is commonly present in various microorganisms [34]. *L. paraplantarum* L-ZS9 contains one intact and one questionable region with a size of 15.9 Kb and 39.6 Kb, respectively. However, *L. paraplantarum* AS-7 harbors only one intact prophage region at a size of 50.9 Kb. Other prophages predicted were Lactob_KC5a, Lactoc_bIL309, Oenoco_phiS1, Paenib_Tripp, Strept_315.2, and Oenoco_phiS13. This perhaps indicates the plasticity of all three strains, of which DSM 10667 ranks the highest. The high plasticity of the DSM 10667 genome could be due to the liquid matrices of the beer system from which this strain was isolated. We speculate that both L-ZS9 and AS-7 isolation sources were closer to solid matrices, which perhaps reduces the infectivity of the (pro)phages to resident organisms in that environment. We'd anticipate a different trend of phage infectivity between dairy and meat starter cultures, where dairy cultures are more prone to phage attacks and perhaps prophage attachments due to the liquid structure of milk, whereas meat matrices provide protection to bacterial strains.

OrthoANI (Orthologous Average Nucleotide Identity) represents the similarity ratios of orthologous regions between multiple genomes. According to OrthoANI results, the most closely related strains are *L. paraplantarum* AS-7 and DSM 10667, with a similarity percentage of 99.749% achieved. This close similarity could perhaps be linked to the origin of those strains, which is a plant-based environment. L-ZS9 was segregated from the other two strains based on a remarkably lower similarity rate of 97.49% and 97.86% against AS-7 and DSM 10667, respectively. In a parallel manner, synteny plots achieved across three strains revealed that DSM 10667 and AS-7 have better synteny compared to DSM 10667 vs. L-ZS9 or AS-7 vs. L-ZS9. (Supplementary Figures 1, 2, and 3).

Table 3. The predicted prophage regions of *Lactiplantibacillus paraplantarum* strains

Strain name	Region Length	Region Position	Completeness	Most Common Phage	Total Protein	G-C % Content
	48.3Kb	966930-1015283	Intact	Lactob_Sha1	64	41.22%
	24.9Kb	1773553-1798540	Intact	Lactob_KC5a	31	43.89%
DSM 10667	41.9Kb	1801471-1843386	Intact	Lactob_KC5a	54	42.15%
	59.2Kb	2119916-2179148	Intact	Oenoco_phiS1	69	41.45%
	8.1Kb	2315784-2323924	Questionable	Paenib_Tripp_	8	41.09%
	15Kb	1840319-1855322	Incomplete	Lactoc_bIL30	23	39.97%
AS-7	50.9Kb	530636-581617	Intact	Oenoco_phiS13	57	41.68%
	39.6Kb	2210312-2250005	Intact	Oenoco_phiS13	48	41.90%
L-ZS9	15.9Kb	36418-52364	Questionable	Strept_315.2	21	41.83%

Table 4. OrthoANI percentage of *Lactiplantibacillus paraplantarum* DSM 10667, L-ZS9 and AS-7.

Genome 1 Name	Genome 2 Name	ANI 1->2	ANI2->1	AF1->2	AF2->1
AS-7	L-ZS9	97.49	97.495	90.0	86.9
AS-7	DSM 10667	99.74	99.749	95.2	85.2
DSM 10667	L-ZS9	97.86	97.861	79.2	85.6
DSM 10667	AS-7	99.74	99.749	95.2	85.2

The CRISPRviz and CRISPRcas finder results did not identify any confirmed CRISPR or cas regions. However, all three strains are predicted to carry three questionable CRISPR genes.

3.2. Carbohydrate Metabolism

3.2.1. Glycolysis Pathway

The extracellular D-Glucose molecule has been transported into the cytoplasm through the phosphotransferase system, and this glucose is

phosphorylated to D-Glucose 6-phosphate in only DSM 10667, which is further converted to Fructose 6-phosphate by glucose-6-phosphate isomerase [EC 5.3.1.9], Fructose 1,6-phosphate by phosphohexokinase [EC 2.7.1.11], Glyceraldehyde 3-phosphate by fructose-bisphosphate aldolase [EC 4.1.2.13], and Glycerone-phosphate by triose-phosphate isomerase [EC 5.3.1.1] in all three strains. The glyceraldehyde-3 phosphate is converted to Pyruvate, Acetyl CoA, and Lactate. The latter is also being utilized in propanoate metabolism. Arbutin and Salicin sugars are also being imported into the cell through specific PTS [EC2.7.1.69] by which can either feed into glycolysis or the pentose phosphate pathway. However, sausage isolate L-ZS9 and plant isolate AS-7 strains did not carry the [E.C. 2.7.1.69] enzyme, thus extracellular arbutin and salicin cannot be imported as arbutin-6P and salicin-6P into the cytoplasm. We speculate that, most likely, these sugars are being utilized through the EMP pathway (Figures S8, S9, and S10).

3.2.2. Pentose Phosphate Pathway

Five carbon sugars, such as D-Ribose, have been utilized by the pentose phosphate pathway for DSM 10667, part of its heterofermentative utilization of pentoses. Unlike other sugars discussed above, D-Ribose has been phosphorylated by the ribokinase [E.C. 2.7.1.15] enzyme to form D-Ribose 5-P, which can go to purine, pyrimidine, or histidine metabolism by ribose-phosphate diphosphokinase [EC 2.7.6.1], Ribose-5P by spermine oxidase [EC 1.5.3.16], or D-Glyceraldehyde3-P by nicotinamide N-methyltransferase [EC 2.1.1.1], which is later converted to lactate. L-ZS9 is the only strain that has the transaldolase enzyme; therefore, it can convert D-glyceraldehyde-3-phosphate to beta-D-fructose 6-phosphate (Figures S11, S12, and S13).

3.2.3. Putative Sugar Utilizations

In galactose metabolism, there are slight differences between the 3 strains. For example, *L. paraplantarum* AS-7 and L-ZS9 are predicted to convert Galactitol-1P to D-Tagatose-6P by the Galactitol-1-phosphate 5-dehydrogenase enzyme. Although *L. paraplantarum* DSM10667 cannot produce D-Tagatose-6P, all 3 strains can synthesize D-Tagatose-6P from D-Tagatose 1,6 P2 by removing one phosphate group and producing 1 mol ATP molecule. They can also produce raffinose from stachyose oligosaccharides, as shown in Figures S14, S15, and S16 in supplementary.

Genome analysis of *L. paraplantarum* revealed that only DSM10667 could ferment rhamnose sugar, while L-ZS9 and AS-7 cannot. Different from the other two strains, the DSM 10667 strain can make the conversion between L-rhamnose and L-rhamnulose sugars with the help of L-rhamnose isomerase. Only DSM 10667 produces sorbitol and sorbose pathways in fructose metabolism (Figures S17, S18, and S19).

L. paraplantarum DSM 10667 could be converted from extracellular maltose to Maltose-6P by N-phosphohistidine-D-mannose phosphotransferase. Also, it could perform the conversion of isomaltose to D-Glucose to maltose by oligo-1,6-glucosidase. When sucrose is present in fermentation, DSM 10667 could convert sucrose into dextran and fructose by dextransucrase [EC 2.4.1.5]. Dextran is an exopolysaccharide with a chemical structure of complex branched glucan with dextran chains of varying lengths from 3-2000 kDA [35]. The beer contaminant DSM 10667 strain has genes encoding the dextransucrase enzyme, which might cause a slimy structure in beer, which is called a structural defect and could be one of the reasons why this strain is being considered a beer contaminant. Therefore, DSM 10667, an unwanted adventitious lactic acid bacterium, could potentially cause quality defects and economic losses to beer processors (Figures S20, S21, and S22).

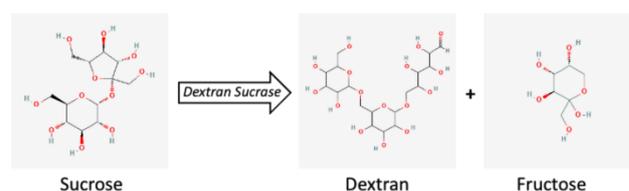


Figure 7. Activity of dextransucrase enzyme which can convert sucrose to dextran and fructose.

3.3. Nitrogen Metabolism

The putative nitrogen metabolism pathway indicated that L-Glutamine and L-Glutamate amino acids can be biosynthesized from ammonia in all three strains. They are also capable of converting L-Aspartate via aspartate ammonia-lyase [EC 4.3.1.1] or L-Asparagine via aspartate ammonia-ligase [EC 6.3.1.1]. The ammonia formation from aspartate can be a stress mechanism for combating the acidic condition of the cytoplasm to maintain a relatively neutral pH value within the cell matrix. Moreover, only L-ZS9 can convert nitrate to nitrite by the nitrate reductase [EC 1.7.99.4] enzyme, which is one of the

technological attributes that is highly desirable in meat starter cultures (Figures S28, S29, and S30).

3.3.1. Histidine and Arginine

The putative histidine metabolism of all three strains predicted that L-Histidine can be biosynthesized from PRPP after a series of conversions. Since no histidine decarboxylase enzyme-encoding gene was found in the genome of strains, L-Histidine cannot be converted to the histamine molecule, which is detrimental to health. However, it was interesting to note that histamine can be converted to 4-Imidazole only by DSM 10667 (Figures S32, S33, and S34).

L. paraplantarum L-ZS9 has quite different in arginine metabolisms compared to the other two strains. For instance, DSM 10667 and AS-7 cannot complete the urea cycle due to the absence of arginase, while *L. paraplantarum* L-ZS9 could convert arginine to urea, also known as arginine decarboxylation, which further increases the pH, a potential acid stress response mechanism for bacteria, by neutralizing the cytoplasm. Interestingly, only DSM 10667 has the capability to produce p-coumaroyl putrescine and feruloyl putrescine in arginine metabolism.

3.4. Pyruvate Metabolism

Pyruvate is an intermediate molecule that is formed upon the cleavage of sugars through glycolysis or the pentose phosphate pathway. DSM 10667 can convert pyruvate to L-lactate by L-lactate dehydrogenase [EC 1.1.1.27] and D-Lactate by D-lactate dehydrogenase [EC 1.1.1.28]. The L-ZS9 strain also has these enzymes and can convert pyruvate to lactate molecules. All three strains can also convert pyruvate to Oxaloacetate by Pyruvate carboxylase [EC 6.4.1.1] or L-Malate by malate dehydrogenase [EC 1.1.1.40], Formate by formate C-acetyltransferase [EC 2.3.1.54], Acetyl Co-A, Acetaldehyde by acetaldehyde dehydrogenase [EC 1.2.1.10], and Acetyl phosphate by pyruvate oxidase [EC 1.2.3.3]. It is interesting to note that there is interconversion in DSM 10667 and AS-7 strains between D-Lactaldehyde and methylglyoxal glyoxylate reductase [EC 1.1.1.79], which is a brown color agent in fermented foods that is not a desirable attribute in most dairy products. Unlike DSM 10667, L-ZS9 does not convert methylglyoxal to D-Lactaldehyde. For example, methylglyoxal formation in cheese is causing a serious economic loss for dairy products in the USA. In a similar manner, the occurrence of these metabolites in beer can also be detrimental to the

appearance of the product and might reduce its quality. There is another interconversion between acetyl Co-A intermediate and acetaldehyde by acetaldehyde dehydrogenase [EC 1.2.1.10], resulting in an aroma compound in some of the fermented dairy products, and this conversion also occurred in the L-ZS9 strain [EC1.2.1.10]. Another avenue from pyruvate is the conversion of Acetyl Co-A to Malonyl Co-A by acetyl-CoA carboxylase [EC 6.4.1.2], which is predicted to be a precursor for fatty acid biosynthesis (Figures S22, S23, and S24). *L. paraplantarum* DSM10667 can be distinguished from L-ZS9 and AS-7 by citric acid metabolism. DSM10667 can synthesize malate from oxaloacetate by the malate dehydrogenase enzyme and the conversion of 1 mol NADH₂ to NAD, then fumarate by the fumarate hydratase enzyme, and finally succinate by the succinate dehydrogenase enzyme, unlike L-ZS9 and AS-7 strains. During succinate production, NADH₂ turns into NAD molecules. L-ZS9 and AS-7 cannot produce succinate because of the absence of the succinate dehydrogenase enzyme, as shown in Figures S23, S24, and S25.

4. Conclusion

We performed a comparative genome analysis of three *L. paraplantarum* strains that were isolated from diverse ecosystems such as beer, sausage, fruit, and vegetables. The total genome sizes achieved were DSM 10667>L-ZS9>AS-7. Phylogenetic analysis based on whole-genome sequence and average nucleotide identity revealed that DSM 10667 and AS-7 were closely related, whereas L-ZS9 differed from both strains. This perhaps relates to the fact that the origins of strains DSM 10667 and AS-7 were relatively similar (i.e., plant-related material) compared to L-ZS9, which has a quite different origin (i.e., fermented sausage). These differences based on the origin of isolation are also evident in the unique genes found in each genome. For example, DSM 10667 harbors 438 unique genes, L-ZS9 harbors 248 unique genes, and AS-7 harbors 35 unique genes. Arginine, which is an abundant amino acid in sausage composition, can be converted to urea only by L-ZS9. Neither of the other two strains has that capability, perhaps because plant materials are not reliable sources of arginine. Another example is the dextran biosynthesis potential of DSM 10667, which has already been identified as an undesirable strain in beer manufacturing technology. These findings provide a better understanding of the genomic characteristics of *L. paraplantarum* strains for commercial and scientific interest.

Acknowledgment

The authors would like to thank Mr. Ismail Gumustop for valuable support in manuscript formatting.

Funding

There is no funding for this project.

Conflict of Interest

The authors declare no competing interest.

Author Contribution Statement

Conceptualization: FO

Method: FO, RB

Manuscript writing reviewing editing: FO, RB

Supervision: FO

References

- [1]. H. Nishioka, T. Ohno, H. Iwahashi, and M. Horie, "Diversity of Lactic Acid Bacteria Involved in the Fermentation of Awa-bancha," *Microbes Environ*, vol. 36, 2021. [Online]. Available: <https://doi.org/10.1264/jsme2.ME21029>.
- [2]. J. Zheng, S. Wittouck, E. Salvetti, et al., "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*," *Int. J. Syst. Evol. Microbiol.*, vol. 70, 2020. [Online]. Available: <https://doi.org/10.1099/ijsem.0.004107>.
- [3]. Reale, T. Zotta, R. G. Ianniello, et al., "Selection criteria of lactic acid bacteria to be used as starter for sweet and salty leavened baked products," *LWT*, vol. 133, p. 110092, 2020. [Online]. Available: <https://doi.org/10.1016/j.lwt.2020.110092>.
- [4]. S. Akbulut, M. O. Baltacı, G. Adıguzel, and A. Adıguzel, "Identification and biotechnological characterization of Lactic Acid Bacteria Isolated from White Cheese Samples," Research Square, 2021.
- [5]. S. Nishida, M. Ishii, Y. Nishiyama, et al., "Lactobacillus paraplantarum 11-1 Isolated from Rice Bran Pickles Activated Innate Immunity and Improved Survival in a Silkworm Bacterial Infection Model," *Front. Microbiol.*, vol. 8, 2017.
- [6]. W. J. Park, K. H. Lee, J. M. Lee, et al., "Characterization of pC7 from *Lactobacillus paraplantarum* C7 derived from Kimchi and development of lactic acid bacteria--*Escherichia coli* shuttle vector," *Plasmid*, vol. 52, pp. 84–88, 2004. [Online]. Available: <https://doi.org/10.1016/j.plasmid.2004.05.001>.
- [7]. W. E. Hussein, E. Huang, I. Ozturk, and A. E. Yousef, "Draft Genome Sequence of *Lactobacillus paraplantarum* OSY-TC318, a Producer of the Novel Lantibiotic Paraplantaracin TC318," *Microbiol. Resour. Announc.*, vol. 8, p. e00274-19, 2019. [Online]. Available: <https://doi.org/10.1128/MRA.00274-19>.
- [8]. K. Sharma, N. Sharma, S. Handa, and S. Pathania, "Purification and characterization of novel exopolysaccharides produced from *Lactobacillus paraplantarum* KM1 isolated from human milk and its cytotoxicity," *J. Genet. Eng. Biotechnol.*, vol. 18, p. 56, 2020. [Online]. Available: <https://doi.org/10.1186/s43141-020-00063-5>.
- [9]. S. M. Devi, N. K. Kurrey, and P. M. Halami, "In vitro anti-inflammatory activity among probiotic *Lactobacillus* species isolated from fermented foods," *J. Funct. Foods*, vol. 47, pp. 19–27, 2018. [Online]. Available: <https://doi.org/10.1016/j.jff.2018.05.036>.
- [10]. R. J. da Costa, F. L. S. Voloski, R. G. Mondadori, et al., "Preservation of Meat Products with Bacteriocins Produced by Lactic Acid Bacteria Isolated from Meat," *J. Food Qual.*, vol. 2019, p. e4726510, 2019. [Online]. Available: <https://doi.org/10.1155/2019/4726510>.
- [11]. L. Liu and P. Li, "Complete genome sequence of *Lactobacillus paraplantarum* L-ZS9, a probiotic starter producing class II bacteriocins," *J. Biotechnol.*, vol. 222, pp. 15–16, 2016. [Online]. Available: <https://doi.org/10.1016/j.jbiotec.2016.02.003>.
- [12]. "IMG," [Online]. Available: https://img.jgi.doe.gov/cgi-bin/m/main.cgi?section=TaxonDetail&page=taxonDetail&taxon_oid=2922794270. Accessed: Sep. 4, 2023.

- [13]. L. C. Reimer, A. Vetcinina, J. S. Carbasse, et al., "Bac Dive in 2019: bacterial phenotypic data for High-throughput biodiversity analysis," *Nucleic Acids Res.*, vol. 47, pp. D631–D636, 2019. [Online]. Available: <https://doi.org/10.1093/nar/gky879>.
- [14]. H. Nordberg, M. Cantor, S. Dusheyko, et al., "The genome portal of the Department of Energy Joint Genome Institute: 2014 updates," *Nucleic Acids Res.*, vol. 42, pp. D26-31, 2014. [Online]. Available: <https://doi.org/10.1093/nar/gkt1069>.
- [15]. J.-Y. Chun, W.-J. Jeong, J.-S. Kim, et al., "Hydrolysis of Isoflavone Glucosides in Soymilk Fermented with Single or Mixed Cultures of *Lactobacillus paraplantarum* KM, *Weissella* sp. 33, and *Enterococcus faecium* 35 Isolated from Humans," *J. Microbiol. Biotechnol.*, vol. 18, pp. 573–578, 2008.
- [16]. R. Kant, J. Blom, A. Palva, et al., "Comparative genomics of *Lactobacillus*," *Microb. Biotechnol.*, vol. 4, pp. 323–332, 2011. [Online]. Available: <https://doi.org/10.1111/j.1751-7915.2010.00215>.
- [17]. M. Schmid, J. Muri, D. Melidis, et al., "Comparative Genomics of Completely Sequenced *Lactobacillus helveticus* Genomes Provides Insights into Strain-Specific Genes and Resolves Metagenomics Data Down to the Strain Level," *Front. Microbiol.*, vol. 9, p. 63, 2018. [Online]. Available: <https://doi.org/10.3389/fmicb.2018.00063>.
- [18]. S. Wuyts, C. N. Allonsius, S. Wittouck, et al., "Comparative genome analysis of *Lactobacillus mudanjiangensis*, an understudied member of the *Lactobacillus plantarum* group," *Microb. Genomics*, vol. 5, p. e000286, 2019. [Online]. Available: <https://doi.org/10.1099/mgen.0.000286>.
- [19]. M. C. Curk, J. C. Hubert, and F. Bringel, "*Lactobacillus paraplantarum* sp. nov., a new species related to *Lactobacillus plantarum*," *Int. J. Syst. Bacteriol.*, vol. 46, pp. 595–598, 1996. [Online]. Available: <https://doi.org/10.1099/00207713-46-2-595>.
- [20]. S. Mukherjee, D. Stamatis, J. Bertsch, et al., "Genomes OnLine Database (GOLD) v.8: overview and updates," *Nucleic Acids Res.*, vol. 49, pp. D723–D733, 2021. [Online]. Available: <https://doi.org/10.1093/nar/gkaa983>.
- [21]. N.-F. Alikhan, N. Petty, N. Ben Zakour, and S. Beatson, "BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons," *BMC Genomics*, vol. 12, p. 402, 2011. [Online]. Available: <https://doi.org/10.1186/1471-2164-12-402>.
- [22]. Z. Zhang, S. Schwartz, L. Wagner, and W. Miller, "A greedy algorithm for aligning DNA sequences," *J. Comput. Biol. J. Comput. Mol. Cell Biol.*, vol. 7, pp. 203–214, 2000. [Online]. Available: <https://doi.org/10.1089/10665270050081478>.
- [23]. R. Wattam, D. Abraham, O. Dalay, et al., "PATRIC, the bacterial bioinformatics database and analysis resource," *Nucleic Acids Res.*, vol. 42, pp. D581–D591, 2014.
- [24]. de Jong, S. A. F. T. van Hijum, J. J. E. Bijlsma, et al., "BAGEL: a web-based bacteriocin genome mining tool," *Nucleic Acids Res.*, vol. 34, pp. W273–W279, 2006. [Online]. Available: <https://doi.org/10.1093/nar/gkl237>.
- [25]. J. Zhang and T. L. Madden, "PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation," *Genome Res.*, vol. 7, pp. 649–656, 1997. [Online]. Available: <https://doi.org/10.1101/gr.7.6.649>.
- [26]. D. Arndt, J. R. Grant, A. Marcu, et al., "PHASTER: a better, faster version of the PHAST phage search tool," *Nucleic Acids Res.*, vol. 44, pp. W16-21, 2016. [Online]. Available: <https://doi.org/10.1093/nar/gkw387>.
- [27]. G. Marçais, A. L. Delcher, A. M. Phillippy, et al., "MUMmer4: A fast and versatile genome alignment system," *PLOS Comput. Biol.*, vol. 14, p. e1005944, 2018. [Online]. Available: <https://doi.org/10.1371/journal.pcbi.1005944>.
- [28]. I.-M. A. Chen, K. Chu, K. Palaniappan, et al., "The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities," *Nucleic Acids Res.*, vol. 49, pp. D751–D763, 2021. [Online]. Available: <https://doi.org/10.1093/nar/gkaa939>.
- [29]. D. Couvin, A. Bernheim, C. Toffano-Nioche, et al., "CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins," *Nucleic Acids Res.*, vol. 46, pp. W246–W251, 2018. [Online]. Available: <https://doi.org/10.1093/nar/gky425>.
- [30]. H. Zhang, T. Yohe, L. Huang, et al., "dbCAN2: a meta server for automated carbohydrate-active enzyme annotation," *Nucleic Acids Res.*, vol. 46, pp. W95–W101, 2018. [Online]. Available: <https://doi.org/10.1093/nar/gky418>.

- [31]. S. Mukherjee, R. Seshadri, N. J. Varghese, et al., "1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life," *Nat. Biotechnol.*, vol. 35, pp. 676–683, 2017. [Online]. Available: <https://doi.org/10.1038/nbt.3886>.
- [32]. W. E. Hussein, E. Huang, I. Ozturk, et al., "Genome-Guided Mass Spectrometry Expedited the Discovery of Paraplantaricin TC318, a Lantibiotic Produced by *Lactobacillus paraplantarum* Strain Isolated From Cheese," *Front. Microbiol.*, vol. 11, p. 1381, 2020. [Online]. Available: <https://doi.org/10.3389/fmicb.2020.01381>.
- [33]. S.-C. Yang, C.-H. Lin, C. T. Sung, and J.-Y. Fang, "Antibacterial activities of bacteriocins: application in foods and pharmaceuticals," *Front. Microbiol.*, vol. 5, 2014.
- [34]. E. Evanovich, P. J. de Souza Mendonça Mattos, and J. F. Guerreiro, "Comparative Genomic Analysis of *Lactobacillus plantarum*: An Overview," *Int. J. Genomics*, vol. 2019, p. e4973214, 2019. [Online].
- [35]. Al-Farga A, Abed S. Production of dextrans and their applications in human health and nutrition–Review. *Eur Acad Res* 29. (2016)
- [36]. "Type Strain Genome Server," *Dsmz.de*, 2019. <https://tygs.dsmz.de/> (accessed Aug. 07, 2023).