microorganisms

Article

Genomic Insight into *Pediococcus acidilactici* HN9, a Potential Probiotic Strain Isolated from the Traditional Thai-Style Fermented Beef Nhang

Komwit Surachat [1,2,*], Duangporn Kantachote [3], Panchalika Deachamag [3] and Monwadee Wonglapsuwan [3]

1   Division of Computational Science, Faculty of Science, Prince of Songkla University, Hatyai, Songkhla 90110, Thailand
2   Molecular Evolution and Computational Biology Research Unit, Faculty of Science, Prince of Songkla University, Hatyai, Songkhla 90110, Thailand
3   Division of Biological Science, Faculty of Science, Prince of Songkla University, Hatyai, Songkhla 90110, Thailand; duangporn.k@psu.ac.th (D.K.); passanee.d@psu.ac.th (P.D.); monwadee.wo@psu.ac.th (M.W.)
*   Correspondence: komwit.s@psu.ac.th

Abstract: *Pediococcus acidilactici* HN9 is a beneficial lactic acid bacterium isolated from Nhang, a traditional Thai-style fermented beef. In this study, the molecular properties of *P. acidilactici* HN9 were characterized to provide insights into its potential probiotic activity. Specifically, this work sought to report the complete genome of *P. acidilactici* HN9 and perform a comparative genome analysis with other bacterial strains belonging to the genus *Pediococcus*. Genomic features of HN9 were compared with those of all other bacterial *Pediococcus* strains to examine the adaptation, evolutionary relationships, and diversity within this genus. Additionally, several bioinformatic approaches were used to investigate phylogenetic relationships, genome stability, virulence factors, bacteriocin production, and antimicrobial resistance genes of the HN9 strain, as well as to ensure its safety as a potential starter culture in food applications. A 2,034,522 bp circular chromosome and two circular plasmids, designated pHN9-1 (42,239-bp) and pHN9-2 (30,711-bp), were detected, and used for pan-genome analysis, as well as for identification of bacteriocin-encoding genes in 129 strains belonging to all *Pediococcus* species. Two CRISPR regions were identified in *P. acidilactici* HN9, including type II-A CRISPR/CRISPR-associated (Cas). This study provides an in-depth analysis on *P. acidilactici* HN9, facilitating a better understanding of its adaptability to different environments and its mechanism to maintain genome stability over time.

Keywords: beneficial lactic acid bacteria; pediococci; *Pediococcus acidilactici*; comparative genome analysis; fermented food

1. Introduction

Pediococcus, a genus of gram-positive lactic acid bacteria (LAB), is classified as part of the *Lactobacillaceae* family [1]. These bacteria, which are commonly found in fermented vegetables, meats, and dairy products, play an important physiological role in fermentation [2–4]. Several species of *Pediococcus* are used in the food industry, owing to their inhibitory effects on other microorganisms via antimicrobial peptide production [5,6]. In addition, there are many reports of bacteriocins produced by *Pediococcus* strains, including Pediocin F [7], Pediocin PA-1 [8], and Pediocin SA-1 [9]. *Pediococcus acidilactici* is a potent producer of bacteriocins that has been used in the food industry for many years [3,4,10,11]. Moreover, several studies have reported that *P. acidilactici* can inhibit the growth of pathogens during the fermentation process and food storage [4,6,12,13].

Pediococcus acidilactici strains have been widely studied in food and feed applications. For instance, *P. acidilactici* SMVDUDB2 [14] was found to be capable of phytate degradation and to possess probiotic traits, which can be useful in food and feed applications.

Interestingly, probiotic *P. acidilactici* MA18/5M reportedly modulates antiviral responses in Atlantic salmon when used as a dietary supplement [15], while strain CNCM I-4622 (trade name: Bactocell®) has been authorized by The Panel on Additives and Products or Substances used in Animal Feed (FEEDAP) as a feed additive for all fish, shrimps, and crustaceans [16]. Additionally, *P. acidilactici* R037 has been reported to induce interleukin (IL)-10-producing regulatory T cells and to have demonstrated efficacy in treating multiple sclerosis via oral administration in mice [17].

It has long been known that the use of starter culture in fermented foods in order to improve the quality and produce safe food compared with spontaneous fermentation. *P. acidilactici* strains have key enzymes involved in homo-fermentation including D- and L- lactate dehydrogenases [14]. Therefore, they can quickly produce lactic acid to reduce fermentation time. Consequently, *P. acidilactici* strains are used generally as the starter culture in fermented meat products including *P. acidilactici* ATCC 8042 [18], *P. acidilactici* MCH14 [19] since they can grow and survive at different temperatures, salt concentrations, and pH values [20–22]. These physiological traits contribute and play an important role in beef fermentation. Also, these strains can inhibit the growth of the foodborne pathogens *Listeria monocytogenes* and *Clostridium perfringens* [19]. Additionally, to gain more benefits from starter culture, its probiotic properties should be considered by testing its property to survive in the gastrointestinal tract for promoting health benefits to the host. Based on this property, LAB including *P. acidilactici* strains are potential candidates to be probiotics, as their resistance in acidity. Hence, fermented foods, e.g. dairy products, and fermented meat products are important probiotic sources for humans and animals.

However, only a few studies were focused on the comparative genomics of *P. acidilactici* strains. For example, Ranjan et al. [23] proposed the genome sequence of *P. acidilactici* NRCC1 isolated from dromedary camel to compare genomic features with reference strain (*P. acidilactici* DSM 20284). Also, Snauwaert et al. [24] presented the draft genome of *P. damnosus* LMG 28219 and performed a comparative analysis against several genomes, including *P. acidilactici*, to underly the mechanisms of its adaptation to the beer niche. In particular, pan-genome analysis for the exploration of genus and species diversity, as well as the ability of these bacteria to produce bacteriocins, is required to gain a comprehensive understanding of these different aspects, including their associated safety when present in food products.

In current practice, before the application of any potent microbial strain as a starter culture for producing fermented food/beverage, its safety profile must be investigated—particularly drug resistance genes and virulence factors. To date, performing whole genome analyses presents the most effective method for evaluating the safety of any strain. This approach also provides insights into the presence of beneficial genes that encode bacteriocins and oligopeptides, etc., which may offer additional value to the fermented food, allowing them to serve as functional food. Therefore, the aims of this study were to determine the complete genome of *P. acidilactici* HN9 isolated from the traditional Thai-style fermented beef Nhang and to perform comparative genomic analysis of all bacterial strains in the genus *Pediococcus*. Furthermore, to display the diversity within this genus, the relationships among all strains were analyzed using the phylogenetic tree generated with all core genes in this genus. Specifically, we used 129 assemblies downloaded from the NCBI database to perform pan-genome analysis. This is the first study, to our knowledge, to provide a meta-analysis for all bacterial strains in the genus *Pediococcus* and to identify bacteriocins produced. In addition, the safety of using *P. acidilactici* HN9 in food and beverage applications was evaluated using genomic data.

2. Materials and Methods

2.1. Bacterial Strain and DNA Extraction

Pediococcus acidilactici HN9 was isolated from fermented beef using the de Man, Rogosa, and Sharp (MRS, Merck, Darmstadt, Germany) broth at the Faculty of Science, Prince of Songkla University. This strain was selected based on its antibacterial activity

against foodborne pathogens and usability as a starter culture for producing fermented meat products. A single colony of *P. acidilactici* HN9 was cultivated in the MRS broth leaving a headspace in the culture tube to achieve a microaerophilic condition at 37 °C for 24 h. Genomic DNA was then extracted and purified using DNeasy extraction kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions as described previously [25].

2.2. Whole Genome Library Preparation and Sequencing

Two whole genome libraries were prepared to perform both short-read and long-read sequencing. High-quality genomic DNA of *P. acidilactici* HN9 was extracted for use in both libraries. For long-read sequencing, purified genomic DNA was sequenced with a PacBio RSII sequencer (Pacific Biosciences, Menlo Park, CA, USA). After constructing the RSII library, sequences were generated using P4-C2 chemistry in a single-molecule real-time (SMRT) cell [26], yielding 139,184 subreads with an average read length of approximately 4.6 kbp. Short-read sequencing was performed in-house at Prince of Songkla University using a NextSeq 550 sequencer (Illumina, Inc., San Diego, CA, USA). Libraries were prepared using Nextera XT DNA Library Prep Kit (Illumina) according to the manufacturer's instructions. All reads generated by the sequencer were 150-bp paired-end reads.

2.3. Genome Assembly

SMRTbell adapters and short polymerase reads were removed from the raw reads of PacBio sequencing, followed by removal of low-quality reads (lower than 0.85) and short subreads (<1000 bp) using Canu [27] with -correct parameter to replace noisy read sequences that had representative sequences computed on the basis of overall overlapping reads. The raw reads were trimmed with -trim parameter to select high-quality sequences for further analysis. We obtained 126,932 raw reads from the PacBio sequencer, with a total yield of 771 Mbp. The initial draft of the genome assembly was generated using the -assemble parameter with the genome size set at 2.6 Mbp. The draft assembly was circularized using Circlator [28] to identify and trim the overhangs on chromosomes and plasmids using the corrected reads and the draft assembly obtained using Canu software.

2.4. Small Plasmid Identification and Assembly

PacBio reads are long and may exceed the lengths of small plasmid sequences with a size smaller than 10 kb [29]. During assembly, a parameter setting on the seed read length cut-off that is greater than actual plasmid size in the hierarchical genome assembly process (HGAP) software [30] may miss small plasmids [31]. In order to scan a small plasmid, all short-reads from NextSeq 550 sequencer were aligned to the draft assembly using BWA-mem [32] and the unmapped Illumina reads were extracted from the alignment file (.bam) using SAMtools [33]. The unmapped reads were assembled with Spades version 3.13.0 [34]. We only selected contigs with high coverage to perform trimming and fix the orientation of the plasmids. A blast database was created using the high coverage contigs with makeblastdb command in the NCBI-blast suite program to perform trimming of small plasmids [35]. To ensure that the selected sequences were indeed the plasmid, the PlasmidFinder v. 2.1 [36] was used to scan and exclude redundant sequences including ribosomal DNA or transposons from analysis.

Then, the initial 5 lines of the fasta sequences of the contigs were extracted for searching against all assemblies. If the search returned more than two hits (the first hit should come from the start of the sequence), the overhang of the plasmid could be identified. The overhanging bases were trimmed at the position before the start of the second hit. Finally, the plasmid contig orientation was rearranged. The bacterial chromosome was oriented at the gene encoding DnaA, while plasmids were usually oriented at the replication gene [37]. The Circlator tool was then used to fix the orientation of the plasmid using the fixstart tool. Thus, complete plasmid sequences were finally obtained.

2.5. Assembly Polishing and Evaluation

The assembly was polished to correct the draft assembly of long-read sequences using short-read sequences. First, an alignment file was generated using Illumina paired-end reads to align back to the draft PacBio assembly generated using BWA-mem. Then, Pilon [38], a software for correcting incorrect bases, was applied to polish the sequences of the chromosome and all plasmids of *P. acidilactici* HN9. Finally, the quality of the assembly was evaluated using Busco software version 4.1.2 [39,40]. Busco can provide information regarding annotation completeness based on evolutionarily informed expectations of gene content.

2.6. Genome Annotation and Visualization

Gene finding was performed using Glimmer 3.0 [41]. Functional annotations were predicted with Rapid Annotations using the Subsystems Technology (RAST) [42–44] and the NCBI Prokaryotic Genome Annotation Pipeline [45] in our local server. We identified all tRNAs and rRNAs with tRNAscan-SE [46] and RNAmmer [47], respectively. The clustered regularly interspaced short palindromic repeats (CRISPRs) and tandem repeats were identified using CRISPRFinder [48] and Tandem Repeat Finder [49], respectively. Finally, the genomic content was visualized with the CGView server [50–52].

2.7. Comparative Genomic Analysis

All 129 genome assemblies of the genus *Pediococcus* were compared, and pan-genome analysis was performed, including all complete genome, scaffold, and contig level sequences. Further, we performed a pan-genome analysis at the species level of *P. acidilactici* with 16 complete genome, 11 scaffold, and 21 contig level sequences of 54 strains from the GenBank database (accessed date: 3 September 2020). The assembly details and accession numbers are illustrated in Supplementary Table S1 and S2. The annotation files of all assemblies downloaded from the GenBank database were regenerated with the same protocol to reduce biases arising due to different gene calling methods. Prokka version 1.14.641 [53] was used to annotate all assemblies for creating the General Feature Format (GFF) files. The pan-genome analyses were performed using Roary version 3.13.0 [54] with a minimum BLASTp identity of 60% and a threshold of 97% of isolates a gene must be in to be defined as a conserved gene. Two analyses for pan-genome were identified by retrieving the group of genes shared by all strains in the clade of interest (core genome) and the group of genes that are not present in all the strains (accessory genome). The pan-genome is a combination of all the genes that are found in the clade of interest. Therefore, we first compared the genome assemblies of all strains in the genus *Pediococcus* (129 strains). Second, we used only strains at the species level (45 strains) to compute the pan-genome. All related genes, including core, accessory, and unique gene sets, were queried using Roary scripts.

2.8. Phylogenetic Analyses

The single-copy marker genes computed using Roary were concatenated and aligned using MUSCLE [55]. A phylogenetic tree was then constructed using RaxMLHPC [56] with the neighbor-joining method. The number of bootstrap repetitions was set to 500 to test tree reliability. Tree visualization was illustrated in Geneious R10.2.6 software (Biomatters Ltd., Auckland, New Zealand).

2.9. Multiple Genome Alignments Analysis

To further analyze the evolutionary relationships among the species, we selected the complete genomes in the same clade of the phylogenetic tree to construct alignments using the MAUVE [57] plugin in Geneious software [58].

2.10. Identification of Bacteriocin-Encoding Genes

Bacteriocin-encoding genes and gene clusters were predicted and illustrated using Bagel4 web server [59,60]. To determine the presence of bacteriocin operons in all strains used in this study, BLASTp was performed against the BAGEL database [60].

2.11. Identification of Carbohydrate-Active enZyme (CAZyme)

Genes associated with Carbohydrate-active enzymes (CAZymes) families were identified by searching against the CAZy database [61]. All protein sequences annotated from the genome annotation process were scanning using the dbCAN server [62] with HMMER v. 3.3.2 [63] against CAZy database. An E-value of 1e-15 and coverage of 0.35 were set as the cutoff threshold for identifying CAZyme Class [64].

2.12. In Silico Comparison of Genes Associated with Meat Fermentation

Genes associated with meat fermentation including *asnA2, pgm, uppS, asnS, tnpA1, ctsR, ldhL, ldhD, ctsR* and *rrp-3*, were downloaded from the NCBI server to create a local database for performing BLAST analysis. Then, every genome was then searched against the created database using BLASTn with an E-value cutoff at 1e-15 to determine the presence/absence of the gene. The visualization of the result was presented in the heatmap using R script.

2.13. Identification of Antibiotic Resistance Genes and Virulence Factors

Pathogenic genes, antibiotic resistance genes, and virulence factors were identified to assess the safety of *P. acidilactici* HN9 use. The genomic sequences were searched against several databases, including the Comprehensive Antibiotic Resistance Database (CARD) [65], ResFinder [66] and Virulence Factors Database (VFDB) [67]. Resistance Gene Identifier (RGI) was used against the CARD database to predict resistomes from proteins based on homology and SNP models [65]. The criteria of RGI identification were classified into three types: "Perfect hit" (100% identical with reference), "String hit" (Bit score > 450 and not identical), and "Loose hit" (Bit score < 450, matched in some regions). In addition, ResFinder 4.0 was used to scan for antimicrobial resistance (AMR) genes with a 90% identity threshold and 60% minimum length. The probability of being a human pathogen was determined using the PathogenFinder [68]. Additionally, CRISPRs, prophage regions, and insertion sequences were predicted using CRISPRFinder [48], PHASTER [69], and ISfinder [70], respectively.

3. Results

3.1. General Genome Characteristics and Functional Annotation of P. acidilactici HN9

The genome of *P. acidilactici* HN9 comprises one 2,034,522-bp circular chromosome and two circular plasmids, designated as pHN9-1 (42,239-bp) and pHN9-2 (30,711-bp). The circular representations of *P. acidilactici* HN9 chromosome and plasmids are visualized as shown in Figure 1. General information of *P. acidilactici* HN9 genome and its genomic features, including prophages, CRISPR region, and AMR gene, are given in Table 1. For functional annotation by RAST subsystems, we identified 2013 coding sequences assigned in 200 subsystems in the chromosome. Only 590 (29%) of all coding sequences (CDS) were hypothetical or unknown. Most assigned subsystems were Carbohydrates (169); Protein Metabolism (119); and Cofactors, Vitamins, Prosthetic Groups, and Pigments (66), respectively. Interestingly, 12 CDS were identified in the Fermentation subcategory including lactate (8), as well as acetoin and butanediol metabolism (4). Also, we found 5 CDS of Restriction-Modification (R-M) System including *HsdR, HsdM*, and *HsdS*.
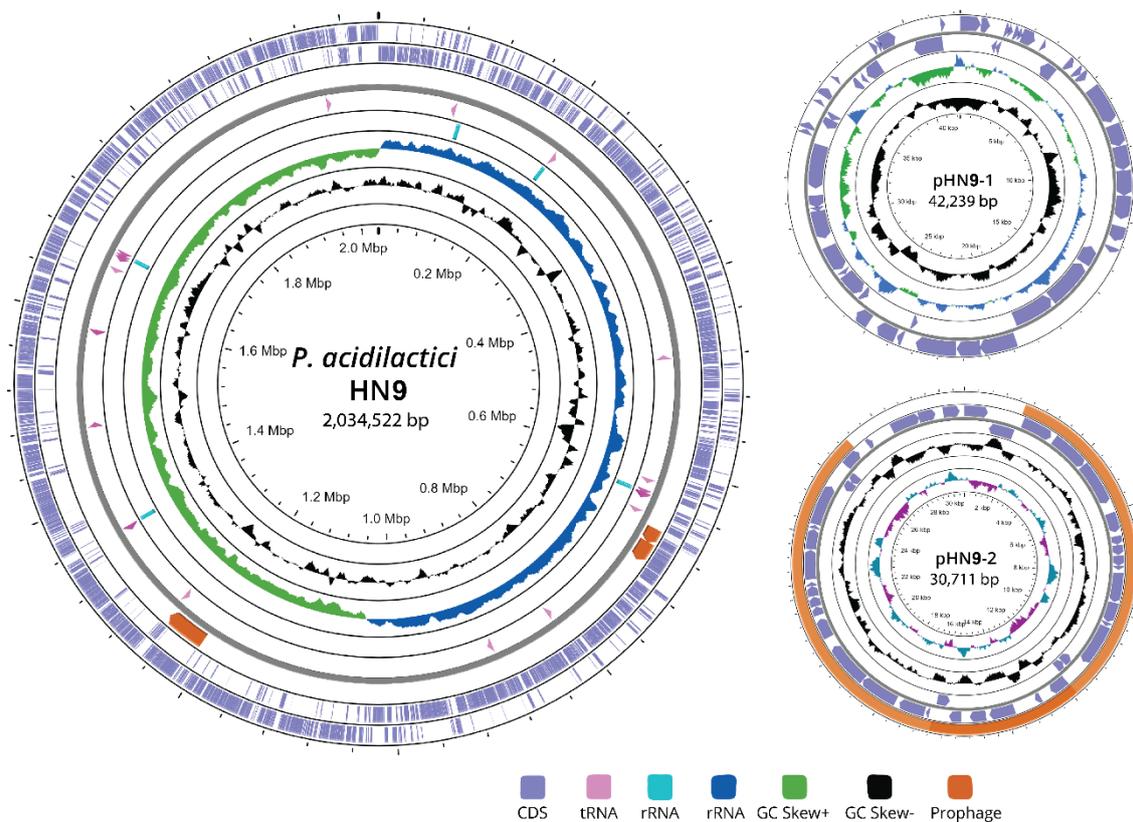
Figure 1. Circular genome and plasmid representations of *P. acidilactici* HN9. The circular illustrations were visualized using the CGView server beta version and contain seven rings. The outermost circle and the second circle show the positions of the CDSs in forward and reverse strand directions, respectively. The next rings show the following information: prophage regions, tRNA, rRNA, GC content, GC skew+, and GC skew-, respectively.

Table 1. Genomic features of *P. acidilactici* HN9.

	Chromosome	Plasmids	
		pHN9-1	pHN9-2
Size (bp)	2,034,522	42,239	30,711
DNA G+C content (%)	42.1	42.6	36.0
Number of CDS	2013	54	42
rRNA	15	0	0
tRNA	56	0	0
RAST subsystems	200	2	2
CRISPR	2	0	0
Prophages	3	0	2
Integrative and conjugative elements (ICEs)	0	1	0
Antimicrobial resistance (AMR) gene	0	0	0
Probability of being a human pathogen	0.124	0.209	0.174

In the case of the plasmids, a total of 54 and 42 CDS were identified in pHN9-1 and pHN9-2, respectively. In pHN9-1, most of the CDS identified were hypothetical proteins (11), mobile element proteins (8), transposases (6), and oligopeptide ABC transporters (5). These CDS were assigned into only two subsystems: Stress Response (1) and Carbohydrates (3) by RAST server. In pHN9-2, most of the CDS identified were hypothetical proteins (15), mobile element proteins (2), plasmid replication initiation protein (2) and sucrose permease (2). Also, only two subsystems were predicted in pHN9-2; these were Sulfur Metabolism (2) and Carbohydrates (6). Detailed informa-

tion on the subsystems is illustrated in Figure S1. In addition, Microbial Nucleotide BLAST analysis against bacterial database was performed in the NCBI server to find closely associated genomes. As expected, the top-five hits of the pHN9-1 search belonged to plasmids of *P. acidilactici* with 99% identity and 70% coverage; these were *P. acidilactici* SRCM102732 (NZ_CP028250), SRCM102731 (NZ_CP028248), SRCM100313 (NZ_CP021488.1), SRCM100424 (NZ_CP021485), and SRCM101189 (NZ_CP021530), respectively. The sequence of pHN9-1 had higher similarities to plasmids of *P. pentosaceus* ATCC 25745 (NC_008525) and *Lactiplantibacillus plantarum* ATCC 8014 (NZ_CP024413) with identities over 99%, although coverages of the sequences were only 30%. BLAST identification of pHN9-2 revealed that the sequence was more similar to that of the *Lactiplantibacillus* group, with over 99% identity and 30% coverage.

3.2. Core and Pan-Genomes of P. acidilactici Strains

We searched the complete genome of *P. acidilactici* HN9 against the microbial database in the NCBI server. The hits from the BLAST search with the highest similarities had 99% identity and 90% coverage. The strains sharing the highest similarities were *P. acidilactici* SRCM103367, SRCM103444, and JQII-5. The genome diversity among related bacterial strains can be evaluated using pan-genome analysis. Interestingly, we found the ratios of core genes, soft-core genes, shell genes, and cloud genes at the genus versus the species level to be (226: 428), (123: 503), (2937: 1677), and (15,270: 4124), respectively, as shown in Figure 2. Core genes were found in >99% of the genomes, soft core genes were found in 95 to 99%, shell genes were found in 15 to 95%, and cloud genes were found in <15%.

3.3. Unique Genetic Content of HN9

The proportion of strain-specific genes found only in HN9 was around 6.16%. Among the 124 unique genes, 31 and 93 encoded proteins identified with known functions and hypothetical proteins, respectively. Strain HN9 carried several genes involved in the transport of sucrose, including *cscB_1* and *cscB_2*. These genes are related to the transport of maltose, fructose, and lactulose into the cell [71]. Further, we detected Sucrose-6-phosphate hydrolase (*sacA_1* and *sacA_2*) with 100% coverage and 95% identity of amino acid sequence to the reference sequence (WP_173819270), which enables this strain to metabolize sucrose as the sole carbon source. The other remaining genes were insertion sequence elements, IS30 family transposase ISLpl1, and IS3 family transposase IS1223, which are commonly found in the *Lactobacillus* group and other lactic acid bacteria. These genes may have been transferred among lactic acid bacteria during their evolution.

3.4. Phylogenetic Analysis and Comparison

The phylogenetic trees were constructed based on the pan and core genes (Figure 3 and Figure S2) at the genus level, with a total of 129 strains obtained from the NCBI database. Figure 3 shows that *P. acidilactici* HN9 is grouped together with *P. acidilactici* members and explicitly separated from other groups. However, the genus level phylogenetic tree could not distinguish among the members of the *P. acidilactici* group. Therefore, we constructed another phylogenetic tree based on the core genes of all strains found in the *P. acidilactici* group to further analyze the phylogenetic relationship among the *P. acidilactici* strains. The tree was built based on 426 core genes sharing a similarity of over 60% of amino acid sequences, as shown in Figure S3. HN9 showed the closest relationship with IRZ12B, with the bootstrap value of each node being 100 in this clade. IRZ12B was also isolated from food in Iran under BioSample number SAMN15337232. IRZ12B and HN9 shared 1653 core genes (85%) and contained the bacteriocin-encoding gene, Enterolysin A. IRZ12B contained over 300 unique genes compared to HN9, from which 195 were identified as uncharacterized proteins. Moreover, HN9 was close to other strains, including JQII-5, S7, SRCM100313, SRCM100320, SRCM101189, and SRCM100424, which formed the nearest clade.
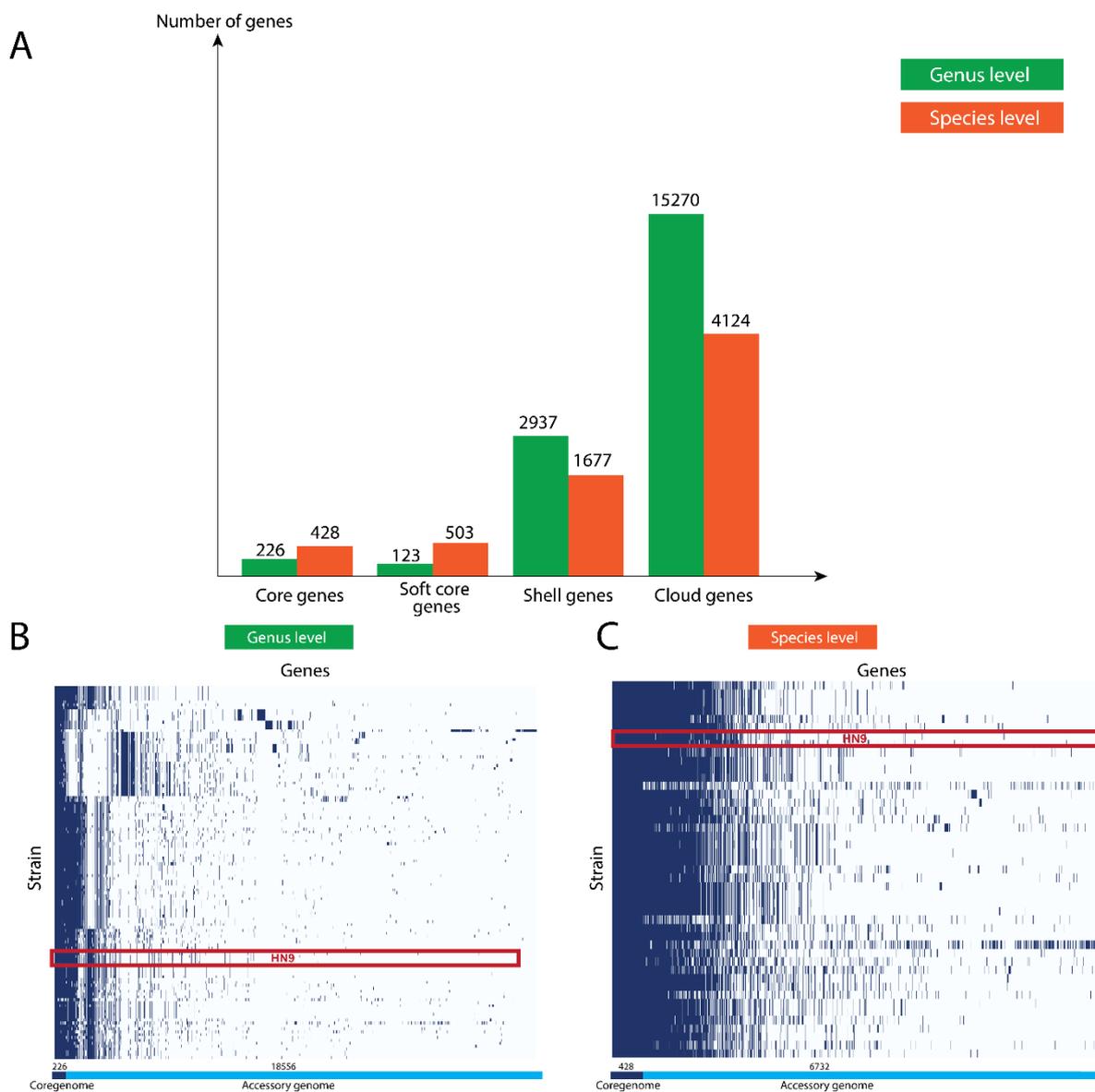
Figure 2. Pan-genome analyses of *P. acidilactici* HN9 at species and genus levels determined using Roary software. 45 and 129 strains were compared at the species and genus levels, respectively. (**A**) The summary of pan-genome analyses at both levels. (**B**,**C**) plots of pan-genome matrices-based genus and species levels, respectively. The presence of each gene in each strain is indicated by the blue color, while the white color indicates the absence of the gene from the strain. The red box represents a set of genes of *P. acidilactici* HN9.

3.5. Multiple Genome Alignments Analysis

Multiple genome alignments represent evolutionary changes in the nucleotides by aligning homologous regions of sequences [57]. Therefore, four genomes that were closest to that of HN9 were selected based on their locations in the phylogenetic tree and the level of completeness of their assembly, as shown in Figure 4. Only complete genomes were selected for multiple genome alignments; the following strains were included: JQII-5, SRCM100313, SRCM101189, and SRCM100424. Multiple genome alignments showed that the five genomes shared three main regions Strain JQII-5, SRCM100313, SRCM101189, and SRCM100424, shared certain similarities. However, strain HN9 was different from these strains, as shown in the lime colored region of Figure 4; the region of similarity among the four other strains was replaced by various phage proteins, such as phage major capsid protein, phage portal, and phage terminase, in HN9. This illustrates how this strain

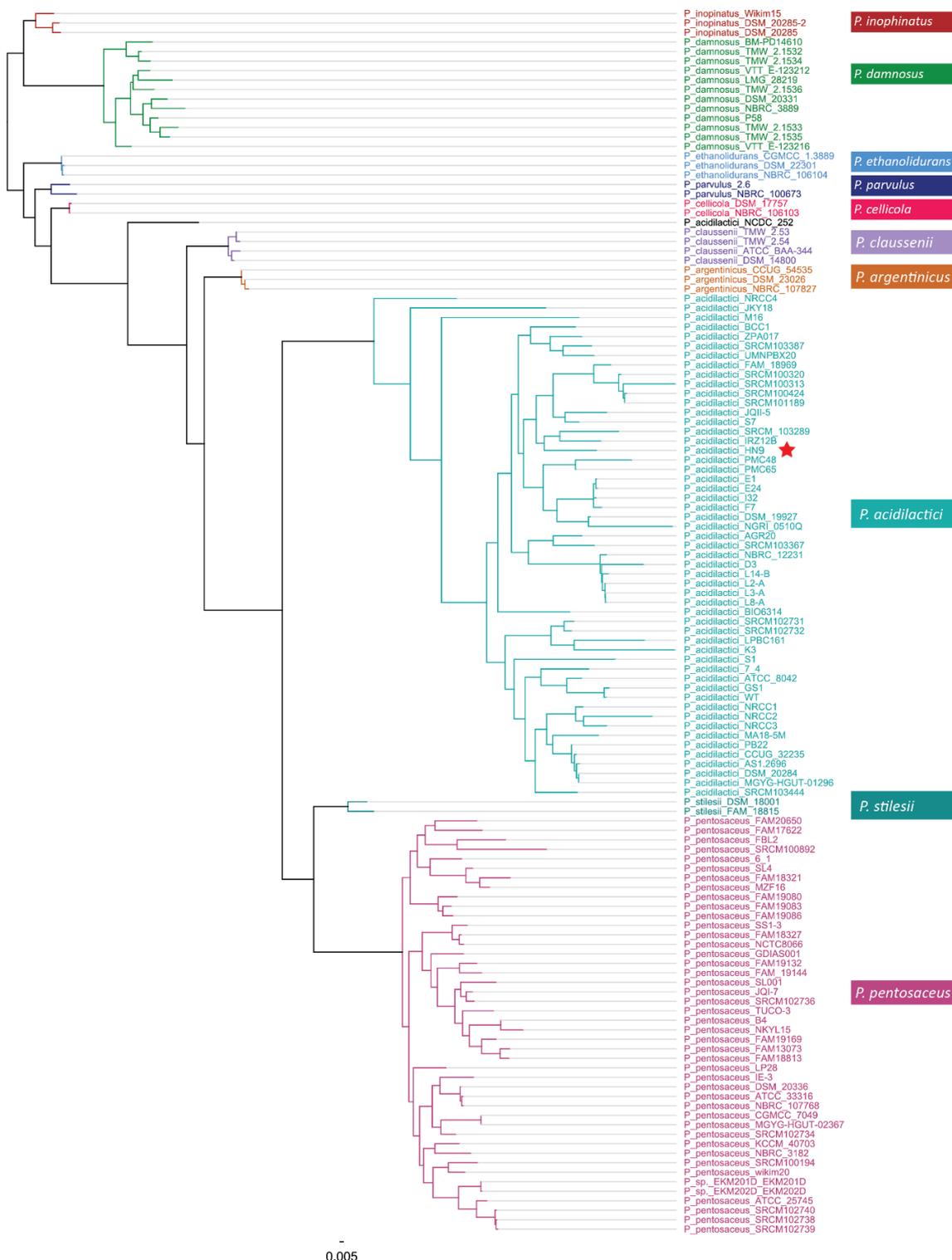diverges from the other strains and provides important evidence of HN9 reconstruction during its evolution over time.

Figure 3. Phylogenetic analysis of *Pediococcus acidilactici* HN9 and other strains at the genus level, based on 33,568 pan genes from 129 strains. The tree was constructed using the neighbor-joining method with a bootstrap of 500 replications by concatenating the single-copy marker genes computed using Roary. The scale bar represents an evolutionary distance. The red star represents the location of *P. acidilactici HN9*.
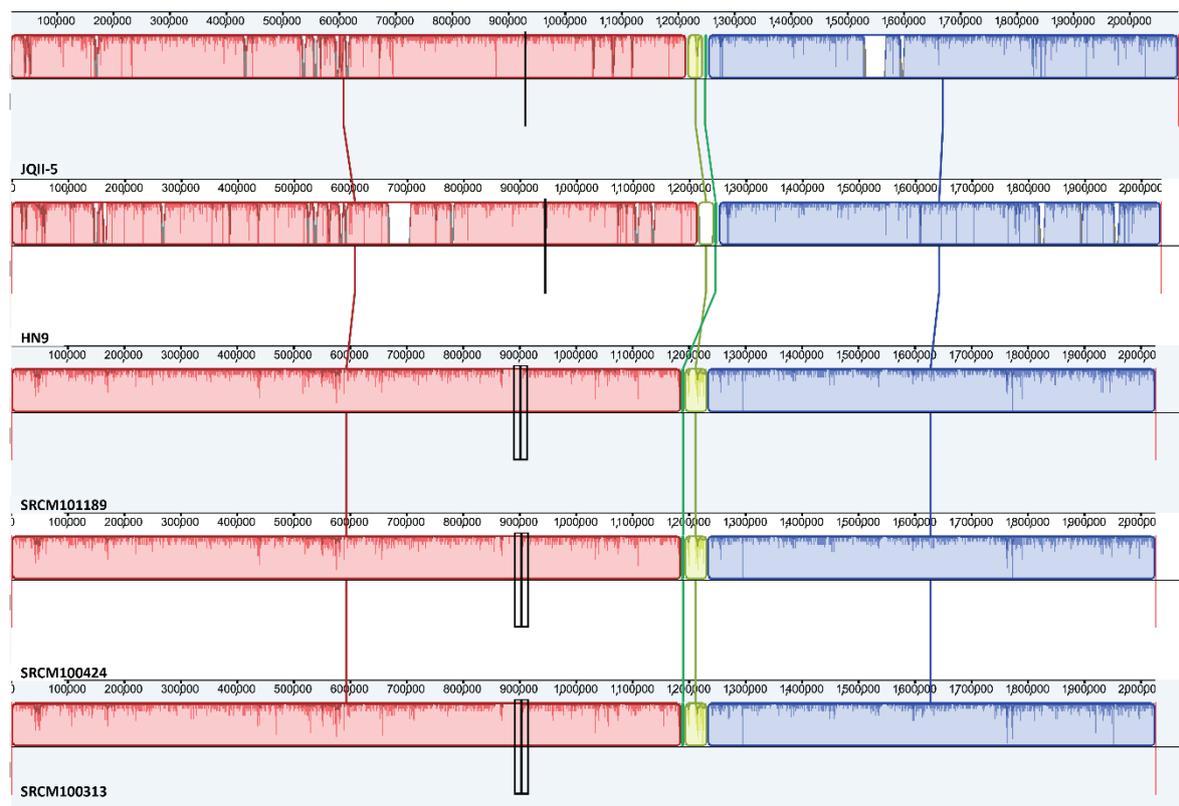
Figure 4. Multiple genome alignment of *Pediococcus acidilactici* strains, including HN9, JQII-5, SRCM100313, SRCM101189, and SRCM100424. Visualization of alignment is arranged into one horizontal panel per genome sequence, with the label of the genome sequence name on the bottom-left of each panel. The homologous blocks are represented with the same color and are connected within the genome by lines.

3.6. Genome Stability

To evaluate genome stability, we investigated the presence of prophages, mobile genetic elements, and insertion elements. In HN9, we found three prophage regions predicted by the PHASTER server, including one incomplete and two intact prophage regions. The information of each region is given in Table 2. Most of the components in every phage region are hypothetical proteins and phage-related proteins from various genera, including *Lactobacillus*, *Streptococcus*, and *Bacillus*.

Table 2. Prophage information of *Pediococcus acidilactici* HN9.

Region	Region Length	Completeness	TotalCDS	Region Position	Most Common Phage	GC %
1	16.6Kb	incomplete	30	665577- 682211	PHAGE_Lactob_Lc_Nu_NC_007501(3)	37.16
2	21.9Kb	intact	24	681401-703303	PHAGE_Strept_Abc2_NC_013645(6)	39.82
3	43.5Kb	intact	56	1210062-1253564	PHAGE_Lactob_iLp1308_NC_028911(10)	39.46

Insertion sequences (IS) in HN9 were identified using ISfinder. Most IS originated in the genera *Lactobacillus* (*L. plantarum*, *L. fructivorans*, *L. johnsonii*, and *L. sakei*), and *Pediococcus* (*P. pentosaceus*), including ISLpl1, ISPp1, ISLfr1, ISLjo2, and ISLjo2. We also found IS from the species *Enterococcus hirae* (IS1310), *Bacillus megaterium* (ISBame1), *Leuconostoc mesenteroides* (IS1165), and *Weissella cibaria* (ISWci1), revealing the mechanism of HN9 adaptation for survival via integration of these elements to its genome. In addition, two CRISPR loci were identified in the genome of *P. acidilactici* HN9, including type II-A of CRISPR/CRISPR-associated (Cas). We found two candidates of the CRISPR

region (1 confirmed, 1 questionable) in the HN9 genome, which belonged to the type II-A CRISPR/CRISPR-associated (Cas) system that includes *cas1, cas2, cas9, and csn2*.

3.7. Identification of Antimicrobial Peptides

To identify the bacteriocin loci, BAGEL4 server was used to perform genome mining and locate the gene encoding the antimicrobial compound. A bacteriocin-encoding locus was identified in the genome of *P. acidilactici* HN9, as shown in Figure 5. Enterolysin A was identified as a core protein in the Area of Interest (AOI) with 100% coverage and 68.7% amino acid sequence identity to the reference sequence (WP_011673272) [72], as shown in Figure S4. It was surrounded by several ORFs, including Antitoxin SocA, Autolysin, and an uncharacterized protein. Enterolysin A is a cell wall-degrading bacteriocin found in several bacterial strains, including *Enterococcus*, *Lactobacillus*, *Streptococcus* [73], and *Pediococcus* [74]. This bacteriocin shows potential for inhibition of enterococci, pediococci, lactococci, and lactobacilli growth. Furthermore, we constructed the phylogenetic tree from Enterolysin A identified in *P. acidilactici* HN9 with other 14 strains as shown in Figure 5B. The lengths and the similarities among these genes varied from 1140 to 3006 bp and 30 to 100%, respectively. The sequence lengths of Enterolysin A from *P. damnosus* and *P. stilesii* are in range of 1–1.8 kbp while the sizes of *P. pentosaceus* and *P. acidilactici* are around 2.9 kbp.

In addition, the genomic data of all strains in the genus *Pediococcus* were used to identify bacteriocin genes using BLASTX against the BAGEL4 database. The BLASTX parameters used for the identification included an E-value cutoff-point at 1e-20, and identity percentage and minimum coverage at 70%. Among the 130 strains (including *P. acidilactici* HN9), 40 contained at least one bacteriocin-encoding gene in class II, and 22 strains contained at least one bacteriocin-encoding gene in class III, as shown in Figure 6. *P. Pentosaceus* FAM13073, FAM17622, FAM19086, and GDIAS001 were the most productive bacteriocin producer, encoding 10 different types of bacteriocin. The total number of distinct bacteriocins in class II encoded by the genus *Pediococcus* was 26, with Penocin A, Pediocin, and Pediocin PA being the most common bacteriocins found in this genus, which were produced by 26, 22, and 22 strains, respectively. In contrast, only 4 different class III bacteriocins were identified in the *Pediococcus* group. Enterolysin A was the most common bacteriocin in this genus, which was identified, in 95% of strains encoded class III bacteriocin. In addition, we changed the parameters to scan a wider result by setting the cutoff-point to 1e-10 and identity percentage and minimum coverage to 50%. As a result, we identified 55 strains contained at least one bacteriocin-encoding gene in class II, and 83 contained at least one bacteriocin-encoding gene in class III, as shown in Figure S5.

3.8. Identification of Carbohydrate-Active enZyme (CAZyme) in the P. acidilactici HN9 Genome

The CAZyme is a sequence-based classification of enzymes that enable to synthesize or breakdown complex carbohydrates and glycoconjugates [75]. We identified 67 genes in five families of CAZymes including Glycoside Hydrolase (GH), Glycosyl Transferase (GT), Carbohydrate Esterase (CE), Auxiliary Activity (AA), and Carbohydrate-Binding Module (CBM). The largest family found in the HN9 genome was the GH, which encoded 33 genes in 17 different families (Figure 7).

3.9. In Silico Comparison of Genes Associated with in Meat Fermentation

There was a report that showed some genes highly expressed during meat fermentation including L-Asparaginase 2 (*asnA2*), Phosphoglucomutast (*pgm*), Undecaprenyl pyrophosphate synthetase (*uppS*), Asparaginyl-tRNA synthase (*asnS*), Transposase of IS1520 orfA (*tnpA1*-IS1520), Regulator of class III heat shock genes (*ctsR*), and Response regulator, two-component system (rrp-3) [76]. Therefore, we compared those genes among all bacteria in the *Pediococcus* strains. Also, R-M system-related genes, the presence of the bacteriocin-encoding genes, were included in this analysis to distinguish the ability in the fermentation of each strain, as shown in Figure 8. Not surprisingly, we found that all strains encode the key genes in fermentation, including *ldhL*, *ldhD*. Moreover, every bacterial strain

contained various genes such as *pgm*, *uppS*, *asnS*, *ctsR*, *rrp-3,* and several unknown functions of hypothetical proteins. Only *asnA* and tnpA were present in some bacteria. Surprisingly, we found that only 46 strains had *hsdM*, *hsdR*, and *hsdS* which are important genes in the R-M system. We found those three genes in the HN9 genome.
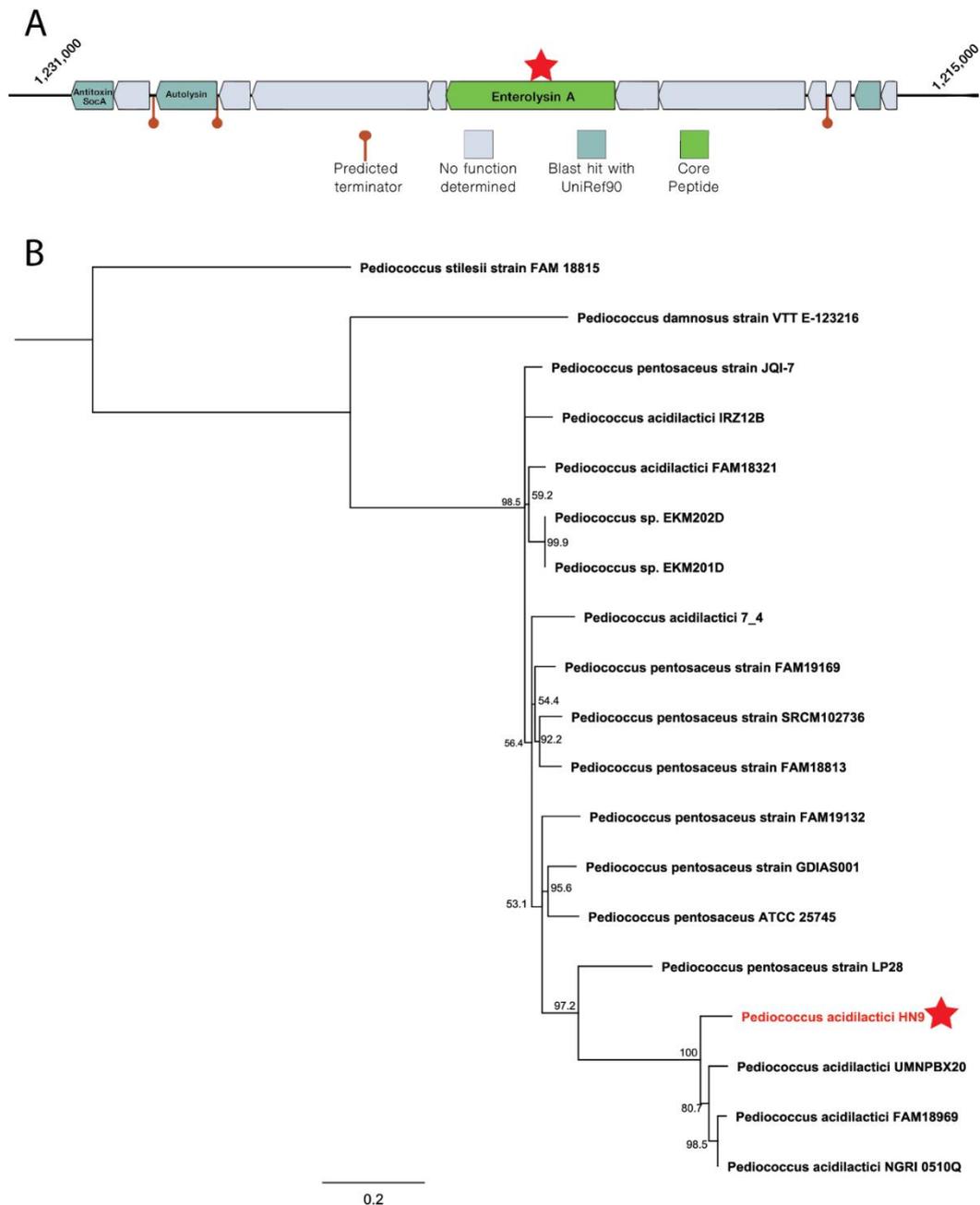

Figure 5. Antimicrobial peptides identified in the HN9 and its phylogenetic relation with other strains. (**A**) Area of Interest (AOI) of *Pediococcus acidilactici* in the genome within a range of 20 kb. Enterolysin A was found as a core protein surrounded by orf00004 (Antitoxin SocA), orf00007 (Autolysin), and orf00023 (uncharacterized protein). The location of Enterolysin A is marked with the red star. (**B**) Phylogenetic analysis of Enterolysin A sequences from the HN9 and other 19 strains. The red star represents the location of *P. acidilactici HN9* in the phylogenetic tree.

Figure 6. Identification of bacteriocin-encoding genes from all bacterial strains in the genus *Pediococcus* using BLASTX analysis against BAGEL4 database in (**A**) class II and (**B**) class III. The X-axis represents the name of the bacterial strain. The Y-axis represents the total number of bacteriocin identified in each strain. The different colors in each bar represent the different bacteriocins found in each strain, and the rectangle boxes represent the diversity of bacteriocin encoded in different genus. The red star indicates the location of *P. acidilactici HN9*.

3.10. Safety Evaluation and Identification of Antibiotic Resistance Genes

The genomic information of HN9 was used to evaluate the safety of using this strain in food applications. Several tools and databases were used to search for AMR genes, virulence factors, and pathogenic genes. Using the parameters "Perfect hit and Strict hit only" and "High-quality/coverage," we found no hit in both chromosome and plasmids. Also, the result from the ResFinder showed no hits. However, when the parameter was changed to "Perfect hit, Strict hit and Loose hits", 122 hits were identified as AMR genes with an identity of matching region range of 21–70 % and coverage of 27–307%. By easing the searching criteria, most of these observed hits were not AMR genes such as *fusA*, *rpoB*, or *gyrB*. Further investigation is warranted to evaluate the antimicrobial susceptibility of this organism. Additionally, the virulence factors were screened using the VFDB database

with VFanalyzer [77]. Furthermore, the probability of HN9 being a pathogen was assessed by pathogenicity factor analysis using PathogenFinder. The risk scores of the HN9 genome, plasmid pHN9-1, and pHN9-2 were 0.124, 0.209, and 0.174, respectively; these values are low for being considered as human pathogen compared to known probiotic strains such as *Lactobacillus johnsonii* ZLJ010 (0.193) [78], *Lactobacillus casei* DSM 20011 (0.198) [79], *Lactobacillus reuteri* PNW1 (0.217) [80], and known pathogens such as *Kosakonia radicincitans* DSM 107547 (0.71) [81] and *Enterobacter* sp. C6 (0.755) [82]. Moreover, we found no match of proteins from pathogenic families in HN9. Altogether, the results of our investigations on genomic information of *P. acidilactici* HN9 indicate the safety of HN9 for use in food and beverage applications.
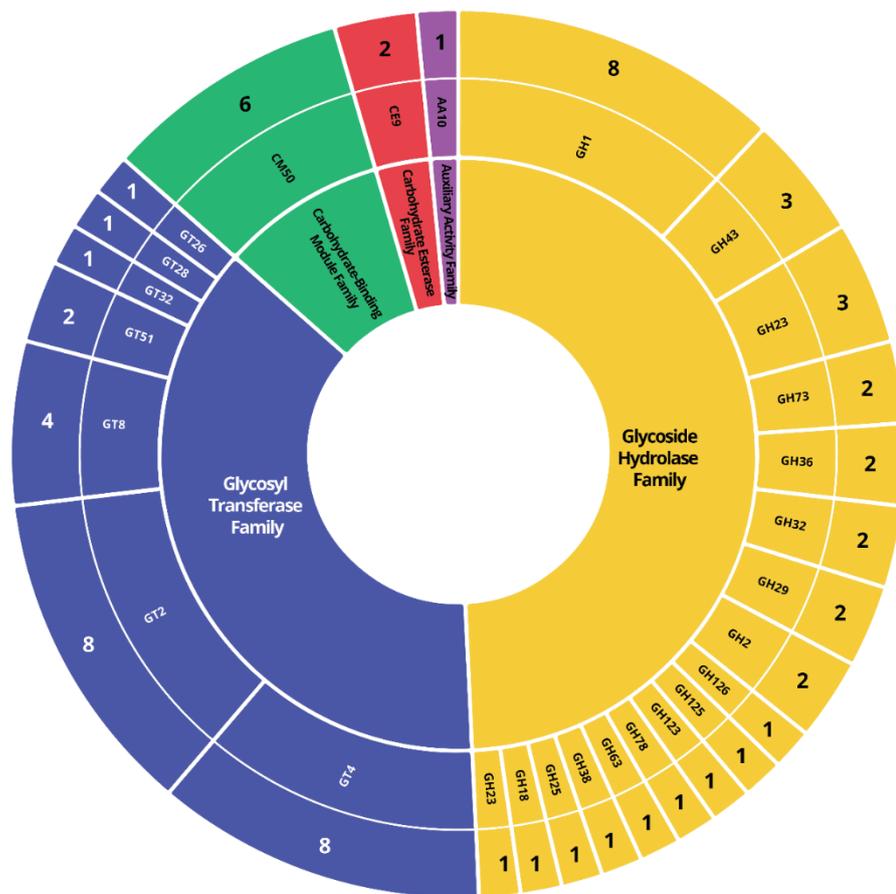
Figure 7. CAZymes distribution in HN9 genome. CAZymes were assigned by searching against the CAZy database using the dbCAN webserver. Different colors represent different classes of CAZymes found in the genome. The representation from the inner to outer rings are, CAZyme classes, CAZyme families, and the number of genes identified in each family, respectively.
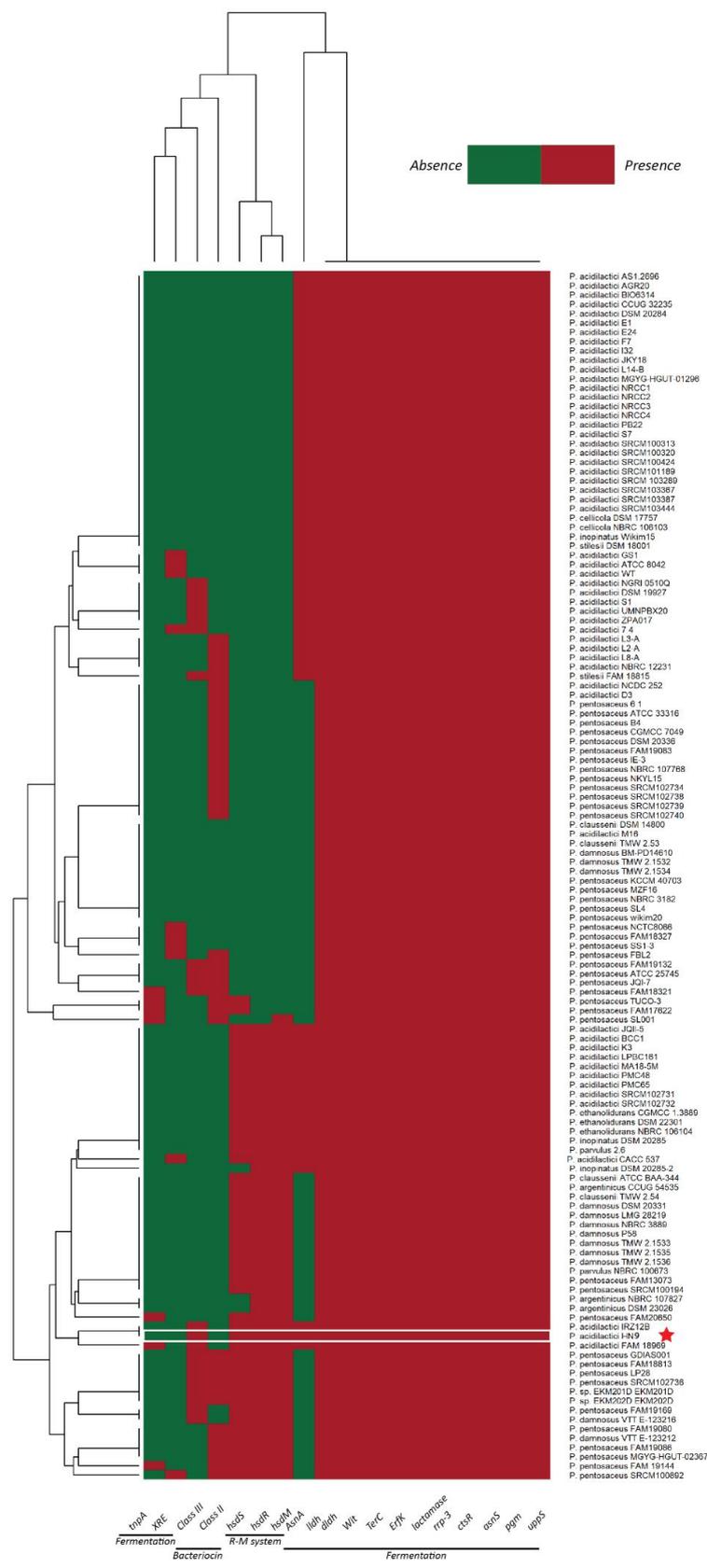
Figure 8. Gene presence-absence heatmap of fermentation related genes. 19 genes were selected to compare the ability in fermentation of *Pediococcus* 129 strains. Red color represents the presence of the gene and green color indicates the absence of the gene in the genome. The red star indicates the location of *P. acidilactici HN9*.

4. Discussion

In this study, the complete genome of *P. acidilactici* HN9 and comparative genomic analyses among related species were carried out. The *acidilactici* strains are commonly found in various fermented foods, including vegetables, dairy products, and meat. It has been previously investigated and used as a probiotic to treat constipation [82], diarrhea [5], relieve stress [12], and improve the immune response in human and animals [13]. The functional annotation of HN9 was annotated based on the predicted genes. The subsystems of carbohydrates and protein metabolism were the major subsystem identified in the HN9, which were involved directly in the biological characteristics of this strain.

Carbohydrate-Active enZymes (CAZymes) are directly involved in biosynthesis, binding, and catabolism of carbohydrates. These enzymes, therefore, play a crucial role in sugar metabolism [75]. The highest number of CAZymes found in the HN9 genome belonged to the GH class. These enzymes are primarily involved in carbohydrate metabolism such as hydrolyzing complex carbohydrates [83,84] and regulating mass degradation during fermentation [85]. In addition, we identified the key enzymes associated with homo-fermentation, including D- and L- lactate dehydrogenases (LDH-D and LDH-L, respectively). These enzymes produce D- and L-lactate, which are key products in the fermentation process of the Embden-Meyerhof-Parnas (EMP) pathway. This suggests that the HN9 can produce important fermentation end-products for producing fermented beef products [86].

The pan-genome analysis was performed to reveal the genomic diversity among *Pediococcus* strains at both genus and species level. Interestingly, we found that the bacterial strains in this genus shared only ~0.7% core genes (226 from 33,568) and contained a very large portion of cloud genes (~45%). Meanwhile, the number of core and cloud genes in only *P. acidilactici* strains were ~6% and ~60%, respectively. The large proportion of cloud genes reveals a remarkably high level of genetic diversity at both genus and species level. Not surprisingly, we found that many bacterial strains contain various mobile genetic elements including plasmids and prophage regions. This can further suggest that HGT could be a key to genomic diversity in this genus [24,74]. Furthermore, the pan-genome analysis also reveals that the trend of pan genes is still increasing and tend to be the "open" genome in both genus and species level as shown in Figure S6. This can indicate that genomes of *Pediococcus* strains are still needed to be sequenced to explore more undiscovered genes and study the genetic diversity among these bacterial strains in this genus.

Bacteriocins are antimicrobial peptides produced by various bacterial species in their ecological niches to inhibit the growth of similar or closely related bacterial strains [87]. Most bacteriocins are produced by gram-positive bacteria, although few are produced by gram-negative bacterial strains [88]. The meta-analysis showed that *Pediococcus* members are potential bacteriocin-producing strains. Most strains in *P. pentosaceus* encode class II and III bacteriocin, while *P. acidilactici* strains mostly encode class III bacteriocin. Class II bacteriocins are small non-lantibiotics peptides comprising approximately 30–70 amino acid residues. Meanwhile, those in class III are large, protein size, bacteriocins [89]. In HN9, we identified a class III bacteriocin, Enterolysin A, which can inhibit the growth of various bacteria including enterococci, pediococci, lactococci, and lactobacilli by degrading the cell wall [73]. Enterolysin A was found in only 20 strains (15.75%) from 129 strains as shown in Figure 6B. In the phylogenetic analysis of strains harboring Enterolysin A, *P. acidilactici* HN9 diverged from other species including *P. pentosaceus*, *P. damnosus*, and *P. stilesii* to form a distinct clade together with three other *P. acidilactici* strains as shown in Figure 5B. In addition, some of *P. acidilactici* strains were found in the same clade as *P. pentosaceus*. This result shows how close these two species are in terms of evolution. However, the diversity of gene content of all bacterial strains is extremely high since different species encode various sizes of the bacteriocin (1–3 kbp).

Various molecular mechanisms, such as horizontal gene transfer, genome rearrangement, and the presence of mobile genetic elements, could alter the genome stability in

bacteria [90,91]. A prophage is a DNA region inserted and integrated into the bacterial DNA chromosome or plasmid [92]. This genetic component is commonly found in various bacterial strains, including *Pediococcus* species [24,74,93]. Interestingly, we found two phage lysins integrated into the prophage region of the genome. Phage lysin has been reported to inhibit pathogenic bacteria [94] and is used in dairy production [95]. The presence of this genetic element could elucidate the mechanism of maintaining stability and adapting to environmental changes.

Bacteriophage infection is a serious problem with fermentation failure. It can easily disrupt commercial fermentation by inhibiting or reducing starter culture activities. Some bacteria have defense mechanisms including Restriction-Modification (R-M) and CRISPRs System to prevent themselves against foreign DNA from other organisms. For *P. acidilactici* HN9, we found two R subunits, two M subunits, and one S subunit of the R-M system. The S subunit is a key genetic element to specify and recognize the DNA sequence. The R subunit cleavages the recognized phage DNA and the M subunit plays an important role in methylation reaction. Interestingly, we found both the R-M system and bacteriocin-encoding gene in the HN9. Among *Pediococcus* strains, we found only 12.4% (16 strains) that have these two properties in the genome. Hence, the HN9 could prevent fermentation failure from the bacteriophage and pathogen infection during the fermentation process.

In conclusion, our investigation indicates that many *Pediococcus* strains are good candidates for use as a starter culture in the meat/beverage fermentation industry. They encode various genetic elements that are very beneficial and could promote the fermentation process such as the R-M system, CRISPR/Cas system, bacteriocin-encoding genes. These systems play key roles in preventing the integration of virulence factors, pathogenic genes, antimicrobial resistance (AMR) genes via horizontal gene transfer during its evolution. Also, they can inhibit and prevent the growth of spoilage bacteria and foodborne pathogens. For the HN9 genome, we found no match of proteins from pathogenic families and antibiotic resistance genes and contain many beneficial genes that help promote the meat fermentation for producing functional fermented meats. These results of our investigation suggest that *P. acidilactici* HN9 could be a great potential starter culture for food and beverage fermentation applications.

References

1. Baros, R.R.; Carvalho, M.D.S.; Peralta, J.M.; Facklam, R.R.; Teixeira, L.M. Phenotypic and genotypic characterization of *Pediococcus* strains isolated from human clinical sources. *J. Clin. Microbiol.* **2001**, *39*, 1241–1246. [CrossRef] [PubMed]
2. Cai, Y.; Kumai, S.; Ogawa, M.; Benno, Y.; Nakase, T. Characterization and identification of *Pediococcus* species isolated from forage crops and their application for silage preparation. *Appl. Environ. Microbiol.* **1999**, *65*, 2901–2906. [CrossRef] [PubMed]
3. Barbosa, J.; Borges, S.; Teixeira, P. *Pediococcus acidilactici* as a potential probiotic to be used in food industry. *Int. J. Food Sci. Tech.* **2015**, *50*, 1151–1157. [CrossRef]
4. Kaya, H.I.; Simsek, O. Characterization of *Pediococcus acidilactici* PFC69 and *Lactococcus lactis* PFC77 Bacteriocins and Their Antimicrobial Activities in Tarhana Fermentation. *Microorganisms* **2020**, *8*, 1083. [CrossRef]
5. Balgir, P.P.; Kaur, B.; Kaur, T.; Daroch, N.; Kaur, G. In Vitro and In Vivo Survival and Colonic Adhesion of *Pediococcus acidilactici* MTCC5101 in Human Gut. *BioMed Res. Int.* **2013**, *2013*, 583850. [CrossRef]
6. Song, D.H.; Lee, J.M.; Chung, K.H.; An, J.H. Penicillin Binding Protein from *Pediococcus acidilactici* Isolated from Nuruk for Food Biopreservative. *Iran. J. Public Health* **2018**, *47*, 1653–1659.
7. Papagianni, M.; Anastasiadou, S. Pediocins: The bacteriocins of Pediococci. Sources, production, properties and applications. *Microb. Cell Factories* **2009**, *8*, 3. [CrossRef]
8. Beaulieu, L.; Tolkatchev, D.; Jette, J.F.; Groleau, D.; Subirade, M. Production of active pediocin PA-1 in *Escherichia coli* using a thioredoxin gene fusion expression approach: Cloning, expression, purification, and characterization. *Can. J. Microbiol.* **2007**, *53*, 1246–1258. [CrossRef]
9. Anastasiadou, S.; Papagianni, M.; Filiousis, G.; Ambrosiadis, I.; Koidis, P. Pediocin SA-1, an antimicrobial peptide from *Pediococcus acidilactici* NRRL B5627: Production conditions, purification and characterization. *Bioresour. Technol.* **2008**, *99*, 5384–5390. [CrossRef]
10. Abbasiliasi, S.; Tan, J.S.; Bashokouh, F.; Ibrahim, T.A.T.; Mustafa, S.; Vakhshiteh, F.; Sivasamboo, S.; Ariff, A.B. In vitro assessment of *Pediococcus acidilactici* Kp10 for its potential use in the food industry. *BMC Microbiol.* **2017**, *17*, 121. [CrossRef]
11. Othman, M.; Ariff, A.B.; Wasoh, H.; Kapri, M.R.; Halim, M. Strategies for improving production performance of probiotic *Pediococcus acidilactici* viable cell by overcoming lactic acid inhibition. *AMB Express* **2017**, *7*, 215. [CrossRef] [PubMed]
12. Castex, M.; Lemaire, P.; Wabete, N.; Chim, L. Effect of probiotic *Pediococcus acidilactici* on antioxidant defences and oxidative stress of *Litopenaeus stylirostris* under *Vibrio nigripulchritudo* challenge. *Fish Shellfish Immunol.* **2010**, *28*, 622–631. [CrossRef] [PubMed]
13. Neissi, A.; Rafiee, G.; Nematollahi, M.; Safari, O. The effect of *Pediococcus acidilactici* bacteria used as probiotic supplement on the growth and non-specific immune responses of green terror, Aequidens rivulatus. *Fish Shellfish Immunol.* **2013**, *35*, 1976–1980. [CrossRef] [PubMed]
14. Bhagat, D.; Raina, N.; Kumar, A.; Katoch, M.; Khajuria, Y.; Slathia, P.S.; Sharma, P. Probiotic properties of a phytase producing *Pediococcus acidilactici* strain SMVDUDB2 isolated from traditional fermented cheese product, Kalarei. *Sci. Rep.-Uk* **2020**, *10*, 1926. [CrossRef] [PubMed]
15. Jaramillo-Torres, A.; Rawling, M.D.; Rodiles, A.; Mikalsen, H.E.; Johansen, L.H.; Tinsley, J.; Forberg, T.; Aasum, E.; Castex, M.; Merrifield, D.L. Influence of Dietary Supplementation of Probiotic *Pediococcus acidilactici* MA18/5M During the Transition From Freshwater to Seawater on Intestinal Health and Microbiota of Atlantic Salmon (*Salmo salar* L.). *Front. Microbiol.* **2019**, *10*, 2243. [CrossRef]
16. Bampidis, V.; Azimonti, G.; Bastos, M.D.; Christensen, H.; Dusemund, B.; Kouba, M.; Durjava, M.K.; Lopez-Alonso, M.; Puente, S.L.; Marcon, F.; et al. Assessment of the application for renewal of authorisation of Bactocell (R) (*Pediococcus acidilactici* CNCM I-4622) as a feed additive for all fish and shrimps and its extension of use for all crustaceans. *EFSA J.* **2019**, *17*, e05691. [CrossRef]
17. Takata, K.; Kinoshita, M.; Okuno, T.; Moriya, M.; Kohda, T.; Honorat, J.A.; Sugimoto, T.; Kumanogoh, A.; Kayama, H.; Takeda, K.; et al. The Lactic Acid Bacterium *Pediococcus acidilactici* Suppresses Autoimmune Encephalomyelitis by Inducing IL-10-Producing Regulatory T Cells. *PLoS ONE* **2011**, *6*, e27644. [CrossRef]
18. Lee, E.; Nam, K.T.; Lee, K.W.; Lee, S.R. *Pediococcus* spp.-fermented chicken meat for dogs. *J. Anim. Sci. Technol.* **2020**, *62*, 84–93. [CrossRef]
19. Nieto-Lozano, J.C.; Reguera-Useros, J.I.; Pelaez-Martinez, M.D.; Sacristan-Perez-Minayo, G.; Gutierrez-Fernandez, A.J.; de la Torre, A.H. The effect of the pediocin PA-1 produced by *Pediococcus acidilactici* against *Listeria monocytogenes* and *Clostridium perfringens* in Spanish dry-fermented sausages and frankfurters. *Food Control* **2010**, *21*, 679–685. [CrossRef]
20. Ammor, M.S.; Mayo, B. Selection criteria for lactic acid bacteria to be used as functional starter cultures in dry sausage production: An update. *Meat Sci.* **2007**, *76*, 138–146. [CrossRef]
21. Laranjo, M.; Elias, M.; Fraqueza, M.J. The Use of Starter Cultures in Traditional Meat Products. *J. Food Qual.* **2017**, *2017*, 9546026. [CrossRef]
22. Deibel, R.H.; Wilson, G.D.; Niven, C.F., Jr. Microbiology of meat curing. IV. A lyophilized *Pediococcus cerevisiae* starter culture for fermented sausage. *Appl. Microbiol.* **1961**, *9*, 239–243. [CrossRef] [PubMed]

23. Ranjan, R.; Pandit, R.J.; Duggirala, S.M.; Joshi, C.G.; Sharma, S.; Patil, N.V. Genome sequencing of *Pediococcus acidilactici* (NRCC1), a novel isolate from dromedary camel (*Camelus dromedarius*) rumen fluid. *Ann. Microbiol.* **2018**, *68*, 103–110. [CrossRef]

24. Snauwaert, I.; Stragier, P.; De Vuyst, L.; Vandamme, P. Comparative genome analysis of *Pediococcus damnosus* LMG 28219, a strain well-adapted to the beer environment. *BMC Genom.* **2015**, *16*, 267. [CrossRef]

25. Surachat, K.; Sangket, U.; Deachamag, P.; Chotigeat, W. In silico analysis of protein toxin and bacteriocins from *Lactobacillus paracasei* SD1 genome and available online databases. *PLoS ONE* **2017**, *12*, e0183548. [CrossRef]

26. Korlach, J.; Bjornson, K.P.; Chaudhuri, B.P.; Cicero, R.L.; Flusberg, B.A.; Gray, J.J.; Holden, D.; Saxena, R.; Wegener, J.; Turner, S.W. Real-Time DNA Sequencing from Single Polymerase Molecules. *Methods Enzymol.* **2010**, *472*, 431–455. [CrossRef]

27. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **2017**, *27*, 722–736. [CrossRef]

28. Hunt, M.; De Silva, N.; Otto, T.D.; Parkhill, J.; Keane, J.A.; Harris, S.R. Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* **2015**, *16*, 294. [CrossRef]

29. Chokesajjawatee, N.; Santiyanont, P.; Chantarasakha, K.; Kocharin, K.; Thammarongtham, C.; Lertampaiporn, S.; Vorapreeda, T.; Srisuk, T.; Wongsurawat, T.; Jenjaroenpun, P.; et al. Safety Assessment of a Nham Starter Culture *Lactobacillus plantarum* BCC9546 via Whole-genome Analysis. *Sci. Rep.-Uk* **2020**, *10*, 10241. [CrossRef]

30. Chin, C.S.; Alexander, D.H.; Marks, P.; Klammer, A.A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E.E.; et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **2013**, *10*, 563–569. [CrossRef]

31. Arredondo-Alonso, S.; Willems, R.J.; van Schaik, W.; Schurch, A.C. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb. Genom.* **2017**, *3*, e000128. [CrossRef] [PubMed]

32. Houtgast, E.J.; Sima, V.M.; Bertels, K.; Al-Ars, Z. Hardware acceleration of BWA-MEM genomic short read mapping for longer read lengths. *Comput. Biol. Chem.* **2018**, *75*, 54–64. [CrossRef] [PubMed]

33. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Proc, G.P.D. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]

34. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef]

35. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST plus: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef]

36. Carattoli, A.; Hasman, H. PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS). *Methods Mol. Biol.* **2020**, *2075*, 285–294. [CrossRef]

37. Hyatt, D.; Chen, G.L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [CrossRef]

38. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.D.; Wortman, J.; Young, S.K.; et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef]

39. Waterhouse, R.M.; Seppey, M.; Simao, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **2018**, *35*, 543–548. [CrossRef]

40. Simao, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef]

41. Delcher, A.L.; Bratke, K.A.; Powers, E.C.; Salzberg, S.L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **2007**, *23*, 673–679. [CrossRef] [PubMed]

42. Brettin, T.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Olsen, G.J.; Olson, R.; Overbeek, R.; Parrello, B.; Pusch, G.D.; et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.-Uk* **2015**, *5*, 8365. [CrossRef] [PubMed]

43. Overbeek, R.; Olson, R.; Pusch, G.D.; Olsen, G.J.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Parrello, B.; Shukla, M.; et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **2014**, *42*, D206–D214. [CrossRef] [PubMed]

44. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST server: Rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75. [CrossRef] [PubMed]

45. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **2016**, *44*, 6614–6624. [CrossRef]

46. Schattner, P.; Brooks, A.N.; Lowe, T.M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **2005**, *33*, W686–W689. [CrossRef]

47. Lagesen, K.; Hallin, P.; Rodland, E.A.; Staerfeldt, H.H.; Rognes, T.; Ussery, D.W. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **2007**, *35*, 3100–3108. [CrossRef]

48. Grissa, I.; Vergnaud, G.; Pourcel, C. CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **2007**, *35*, W52–W57. [CrossRef]

49. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [CrossRef]
50. Stothard, P.; Grant, J.R.; Van Domselaar, G. Visualizing and comparing circular genomes using the CGView family of tools. *Brief. Bioinform.* **2019**, *20*, 1576–1582. [CrossRef]
51. Grant, J.R.; Stothard, P. The CGView Server: A comparative genomics tool for circular genomes. *Nucleic Acids Res.* **2008**, *36*, W181–W184. [CrossRef] [PubMed]
52. Stothard, P.; Wishart, D.S. Circular genome visualization and exploration using CGView. *Bioinformatics* **2005**, *21*, 537–539. [CrossRef] [PubMed]
53. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef] [PubMed]
54. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [CrossRef] [PubMed]
55. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]
56. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [CrossRef]
57. Darling, A.C.E.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [CrossRef]
58. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]
59. van Heel, A.J.; de Jong, A.; Song, C.X.; Viel, J.H.; Kok, J.; Kuipers, O.P. BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* **2018**, *46*, W278–W281. [CrossRef]
60. de Jong, A.; van Heel, A.J.; Kok, J.; Kuipers, O.P. BAGEL2: Mining for bacteriocins in genomic data. *Nucleic Acids Res.* **2010**, *38*, W647–W651. [CrossRef]
61. Lombard, V.; Ramulu, H.G.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [CrossRef] [PubMed]
62. Yin, Y.B.; Mao, X.Z.; Yang, J.C.; Chen, X.; Mao, F.L.; Xu, Y. dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **2012**, *40*, W445–W451. [CrossRef] [PubMed]
63. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef] [PubMed]
64. Zhang, H.; Yohe, T.; Huang, L.; Entwistle, S.; Wu, P.Z.; Yang, Z.L.; Busk, P.K.; Xu, Y.; Yin, Y.B. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **2018**, *46*, W95–W101. [CrossRef]
65. Alcock, B.P.; Raphenya, A.R.; Lau, T.T.Y.; Tsang, K.K.; Bouchard, M.; Edalatmand, A.; Huynh, W.; Nguyen, A.V.; Cheng, A.A.; Liu, S.; et al. CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **2020**, *48*, D517–D525. [CrossRef]
66. Bortolaia, V.; Kaas, R.S.; Ruppe, E.; Roberts, M.C.; Schwarz, S.; Cattoir, V.; Philippon, A.; Allesoe, R.L.; Rebelo, A.R.; Florensa, A.F.; et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **2020**, *75*, 3491–3500. [CrossRef]
67. Liu, B.; Zheng, D.D.; Jin, Q.; Chen, L.H.; Yang, J. VFDB 2019: A comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* **2019**, *47*, D687–D692. [CrossRef]
68. Cosentino, S.; Voldby Larsen, M.; Moller Aarestrup, F.; Lund, O. PathogenFinder–distinguishing friend from foe using bacterial whole genome sequence data. *PLoS ONE* **2013**, *8*, e77302. [CrossRef]
69. Arndt, D.; Grant, J.R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.J.; Wishart, D.S. PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21. [CrossRef]
70. Siguier, P.; Perochon, J.; Lestrade, L.; Mahillon, J.; Chandler, M. ISfinder: The reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **2006**, *34*, D32–D36. [CrossRef]
71. Peng, Y.; Kumar, S.; Hernandez, R.L.; Jones, S.E.; Cadle, K.M.; Smith, K.P.; Varela, M.F. Evidence for the Transport of Maltose by the Sucrose Permease, CscB, of *Escherichia coli*. *J. Membr. Biol.* **2009**, *228*, 79–88. [CrossRef] [PubMed]
72. Doi, K.; Mori, K.; Tashiro, K.; Fujino, Y.; Nagayoshi, Y.; Hayashi, Y.; Kuhara, S.; Ohshima, T. Draft Genome Sequence of *Pediococcus lolii* NGRI 0510Q(T) Isolated from Ryegrass Silage. *Genome Announc.* **2013**, *1*, e00156-12. [CrossRef] [PubMed]
73. Nilsen, T.; Nes, I.F.; Holo, H. Enterolysin A, a cell wall-degrading bacteriocin from *Enterococcus faecalis* LMG 2333. *Appl. Environ. Microb.* **2003**, *69*, 2975–2984. [CrossRef] [PubMed]
74. Jiang, J.; Yang, B.; Ross, R.P.; Stanton, C.; Zhao, J.X.; Zhang, H.; Chen, W. Comparative Genomics of *Pediococcus pentosaceus* Isolated From Different Niches Reveals Genetic Diversity in Carbohydrate Metabolism and Immune System. *Front. Microbiol.* **2020**, *11*, 253. [CrossRef]
75. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Res.* **2009**, *37*, D233–D238. [CrossRef]
76. Liu, M.; Li, X.B.; Xie, Y.Z.; Bi, D.X.; Sun, J.Y.; Li, J.; Tai, C.; Deng, Z.X.; Ou, H.Y. ICEberg 2.0: An updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* **2019**, *47*, D660–D665. [CrossRef]

77. Zhang, W.; Wang, J.; Zhang, D.; Liu, H.; Wang, S.; Wang, Y.; Ji, H. Complete Genome Sequencing and Comparative Genome Characterization of *Lactobacillus johnsonii* ZLJ010, a Potential Probiotic With Health-Promoting Properties. *Front. Genet.* **2019**, *10*, 812. [CrossRef]

78. Toh, H.; Oshima, K.; Nakano, A.; Takahata, M.; Murakami, M.; Takaki, T.; Nishiyama, H.; Igimi, S.; Hattori, M.; Morita, H. Genomic Adaptation of the *Lactobacillus casei* Group. *PLoS ONE* **2013**, *8*, e75073. [CrossRef]

79. Alayande, K.A.; Aiyegoro, O.A.; Nengwekhulu, T.M.; Katata-Seru, L.; Ateba, C.N. Integrated genome-based probiotic relevance and safety evaluation of *Lactobacillus reuteri* PNW1. *PLoS ONE* **2020**, *15*, e0235873. [CrossRef]

80. Mertschnigg, T.; Patz, S.; Becker, M.; Feierl, G.; Ruppel, S.; Bunk, B.; Sproer, C.; Overmann, J.; Zarfel, G. First Report of *Kosakonia radicincitans* Bacteraemia from Europe (Austria)—Identification and Whole-Genome Sequencing of Strain DSM 107547. *Sci. Rep.-Uk* **2020**, *10*, 1948. [CrossRef]

81. Zhang, S.M.; Chaluvadi, S.R.; Bennetzen, J.L. Genome Sequence Analysis of *Enterobacter* sp. C6, Found in the Pitcher Fluids of *Sarracenia rosea*. *Microbiol. Resour. Ann.* **2020**, *9*, e01214–e01219. [CrossRef] [PubMed]

82. Herndon, C.C.; Wang, Y.P.; Lu, C.L. Targeting the gut microbiota for the treatment of irritable bowel syndrome. *Kaohsiung J. Med. Sci.* **2020**, *36*, 160–170. [CrossRef] [PubMed]

83. Stutz, A.E.; Wrodnigg, T.M. Imino Sugars and Glycosyl Hydrolases: Historical Context, Current Aspects, Emerging Trends. *Adv. Carbohydr. Chem. Biochem.* **2011**, *66*, 187–298. [CrossRef] [PubMed]

84. Zhao, G.Z.; Ding, L.L.; Pan, Z.H.; Kong, D.H.; Hadiatullah, H.; Fan, Z.C. Proteinase and glycoside hydrolase production is enhanced in solid-state fermentation by manipulating the carbon and nitrogen fluxes in *Aspergillus oryzae*. *Food Chem.* **2019**, *271*, 606–613. [CrossRef] [PubMed]

85. Papagianni, M. Metabolic engineering of lactic acid bacteria for the production of industrially important compounds. *Comput. Struct. Biotechnol. J.* **2012**, *3*, e201210003. [CrossRef] [PubMed]

86. Meade, E.; Slattery, M.A.; Garvey, M. Bacteriocins, Potent Antimicrobial Peptides and the Fight against Multi Drug Resistant Species: Resistance Is Futile? *Antibiotics (Basel)* **2020**, *9*, 32. [CrossRef]

87. Lopez-Cuellar, M.D.; Rodriguez-Hernandez, A.I.; Chavarria-Hernandez, N. LAB bacteriocin applications in the last decade. *Biotechnol. Biotechnol. Equip.* **2016**, *30*, 1039–1050. [CrossRef]

88. Johnson, E.M.; Jung, Y.G.; Jin, Y.Y.; Jayabalan, R.; Yang, S.H.; Suh, J.W. Bacteriocins as food preservatives: Challenges and emerging horizons. *Crit. Rev. Food Sci.* **2018**, *58*, 2743–2767. [CrossRef]

89. Darmon, E.; Leach, D.R. Bacterial genome instability. *Microbiol. Mol. Biol. Rev.* **2014**, *78*, 1–39. [CrossRef]

90. Laureti, L.; Matic, I.; Gutierrez, A. Bacterial Responses and Genome Instability Induced by Subinhibitory Concentrations of Antibiotics. *Antibiotics (Basel)* **2013**, *2*, 100–114. [CrossRef]

91. Canchaya, C.; Proux, C.; Fournous, G.; Bruttin, A.; Brussow, H. Prophage genomics. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 238–276. [CrossRef] [PubMed]

92. Kareem, R.A.; Razavi, S.H. Plantaricin bacteriocins: As safe alternative antimicrobial peptides in food preservation—A review. *J. Food Saf.* **2020**, *40*, e12735. [CrossRef]

93. Fenton, M.; Ross, P.; McAuliffe, O.; O'Mahony, J.; Coffey, A. Recombinant bacteriophage lysins as antibacterials. *Bioeng. Bugs.* **2010**, *1*, 9–16. [CrossRef] [PubMed]

94. O'Sullivan, L.; Bolton, D.; McAuliffe, O.; Coffey, A. The use of bacteriophages to control and detect pathogens in the dairy industry. *Int. J. Dairy Technol.* **2020**, *73*, 1–11. [CrossRef]

95. Horvath, P.; Barrangou, R. CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science* **2010**, *327*, 167–170. [CrossRef]