Computational Molecular Bioscience, 2012, 2, 61-77
http://dx.doi.org/10.4236/cmb.2012.23006 Published Online September 2012 (http://www.SciRP.org/journal/cmb)

Amino Acid Biosynthesis and Proteolysis in *Lactobacillus bulgaricus* Revisited: A Genomic Comparison

Enuo Liu[1,2], Pei Hao[3], Tomonobu Konno[2], Yao Yu[3], Munehiro Oda[1], Huajun Zheng[4,5*], Zai-Si Ji[2,5*]

[1]Graduate School of Bioresource Sciences, Nihon University, Fujisawa, Japan
[2]Division of Research and Development, Meiji Co., Ltd., Odawara, Japan
[3]Shanghai Institutes for Biological Science, Chinese Academy of Sciences, Shanghai, China
[4]Chinese National Human Genome Center at Shanghai, Shanghai, China
[5]WHO Collaborating Centre for Research in Human Reproduction, Shanghai, China
Email: [*]zai-si_ji@kta.biglobe.ne.jp, zhenghj@chgc.sh.cn

Received April 29, 2012; revised June 5, 2012; accepted June 15, 2012

ABSTRACT

The amino acid biosynthesis and proteolytic system of *Lactobacillus bulgaricus* (*L. bulgaricus*) is important for its growth in niche-specific environments, as well as for flavour formation in the food industry. Comparative analyses of 4 completed sequences of the *L. bulgaricus* strain genome on a genomic scale revealed that genes involved in amino acids synthesis were undergoing reductive evolution. However, the selected industrial strains, namely, *L. bulgaricus* 2038 and *L. bulgaricus* ND02, retained more complete genes in the amino acid synthesis and proteolytic system category than the laboratory strains, and have some unique genes and pathways for obtaining amino acids that enable these bacteria to adapt to their various environmental niches.

Keywords: Amino Acid; Biosynthesis; Proteolysis; *L. bulgaricus*

1. Introduction

The most important application of *Lactobacillus bulgaricus* (*L. bulgaricus)* is as a starter in the manufacture of various fermented dairy products. Its biochemical activity is not only capable of lactic acid production but is also responsible for protein hydrolysis and amino acid biosynthesis, which generates peptides and amino acids for bacterial growth and produces metabolites that contribute to flavor formation in fermented products [1]. The amino acid catabolism system of *L. bulgaricus* functions to balance the bacterium's requirement for amino acids, and its proteolytic system plays a key role in its growth [2,3]. In particular, the functions of cell-wall-bound proteinases and peptidases are the most important functions for cell growth under conditions containing different types of nitrogen.

Several reports describing the proteolytic system of lactic acid bacterial (LAB) with respect to their biochemical and genetic aspects [4,6] have included little information specific to *L. bulgaricus*. However, putative genetic mechanisms underlying amino acid biosynthesis and proteolysis in this bacterial species have not been studied in detail. In the past few years, 4 *L. bulgaricus* genome strains (*L. bulgaricus* ATCC11842, *L. bulgaricus* ATCC BAA365, *L. bulgaricus* 2038, and *L. bulgaricus* ND02) have been completely sequenced; in addition, some strains have been incompletely sequenced to yield information regarding the proteolytic system, which has now allowed a thorough comparative analysis of their amino acid biosynthesis pathway and proteolytic systems on a genomic scale. Furthermore, the available genomic information could provide new insights into the genetic aspects of amino acid biosynthesis and proteolysis within *L. bulgaricus* through the identification of different genetic events at the genomic level. Comparative genomics has revealed some differences between the amino acid biosynthesis and proteolytic systems within Lactobacillus, the differences that are thought to reflect the various environmental niches that these bacteria occupy [7].

In a preliminary study, we described an in silico analysis of the *L. bulgaricus* 2038 amino acid biosynthesis and proteolytic system from its complete sequence, which will be publicly available in 2012 [8]. In this study, we describe an in-depth bioinformatics analysis in which we have systematically explored the diversity of the amino acid biosynthesis and proteolytic system in 4 completely sequenced *L. bulgaricus* strains. Based on our results, we have predicted horizontally unique genes in these 4 completely sequenced *L. bulgaricus* strains, with a focus on the genes required for bacterial growth, and also those that are niche-specific. Distinctions among the

[*]Corresponding author.

bacterial amino acid biosynthesis pathways, cell-wall-bound proteinases and peptidases, peptide/amino acid transport systems, and intracellular proteinases and peptidases are described in detail as examples. Furthermore, results of comparative genomics analysis were used to explore the diversity of members of the proteolytic system in 4 *L. bulgaricus* strains by using pan-genome CGH analysis.

2. Computational Methods

2.1. Comparative Genomic Analyses

Complete genome sequences of *L. bulgaricus* (*L. bulgaricus* ATCC11842, *L. bulgaricus* ATCC BAA365, *L. bulgaricus* ND02, and *L. bulgaricus* 2038) were obtained from the NCBI microbial genome database (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html). Comparative analysis was performed using progressive MAUVE [9]. Peptidase classification was performed using MEROPS Blast Server [10] with E value 0.0001.Metabolic pathway analysis was performed based on KEGG databse [11] through bi-directional best hit method. Extracellular and transmembrane proteins were determined by SignalP3.0 [12], ConPred II [13], and PSORTb v.2.0 [14]. COG assignment was performed using RPS-BLAST againt CDD (conserved domain database) [15] with E-value 1E-3. Protein homology was determined using BlastP with identity > 20% and length coverage > 30%, and alignment of protein sequences was performed using ClustalW.

2.2. Identification of Orthologous Groups

Orthologous groups in the 4 genomes were defined using the MBGD database [16]. The phylogenetic position of *L. bulgaricus* was determined on the basis of ortholog proteins. Concatenated protein sequences were first aligned using ClustalW, and the conserved alignment blocks were then extracted using the Gblocks program [17]. A maximum-likelihood tree was built using PHYML [18] with the following parameters: 100 replications for bootstrap analysis, "JTT" for the substitution model, "estimated" for the proportion of invariable sites, "estimated" for gamma distribution parameters, "4" for the number of substitution categories, "yes" to optimize tree topology, and "BIONJ" for starting tree(s).

3. Results

3.1. Genome Features of Complete Sequenced *L. bulgaricus* Genomes

Regarding the 4 completed genomes of *L. bulgaricus,* strains ATCC11842 and ATCC BAA365 are laboratory strains, and strains 2038 and ND02 are used as industrial strains. Compared to the other 3 strains having a genome size ranging from 1856 kb to 1872 kb with no plasmid, strain ND02 exhibited the greatest genome size at 2,125,753 bp, which was approximately 250 kb larger than the other genomes, and was also accompanied by a 6223-bp plasmid. Comparative genomic analysis showed that the 4 genomes are conserved in their genome

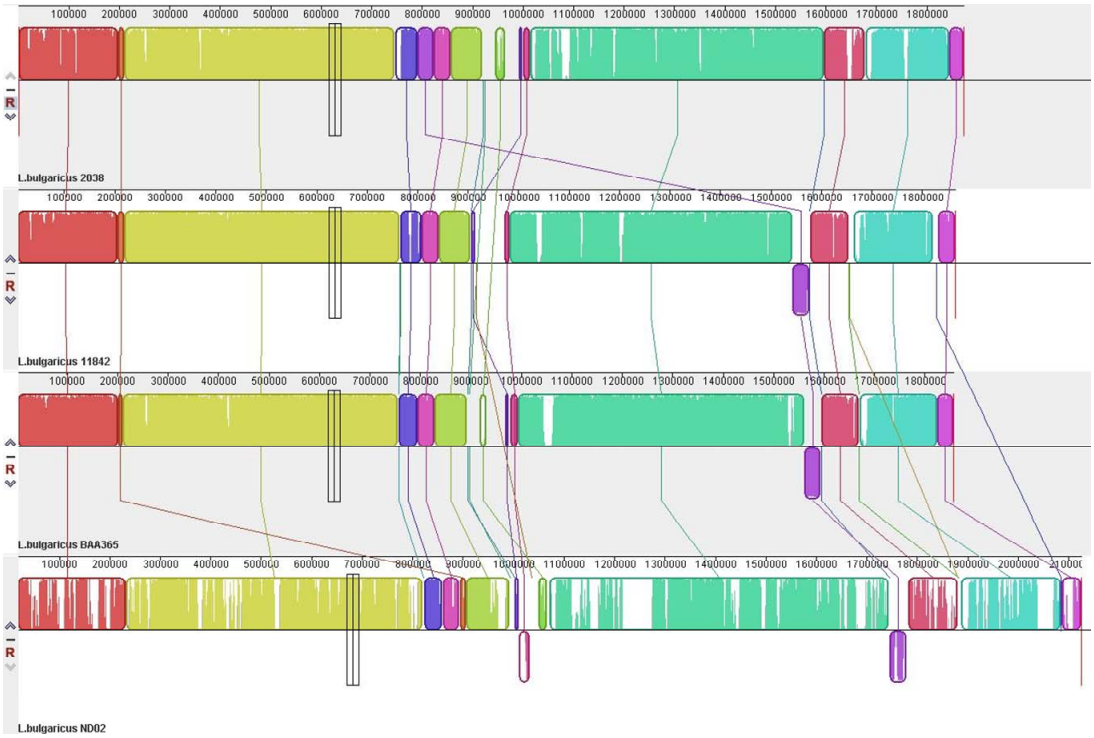

Figure 1. Alignment of the 4 *L. bulgaricus* complete genomes.

structure (**Figure 1**). There are a total of 1,466,078-bp sequences distributed in 451 blocks, which might be considered the "core genome" of *L. bulgaricus*. Meanwhile, 531,087-bp sequences were deemed strain ND02-unique; 136,451-bp sequences as strain 2038-unique; 122,623-bp sequences as strain ATCC 11842-unique; and 118,323-bp sequences as strain ATCC BAA365-unique.

Detailed analysis of the enzymes involved in de novo amino acid biosynthesis revealed 39 proteins in strain 2038, 38 proteins in strain ND02, 21 proteins in strain ATCC BAA365, and 18 proteins in strain ATCC 11842 (**Table S1**). Regarding the proteolytic systems compared among these 4 genomes, strain ND02 possessed 81 proteases and peptidases, while the number in strains 2038, ATCC 11842, and ATCC BAA365 were 72, 64, and 67, respectively (**Table S2**, **S3**). In addition, no unique protease/peptidases in each strain were found according to the cluster of ortholog (COG) among these 4 genomes, although the number of proteases/peptidases varied among different strains. For peptide/amino acid transport systems, strain 2038 possessed the highest number, which encoded 90 proteins involved in this system, while the strains ND02, ATCC BAA365, and ATCC 11842 encoded 75, 61, and 60 proteins, respectively. The highest number of proteins participating in peptide transporters in strain 2038 was mainly due to 24 genes encoding ABC-type oligopeptide transport system substratebinding protein, which had only 18 genes in ND02, 6 genes in ATCC 11842, and 5 genes in ATCC BAA365 strains.

3.2. Gene Gain or Loss in These *L. bulgaricus* Strains

A phylogenetic tree based on proteins of amino acid synthesis and proteolysis systems was constructed (**Figure 2**) for these 4 strains. The genes involved in these 2 systems might have evolved with adaption to different growth environments and passed through different selection processes, and may have played a key role in the growth of the *L. bulgaricus* strain. Strain ND02 was located on a separate branch compared to the other 3 strains, coinciding with its larger genome size and its plasmid. Strain 2038 retained more genes related to amino acid synthesis and proteolysis systems from a common ancestor genome compared to the other 2 laboratory strains.

A total of 1682 ortholog groups were identified within these 4 *L. bulgaricus* strains, in which 1232 ortholog groups existed in all 4 genomes and might be considered the "core genes" of *L. bulgaricus*. Another 268 ortholog groups were revealed in strain ND02 and in at least 1 of the other 3 strains that might be considered "dispensable" genes; these may have existed in the ancestor genome, but could have been subsequently lost during the divergence process that produced strain 2038 and/or the other 2 laboratory strains. Meanwhile, the ortholog groups among these 4 strains were analyzed, and the genes gained or lost during the evolutionary process of each strain were identified. Strain 2038 lost 67 genes after it diverged from strain ND02, while strain ATCC11842 lost over 200 genes and ATCC BAA365 lost 150 genes. Among the 67 genes lost in strain 2038, only 2 encoded for proteases, while the others had nothing to do with amino acid synthesis or proteolysis systems. However, for the 2 laboratory strains, 15 genes involved in amino acid synthesis, 12 genes encoding proteases, and 16 genes participating in transport systems were lost (**Table S4**).

3.3. Unique Genes and Pathways Involved in Amino Acid Synthesis in *L. bulgaricus*

The distribution of enzymes involved in the amino acid biosynthesis pathway in these 4 completely sequenced *L. bulgaricus* genomes is given in **Table S1**. Based on the analysis of de novo amino acid synthesis ability (**Table 1**),

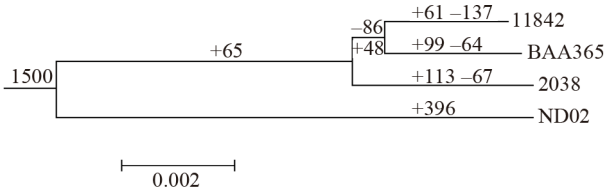

Figure 2. Phylogenetic tree showing evolutionary divergence in the 4 L. bulgaricus genomes based on the results of analyses of amino acid synthesis and paroteolysis systems. The number on each branch represented gene family number gained (+) or lost (−) during evolution.

Table 1. Statistics of amino acid synthesis in *L. bulgaricus* genomes.

	2038	ATCC11842	ATCC BAA-365	ND02
Alanine*	+	+	+	+
Aspartate*	+	+	+	+
Asparagine	+	+	+	+
Glutamate*	+	+	+	+
Glutamine	+	+	+	+
Glycine	−	−	−	−
Serine #	+	−	−	+
Threonine	+	+	+	+
Cysteine	+	−	−	−
Methionine	−	−	−	−
Valine	−	−	−	−
Leucine	−	−	−	−
Isoleucine	−	−	−	−
Lysine	+	−	−	+
Arginine	−	−	−	−
Proline	−	−	−	+
Histidine	−	−	−	−
Phenylalanine	−	−	−	−
Tyrosine	−	−	−	−
Tryptophan	−	−	−	−

*It is not a de novo synthesis pathway; #Postulated.

threonine, asparagine, and glutamine could be synthesized from l-aspartate by all 4 of these *L. bulgaricus* strains. Meanwhile, alanine could be converted from cysteine and glutamate could be transformed from aspartate, and vice versa (**Table S1**).

Proline could be biosynthesized from glutamate only in *L. bulgaricus* ND02 via 3 enzyme-catalyzed reactions. On the other hand, in the other 3 *L. bulgaricus* strains, all 3 genes encoding for the 3 enzymes were lost, except for gamma-glutamyl kinase (EC: 2.7.2.11), which was encoded by strain 2038 (LBU0712). Cysteine can only be synthesized from serine through 2 enzyme-catalyzed reactions in *L. bulgaricus* 2038. The first step is mediated by serine *O*-acetyltransferase (*cysE*, EC: 2.3.1.30, encoded by LBU1138), which was lost in the other 3 *L. bulgaricus* strains. The second step is catalyzed by cysteine synthase A (*cysK*, EC: 2.5.1.47, encoded by LBU1136 or LBU1253).

Lysine could be synthesized in the 2 industrial strains ND02 and 2038 by 7 genes encoding enzymes from dihydrodipicolinate synthase (EC: 4.2.1.52) to diaminopimelate epimerase (EC: 5.1.1.7); the genes are clustered together, and are missing in 2 laboratory strains. Most likely, an acetyl rather than a succinyl-catalyzed reaction took place with regard to enzyme DapD (EC: 2.3.1.117). The other 2 genes encoding Asd (EC: 1.2.1.11) and LysA (EC: 4.1.1.20) were positioned outside the cluster.

Another pathway unique to both industrial strains is serine biosynthesis. This pathway, which starts from 3-phosphoglycerate, includes 3 enzymes; genes encoding the first 2 enzymes, namely, phosphoglycerate dehydrogenase (*serA*, EC: 1.1.1.95) and phosphoserine aminotransferase *(serC*, EC: 2.6.1.52), were found in both strains (**Table S1**). Although the gene encoding the latter enzyme phosphoserine phosphatase (*serB*, EC: 3.1.3.3) was not found in both genomes, 2 homologous genes (LBU0835*/LDBND*_0867 and LBU0857*/LDBND*_0902) were found that might encode proteins performing a function similar to phosphoserine phosphatase.

A unique pathway responsible for amino acid synthesis exists in *L. bulgaricus*; this pathway is different from that found in any other microorganism: alanine is normally produced from pyruvate through alanine transaminase *(alt*, EC: 2.6.1.2), alanine dehydrogenase (*ald*, EC: 1.4.1.1), or aspartate 4-decarboxylase (*asdA*, EC: 4.1.1.12), but no such genes or domains are found in *L. bulgaricus.* There are 2 cysteine desulfurase genes (*nifS*, EC: 2.8.1.7) in all 4 *L. bulgaricus* genomes, suggesting that alanine might be produced from cysteine through a reaction catalyzed by cysteine desulfurase.

3.4. Sequence Comparison to Distinguish Cell-Wall-Bound Proteases and Peptidases

Six extracellular proteases are conserved among all 4 *L. bulgaricus* strains, including 2 serine proteases and 4 metalloproteases; meanwhile, both strains BAA365 and ND02 encode another unique membrane protease (**Table S5**). The protein sequences of the most important proteinase in *L. bulgaricus* proteolysis, namely, PrtB (EC: 3.4.21.83), are highly conserved among the 4 strains. This protein also has identical amino acids around the catalytic sites (Asp-30, His-94, Asn-189, Ser-425) and substrate-binding sites (Gly-131, Val-159, Gly-728, Thr-729). However, PrtB lacked a fragment of 125 amino acids (from residue 1450 to 1574 in mature protein) at the C-terminus in *L. bulgaricus* 2038 and ATCC BAA365, which is rich in Asp and Lys and located 100 amino acids ahead of the sorting signal (LPKKT) (**Figure 3**). This region contains an alpha-helix structure of H domain and the first 45 amino acids of transwall domain (W domain). As is known that H domain is able to position the N-terminal of PrtB outside the cell wall, and W domain spans the cell wall [19]. The loss of this region might affect the attachment to cell wall and folding of PrtB, thus changing the cleavage pattern on substrate.

The maturation of PrtB depends upon PrtM, which could not be identified in the 4 *L. bulgaricus* strains. Based on a search for homologs, the foldase protein PrsA (EC: 5.2.1.8) was identified in all 4 strains—PrsA is known to be an extracellular chaperon involved in extracellular protease maturation, as well as the folding and stability of subtilisins in *Bacillus subtilis* [20]. In comparison with known PrtM proteins, PrsA from the 4 *L. bulgaricus* strains was highly homologous to PrtM (**Figure 4**). The conserved domain of PrtM and PrsA had little variation, supporting the possibility that PrsA take up the role of PrtM.

Two extracellular peptidases involved in peptide degradation varied greatly among the 4 *L. bulgaricus* strains. The endopeptidase EnlA, encoded by LBU*1040*, had no ortholog in strain ATCC BAA365, and the dipeptidase PepD4, EC: 3.4.-.-, encoded by LBU1705, had no ortholog in ATCC 11842. Through protein sequence alignment, the strains 2038, ATCC 11842, and ND02 showed a difference at the C-terminus of EnlA, and no domain was observed in this variable region. For PepD4, strain ND02 lost the largest proportion of the 396-amino acid-long domain PF03577 (from the 24th to the 422nd cordon in LBU1705); thus, its peptidase activity might have also been lost, although it encoded an extracellular homolog of PepD4 (**Figure 5**).

A cell surface housekeeping protease, HtrA (EC: 3.4. 21.-), was identified in the 4 *L. bulgaricus* strains, and is known to be involved in protein maturation and turnover [21]. A comparison of the HtrA proteins within all 4 *L. bulgaricus* strains revealed high similarities among the groups (**Figure 6**).

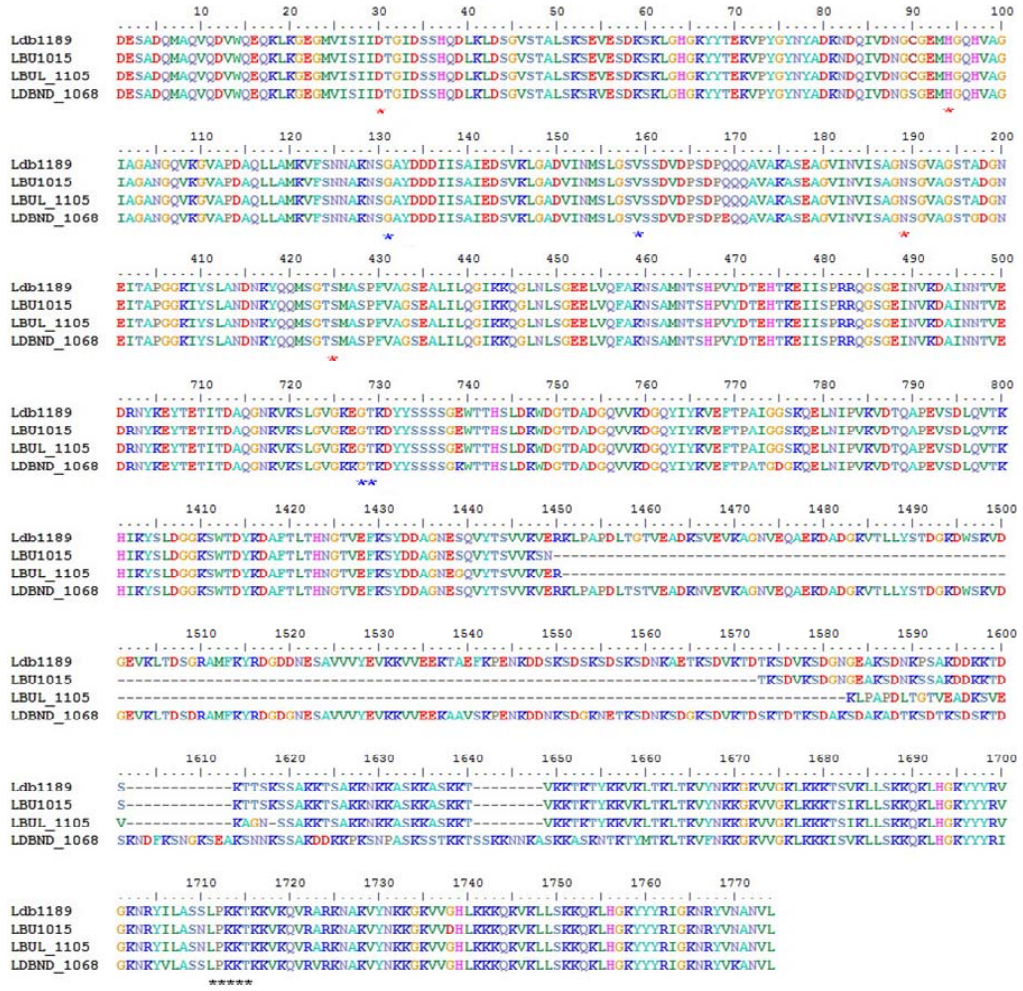

Figure 3. Sequence alignment of mature PrtB. The red star indicates conserved catalytic sites, blue star indicates substrate-binding site and black star indicates the sorting signal.

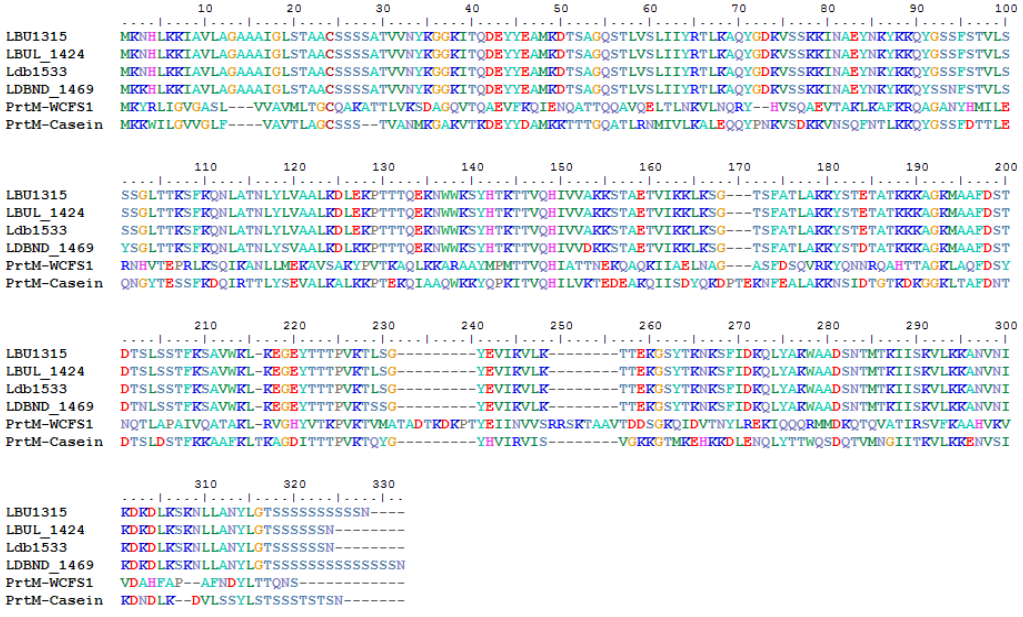

Figure 4. Alignment and domain comparisons in PrtM and PrsA.

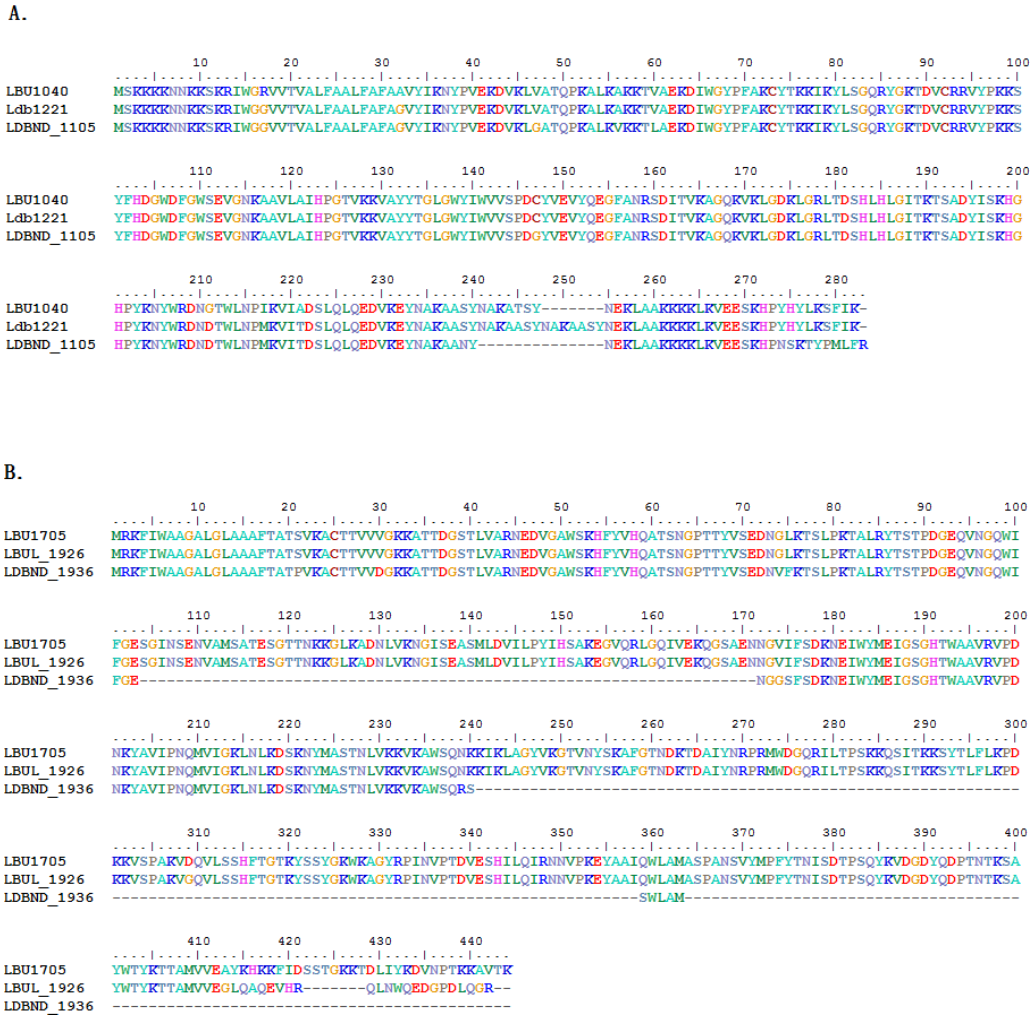

Figure 5. Alignment of the extracellular peptidase En1A (A) and PepD4 (B).

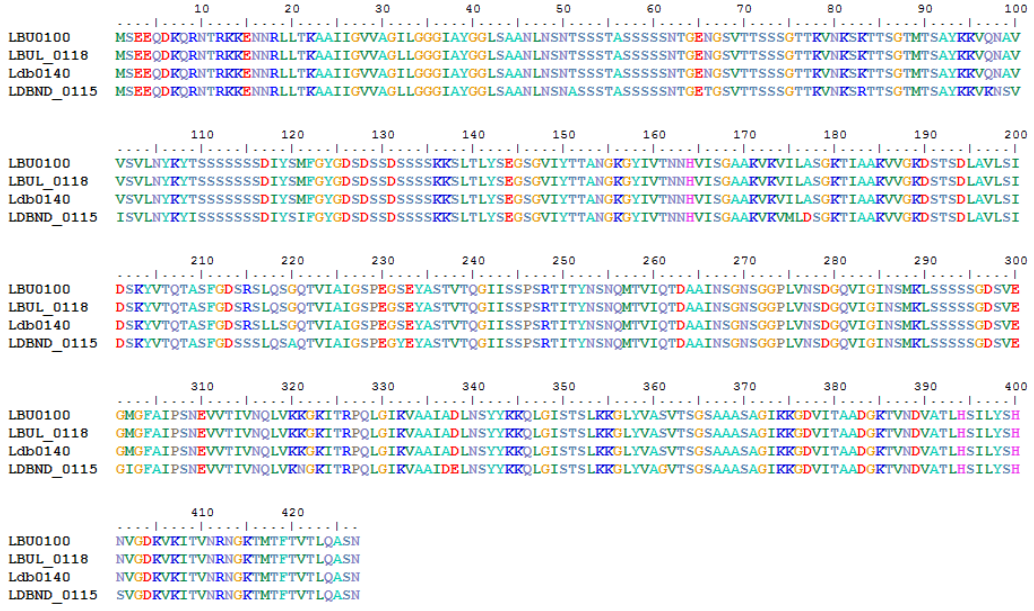

Figure 6. Alignment and domain comparisons in HtrA.

3.5. The Distribution of Intracellular Peptidase in Sequenced *L. bulgaricus* Genomes

Two intracellular proteases were highly conserved, but the intracellular peptidases showed some differences among these 4 *L. bulgaricus* strains. Thirty-three intracellular peptidases were found in *L. bulgaricus* strain 2038. Through comparison, we found that all intracellular protease/peptidases of strain ATCC11842 and ATCC BAA365 had homologs in strain 2038, while strain ND02 encoded 3 unique peptidases, including 2 serine peptidases (LDBND_1003 and LDBND_1120), and 1 metallopeptidase (LDBND_179). Furthermore, strain 2038 encoded 4 peptidases whose homolog was found only in 1 of the other 3 strains, including 1 aminopeptidase, 2 carboxypeptidases, and 1 endopeptidase (**Table S6**). Among them, the homolog of d-aminopeptidase DppA (encoded by LBU0520) was only found in strain ATCC 11842, the homolog of metal-dependent amidase/amino-acylase/carboxypeptidase (encoded by LBU0898) was only found in strain ATCC BAA365, and an amino acid amidohydrolase (encoded by LBU0934) and a metal-loendopeptidase (encoded by LBU1255) had orthologs only in strain ND02.

Two cysteine aminopeptidase genes (PepG1-LBU0223 and PepG2-LBU0224) clustered together to supply all the *L. bulgaricus* strains with cysteine. Alignment studies showed that the 2 peptidases were highly conserved (**Figure 7**).

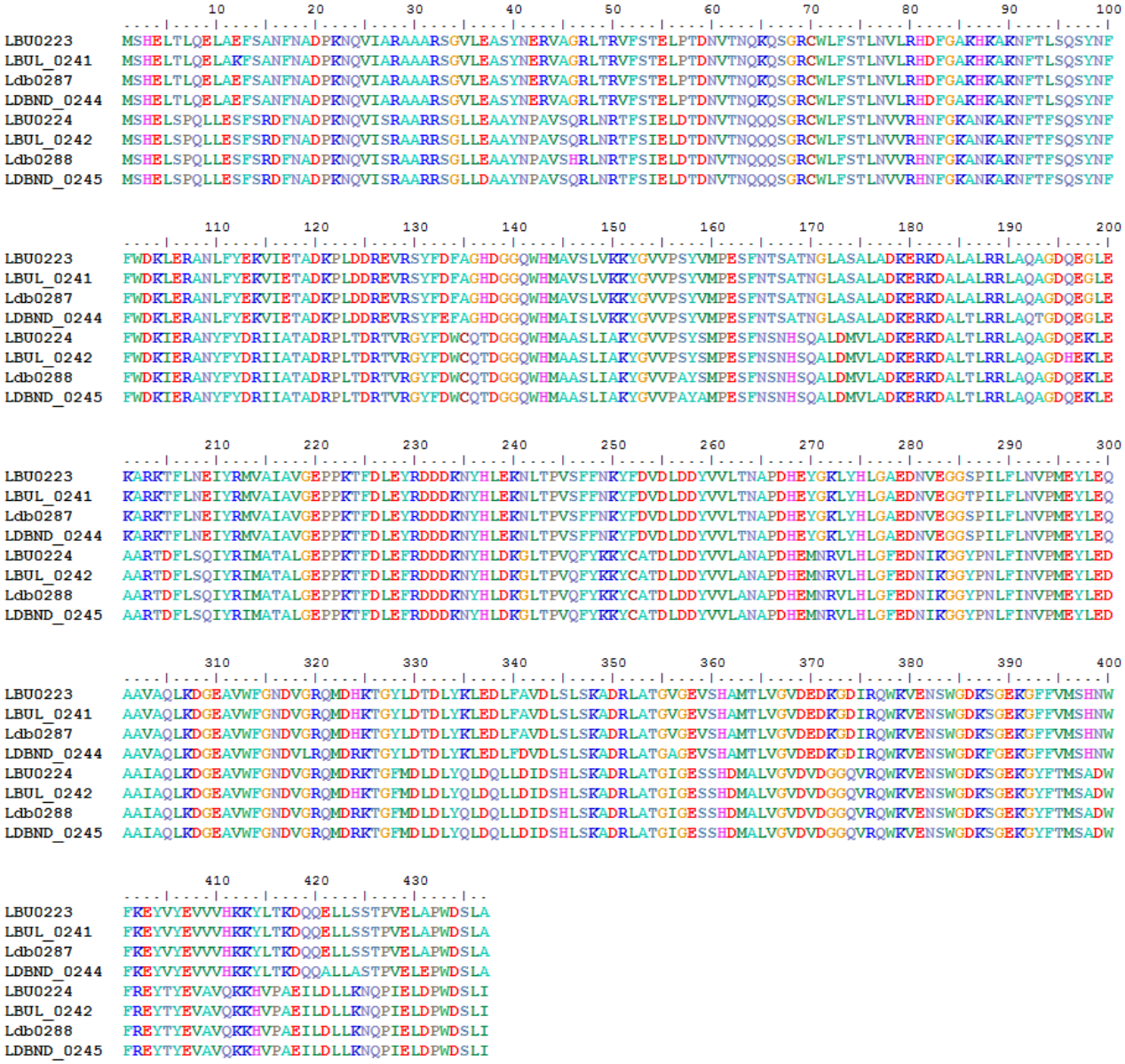

Figure 7. Alignment of cysteine aminopeptidase in the 4 *L. bulgaricus* strains.

3.6. Distribution of the Transport System in Sequenced *L. bulgaricus* Genomes

All 4 *L. bulgaricus* strains possess 2 complete oligoepde transport (Opp) systems (OppABCDF type and oppDFBCA type) [22], each with a very different number and order for substrate-binding proteins (OppA). With respect to the ABC-type transporters and permeases for single amino acids, the strains showed significant differences (**Table S7**).

Whereas 90 genes encoding amino acid/oligopeptide transporters were found in strain 2038, we found 75, 60, and 61 transporter genes, respectively, in the strains ND02, ATCC 11842, and ATCC BAA365 (**Table S3**). This large variation was mainly because of the number of OppA proteins—there were 25 of these in strain 2038, compared to 19 in ND02, and only 6 in ATCC 11842 and 7 in ATCC BAA365 (**Table S7**). In contrast, strains ATCC 11842 and ATCC BAA365 had no unique OppA proteins, while 5 were found in strain ND02, and 6 in strain 2038.

In addition to Opp. systems, 14 amino acid permeases were identified in the 4 *L. bulgaricus* strain genomes (**Table S7**). Five of the amino acid permeases were conserved, and might play roles in importing essential amino acids for *L. bulgaricus* strains. Three amino acid permeases (AppA2, LysP2, and ArcD3) were found only in the 2 industrial strains. Meanwhile, there were 2 unique amino acid permeases (encoded by LBU0206 and LBU0209) in strain 2038 and 1 in strain ND02 (encoded by *LDBND*1404). No unique amino acid permease was found in any of the laboratory strains, possibly illustrating that the laboratory strains had undergone a more extensive reductive evolution.

4. Discussion

In this study, we performed a systematic genome-wide analysis of all the proteins involved in amino acid biosynthesis and proteolysis, from 4 completely sequenced *L. bulgaricus* genomes, including the strains ATCC 11842, ATCC BAA-365, ND02, and 2038. Comparative genomics analysis was conducted to distinguish various subgroups within a protein superfamily, allowing for a highly improved annotation of genes, and clarification of any inconsistent annotations. This information could be used to predict the amino acid catabolism potential for all *L. bulgaricus* strains.

Two pathways have been identified in bacteria to assimilate ammonia into glutamate: 1 is the glutamate dehydrogenase (GDH, EC: 1.4.1.2) pathway, and another is via the glutamine synthase (GS, EC: 6.3.1.2)-glutamate synthase (GOGAT, EC: 1.4.1.13/14) cycle [23]. Both pathways form 1 molecule of glutamate from 1 molecule of ammonia (HN_3) and 1 molecule of 2-oxoglutarate. However, no genes encoding for GDH or GOGAT ware identified in these 4 completely sequenced *L. bulgaricus* genomes. Similar to how *l*-aspartate is synthesized from oxaloacetate and glutamate through the catalysis of aspartate aminotransferase (EC: 2.6.1.1), glutamate could be produced in the same manner from 2-oxoglutarate and l-aspartate. We found 2 genes encoding proteins homologous to aspartate aminotransferase in the 4 *L. bulgaricus* strain genomes (**Table S1**). These results suggest that aspartate aminotransferase catalyzes the formation of both glutamate and aspartate in *L. bulgaricus*.

All the *L. bulgaricus* strains similarly possessed limited anabolic biosynthesis capabilities because of a partial TCA cycle that contained only fumarate reductase (EC: 1.3.99.1) and fumarate hydratase (EC: 4.2.1.2). Six kinds of amino acids (aspartate, asparagine, glutamine, glutamate, threonine, and alanine) could be biosynthesized in all 4 strains. According to the results from the annotation studies, glutamate, aspartate, glutamine, and alanine were related to additional amino acid biosynthesis, nucleoside metabolism, and cell-wall formation. This suggests that *L. bulgaricus* took advantage of the nitrogen-rich environment when they lost the capacity for ammonium assimilation and glutamate biosynthesis.

Only the industrial strains 2038 and ND02 had the ability to biosynthesize lysine and serine by using a separate pathway for each amino acid. Meanwhile cysteine could only be synthesized by 2038 and proline only by ND02 through unique pathways for each. In addition to losing genes, each strain evolved new genes or retained ancestor genes after divergence. Among the 396 unique genes for strain ND02, 24 (6%) were associated with the amino acid synthesis/proteolysis system, whereas in 113 unique genes for strain 2038, 10 (8.8%) were inlved in its amino acid synthesis/ roteolysis system. Except for the 2 ATCC strains, only 5 unique genes encoding proteases or transporters were found in all the strains studied (**Table S8**). Based on the presence of genes encoding protease and transport systems, the 2 industrial stains were more capable of utilizing extrinsic resources than the 2 laboratory strains, and were likely influenced by industrial cultivation [2].

For the 4 genomes of the *L. bulgaricus* strains, the genes encoded a relatively lower number and variety of amino acid biosynthesis components, which lost some biosynthetic capability, as the strains continued to develop in different environments. Protein homolog analysis showed that only strains ND02 and 2038 had unique genes that are involved in amino acid synthesis in the 4 *L. bulgaricus* strains (**Table S4**). Strain ND02 had 6 genes encoding for proline synthesis (LDBND_0093, LDBND_0094, LDBND_0755, LDBND_0756, LDBND_0788, and LDBND_1721), which was the highest number for

this function among the strains studied. Strain 2038 possessed a unique gene, LBU1138, which encoded serine *o*-acetyltransferase, and therefore, endowed this strain with the ability to synthesize cysteine. Although strain ND02 possessed the biggest genome size and the largest unique region, it did not show significantly improved amino acid catabolism and proteolysis compared to strain 2038.

Strain 2038 had a growth advantage relative to the other strains because it could potentially synthesize cysteine, which is difficult to obtain from exogenous casein and can be used as a precursor of alanine. The expression of 2 concatenated cysteine aminopeptidase genes might increase to satisfy the demand for cysteine in *L. bulgaricus* cells or to carry out the same function under different conditions. Strain ND02 could synthesize proline, which might afford it a growth-related advantage over the other strains. These features explain why strains 2038 and ND02 were chosen as the industrial strains; in addition, they possess more genes associated with nitrogen utilization compared to the other strains.

Evolutionary differences between the 2 industrial strains, as illustrated in the phylogenetic tree that was generated from analyzing amino acid synthesis and proteolysis genes, could be attributed to the fact that each industrial strain was isolated from a different geographical area. Therefore, differences between using cow milk from Europe and sheep milk from North Asia, and lower cysteine concentrations in cow milk compared to sheep milk, and differences between these 2 areas feeding environments, may have all contributed to the observed evolutionary divergence between the 2 strains.

Genomic sequence analysis revealed the presence of at least 45 genes encoding putative proteases or peptidases. From this group, the most significant among the genes encoding cell surface proteases is PrtB, which is located downstream of the *lac* operon in *L. bulgaricus* 2038 [24]. A comparison of PrtB among the 4 complete *L. bulgaricus* sequences showed that the protease catalytic sites and substrate-binding sites are conserved, while a 125-amino acid-long deletion in C-terminal was revealed in *L. bulgaricus* 2038 and might affect the folding of the enzyme. These results imply that different *L. bulgaricus* strains might have different substrate specificities and release different peptides.

The foldase protein PrsA was identified in the genomes of all 4 *L. bulgaricus* strains. Because this protein was very highly homologous to PrtM and had the same EC number, and no PrtM was found in *L. bulgaricus* until now, we suggest that it may play role in the maturation of PrtB instead of PrtM; this may explain the maturation and turnover rates of PrtB in all *L. bulgaricus* strains. The cell surface housekeeping protease HtrA was also identified in all 4 *L. bulgaricus* strain genomes. Considing that only 2 cell surface peptidases were identified in all 4 strains, it is reasonable to speculate that these 2 cell surface peptidases may perform a single function.

Two putative cell surface peptidases that are not heat-shock proteinases—the membrane protein related to metalloendopeptidase EnlA and the dipeptidase PepD4—are either missing or were found as a pseudo gene either in strain ATCC BAA-365, or in strains ATCC 11842 and ND02 without a signal peptide in their C-termini. Only *L. bulgaricus* 2038 had these 2 completed genes with a signal. Considering that some intracellular peptidases were identified only in strain 2038, *L. bulgaricus* 2038 might potentially have a more powerful exterior protein degradation capability than other *L. bulgaricus* strains, which could be advantageous, as it would produce more free amino acids than the other strains. These amino acids could then be transported into the cell via 8 ATP-binding ABC-type amino acid transport systems or at least 6 permeases.

Two distinct *opp* operons were associated with multiple OppA proteins in the 4 *L. bulgaricus* strain genomes, but these 2 *opp* operons in strain 2038 are very interesting in their distinct operon structures; PepG1 and PepG2 (LBU0223 and LBU0224) encoding aminopeptidases are located directly adjacent to the OppF gene. There are also more OppA genes existing in strain 2038 compared to the other 3 strains. This might explain the important role of the Opp system in this industrial strain; different OppA proteins are used for different oligopeptides, especially, in different environments. Several peptide transporters or peptidases fall into larger protein super families: 1) aminopeptidase PepC and PepG belonging to the MEROPS peptidase family C1_B, 2) aminopeptidase PepI and PepL belonging to MEROPS family S33, and 3) aminopeptidase PepM together with dipeptidase PepQ and PepZ belonging to MAROPS family M24.

REFERENCES

[1] J. E. Christensen, E. G. Dudley, *et al*., "Peptidases and Amino Acid Catabolism in Lactic Acid Bacteria," *Antonie Van Leeuwenhoek*, Vol. 76, No. 1-4, 1999, pp. 217-246. doi:10.1023/A:1002001919720

[2] P. Hao, H. Zheng, *et al*., "Complete Sequencing and Pan-Genomic Analysis of *Lactobacillus delbrueckii* subsp. *bulgaricus* Reveal Its Genetic Basis for Industrial Yogurt Production," *PLOS One*, Vol. 6, No. 1, 2011, p. e15964. doi:10.1371/journal.pone.0015964

[3] M. van de Guchte, S. Penaud, *et al*., "The Complete Genome Sequence of *Lactobacillus bulgaricus* Reveals Extensive and Ongoing Reductive Evolution," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 103, No. 24, 2006, pp. 9274-9279. doi:10.1073/pnas.0603024103

[4] M. Liu, J. R. Bayjanov, *et al*., "The Proteolytic System of Lactic Acid Bacteria Revisited: A Genomic Comparison,"

BMC Genomics, Vol. 11, No.1, 2010, p. 36. doi:10.1186/1471-2164-11-36

[5] M. Liu, A. Nauta, *et al*., "Comparative Genomics of Enzymes in Flavor-Forming Pathways from Amino Acids in Lactic Acid Bacteria," *Applied and Environmental Microbiology*, Vol. 74, No. 15, 2008, pp. 4590-4600. doi:10.1128/AEM.00150-08

[6] M. Zhou, D. Theunissen, *et al*., "LAB-Secretome: A Genome-Scale Comparative Analysis of the Predicted Extracellular and Surface-Associated Proteins of Lactic Acid Bacteria," *BMC Genomics*, Vol. 11, No. 1, 2010, p. 651. doi:10.1186/1471-2164-11-651

[7] J. Boekhorst, R. J. Siezen, *et al*., "The Complete Genomes of *Lactobacillus plantarum* and *Lactobacillus johnsonii* Reveal Extensive Differences in Chromosome Organization and Gene Content," *Microbiology*, Vol. 150, No. 11, 2004, pp. 3601-3611. doi:10.1099/mic.0.27392-0

[8] H.-J. Zheng, E-N. Liu, *et al*., "*In Silico* Analysis of Amino Acid Biosynthesis and Proteolysis in *Lactobacillus delbrueckii* subsp. *bulgaricus* 2038 and the Implications for Bovine Milk Fermentation," *Biotechnology Letters*, Vol. 34, No. 8, 2012, pp. 1545-1551. doi:10.1007/s10529-012-1006-4

[9] A. E. Darling, B. Mau, *et al*., "ProgressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement," *PLOS One*, Vol. 5, No. 6, 2010, p. e11147. doi:10.1371/journal.pone.0011147

[10] N. D. Rawlings, A. J. Barrett, *et al*., "MEROPS: The Database of Proteolytic Enzymes, Their Substrates and Inhibitors," *Nucleic Acids Research*, Vol. 40, D. 1, 2012, pp. D343-D350. doi:10.1093/nar/gkr987

[11] M. Kanehisa, S. Goto, *et al*., "The KEGG Resource for Deciphering the Genome," *Nucleic Acids Research*, Vol. 32, Suppl. 1, 2004, D277-D280. doi:10.1093/nar/gkh063

[12] J. D. Bendtsen, H. Nielsen, *et al*., "Improved Prediction of Signal Peptides: SignalP 3.0," *Journal of Molecular Biology*, Vol. 340, No. 4, 2004, pp. 783-795. doi:10.1016/j.jmb.2004.05.028

[13] M. Arai, H. Mitsuke, *et al*., "ConPred II: A Consensus Prediction Method for Obtaining Transmembrane Topology Models with High Reliability," *Nucleic Acids Research*, Vol. 32, Suppl. 2, 2004, pp. W390-W393. doi:10.1093/nar/gkh380

[14] J. L. Gardy, C. Spencer, *et al*., "PSORT-B: Improving Protein Subcellular Localization Prediction for Gram-Negative Bacteria," *Nucleic Acids Research*, Vol. 31, No. 13, 2003, pp. 3613-3617. doi:10.1093/nar/gkg602

[15] A. Marchler-Bauer, S. Lu, J. B. Anderson, *et al*., "CDD: A Conserved Domain Database for the Functional Annotation of Proteins," *Nucleic Acids Research*, Vol. 39, Suppl. 1, 2011, pp. D225-D229. doi:10.1093/nar/gkq1189

[16] I. Uchiyama, T. Higuchi, *et al*., "MBGD Update 2010: Toward a Comprehensive Resource for exploring Microbial Genome Diversity," *Nucleic Acids Research*, Vol. 38, Suppl. 1, 2010, pp. D361-365. doi:10.1093/nar/gkp948

[17] G. Talavera and J. Castresana, "Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments," *Systematic Biology*, Vol. 56, No. 4, 2007, pp. 564-577. doi:10.1080/10635150701472164

[18] S. Guindon, J. F. Dufayard, *et al*., "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0," *Systematic Biology*, Vol. 59, No. 3, 2010, pp. 307-321. doi:10.1093/sysbio/syq010

[19] J. E. Germond, M. Delley, *et al*., "Determination of the Domain of the *Lactobacillus delbrueckii* subsp. *Bulgaricus* Cell Surface Proteinase PrtB Involved in Attachment to the Cell Wall after Heterologous Expression of the PrtB Gene in *Lactococcus lactis*," *Applied Environmental Microbiology*, Vol. 69, No. 6, 2003, pp. 3377-3384. doi:10.1128/AEM.69.6.3377-3384.2003

[20] M. Jacobs, J. B. Andersen, *et al*., "*Bacillus subtilis* PrsA Is Required *in Vivo* as an Extracytoplasmic Chaperone for Secretion of Active Enzymes Synthesized either with or without Pro-Sequences," *Molecular Microbiology*, Vol. 8, No. 5, 1993, pp. 957-966. doi:10.1111/j.1365-2958.1993.tb01640.x

[21] N. Vermeulen, M. Pavlovic, *et al*., "Functional Characterization of the Proteolytic System of *Lactobacillus sanfranciscensis* DSM 20451T during Growth in Sourdough," *Applied Environmental Microbiology*, Vol. 71, No. 10, 2005, pp. 6260-6266. doi:10.1128/AEM.71.10.6260-6266.2005

[22] S. Tynkkynen, G. Buist, *et al*., "Genetic and Biochemical Characterization of the Oligopeptide Transport System of *Lactococcus lactis*," *Journal of Bacteriology*, Vol. 175, No. 23, 1993, pp. 7523-7532.

[23] M. Romero, S. Guzman-Leon, *et al*., "Pathways for Glutamate Biosynthesis in the Yeast *Kluyveromyces lactis*," *Microbiology*, Vol. 146, No. 1, 2000, pp. 239-245.

[24] C. Gilbert, D. Atlan, *et al*., "A New Cell Surface Proteinase: Sequencing and Analysis of the PrtB Gene from *Lactobacillus delbruekii* subsp. *bulgaricus*," *Journal of Bacteriology*, Vol. 178, No. 11, 1996, pp. 3059-3065.

Appendix

Table S1. Genes involved in amino acid biosynthesis for each *L. bulgaricus* strain.

			Locus in each strain			
Synthesis Pathway	KO	EC No	2038	ATCC 11842	ATCC BAA-365	ND02
Asparagine	K01914	EC:6.3.1.1	LBU1019	Ldb1194	LBUL_1110	LDBND_1073
	K01953	EC:6.3.5.4	LBU0163/LBU0807	Ldb0209	LBUL_0183	LDBND_0200/LDBND_0839
Glutamine	K01915	EC:6.3.1.2	LBU1261	Ldb1472	LBUL_1368	LDBND_1408
Proline	K00931	EC:2.7.2.11	LBU0712			LDBND_0757
	K00147	EC:1.2.1.41				LDBND_0756
	K00286	EC:1.5.1.2				LDBND_0755
Arginine	K00611	EC:2.1.3.3				LDBND_1721
	K01940	EC:6.3.4.5				LDBND_0788
	K01755	EC:4.3.2.1	LBU0740			LDBND_0789
Cysteine	K00640	EC:2.3.1.30	LBU1138			
	K01738	EC:2.5.1.47	LBU1136/LBU1253	Ldb1325/Ldb1458	LBUL_1235/LBUL_1354	LDBND_1271/LDBND_1389
	K01740	EC:2.5.1.49	LBU0307			LDBND_0325
Methionine	K00928	EC:2.7.2.4	LBU0936/LBU1158	Ldb1349	LBUL_1258	LDBND_1033
	K00133	EC:1.2.1.11	LBU1157	Ldb1348	LBUL_1257	LDBND_1291
	K00003	EC:1.1.1.3	LBU1159	Ldb1350	LBUL_1259	LDBND_1293
	K00651	EC:2.3.1.46	LBU0072/LBU1006/LBU1252		LBUL_0089/LBUL_1353	LDBND_1060/LDBND_1388
	K01760	EC:4.4.1.8	LBU1137	Ldb1326	LBUL_1236	LDBND_1272
	K14155	EC:4.4.1.8	LBU1014	Ldb1187	LBUL_1103	
	K00547	EC:2.1.1.10	LBU0247	Ldb0311	LBUL_0268	LDBND_0272
	K00549	EC:2.1.1.14	LBU0082/LBU1132	Ldb0116	LBUL_1232	LDBND_0086
Serine	K00058	EC:1.1.1.95	LBU0880			LDBND_0927
	K00831	EC:2.6.1.52	LBU0882			LDBND_0928
Threonine	K00928	EC:2.7.2.4	LBU0936/LBU1158	Ldb1349	LBUL_1258	LDBND_1033
	K00133	EC:1.2.1.11	LBU1157	Ldb1348	LBUL_1257	LDBND_1291
	K00003	EC:1.1.1.3	LBU1159	Ldb1350	LBUL_1259	LDBND_1293
	K00872	EC:2.7.1.39	LBU1495	Ldb1760	LBUL_1630	LDBND_1655
	K01733	EC:4.2.3.1	LBU1496	Ldb1761	LBUL_1631	LDBND_1656
Lysine	K00928	EC:2.7.2.4	LBU0936/LBU1158	Ldb1349	LBUL_1258	LDBND_1033
	K00133	EC:1.2.1.11	LBU1157	Ldb1348	LBUL_1257	LDBND_1291
	K00003	EC:1.1.1.3	LBU1159	Ldb1350	LBUL_1259	LDBND_1293
	K01714	EC:4.2.1.52	LBU0933			LDBND_1029
	K00215	EC:1.3.1.26	LBU0932			LDBND_1028
	K00674	EC:2.3.1.117	LBU0935			LDBND_1031
	K00841	EC:2.6.1.-	LBU0931/LBU1079	Ldb1263	LBUL_1180	LDBND_1027/LDBND_1147
	K05823	EC:3.5.1.47	LBU0934			LDBND_1030
	K01778	EC:5.1.1.7	LBU0937			LDBND_1034
	K01586	EC:4.1.1.20	LBU1080		LBUL_1181	LDBND_1148
Tryptophan	K03785	EC:4.2.1.10				LDBND_0094
	K00014	EC:1.1.1.25				LDBND_0093
	K00800	EC:2.5.1.19	LBU1316/LBU1317			LDBND_1470
Phenylalanine/Tyrosine	K04516	EC:5.4.99.5	LBU0249	Ldb0313	LBUL_0271	
	K06209	EC:5.4.99.5	LBU1160	Ldb1351	LBUL_1260	LDBND_1294
	K00832	EC:2.6.1.57	LBU0363	Ldb0441	LBUL_0392	LDBND_0387
Alanine	K04487	EC:2.8.1.7	LBU0617/LBU0646	Ldb0724/Ldb0753	LBUL_0656/LBUL_0686	LDBND_0657/LDBND_0688
Aspartate/Glutamate	K00813/K00832	EC:2.6.1.1	LBU0363/LBU1079	Ldb0441/Ldb1263	LBUL_0392/LBUL_1180	LDBND_0387/LDBND_1147

Table S2. Genes encoding proteases and peptidases in each *L. bulgaricus* strain.

	Number of genes in each strain			
COG	2038	ATCC 11842	ATCC BAA-365	ND02
COG0006	3	3	3	3
COG0024	1	1	1	1
COG0265	1	1	1	1
COG0308	1	1	1	1
COG0319	1	1	1	1
COG0465	1	1	1	1
COG0481	1	1	1	1
COG0501	2	2	2	2
COG0533	1	1	1	1
COG0542	3	3	3	3
COG0596	3	3	3	4
COG0597	1	1	1	1
COG0612	2	2	2	2
COG0624	1	1	1	1
COG0681	1	1	1	2
COG0705	1	1	1	1
COG0739	0	2	0	2
COG0740	1	1	1	2
COG0744	2	2	2	2
COG0750	1	1	1	1
COG0791	7	6	5	8
COG1164	1	1	1	1
COG1214	1	1	1	1
COG1219	1	1	1	1
COG1220	1	1	1	1
COG1363	1	1	1	1
COG1404	1	1	1	1
COG1473	3	1	3	1
COG1506	2	2	3	2
COG1686	1	1	1	1
COG1974	1	1	1	3
COG2039	1	1	1	1
COG2195	2	2	2	2
COG2362	1	1	0	0
COG2936	1	1	1	1
COG3091	1	1	1	1
COG3191	1	0	1	2
COG3480	1	1	1	1
COG3579	3	3	3	3
COG3590	2	1	1	2
COG3764	1	1	1	3
COG4690	6	2	3	5
COG5405	1	1	1	1
COG5549	1	1	1	1
-	2	1	3	4
Total	72	64	67	81

Table S3. Genes encoding peptide or amino acid transporters in each *L. bulgaricus* strain.

	Number of genes in each strain			
COG	2038	ATCC 11842	ATCC BAA-365	ND02
COG0004	1	1	1	0
COG0410	1	0	0	1
COG0444	2	2	2	2
COG0531	5	4	2	4
COG0601	2	2	2	2
COG0687	1	1	2	1
COG0747	2	2	2	2
COG0765	8	8	7	8
COG0833	2	0	0	1
COG0834	7	5	4	6
COG1113	5	2	2	4
COG1114	1	1	1	1
COG1122	2	2	2	2
COG1123	2	2	2	2
COG1125	1	1	1	1
COG1126	7	6	5	7
COG1136	8	8	8	5
COG1173	2	1	2	2
COG1176	1	1	2	1
COG1177	1	1	2	1
COG1732	1	1	1	1
COG3842	1	1	2	1
COG4166	25	6	7	19
-	2	2	2	1
Total	90	60	61	75

Table S4. Genes involved in proteolysis and amino acid biosynthesis of *L. bulgaricus* ND02 but lost in other ATCC strains and 2038.

Gene of strain ND02	Lost in other strains	Class	Gene of strain ND02	Lost in other strains	Class
LDBND_1148	ATCC 11842	Amino acid synthesis	LDBND_1392	ATCC*	Protease
LDBND_1060	ATCC*	Amino acid synthesis	LDBND_1030	ATCC*	Protease
LDBND_1470	ATCC*	Amino acid synthesis	LDBND_1315	ATCC*	Protease
LDBND_0325	ATCC*	Amino acid synthesis	LDBND_1747	ATCC*	Protease
LDBND_0789	ATCC*	Amino acid synthesis	LDBND_1321	ATCC BAA-365	Transporter
LDBND_0927	ATCC*	Amino acid synthesis	LDBND_0265	ATCC 11842	Transporter
LDBND_0757	ATCC*	Amino acid synthesis	LDBND_0266	ATCC 11842	Transporter
LDBND_0839	ATCC*	Amino acid synthesis	LDBND_0267	ATCC 11842	Transporter
LDBND_0928	ATCC*	Amino acid synthesis	LDBND_0492	ATCC 11842	Transporter
LDBND_1027	ATCC*	Amino acid synthesis	LDBND_1631	ATCC*	Transporter
LDBND_1028	ATCC*	Amino acid synthesis	LDBND_1762	ATCC*	Transporter
LDBND_1029	ATCC*	Amino acid synthesis	LDBND_0271	ATCC*	Transporter
LDBND_1031	ATCC*	Amino acid synthesis	LDBND_0382	ATCC*	Transporter
LDBND_1033	ATCC*	Amino acid synthesis	LDBND_0623	ATCC*	Transporter
LDBND_1034	ATCC*	Amino acid synthesis	LDBND_0840	ATCC*	Transporter
LDBND_1105	ATCC BAA-365	Protease	LDBND_1346	ATCC*	Transporter
LDBND_0771	ATCC BAA-365	Protease	LDBND_1363	ATCC*	Transporter
LDBND_1859	ATCC BAA-365	Protease	LDBND_1586	ATCC*	Transporter
LDBND_1861	ATCC BAA-365	Protease	LDBND_1717	ATCC*	Transporter
LDBND_1480	ATCC BAA-365	Protease	LDBND_1763	ATCC*	Transporter
LDBND_0561	ATCC 11842	Protease	LDBND_1856	2038	Protease
LDBND_1857	ATCC 11842	Protease	LDBND_1480	2038	Protease
LDBND_1856	ATCC 11842	Protease			

*ATCC represents both ATCC BAA-365 and ATCC 11842.

Table S5. Genes encoding extracellular proteases or peptidases in each *L. bulgaricus* strain.

		Locus in each strain					
COG	Gene	2038	ATCC BAA-365	ATCC 11842	ND02	Product	Localization
COG0265	HtrA	LBU0100	LBUL_0118	Ldb0140	LDBND_0115	serine protease HtrA	Cellwall
COG1404	PrtB	LBU1015	LBUL_1105	Ldb1189	LDBND_1068	cell wall-associated serine proteinase	Cellwall
COG0501	HtpX1	LBU0117	LBUL_0136	Ldb0160	LDBND_0132	putative protease HtpX-like protein	CytoplasmicMembrane
COG0501	HtpX2	LBU1013	LBUL_1101	Ldb1185	LDBND_1067	Zn-dependent protease	CytoplasmicMembrane
COG0750		LBU1148	LBUL_1248	Ldb1339	LDBND_1282	putative membrane-associated Zn-dependent proteases	CytoplasmicMembrane
		LBU1759	LBUL_1994	Ldb2160	LDBND_2001	CAAX amino terminal protease family protein	CytoplasmicMembrane
	EnlA	LBU1040		Ldb1221	LDBND_1105	putative enterolysin A (Peptidase M23)	Extracellular
COG4690	PepD4	LBU1705	LBUL_1926		LDBND_1936	dipeptidase	Extracellular
			LBUL_1924			membrane protease family stomatin/prohibitin-like protein	Cellwall
					LDBND_1934	membrane protease subunit, stomatin/prohibitin family	Cellwall

Table S6. Genes encoding intracellular peptidases in each *L. bulgaricus* strain.

				Locus in each strain				
MEROPS ID	Family	COG	gene	2038	ATCC BAA-365	ATCC 11842	ND02	Product
-	-	COG3091		LBU0380	LBUL_0409	Ldb0461	LDBND_0405	putative Zinc-dependent metalloprotease
MER084977	M10X	COG5549		LBU1567	LBUL_1785	Ldb1920	LDBND_1766	putative Zn-dependent protease
MER002163	C69	COG4690	PepD1	LBU0421	LBUL_0452	Ldb0511	LDBND_0454	putative dipeptidase A
MER000443	S15	COG2936	PepX	LBU1250	LBUL_1351	Ldb1455	LDBND_1386	X-prolyl dipeptidyl peptidase
MER002038	M24B	COG0006	PepQ	LBU1357	LBUL_1474	Ldb1594	LDBND_1503	Xaa-Pro aminopeptidase
MER002038	M24B	COG0006	PepZ	LBU1450	LBUL_1575	Ldb1700	LDBND_1594	Xaa-Pro aminopeptidase
MER001361	M20A	COG0624	PepV	LBU1488	LBUL_1619	Ldb1746	LDBND_1630	putative dipeptidase V
MER001322	C40	COG0791	Spr	LBU1516	LBUL_1667	Ldb1795	LDBND_1681	putative dipeptidyl-peptidase VI
MER002163	C69	COG4690	PepD3	LBU1679	LBUL_1909	Ldb2063	LDBND_1910	dipeptidase
MER105435	M20B	COG2195	PepT1	LBU0301	LBUL_0326	Ldb0371	LDBND_0319	peptidase T
MER001421	M20B	COG2195	PepT2	LBU1036	LBUL_1126	Ldb1217	LDBND_1101	peptidase T
MER003535	C01B	COG3579	PepG1	LBU0223	LBUL_0241	Ldb0287	LDBND_0244	cysteine aminopeptidase
MER003534	C01B	COG3579	PepG2	LBU0224	LBUL_0242	Ldb0288	LDBND_0245	cysteine aminopeptidase
MER072511	M42	COG1363	PepA	LBU0422	LBUL_0453	Ldb0512	LDBND_0455	M42 family peptidase
MER001243	M24A	COG0024	PepM	LBU0464	LBUL_0496	Ldb0556	LDBND_0497	type I methionine aminopeptidase
MER005922	M55	COG2362	DppA	LBU0520		Ldb0617		D-aminopeptidase
MER060648	S58	COG3191	DmpA	LBU0521	LBUL_0553		LDBND_0560/ LDBND_0561	L-aminopeptidase/D-esterase
MER005462	M24B	COG0006		LBU1225	LBUL_1324	Ldb1429	LDBND_1357	Xaa-Pro aminopeptidase
MER000728	C01B	COG3579	PepC	LBU1473	LBUL_1604	Ldb1730	LDBND_1615	aminopeptidase C
MER000436	S33	COG0596	PepI	LBU1543	LBUL_1762	Ldb1896	LDBND_1743	prolyl aminopeptidase
MER000437	S33	COG0596	pip	LBU1556	LBUL_1775	Ldb1909	LDBND_1755	prolyl aminopeptidase
MER001005	M01	COG0308	PepN	LBU1702	LBUL_1922	Ldb2080	LDBND_1931	aminopeptidase N
MER001380	S33	COG0596	PepL				LDBND_1003	leucyl aminopeptidase
MER074500	M20D	COG1473	Amd	LBU0207	LBUL_0225	Ldb0265		carboxypeptidase
MER081890	M20D	COG1473		LBU0898	LBUL_0961			carboxypeptidase
MER084057	M20D	COG1473		LBU0934			LDBND_1030	amino acid amidohydrolase
MER072517	C15	COG2039	Pcp	LBU1004	LBUL_1093	Ldb1175	LDBND_1058	pyrrolidone-carboxylate peptidase
MER019832	M13	COG3590	PepO	LBU0172	LBUL_0191	Ldb0218	LDBND_0208	endopeptidase O
-	-	COG0542	ClpC	LBU0310	LBUL_0339	Ldb0383	LDBND_0328	endopeptidase Clp ATP-binding chain C
MER072513	M16B	COG0612	PqqL	LBU0498	LBUL_0531	Ldb0595	LDBND_0538	putative metalloendopeptidase
MER019832	M13	COG3590	PepO	LBU1255			LDBND_1392	oligopeptidase O

Continued

MER166506	M22	COG0533	Gcp	LBU1382	LBUL_1501	Ldb1621	LDBND_1530	O-sialoglycoprotein endopeptidase
MER052966	M22	COG1214	Gpp	LBU1384	LBUL_1503	Ldb1623	LDBND_1532	glycoprotein endopeptidase
MER072526	M03B	COG1164	PepF	LBU1650	LBUL_1880	Ldb2034	LDBND_1876	oligoendopeptidase F
MER232366	S9C	COG1506					LDBND_1120	peptidase s9 prolyl oligopeptidase
MER141586	S9C	COG1506	Aph	LBU0861	LBUL_0918	Ldb1011	LDBND_0905	trilobed protease
-	-	COG0612		LBU0497	LBUL_0530	Ldb0594	LDBND_0537	Zn-dependent peptidase
-	-						LDBND_1579	peptidase m20a dipeptidase

Table S7. Genes involved in oligopeptide or amino acid transport systems in each *L. bulgaricus* strain.

		Locus in each strain				
COG	Gene	2038	ATCC BAA-365	ATCC 11842	ND02	Product
COG1136					LDBND_2037	ABC antimicrobial peptide transporter ATPase
COG1126		LBU1565			LDBND_1763	ABC polar amino acid transporter ATPase
COG1126		LBU0124	LBUL_0144	Ldb0168	LDBND_0140	ABC-type amino acid transport system ATP-binding protein
COG1126		LBU0199	LBUL_0216	Ldb0253	LDBND_0883	ABC-type amino acid transport system ATP-binding protein
COG1126		LBU0241	LBUL_0262		LDBND_0266	ABC-type amino acid transport system ATP-binding protein
COG0765		LBU0197	LBUL_0214	Ldb0251		ABC-type amino acid transport system permease protein
COG0765		LBU0198	LBUL_0215	Ldb0252	LDBND_0882	ABC-type amino acid transport system permease protein
COG0765		LBU0242	LBUL_0263		LDBND_0267	ABC-type amino acid transport system permease protein
COG0765		LBU0125	LBUL_0145	Ldb0169	LDBND_0141	ABC-type amino acid transport system substrate-binding and permease protein
COG0834	YckK	LBU0240	LBUL_0261		LDBND_0265	ABC-type amino acid transport system substrate-binding protein
COG0834		LBU0200	LBUL_0217	Ldb0254	LDBND_0884	ABC-type amino acid transport system substrate-binding protein
COG0410	LivF	LBU1231			LDBND_1363	ABC-type branched-chain amino acid transport system ATP-binding protein
COG1126	GlnQ	LBU0427	LBUL_0458	Ldb0517	LDBND_0460	ABC-type glutamine transport system ATP-binding protein
COG1126	GlnQ2	LBU1110	LBUL_1213	Ldb1298	LDBND_1245	ABC-type glutamine transport system ATP-binding protein
COG1126	GlnQ3	LBU1782	LBUL_2020	Ldb2199	LDBND_2043	ABC-type glutamine transport system ATP-binding protein
COG0765	GlnM	LBU0429	LBUL_0460	Ldb0519	LDBND_0462	ABC-type glutamine transport system permease protein
COG0765	GlnP	LBU0430	LBUL_0461	Ldb0520	LDBND_0463	ABC-type glutamine transport system permease protein
COG0765	GlnP2	LBU1111	LBUL_1214	Ldb1299	LDBND_1246	ABC-type glutamine transport system substrate-binding and permease protein
COG0765	GlnP3	LBU1781	LBUL_2019	Ldb2198	LDBND_2042	ABC-type glutamine transport system substrate-binding and permease protein
COG0834	GlnH1	LBU0428	LBUL_0459	Ldb0518	LDBND_0461	ABC-type glutamine transport system substrate-binding protein
COG0834	GlnH2	LBU0431	LBUL_0462	Ldb0521	LDBND_0464	ABC-type glutamine transport system substrate-binding protein
COG0834	GlnH3	LBU1167	LBUL_1267	Ldb1358	LDBND_1301	ABC-type glutamine transport system substrate-binding protein
COG0444	OppD	LBU1192	LBUL_1293	Ldb1386	LDBND_1324	ABC-type Opp system ATP-binding protein
COG0444	OppD2	LBU0221	LBUL_0238	Ldb0284	LDBND_0242	ABC-type Opp system ATP-binding protein
COG1123	Oppf	LBU1191	LBUL_1292	Ldb1385	LDBND_1323	ABC-type Opp system ATP-binding protein
COG1123	OppF2	LBU0222	LBUL_0239	Ldb0285	LDBND_0243	ABC-type Opp system ATP-binding protein
COG0601	OppB	LBU1190	LBUL_1291	Ldb1384	LDBND_1322	ABC-type Opp system permease protein
COG0601	OppB2	LBU0219	LBUL_0236	Ldb0282	LDBND_0240	ABC-type Opp system permease protein
COG1173	OppC	LBU1189		Ldb1383	LDBND_1321	ABC-type Opp system permease protein
COG1173	OppC2	LBU0220	LBUL_0237	Ldb0283	LDBND_0241	ABC-type Opp system permease protein
COG4166	OppA1	LBU0160	LBUL_0180	Ldb0206		ABC-type Opp system substrate-binding protein

Continued

COG4166	OppA1	LBU0365				ABC-type Opp system substrate-binding protein
COG4166	OppA11	LBU0588			LDBND_0623	ABC-type Opp system substrate-binding protein
COG4166	OppA12	LBU0589			LDBND_0625	ABC-type Opp system substrate-binding protein
COG4166	OppA13	LBU0676	LBUL_0717	Ldb0784	LDBND_0719	ABC-type Opp system substrate-binding protein
COG4166	OppA14	LBU0808			LDBND_0840	ABC-type Opp system substrate-binding protein
COG4166	OppA15	LBU0809				ABC-type Opp system substrate-binding protein
COG4166	OppA16/17/18/19	LBU0904/ LBU0905/ LBU0962/ LBU0963		Ldb1060	LDBND_1004	ABC-type Opp system substrate-binding protein
COG4166	OppA2	LBU0213			LDBND_0231	ABC-type Opp system substrate-binding protein
COG0747	OppA2	LBU1187	LBUL_1288	Ldb1381	LDBND_1319	ABC-type Opp system substrate-binding protein
COG0747	OppA21	LBU1188	LBUL_1289	Ldb1382	LDBND_1320	ABC-type Opp system substrate-binding protein
COG4166	OppA22	LBU1215			LDBND_1346	ABC-type Opp system substrate-binding protein
COG4166	OppA23	LBU1410				ABC-type Opp system substrate-binding protein
COG4166	OppA24	LBU1411				ABC-type Opp system substrate-binding protein
COG4166	OppA25	LBU1442			LDBND_1586	ABC-type Opp system substrate-binding protein
COG4166	OppA3 /oppA4	LBU0214/ LBU0215/ LBU0216	LBUL_0233	Ldb0276	LDBND_0233/ LDBND_0235	ABC-type Opp system substrate-binding protein
COG4166	OppA5	LBU0217				ABC-type Opp system substrate-binding protein
COG4166	OppA6	LBU0218	LBUL_0235	Ldb0281	LDBND_0238	ABC-type Opp system substrate-binding protein
COG4166	OppA7	LBU0300	LBUL_0325	Ldb0370	LDBND_0318	ABC-type Opp system substrate-binding protein
COG4166	OppA8	LBU0354			LDBND_0382	ABC-type Opp system substrate-binding protein
COG4166	OppA9	LBU0355				ABC-type Opp system substrate-binding protein
COG1732	ProWX	LBU1009	LBUL_1097	Ldb1180	LDBND_1062	ABC-type proline/glycine betaine transport system substrate-binding and permease protein
COG1125	ProV	LBU1010	LBUL_1098	Ldb1181	LDBND_1063	ABC-type proline/glycine betaine transport system ATP-binding protein
COG3842	PotA	LBU0544	LBUL_0578	Ldb0647	LDBND_0582	ABC-type spermidine/putrescine transport system ATP-binding protein
COG1176	PotB	LBU0545	LBUL_0579	Ldb0648	LDBND_0583	ABC-type spermidine/putrescine transport system permease protein
COG1177	PotC	LBU0546	LBUL_0580	Ldb0649	LDBND_0584	ABC-type spermidine/putrescine transport system permease protein
COG0687	PotD	LBU0547	LBUL_0581	Ldb0650	LDBND_0585	ABC-type spermidine/putrescine transport system substrate-binding protein
COG1136	YhcA	LBU1746	LBUL_1979	Ldb2142	LDBND_1970	ABC-type transport system ATP-binding and permease protein
COG1122	YkoD	LBU0108	LBUL_0126	Ldb0151	LDBND_0124	ABC-type transport system ATP-binding protein
COG1136	YxeB	LBU1688	LBUL_1912	Ldb2069	LDBND_1922	ABC-type transport system ATP-binding protein
COG1136		LBU0014	LBUL_0014	Ldb0014		ABC-type transport system ATP-binding protein
COG1136		LBU0793	LBUL_1700	Ldb1828		ABC-type transport system ATP-binding protein
COG1136		LBU0924/ LBU0943	LBUL_0989/ LBUL_1019	Ldb1088/ Ldb1119	LDBND_1020	ABC-type transport system ATP-binding protein
COG1136		LBU0925/ LBU0942	LBUL_0990/ LBUL_1018	Ldb1089/ Ldb1118	LDBND_1021	ABC-type transport system ATP-binding protein
COG1122		LBU1459	LBUL_1583	Ldb1709	LDBND_1600	ABC-type transport system ATP-binding protein
	yxeA	LBU1689	LBUL_1913	Ldb2070	LDBND_1923	ABC-type transport system permease protein
COG1113	aapA1	LBU1462	LBUL_1586	Ldb1712	LDBND_1605	amino acid permease
COG1113	aapA2	LBU1489/ LBU1490		LDBND_1631	Amino acid permease	COG1113
COG0833	lysP2	LBU0246			LDBND_0271	amino acid permease
COG0531	potE	LBU0206				amino acid permease

Continued

COG1113	YdgF	LBU1517	LBUL_1668	Ldb1796	LDBND_1682	amino acid permease
COG0531	YlcA	LBU0459	LBUL_0491		LDBND_0492	amino acid permease
COG0531		LBU0450	LBUL_0481	Ldb0539	LDBND_0481	amino acid permease
COG0531		LBU0567	LBUL_0604	Ldb0672	LDBND_0601	amino acid permease
COG0531		LBU1506	LBUL_1646			amino acid permease
COG0833	LysP1	LBU0209				amino acid transport protein
COG0004	AmtB	LBU1417	LBUL_1541	Ldb1663		ammonium transporter
COG0531					LDBND_1404	APC family amino acid-polyamine-organocation transporter
COG1113	ArcD3	LBU0784			LDBND_1717	arginine/ornithine antiporter
COG1114	BrnQ	LBU0400	LBUL_0431	Ldb0483	LDBND_0429	branched-chain amino acid transport system II carrier protein
COG0834		LBU1563/ LBU1564			LDBND_1762	amino acid ABC transporter ATP-binding protein
COG0765					LDBND_1764	glutamine ABC transporter permease component
COG4166					LDBND_0236	lipoprotein, peptide binding protein oppA-like protein
COG4166					LDBND_1578	lipoprotein, peptide binding protein oppA-like protein
COG4166		LBU0366	LBUL_0397	Ldb0446	LDBND_0390	oligopeptide ABC superfamily ATP binding cassette transporter, binding protein
		LBU0792	LBUL_1701	Ldb1829		peptide ABC transporter permease
COG4166					LDBND_0237	peptide ABC transporter substrate-binding protein
COG4166					LDBND_0961	peptide ABC transporter substrate-binding protein
COG4166					LDBND_0962	peptide ABC transporter substrate-binding protein
COG3842				Ldb2180		spermidine/putrescine ABC transporter ATP-binding protein
COG1176				Ldb2178		spermidine/putrescine ABC transporter permease
COG1177				Ldb2179		spermidine/putrescine ABC transporter permease
COG0687				Ldb2181		spermidine/putrescine ABC transporter substrate-binding protein

Table S8. Unique genes of each *L. bulgaricus* strain.

Locus	Class	Locus	Class
LBU1138	Amino acid synthesis	LDBND_0093	Amino acid synthesis
LBU1183	Protease	LDBND_0094	Amino acid synthesis
LBU0206	Transporter	LDBND_0755	Amino acid synthesis
LBU0209	Transporter	LDBND_0756	Amino acid synthesis
LBU0217	Transporter	LDBND_0788	Amino acid synthesis
LBU0355	Transporter	LDBND_1721	Amino acid synthesis
LBU0365	Transporter	LDBND_0101	Protease
LBU0809	Transporter	LDBND_0478	Protease
LBU1410	Transporter	LDBND_1003	Protease
LBU1411	Transporter	LDBND_1120	Protease
LBUL_1924	Protease	LDBND_1197	Protease
Ldb2178	Transporter	LDBND_1236	Protease
Ldb2179	Transporter	LDBND_1579	Protease
Ldb2180	Transporter	LDBND_1934	Protease
Ldb2181	Transporter	LDBND_0236	Transporter
		LDBND_0237	Transporter
		LDBND_0624	Transporter
		LDBND_0934	Transporter
		LDBND_0961	Transporter
		LDBND_0962	Transporter
		LDBND_1404	Transporter
		LDBND_1578	Transporter
		LDBND_1764	Transporter
		LDBND_2037	Transporter

