Food Microbiology 77 (2019) 146–157

Contents lists available at ScienceDirect

Food Microbiology

journal homepage: www.elsevier.com/locate/fm

Genomic and metabolic features of the *Bacillus amyloliquefaciens* group–*B. amyloliquefaciens*, *B. velezensis*, and *B. siamensis*– revealed by pan-genome analysis

Byung Hee Chun, Kyung Hyun Kim, Sang Eun Jeong, Che Ok Jeon*

Department of Life Science, Chung-Ang University, Seoul, 06974, Republic of Korea

ARTICLE INFO

Keywords:
Bacillus amyloliquefaciens
Bacillus velezensis
Bacillus siamensis
Pan-genome
Genomic and metabolic features

ABSTRACT

The genomic and metabolic features of the *Bacillus amyloliquefaciens* group comprising *B. amyloliquefaciens*, *B. velezensis*, and *B. siamensis* were investigated through a pan-genome analysis combined with an experimental verification of some of the functions identified. All *B. amyloliquefaciens* group genomes were retrieved from GenBank and their phylogenetic relatedness was subsequently investigated. Genome comparisons of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* showed that their genomic and metabolic features were similar; however species-specific features were also identified. Energy metabolism-related genes are more enriched in *B. amyloliquefaciens*, whereas secondary metabolite biosynthesis-related genes are enriched in *B. velezensis*. Compared to *B. amyloliquefaciens* and *B. siamensis, B. velezensis* harbors more genes in its core-genome which are involved in the biosynthesis of antimicrobial compounds, as well as genes involved in D-galacturonate and D-fructuronate metabolism. *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* all harbor a xanthine oxidase gene cluster (*xoABCDE*) in their core-genomes that is involved in metabolizing xanthine and uric acid to glycine and oxalureate. A reconstruction of *B. amyloliquefaciens* group metabolic pathways using their individual pan-genomes revealed that the *B. amyloliquefaciens* group strains have the ability to metabolize diverse carbon sources aerobically, or anaerobically, and can produce various metabolites such as lactate, ethanol, acetate, CO_2, xylitol, diacetyl, acetoin, and 2,3-butanediol. This study therefore provides insights into the genomic and metabolic features of the *B. amyloliquefaciens* group.

1. Introduction

The name *Bacillus amyloliquefaciens* as an amylase- and protease-producing bacteria, was first proposed by Fukumoto in 1943 (Fukumoto, 1943), but it was not validated in published form until the 1980s, being just recognized as a close relative of *B. subtilis* or *B. subtilis* subsp. *amyloliquefaciens* because of its similarity to *B. subtilis* in terms of phenotypic properties. In 1987, the name "*Bacillus amyloliquefaciens*" was validated as a new species by comparing the DNA homology with *B. subtilis* and closely related taxa (Fukumoto, 1943; Priest et al., 1987). The name *B. velezensis* was first proposed in 2005 as a phylogenetic lineage distinct from *B. amyloliquefaciens* (Ruiz-Garcia et al., 2005). A previous study based on average nucleotide identity (ANI), *in silico* DNA-DNA hybridization (DDH), and a core-genome-based phylogenetic analysis showed that '*B. methylotrophicus*', '*B. amyloliquefaciens* subsp. *plantarum*', and '*B. oryzicola*' published in 2010, 2011, and 2015, respectively (Borriss et al., 2011; Chung et al., 2015; Madhaiyan et al., 2010), were later heterotypic synonyms of *B. velezensis* (Dunlap et al., 2016), indicating that they all belong to *B. velezensis*. *B. siamensis* was first published in 2010 (Sumpavapol et al., 2010), and the name '*Bacillus vanillea*' was proposed in 2015 (Chen et al., 2015) being a later heterotypic synonym of *B. siamensis* (Dunlap, 2015). Because these three *Bacillus* species, *B. amyloliquefaciens*, *B. velezensis*, and *B. siamensis*, are phylogenetically closely clustered within the *B. subtilis* species complex, they are referred as the "operational group *B. amyloliquefaciens*" (Fan et al., 2017).

Some *B. velezensis* and *B. siamensis* strains have been suggested to be plant pathogen-controlling agents because they produce diverse antimicrobial compounds such as surfactin, iturin, bacilysin, fengycin, macrolactin, bacilliaene, difficidin, plantazolicin, amylocyclicin, bacillibactin, mersacidin, ericin, and LCI (Chen et al., 2007; Fan et al., 2017; He et al., 2013; Jeong et al., 2012; Palazzini et al., 2016; Rückert et al., 2011). Genome-based study of organisms is a very useful and efficient approach to understand their life styles and genomic features

* Corresponding author. Department of Life Science, Chung-Ang University, 84, HeukSeok-Ro, Dongjak-Gu, Seoul, 06974, Republic of Korea.
E-mail address: cojeon@cau.ac.kr (C.O. Jeon).

https://doi.org/10.1016/j.fm.2018.09.001
Received 13 March 2018; Received in revised form 31 July 2018; Accepted 1 September 2018
Available online 05 September 2018
0740-0020/

and the first genome-based study to understand the life style and genomic features of *B. velezensis* species was accomplished in 2007 (Chen et al., 2007). In recent years, with the development of high-throughput and low-cost sequencing technologies, genomic information-based approaches have been used extensively to obtain a comprehensive understanding of the metabolic properties and lifestyle traits of organisms. However, most of the genome-based studies of *B. amyloliquefaciens* group members have focused on taxonomic analysis or profiling or have been limited to the development of plant pathogen-controlling agents by comparing their antimicrobial activities (Kim et al., 2017; Yi et al., 2014).

A pan-genome analysis can provide insights into genomic and metabolic diversities, as well as the lifestyle traits of phylogenetic lineages, because a pan-genome analysis describes all the possible metabolic and physiological repertories of phylogenetic lineages (Chun et al., 2017; Deng et al., 2010; Douillard et al., 2013). *B. amyloliquefaciens* group members have been reported to be mainly responsible for the fermentation of traditional fermented foods such as doenjang (a Korean traditional fermented soybean paste), ganjang (a Korean traditional fermented soybean sauce), douche (a Chinese traditional soybean fermented food), and poo-khem (a Thai traditional fermented salted crab food) (Alvindia and Natsuaki, 2009; Peng et al., 2003; Sumpavapol et al., 2010; Yao et al., 2016), and they have been proposed as starter cultures for the fermentation of various foods because they have high amylase and protease activities and food pathogen-controlling properties (Compaoré et al., 2013; Eom and Choi, 2016; Lee et al., 2017; Meidong et al., 2017; Zhang et al., 2016a, 2017). Therefore, in this study, the phylogenetic and genomic diversities and the genomic and metabolic features of the *B. amyloliquefaciens* group (*B. amyloliquefaciens*, *B. velezensis*, and *B. siamensis*) were investigated via clusters of orthologous groups (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Basic Local Alignment Search Tool (BLAST) analyses using all genomes of the *B. amyloliquefaciens* group members available in the GenBank database. To our knowledge, the genomic and fermentative metabolic features of *B. amyloliquefaciens*, *B. velezensis*, and *B. siamensis* have not yet been comprehensively explored using their pan-genomes, and this study will provide insights into the fermentative metabolic features of the *B. amyloliquefaciens* group members during food fermentations.

2. Materials and methods

2.1. Collection of genomes belonging to the B. amyloliquefaciens group

To collect all the publicly available genomes belonging to the *B. amyloliquefaciens* group (*B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis*), the genomes of all strains with more than 98.0% 16S rRNA gene sequence similarities with their type strains (strains DSM 7^T, KCTC 13613^T, and NRRL B-41580^T, respectively) were retrieved from GenBank. The completeness and contamination rates of the retrieved genomes were assessed using the CheckM program (ver. 1.0.4) (Parks et al., 2015) based on the lineage-specific marker genes of the genus *Bacillus*, and qualified genomes showing ≥ 90.0% completeness and ≤10.0% contamination rates were used for further analyses.

To select genomes phylogenetically belonging only to the *B. amyloliquefaciens* group from the high-quality genomes, genome-based ANI and *in silico* DDH analyses for the type strains of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* were performed, as described previously (Chun et al., 2017). The ANI and *in silico* DDH values were calculated using a stand-alone program (http://www.ezbiocloud.net/sw/oat) (Lee et al., 2016) and the server-based genome-to-genome distance calculator (ver. 2.1, http://ggdc.dsmz.de/distcalc2.php) (Meier-Kolthoff et al., 2013), respectively. The pair-wise values of ANI and *in silico* DDH relatedness among genomes belonging to the *B. amyloliquefaciens* group were visualized as heat-maps and hierarchical clusters using GENE-E (http://www.broadinstitute.org/cancer/software/GENE-E/).

2.2. Phylogenetic analyses based on 16S rRNA gene sequences and core-genome

To infer the evolutionary relationships between all genomes belonging to the *B. amyloliquefaciens* group, phylogenetic analyses based on 16S rRNA gene sequences and core genomes were performed. For a 16S rRNA gene sequence-based phylogenetic analysis, the 16S rRNA gene sequences of all genomes of the *B. amyloliquefaciens* group members were aligned using the Infernal secondary-structure aware aligner, available in the Ribosomal Database Project (http://rdp.cme.msu.edu/) (Nawrocki and Eddy, 2007), and a phylogenetic tree using the neighbor-joining (NJ) algorithm was constructed using the MEGA program (ver. 7) (Kumar et al., 2016). For the core-genome-based phylogenetic analysis, core-genomes of the genomes of all *B. amyloliquefaciens* group members and the genome of *B. subtilis* NCIB 3610^T (as an out-group) were extracted using the USEARCH program (ver. 9.0) (Edgar, 2010), with a 50% sequence identity cut-off, available in a bacterial pan-genome analysis (BPGA) pipeline (ver. 1.3) (Chaudhari et al., 2016). Concatenated amino acid sequences of the core-genomes were aligned using the MUSCLE program (ver. 3.8.31) (Edgar, 2004), and a phylogenetic tree was constructed using the NJ algorithm available in MEGA (Kumar et al., 2016).

2.3. Pan- and core-genome, COG, and KEGG analyses

Pan- and core-genomes of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* were analyzed using the BPGA pipeline, with a 50% sequence identity cut-off. The assignment of functional genes derived from either the pan-genomes or the core-genomes (including the soft core-genome in the case of *B. velezensis*) of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* into COG categories was performed using the USEARCH program query of the COG database within the BPGA pipeline. Predicted proteins derived from the respective genomes of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* were submitted to BlastKOALA (http://www.kegg.jp/blastkoala/) (Kanehisa et al., 2016) for functional annotation based on KEGG Orthology (KO), and their metabolic pathways based on KO numbers were generated using the iPath v2 module (http://pathways.embl.de/iPath2.cgi#) (Yamada et al., 2011). KEGG metabolic pathways for the respective *Bacillus* species were displayed by line thickness based on the numbers of strains harboring genes with the same KO numbers. The core-genomes (including the soft core-genome of *B. velezensis*) of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* were compared, and their species-specific genes were listed after confirming them through a BlastN analysis.

2.4. Genomic and functional analysis of antimicrobial genes

Antimicrobial genes or operons in the genomes of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* were identified by bioinformatics, and their antimicrobial activities were assessed. The presence or absence of antimicrobial genes in the genomes of the respective *Bacillus* species was manually confirmed through a BlastN search against their genomes, using reference antimicrobial gene sequences available from other closely related bacteria deposited in GenBank. The antimicrobial activities of the type strains of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* (DSM 7^T, KCTC 13613^T, and LMG 22478^T, respectively) were quantitatively evaluated against five representative pathogens (*Escherichia coli* O157:H7, *B. cereus* ATCC 27348, *Staphylococcus aureus* KCTC 3881, *Listeria monocytogenes* KCTC 13064, and *Aspergillus flavus* subsp. *flavus* KACC 41809) grown on tryptic soy agar (TSA, Becton Dickinson, USA; for bacteria) and potato dextrose agar (Becton Dickinson, USA; for fungi). To assess antibacterial activities, 5 μl of pathogenic strains grown overnight in tryptic soy broth (TSB, Becton Dickinson, USA) were added to 5 ml of low melting TSA (TSB with 2.0% low melting agar; Bio-Rad, USA) to a final concentration of approximately 10^6 colony-forming units/mL, which was over-layered on

solidified TSA. After the top agars were hardened, 5 μL of cell-free supernatants from strains DSM 7^T, KCTC 13613^T, and LMG 22478^T cultivated in TSB overnight were poured onto 5-mm filter papers (Whatman, USA) on the top agar and the antibacterial activities were evaluated by examining the formation of clear zones around the filter papers after 24 h incubation at 37 °C (Khochamit et al., 2015). The antifungal activities were assessed as described previously (Jeon et al., 2016).

2.5. Genomic and functional analysis of D-galacturonate, D-fructuronate, and D-glucuronate metabolism

The genes involved in the metabolism of D-galacturonate, D-fructuronate, and D-glucuronate from *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* species genomes were examined by bioinformatics and based on this data the metabolic capabilities of the different species were assessed. The genes involved in the metabolism of D-galacturonate, D-fructuronate, and D-glucuronate were identified in the core-genomes of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis*, and manually confirmed through BlastN analyses based on their corresponding reference gene sequences available in closely related bacteria. To assess the metabolic capabilities of the *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* species to use D-galacturonate and D-glucuronate as the sole carbon and energy sources, cells from the representative type strains DSM 7^T, KCTC 13613^T, and LMG 22478^T grown in TSB were inoculated (2%, v/v) into M9 minimal broth (Mathews et al., 2014) containing D-galacturonate (0.2%, w/v), D-glucuronate (0.2%, w/v), or glucose (0.2%, positive control) as the sole carbon and cultivated with shaking (220 rpm) at 37 °C. D-fructuronate was not tested, because D-fructuronate was not commercially available. Growth was monitored by measuring the culture's optical density (OD) at 600 nm. The metabolic capabilities of *B. velezensis* strains KD1 and YJ11-1-4 for D-galacturonate and D-glucuronate were additionally tested using the same procedure.

2.6. Genomic analysis of xanthine metabolic pathways and xanthine oxidase assay

The xanthine metabolic pathways in the genomes of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* were investigated using bioinformatics. Xanthine metabolic genes were identified in the core-genomes from *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis*, and manually confirmed through BlastN analyses based on the reference gene sequences of closely related bacteria. To assess xanthine oxidase activity, cells from the strains DSM 7^T, KCTC 13613^T, and LMG 22478^T, aerobically grown in 5 ml TSB at 37 °C for 24 h, were harvested by centrifugation (12,300 × *g*, 3 min) and disrupted by sonication (interval: 3 s; reaction time: 10 min; amplitude: 23%) in 1 ml of 0.1 M Tris-HCl buffer (pH 7.5). The assay mixture for the xanthine oxidase assay consisted of 2.24 ml of 0.1 M Tris-HCl (pH 7.5), 0.08 ml of 10 mM xanthine in 0.25 M NaOH, 0.08 ml of 1 mM oxonic acid in distilled water, and 0.1 ml of cell lysate as described previously (Masuoka et al., 2015). The production of uric acid from xanthine was monitored by measuring the absorbance at 293 nm over time (Masuoka et al., 2015). The metabolic capabilities of the strains DSM 7^T, KCTC 13613^T, and LMG 22478^T for xanthine as the sole carbon and energy source were evaluated, as described for the metabolism of D-galacturonate and D-glucuronate.

2.7. Phenotypic characterization and reconstruction of metabolic pathways

The metabolic capabilities of the type strains of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* (strains DSM 7^T, KCTC 13613^T, and LMG 22478^T, respectively) to metabolize various carbohydrates were tested using API 50 CHB (bioMèrieux, France), according to the manufacturer's instructions. Anaerobic growth was assessed on TSA and TSA supplemented with external electron acceptors [sodium nitrate (10 mM) or sodium nitrite (2 mM)] under anaerobic conditions (4–10% CO_2) using the GasPak Plus system (BBL) at 37 °C for 21 days. To investigate the metabolic features of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis*, their metabolic pathways for various carbohydrates were reconstructed based on the predicted KEGG pathways and EC numbers. In addition, the presence or absence of the metabolic genes in each *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* species was manually confirmed through BlastP analyses against their genomes, using reference protein sequences available in other closely related bacteria.

3. Results and discussion

3.1. Genome collection of the B. amyloliquefaciens group

At the time of writing (October 2017), a total of 276 genomes sharing more than 98.0% 16S rRNA gene sequence similarities with the type strains of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* (DSM 7^T, KCTC 13613^T, and LMG 22478^T, respectively) were retrieved from GenBank. Four low-quality genomes were removed using the CheckM program based on the criteria that the genome is considered to be near complete (≥90%), with medium contamination (≤10%) (Parks et al., 2015). Eventually, 272 qualified genomes were used for the ANI and *in silico* DDH analyses.

The ANI and *in silico* DDH analyses showed that among the 272 genomes, 93 genomes shared more than 95–96% ANI and 70% *in silico* DDH values, which have been used as the criteria for the delineation of prokaryotic species (Goris et al., 2007; Richter and Rosselló-Móra, 2009; Rosselló-Móra and Amann, 2015), with the type strains of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* (supplementary Table S1). The other genomes shared relatively low ANI (< 86.8%) and *in silico* DDH (< 31.5%) values with the type strains, clearly suggesting that they did not belong to the *B. amyloliquefaciens* group. The 93 genomes belonging to the *B. amyloliquefaciens* group had very high 16S rRNA gene sequence similarities (> 99.2%), high ANI (> 93.4%), and *in silico* DDH (> 52.3%) values (supplementary Fig. S1), suggesting that members of the *B. amyloliquefaciens* group clustered together, with a clear distinction from other *Bacillus* groups, and that they probably evolved from a common ancestor. The ANI and *in silico* DDH analyses based on the type strains DSM 7^T, KCTC 13613^T, and NRRL B-41580^T showed that the 93 genomes separately clustered into three *Bacillus* species, *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis*, with 7, 5, and 81 genomes, respectively. The classification of the *B. amyloliquefaciens* group strains into the three *Bacillus* species was consistent with previous results (Dunlap et al., 2016; Fan et al., 2017). The average genome sizes, gene numbers, and G + C contents of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* strains were 3.9 ± 0.1 Mb, 4016 ± 135 genes, and 45.9 ± 0.2%; 3.9 ± 0.2 Mb, 3915 ± 196 genes, and 46.1 ± 0.3%; and 4.0 ± 0.1 Mb, 3955 ± 127 genes, and 46.4 ± 0.4%, respectively.

3.2. Phylogenetic analyses of B. amyloliquefaciens group members based on 16S rRNA gene sequences and core-genome

To infer the phylogenetic relatedness among the 93 genomes of the *B. amyloliquefaciens* group, a phylogenetic analysis based on the 16S rRNA gene sequences was performed (supplementary Fig. S2A). The phylogenetic tree showed that most of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* strains relatively clustered together into their type strains, although 16S rRNA gene sequence similarities were very high. However, strains SDLI1, JS25R, BH072, and Y2 did not cluster into their type strains. Moreover, a *B. velezensis* strain (W2) formed a phylogenetic lineage clearly distinct from the type strains of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis*. These results confirm again that 16S rRNA gene sequence-based analysis may not be appropriate to classify the *B. amyloliquefaciens* group (Dunlap et al., 2016; Fan et al.,

2017).

To infer genome-based phylogenetic relationships among the 93 genomes, a phylogenetic analysis using the concatenated amino acid sequences of core-genomes (1867 genes) derived from the 93 genomes and *B. subtilis* NCIB 3610^T genome was constructed (supplementary Fig. S2B). The phylogenetic tree showed that the 93 genomes of the *B. amyloliquefaciens* group were clearly separated into three different phylogenetic lineages, suggesting that the members of the *B. amyloliquefaciens* group might have independently speciated into *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis*. These genome classifications into *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* species were consistent with those reported previously (Dunlap et al., 2016; Fan et al., 2017). The ANI and *in silico* DDH values and core-genome-based on phylogenetic analyses suggest that genome-based phylogenetic analyses were necessary for the correct classification of the *B. amyloliquefaciens* group members and many genomes belonging to the *B. amyloliquefaciens* group in GenBank are misnamed.

3.3. Pan- and core-genome analysis of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis*

Pan- and core-genome analysis is an important approach to explore the genomic and metabolic repertories of a phylogenetic lineage in organisms (Chun et al., 2017; Shin et al., 2016). Therefore, pan- and core-genome analysis of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* species was conducted using all their genomes (Fig. 1). In a pan-genome analysis, the number of cumulative genes for an increase of genomes used can be estimated by the Heaps' law ($n = k*N^{-\alpha}$, where n is the expected gene number for an increase of genomes (N) and k is a constant to fit the curve) (Chun et al., 2017; Tettelin et al., 2008). The formula revealed that the pan-genomes of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* species increased with α values of 0.098, 0.125, and 0.158, respectively, suggesting that they all might be open pan-genomes experiencing frequent evolutionary changes through gene gains and losses or lateral gene transfers for efficient environmental adaptations.

Diverse sequencing errors are always generated by base over- or under-call or amplification errors during genome sequencing of organisms, eventually causing erroneous descriptions of normal genes as pseudogenes or non-coding genes in a genome annotation process. Therefore, it is reasonable to assume that some portions of pseudogenes indicated in the genomes presented in supplementary Table S1 may be genes performing normal metabolic and physiological functions. It has been suggested that in a pan- and core-genome analysis, genes identified from 95 to 99% genomes used could be defined as soft core-genome that was considered as common metabolic and functional features of a phylogenetic clade (Bezuidt et al., 2016; Periwal et al., 2015). Therefore, in this study, genes identified in more than 77 B. *velezensis* genomes were considered as common metabolic and functional features of *B. velezensis*, while only genes identified in all *B. amyloliquefaciens* and *B. siamensis* genomes were considered as their common metabolic and functional features (supplementary Fig. S3). The pan-genomes for *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* contained a total of 4,477, 4,397, and 6538 genes consisting of 3,064, 3,066, and 3172 genes in the core-genomes (including soft core-genomes), 885, 613, and 2471 genes in the accessory-genomes (present in more than two strains), and 528, 718, and 895 genes in the unique genomes, respectively, whose differences might be mainly caused by the number of genomes used in the analysis because the number of pan- and core-genes represented a function of the number of genomes used for a pan-genome analysis.

3.4. Metabolic and functional features of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis*

A core-genome that describes common genes shared by all genomes

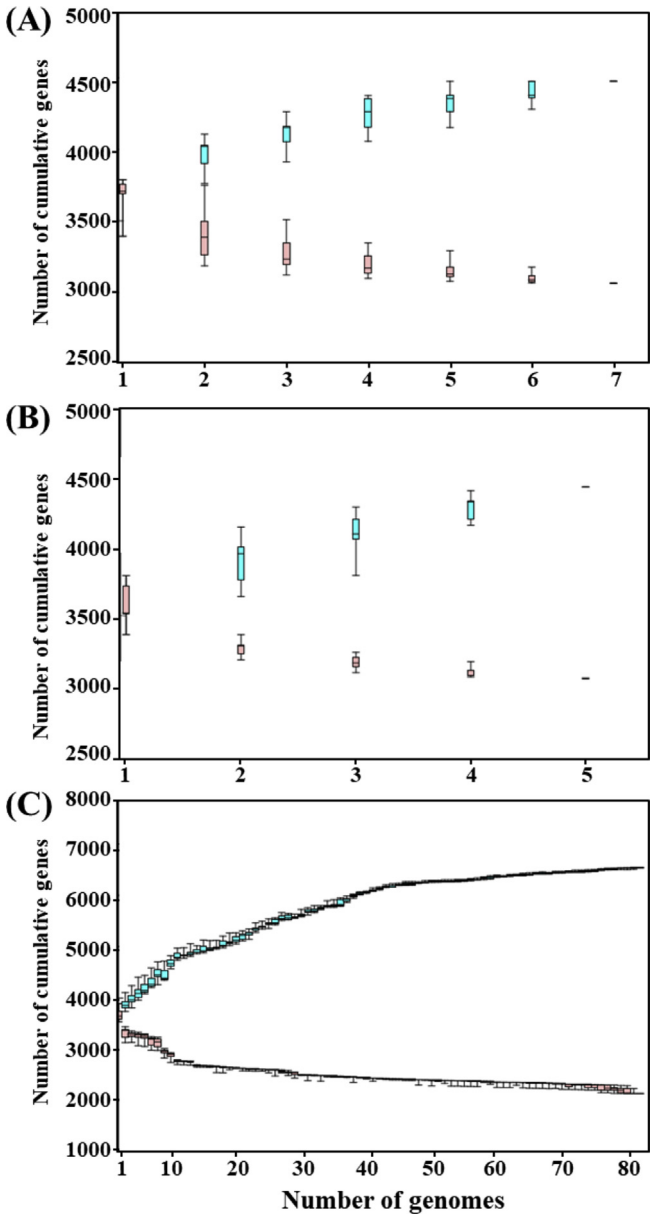


Fig. 1. Pan- and core-genome plots of *Bacillus amyloliquefaciens* (A), *B. siamensis* (B), and *B. velezensis* (C). For each species, randomly ordered genome lists were generated, and 20 sets of the randomly ordered genomes were subjected to pan- and core-genome analyses. The average sizes of the core- (pink color) and pan-genome (cyan color) are plotted, with the error bars representing standard deviations. The pan-genomes represent the total genes present in the genomes of the sampled subsets, and the core-genomes represent the genes shared in common by all genomes in the same subsets. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

of a phylogenetic clade represents common metabolic and functional features of a phylogenetic clade, while a pan-genome that describes total genes existing in all genomes of a phylogenetic clade represents all metabolic and functional repertories in a phylogenetic clade (Chun et al., 2017). To investigate the metabolic and functional features of *B. amyloliquefaciens, B. siamensis*, and *B. velezensis*, the COG distributions of pan- and core-genomes (including soft core-genomes) in each species were analyzed and compared (Fig. 2). The COG distributions of the pan-genome in each species were similar. However, energy metabolism-related genes including energy production and conversion (C), translation, ribosomal structure, and biogenesis (J), and cell envelope

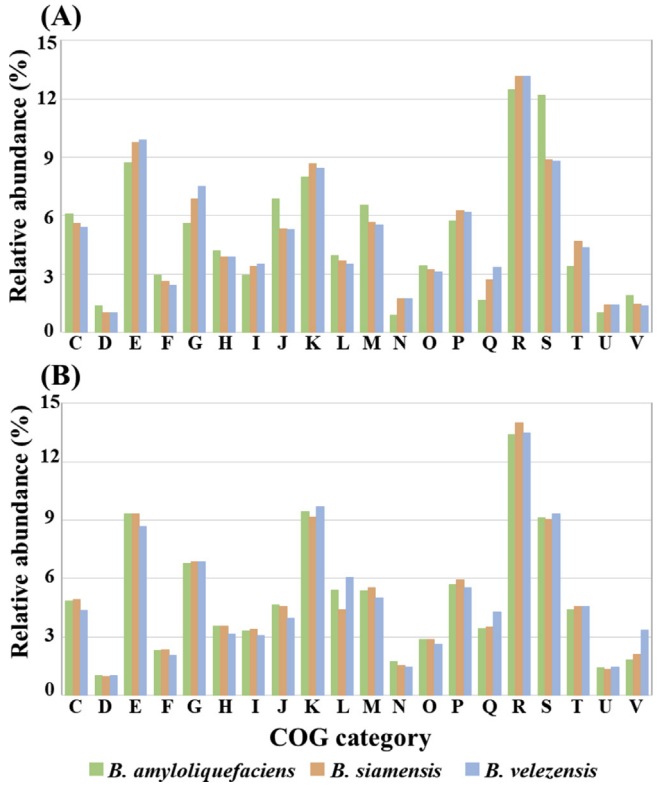


Fig. 2. Comparisons of functional categories in the clusters of orthologous groups (COG) functional categories in the pan- (A) and core- (B) genomes of *Bacillus amyloliquefaciens*, *B. siamensis*, and *B. velezensis*. The alphabetical codes represent the following COG functional categories: C, energy production and conversion; D, cell division and chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme metabolism; I, lipid metabolism; J, translation, ribosomal structure, and biogenesis; K, transcription; L, DNA replication, recombination, and repair; M, cell envelope biogenesis, outer membrane; N, cell motility and secretion; O, post-translational modification, protein turnover, and chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolite biosynthesis, transport, and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms.

biogenesis, outer membrane (M) were enriched in the pan-genome of *B. amyloliquefaciens*, while the COG category genes of carbohydrate transport and metabolism (G) and secondary metabolite biosynthesis, transport, and catabolism (Q) were enriched in the pan-genome of *B. velezensis* (Fig. 2A). The functional genes related to defense mechanisms (V) and secondary metabolite biosynthesis, transport, and catabolism (Q) were enriched in the core-genome of *B. velezensis* (Fig. 2B). These COG analyses of the *B. amyloliquefaciens* group suggest that *B. amyloliquefaciens* strains might have evolved to have more efficient energy metabolisms, while *B. velezensis* strains might have evolved to have more efficient defense systems.

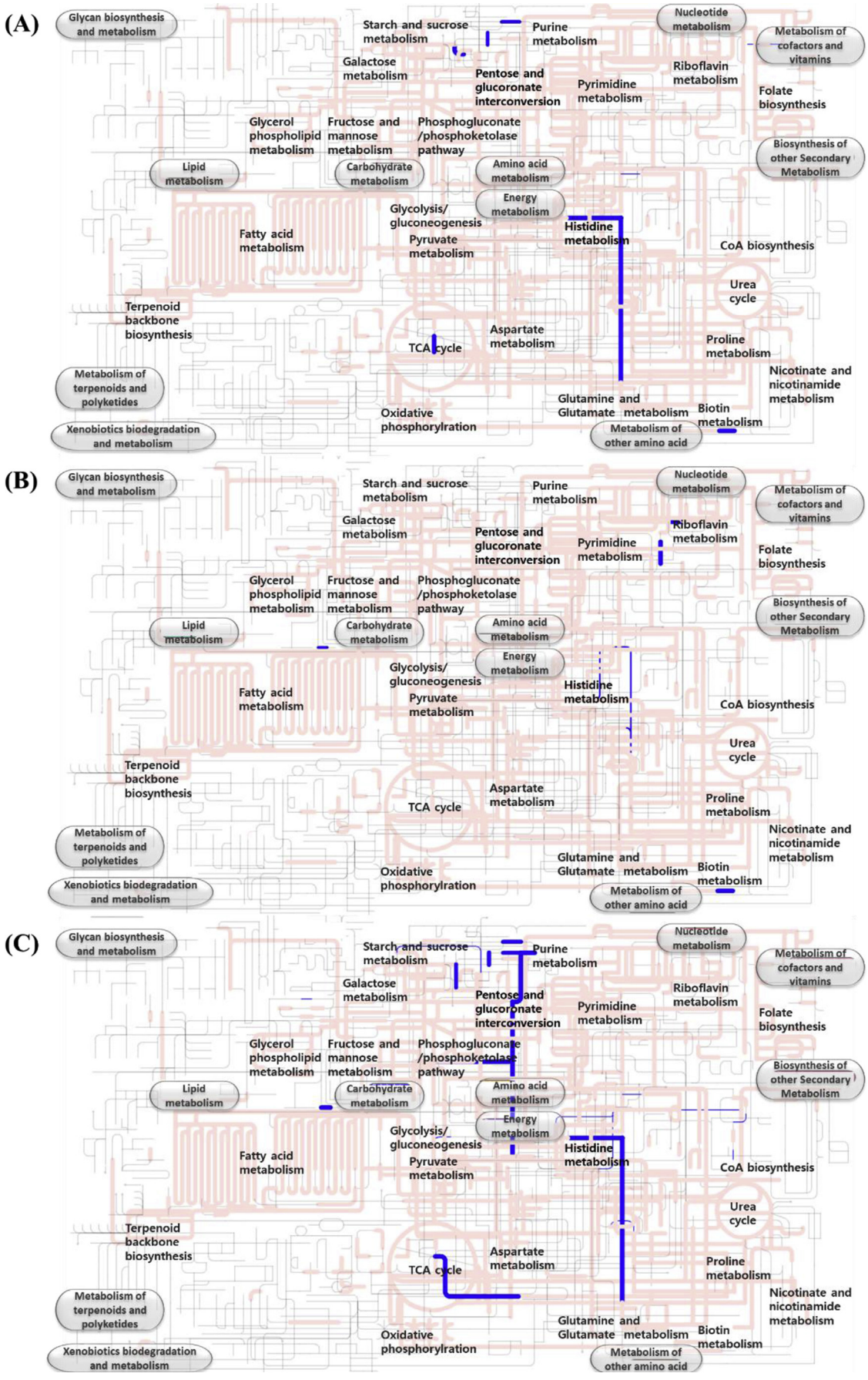
The metabolic features of *B. amyloliquefaciens, B. siamensis*, and *B. velezensis* were also compared through a KEGG pathway analysis using all genomes of each species (Fig. 3). The overall KEGG metabolic pathways of three *Bacillus* species were found to be quite similar. However, the KEGG metabolic pathway analysis showed that a histidine metabolism pathway (K01745, K01712, K01468, and K01479) and a glycerol phospholipid metabolic gene (K01046) were identified from the core genomes of only *B. amyloliquefaciens* and *B. velezensis*. In addition, the analysis showed that genes related to pentose and glucuronate inter-conversions (K01625, K01686, K01685, and K00874) were identified in only *B. velezensis* as core-genome, suggesting that the metabolism can be a species-specific metabolic feature that can be used to differentiate *B. velezensis* from *B. amyloliquefaciens* and *B. siamensis*.

To investigate the species-specific metabolic and functional features of *B. amyloliquefaciens, B. siamensis*, and *B. velezensis*, their core-genomes (including soft core-genomes in *B. velezensis*) were compared, and their species-specific genes are listed (supplementary Table S2). *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* harbored 21, 16, and 13 species-specific genes, respectively, probably contributing to their species-specific metabolic and functional features and possibly useful as markers to differentiate *B. amyloliquefaciens, B. siamensis*, and *B. velezensis* (supplementary Table S2). In particular, *B. velezensis* harbored additional genes responsible for D-galacturonate and D-fructuronate metabolisms (galaturonate isomerase, altronate oxidoreductase, altronate dehydratase, 2-dehydro-3-deoxygluconokinase, deoxyphosphogluconate aldolase, mannoate oxidoreductase, and mannoate dehydratase) as the core-genome. The species-specific gene analysis also showed that only *B. siamensis* harbored a xanthine metabolic pathway including xanthine permease-, xanthine oxidase-, and uric acid oxidase-coding genes in the core-genome. Because functional features coded by species-specific genes can be used as phenotypic biomarkers to differentiate *B. amyloliquefaciens, B. siamensis*, and *B. velezensis*, some species-specific metabolic features were experimentally confirmed.

3.5. Genomic and functional analysis of antimicrobial features in *B. amyloliquefaciens, B. siamensis, and B. velezensis*

Genes or operons associated with antimicrobial properties were bioinformatically investigated in *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* to evaluate their potential antimicrobial activities (Table 1, supplementary Fig. S4). The bioinformatic analysis showed that genes or operons related to the biosynthesis of surfactin, bacillibactin, amylocyclicin, and iturin were identified from all genomes of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis*, which was consistent with previous results (Fan et al., 2017; Palazzini et al., 2016), indicating that their presence may be a common genomic feature of the *B. amyloliquefaciens* group. The analysis also showed that biosynthetic genes or operons capable of producing macrolactin, bacilysin, difficidin, LCI, and fengycin were identified from the core-genome of *B. velezensis*, suggesting that they are common genetic features of *B. velezensis*. However, they (except for LCI) were identified from the only accessory- or unique-genomes of *B. amyloliquefaciens* and *B. siamensis*; a gene encoding LCI was identified from the core-genome of *B. siamensis*, while it was identified from the accessory-genome of *B. amyloliquefaciens*. A biosynthetic operon of fengycin was identified from all *B. amyloliquefaciens* strains, but the operon lacked the genes *fenA*, *fenB*, and *fenC* when compared to those of *B. velezensis* strains, suggesting that the fengycin biosynthetic operon in *B. amyloliquefaciens* strains may be incomplete and hence did not favor the production of fengycin (Borriss et al., 2011). The biosynthetic genes of macrolactin and difficidin were identified from strain RD7-7 only, among the *B. amyloliquefaciens* strains. Biosynthetic genes of bacilysin, macrolactin, and fengycin were identified from strain SDLI1 only, among the *B. siamensis* strains. A biosynthetic gene of LCI was identified from all *B. amyloliquefaciens* strains except for strains K2 and RD7-7, but their sequences had a low sequence homology with those in *B. siamensis* and *B. velezensis* (< 44% nucleotide similarities), which suggests that the biosynthetic gene of LCI in *B. amyloliquefaciens* strains may be derived from a different bacterial ancestor. The bioinformatic analysis showed that only *B. velezensis* strains additionally harbored biosynthetic operons of plantazolicin, mersacidin, and ericin in their accessory- or unique-genome, as described previously (He et al., 2013; Palazzini et al., 2016; Scholz et al., 2011). These results suggest that *B. velezensis* strains may have stronger and broader antimicrobial activities against diverse pathogens compared to *B. amyloliquefaciens* and *B. siamensis* strains.

Many *B. amyloliquefaciens* group strains have been reported as plant-associated bacteria with plant growth-promoting properties (Chen



(caption on next page)

Fig. 3. Metabolic Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in *Bacillus amyloliquefaciens* (A), *B. siamensis* (B), and *B. velezensis* (C). The pathways were generated using the iPath v2 module based on KEGG Orthology numbers of genes identified in all genomes in each species. Line thicknesses are proportional to the number of genomes harboring the metabolic pathways in each species. Metabolic pathways that are common in all three *Bacillus* species are drawn in faint red, and those that are not common within the three *Bacillus* species are depicted in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

et al., 2007, 2009b; Zhang et al., 2016b). On the other hand, *B. amyloliquefaciens* group strains have been reported to play important roles during the fermentation of traditional fermented foods and they have been proposed as starter cultures for the fermentation of various foods (Alvindia and Natsuaki, 2009; Peng et al., 2003; Sumpavapol et al., 2010; Yao et al., 2016). Antimicrobial activity has been known to be one of the important properties of plant growth-promoting bacteria (PGPB) as well as starter cultures of fermented foods because they can repress the growth of plant pathogens or food-driven pathogens (Compaoré et al., 2013; Jeon et al., 2016; Meidong et al., 2017; Zhang et al., 2016a, 2017), which suggests that *B. velezensis* strains can be good candidates of PGPB and starter cultures of fermented foods.

The antimicrobial properties of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* were assessed based on the inhibitory activities of their type strains (DSM 7^T, KCTC 13613^T, and LMG 22478^T, respectively) against five representative pathogenic strains *E. coli*, *B. cereus*, *S. aureus*, *L. monocytogenes*, and *A. flavus* (Table 2). *B. amyloliquefaciens* DSM 7^T did not show evident antimicrobial activities against any of the test strains, while *B. velezensis* LMG 22478^T exhibited inhibitory activity against all test strains, especially *S. aureus*, *L. monocytogenes*, and *A. flavus*. *B. siamensis* KCTC 13613^T showed antimicrobial activity against *E. coli*, *S. aureus*, and *L. monocytogenes*. The presence of antimicrobial genes and the result of antimicrobial tests shown in Tables 1 and 2 suggest that the biosynthesis genes of macrolactin, LCI, and fengycin mainly contribute to the antimicrobial activities of *B. velezensis*.

3.6. Genomic and functional analysis of D-galacturonate, D-fructuronate, and D-glucuronate metabolism

Hexuronates including galacturonate, fructuronate, and glucoronate are sugar acids derived from galactose, fructose, and glucose, respectively, with their sixth carbon atoms oxidized to a carboxylic acid. Bioinformatic analysis of the genomes of the *B. amyloliquefaciens* group showed that the putative genes capable of metabolizing D-galacturonate and D-fructuronate to pyruvate and 3-glyceraldehyde-3-phosphate that were found in *B. velezensis* strains described as *Bacillus amyloliquefaciens* previously (He et al., 2013) were identified in the core-genome of *B. velezensis* (Fig. 4A), but they were not identified in the genomes of *B. amyloliquefaciens* and *B. siamensis* strains.

The D-galacturonate and D-fructuronate metabolic genes were split into two gene clusters in all *B. velezensis* strains (Fig. 4A), the genetic maps of which were identical to that of a *B. velezensis* strain previously described as *B. amyloliquefaciens* subsp. *plantarum* YAU B9601−Y2 (He et al., 2013). One gene cluster included D-galacturonate metabolic genes to metabolize D-galacturonate to 2-keto3-deoxygluconate (KDG), and the other gene cluster included complete D-fructuronate metabolic genes to metabolize D-fructuronate to pyruvate and 3-glyceraldehyde-3-phosphate through KDG (Fig. 4B). However, no gene encoding D-glucuronate isomerase, which converts D-glucuronate to D-fructuronate was identified, suggesting that *B. velezensis* was unable to metabolize D-glucuronate. No genes closely related to the D-galacturonate and D-fructuronate metabolic genes, present as the core-genome of *B. velezensis*, were searched from GenBank by BlastN, suggesting that the D-galacturonate and D-fructuronate metabolic genes of *B. velezensis* uniquely and independently descended from an ancestor.

The capabilities of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* to metabolize D-galacturonate and D-glucuronate were confirmed using their type strains (DSM 7^T, KCTC 13613^T, and LMG 22478^T, respectively) in M9 minimal media supplemented with 0.2% of D-galacturonate or D-glucuronate as a sole carbon and energy source. The growth tests showed that only strain LMG 22478^T grew well on D-galacturonate (Fig. 5), which was in well accordance with the bioinformatics results that a D-galacturonate metabolic pathway was identified in only *B. velezensis*, suggesting that the D-galacturonate metabolism is a species-specific metabolic property of *B. velezensis* in the *B. amyloliquefaciens* group. Similar to the bioinformatics results, the growth tests showed

Table 2
Quantitative antimicrobial activities of *Bacillus amyloliquefaciens* DSM 7^T, *B. siamensis* KCTC 13613^T, and *B. velezensis* LMG 22478^T against five representative pathogens[a].

Pathogenic strain	DSM 7^T	KCTC 13613^T	LMG 22478^T
Escherichia coli O157:H7	–	+	+
Bacillus cereus ATCC 27348	–	–	+
Staphylococcus aureus KCTC 3881	–	+	+ + +
Listeria monocytogenes KCTC 13064	–	+ + +	+ + +
Aspergillus flavus subsp. *flavus* KACC 41809	–	–	+ + +

[a] The antimicrobial activities were assessed based on the diameters (mm) of clear zones formed around paper disks with culture supernatants (against pathogenic bacteria) or colonies (against pathogenic fungi) of *B. amyloliquefaciens* DSM 7^T, *B. siamensis* KCTC 13613^T,and *B. velezensis* LMG 22478^T: , no clear zone; +, 0.01–1.0 mm; + +, 1.0–2.0 mm; + + +, > 2.0 mm.

Table 1
Putative antimicrobial genes or operons identified in the core-genomes of *Bacillus amyloliquefaciens* (*Ba*), *B. siamensis* (*Bs*), and *B. velezensis* (*Bv*) strains. +, identified; –, not identified.

Gene or operon	Product	*Ba*	*Bs*	*Bv*	Potential target organisms
mlnABCDEFGHI	Macrolactin	–	–	+	*Bacteria* (Schneider et al., 2007)
bacABCDEGH	Bacilysin	+	–	+	*Staphylococcus* (Chen et al., 2009a; Perry and Abraham, 1979)
dfnCBXYADEFGHIJMLK	Difficidin	–	–	+	*S. aureus*, *Escherichia coli* (Chen et al., 2009a)
fenABCDE	Fengycin	–	–	+	*Fungi* (Chen et al., 2007)
ituDABC	Iturin	+	+	+	*Fungi* (Xu et al., 2013)
srfABCD	Surfactin	+	+	+	*Bacteria* (Chen et al., 2007)
LCI	LCI	–	+	+	*Xanthomonas* (Gong et al., 2011)
baeBCDEGHIJLMNRS	Bacillaene	+	+	+	*E. coli* (Chen et al., 2007)
PznFKGHIAJCDBEL	Plantazolicin	–	–	–	*Bacillus* (Scholz et al., 2011)
dhbAECBF	Bacillibactin	+	+	+	Siderophore (Scholz et al., 2011)
mrsK2R2FGEAR1DMT	Mersacidin	–	–	–	Gram positive bacteria (He et al., 2013)
LanBTCASIFEGRK	Ericin	–	–	–	Antibacterial (Palazzini et al., 2016)
acnBADEFC	Amylocyclicin	+	+	+	*Bacillus cereus* (Scholz et al., 2014)

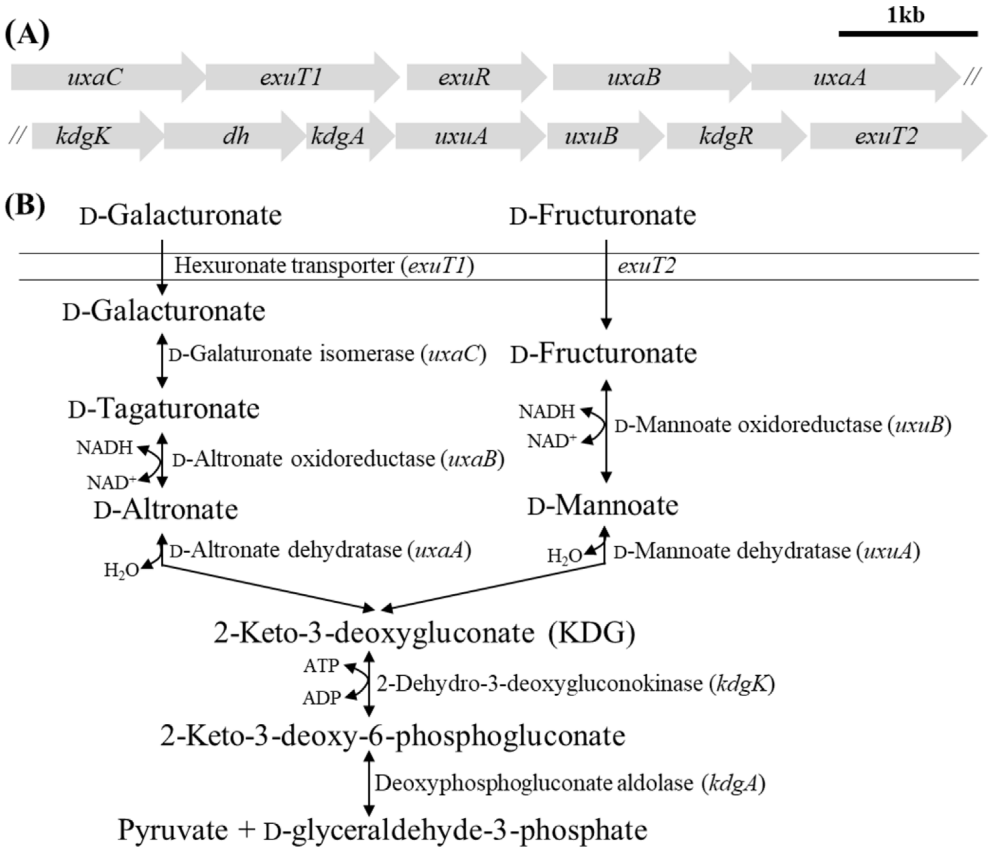


Fig. 4. Physical maps of the putative genes responsible for D-galacturonate and D-fructuronate metabolisms (A) and the proposed metabolic pathways in *Bacillus velezensis*. *exuR*, hexuronate transcription regulator; *kdgR*, 2-keto-3-deoxygluconate transcription regulator; *dh*, putative dehydrogenase.

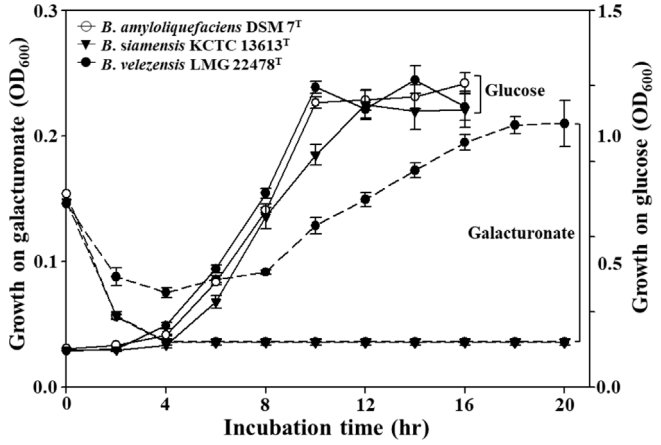


Fig. 5. Growth curves of *Bacillus amyloliquefaciens* DSM 7^T, *B. siamensis* KCTC 13613^T, and *B. velezensis* LMG 22478^T in minimal media supplemented with D-galacturonate and D-glucose as the sole carbon source.

that none of the test strains including strain LMG 22478^T grew on D-glucuronate as a sole carbon source (data not shown), indicating the lack of D-glucuronate metabolic genes. The growth tests showed that two other *B. velezensis* strains KD1 and YJ11-1-4 also had the same metabolic capabilities for D-galacturonate and D-glucuronate (data not shown), as strain LMG 22478^T, further reinforcing our results.

In previous studies, it was reported that *B. velezensis* strains (previously referred to as plant-associated *B. amyloliquefaciens*) might possess advantages to retain their plant-associated features as PGPB compared to *B. amyloliquefaciens* (previously referred to as non-plant-associated *B. amyloliquefaciens*) because they possess the metabolic property of hexuronates, cell wall components (Dunlap et al., 2013; He et al., 2013). D-Galacturonate is a major component of pectin, the most abundant polysaccharide in the plant cell wall (Alazi et al., 2016; Willats et al., 2001), suggesting that the D-galacturonate metabolic ability may be one of the most important properties enabling the dominance of *B. velezensis* in fermented soybean foods such as doenjang, miso, natto, and meju (Jang et al., 2011; Jeon et al., 2016).

3.7. Genomic and functional analysis of xanthine metabolic pathways and xanthine oxidase assay

Xanthine is a purine base found in animals and plants and is subsequently converted to uric acid by xanthine oxidase (Hadwan et al., 2014). All *B. amyloliquefaciens, B. siamensis*, and *B. velezensis* strains harbored a xanthine oxidase (called xanthine dehydrogenase under anaerobic conditions) gene cluster (*xoABCDE*) to convert xanthine to uric acid, indicating that xanthine oxidation was a common feature of the *B. amyloliquefaciens* group. In addition to the xanthine oxidase gene cluster, a complete xanthine and uric acid metabolic gene cluster metabolizing xanthine and uric acid to glycine and oxalureate was identified in the genomes of all *B. siamensis* strains (Fig. 6A). However, no xanthine metabolic gene cluster with the same physical map was identified in *B. amyloliquefaciens* and *B. velezensis*, suggesting that the xanthine metabolism pathway shown in Fig. 6B is a species-specific metabolic feature of *B. siamensis* in the *B. amyloliquefaciens* group.

The genomic analysis of the *B. amyloliquefaciens* group indicated that only *B. siamensis* strains harbored two copies of xanthine oxidase genes (*xoABCDE*), suggesting that *B. siamensis* strains may have a higher xanthine oxidation activity than *B. amyloliquefaciens* and *B. velezensis*, which was confirmed by the xanthine oxidation activity test. The results showed that, as expected, *B. siamensis* KCTC 13613^T oxidized xanthine to uric acid faster than *B. amyloliquefaciens* DSM 7^T and *B. velezensis*

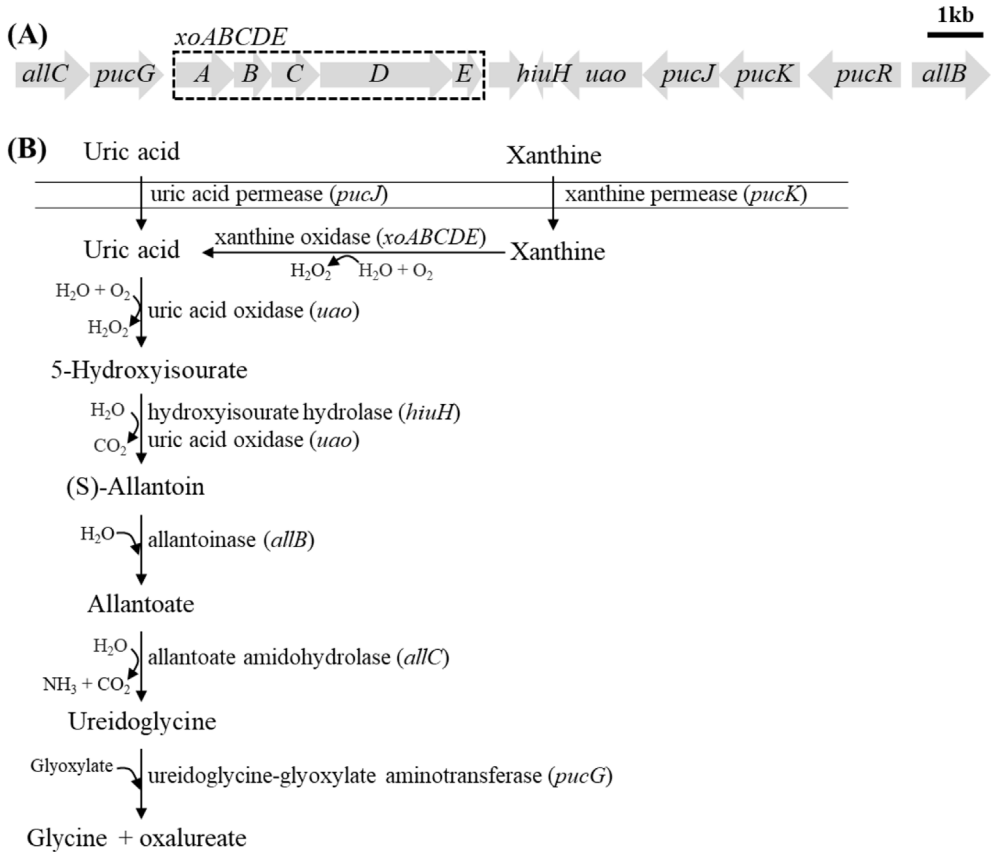


Fig. 6. A physical map of putative genes responsible for xanthine metabolism (A) and a proposed xanthine metabolic pathway in *Bacillus siamensis* (B). Genes encoding xanthine oxidase (*xoABCDE*) are indicated by the dotted box. *pucR*, transcriptional regulator.

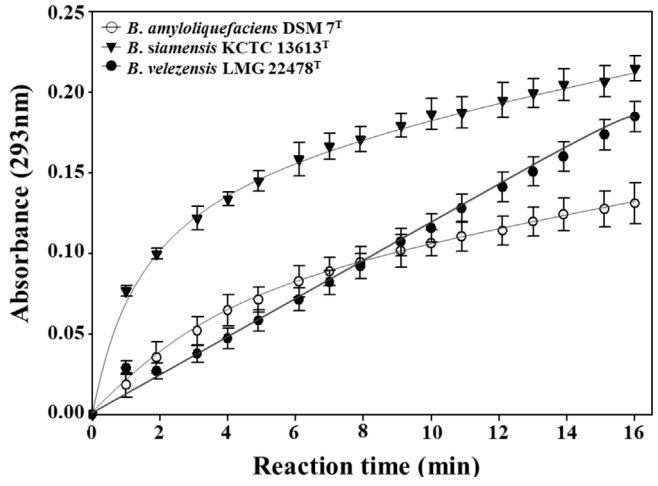


Fig. 7. Xanthine oxidation assay for *Bacillus amyloliquefaciens* DSM 7^T, *B. siamensis* KCTC 13613^T, and *B. velezensis* LMG 22478^T. Xanthine oxidation activity was assessed by measuring uric acid production from xanthine (absorbance at 293 nm).

LMG 22478^T (Fig. 7).

The xanthine oxidase genes present in the xanthine metabolic gene cluster of only *B. siamensis* had very low amino acid identities (< 50%) with the xanthine oxidase genes commonly identified in the *B. amyloliquefaciens* group, indicating that *B. siamensis* might acquire the xanthine (uric acid) metabolic gene cluster via lateral gene transfer from a different group. The BlastN analysis showed that the xanthine (uric acid) metabolic genes in *B. siamensis* shared high sequence similarities with those in *B. subtilis*, (> 90%), suggesting that a lateral gene transfer of the genes might have happened between *B. siamensis* and *B. subtilis*. The growth test performed in minimal media supplemented with xanthine as a sole carbon source showed that no growth occurred in any of the type strains of *B. amyloliquefaciens, B. siamensis*, and *B. velezensis* (data not shown), suggesting that *B. amyloliquefaciens* group members are unable to metabolize xanthine as a carbon source for their growth.

3.8. Phenotypic properties and metabolic pathway reconstruction of *B. amyloliquefaciens, B. siamensis, and B. velezensis*

The metabolic capabilities of *B. amyloliquefaciens, B. siamensis*, and *B. velezensis* for various carbon compounds were investigated using their type strains (strains DSM 7^T, KCTC 13613^T, and LMG 22478^T, respectively) in the API 50 CHB system (supplementary Table S3). Results showed that strains DSM 7^T, KCTC 13613^T, and LMG 22478^T had similar metabolic capabilities to metabolize very diverse carbon compounds including D-glucose, D-fructose, D-lactose, sucrose, D-maltose, D-cellobiose, D-ribose, D-xylose, L-arabinose, mannose, inositol, mannitol, sorbitol, α-methyl- D-glucoside, amygdaline, arbutine, salicin, and glycerol, indicating that *B. amyloliquefaciens* group members were metabolically versatile. The only difference was that *B. velezensis* LMG 22478^T could metabolize trehalose, whereas *B. amyloliquefaciens* DSM 7^T and *B. siamensis* KCTC 13613^T could not.

B. subtilis and some *B. amyloliquefaciens* members like *B. subtilis* have been known to be strictly aerobic (Niu et al., 2018; Priest et al., 1987), and this study also showed that *B. subtilis* NCIB 3610^T used as a control did not grow on TSA under anaerobic conditions, as previously described (Clements et al., 2002). However, many other *Bacillus* species can grow anaerobically, especially in the presence of external electron acceptors (Clements et al., 2002; Nakano and Zuber, 1998). Many *Bacillus* strains including *B. amyloliquefaciens* group members have been

dominantly identified in fermented foods such as fermented soybean foods processed under anaerobic conditions (Yang et al., 2017; Yao et al., 2016; Zhang et al., 2016a). The anaerobic growth test in this study showed that all type strains of *B. amyloliquefaciens, B. siamensis*, and *B. velezensis* displayed anaerobic growth on TSA even without external electron acceptors, although their rates of growth were very slow compared to their aerobic growth rates. In previous studies, it has been reported that the *B. siamensis* type strain and a *B. velezensis* strain could also grow under anaerobic conditions, indicating that it is a facultative anaerobe (Jeon et al., 2016; Sumpavapol et al., 2010). In addition, *B. licheniformis*, which is phylogenetically and physiologically closely related to the *B. amyloliquefaciens* group, is also a facultative anaerobe capable of fermenting glucose under anaerobic conditions (Clements et al., 2002). These results suggest that the members of the *B. amyloliquefaciens* group have the ability to metabolize carbon compounds under anaerobic conditions, and are responsible for the fermentation of traditional soybean foods.

To further investigate the metabolic features of the *B. amyloliquefaciens, B. siamensis*, and *B. velezensis* strains, their metabolic pathways were reconstructed based on a KEGG pathway analysis in combination with a BlastP analysis of their genomes. These bioinformatically constructed metabolic pathways suggested that the *B. amyloliquefaciens, B. siamensis*, and *B. velezensis* strains commonly harbor genes in their core-genomes that encode enzymes capable of metabolizing diverse array of carbon sources including, D-glucose, D-fructose, D-lactose, sucrose, maltose, cellobiose, D-ribose, D-xylose, L-arabinose, mannitol, arbutine, melibiose, raffinose, salicin, sorbitol, and stachyose (Fig. 8), consistent with the phenotypic test results using the API 50 CHB system (Table S3). Two *B. amyloliquefaciens* group strains were reported to have an ability to metabolize methanol (Madhaiyan et al., 2010). The reconstructed metabolic pathway showed that all *B. amyloliquefaciens* group strains harbored all genes associated with methanol metabolism, indicating that methanol metabolism may be a common metabolic feature of the *B. amyloliquefaciens* group. *B. amyloliquefaciens* and *B. velezensis* harbored a complete trehalose metabolic pathway including trehalose phosphotransferase system (PTS) in their core-genomes, but the phenotypic tests showed that only *B. velezensis* LMG 22478^T metabolized trehalose, while *B. amyloliquefaciens* DSM 7^T did not display trehalose metabolic activity (Table S3). Additional analysis of the trehalose metabolic genes revealed that all *B. amyloliquefaciens* strains except for strain DSM 7 ^{T} harbored a gene encoding normal trehalose PTS, while the gene encoding trehalose PTS in *B. amyloliquefaciens* DSM 7 ^{T} was incomplete as a pseudogene, which might explain why strain DSM 7 ^{T} was negative in the trehalose metabolic activity test. A mannose PTS responsible for mannose transport was identified in all *B. amyloliquefaciens* group strains, except the *B. siamensis* strains, and the type strain of *B. siamensis* showed a weak mannose metabolic activity in the phenotypic tests, confirming the absence of the PTS gene in *B. siamensis* strains. The reconstructed metabolic pathways show that the pathways responsible for D-galacturonate and D-fructuronate metabolism are present only in the core-genome of *B. velezensis*, in accordance with the above bioinformatics analysis and experimental results (Figs. 4 and 5).

The reconstructed metabolic pathways show that all of the *B.*

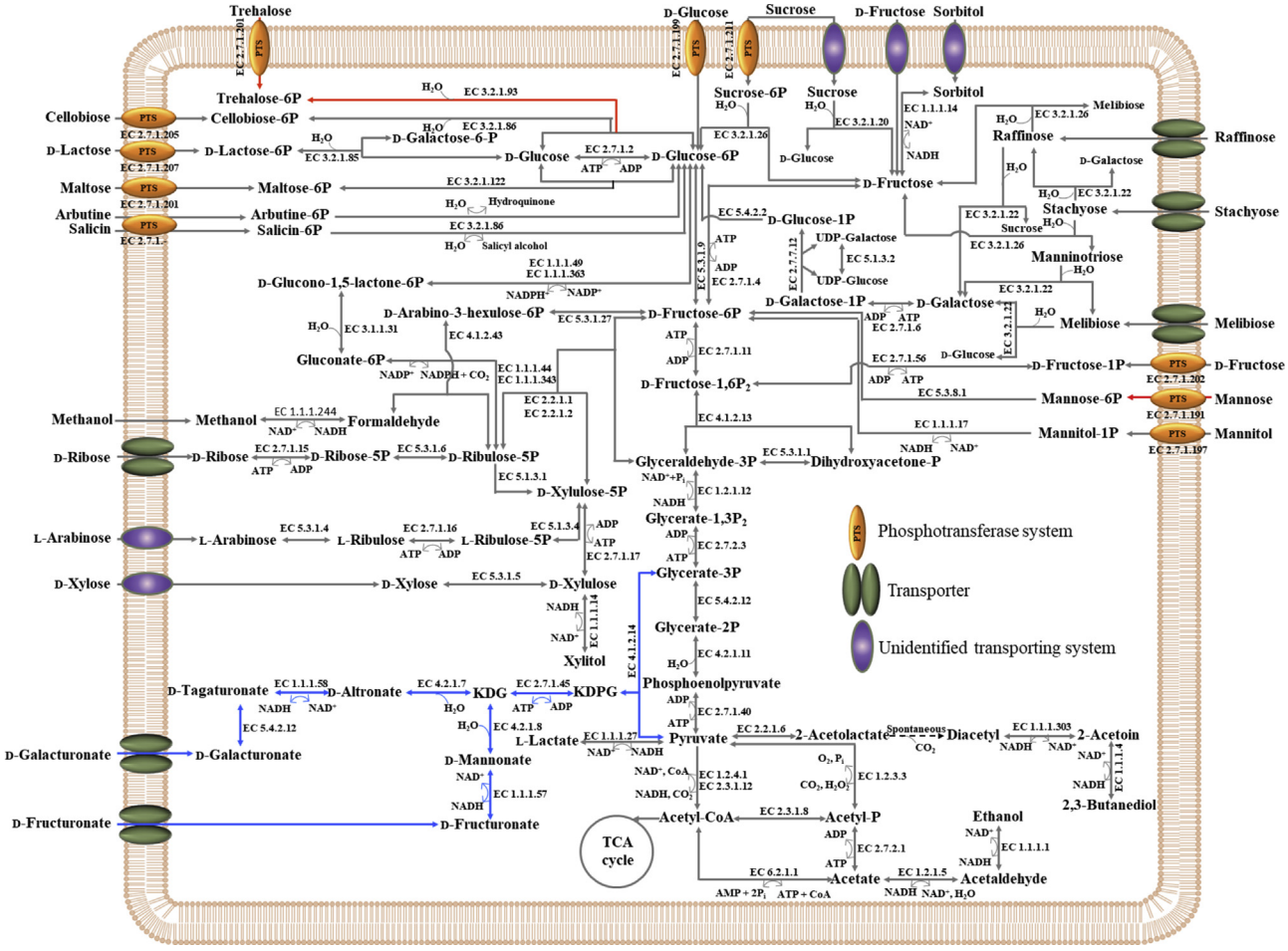


Fig. 8. Proposed metabolic pathways of the *Bacillus amyloliquefaciens* group for carbon compounds. Metabolic pathways that are present in all strains of the *B. amyloliquefaciens* group are depicted in black, and metabolic pathways that are only present in the *B. velezensis* strains or the *B. velezensis* and *B. amyloliquefaciens* strains are depicted in blue and red, respectively. KDG, 2-keto-3-deoxygluconate; KDPG, 2-keto-3-deoxy-6-phosphogluconate.

amyloliquefaciens group strains have glycolysis, the pentose phosphate pathway, and the tricarboxylic acid cycle to metabolize carbon compounds, indicating that they can grow aerobically with metabolic versatility. However, interestingly, the reconstructed metabolic pathways show that they also harbor pathways that allow them to metabolize pyruvate to L-lactate, acetate, and ethanol, which are representative fermentation products, indicating that they are able to grow anaerobically, confirming the result of the anaerobic growth test. Genes encoding acetolactate synthase (Enzyme Commission (EC) 2.2.1.6) and diacetyl reductase (EC 1.1.1.303) that convert pyruvate to diacetyl and diacetyl to acetoin were found in all *B. amyloliquefaciens* group members, suggesting that diacetyl and acetoin, representative cheese flavors, may be important flavoring compounds in fermented soybean foods (Passerini et al., 2013). In addition, 2,3-butanediol dehydrogenase (EC 1.1.1.4), converting 2-acetoin to 2,3-butanediol with the regeneration of NAD^+, was also found in all *B. amyloliquefaciens* group strains, indicating that cheese flavoring can be weakened by this enzyme during fermentation. The reconstructed metabolic pathways also show that pyruvate can be converted into acetate via acetyl-P without the production of NADH by pyruvate oxidase (EC 1.2.3.3) and acetate kinase (EC 2.7.2.1). The reconstructed metabolic pathways also show that all strains of the *B. amyloliquefaciens* group harbor genes to produce xylitol and sorbitol, sugar alcohols, through reduction of their corresponding sugars. The pathways also show that all *B. amyloliquefaciens* group strains have the ability to metabolize sorbitol, but not xylitol, because of the absence of a transport system for xylitol, in accordance with the results of the phenotypic tests shown in Table S3.

4. Conclusions

In this study, the phylogenetic and genomic diversities and metabolic features of the *B. amyloliquefaciens* group (*B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis*) frequently identified in fermented soybean foods were investigated through phylogenomic and pan- and core-genome analyses. The functional and metabolic features of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* strains were compared through COG, KEGG, and BLAST analyses, and some metabolic features were experimentally verified. This study provides useful information about the phylogenetic properties, antimicrobial activities, and metabolic and fermentative features of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* strains, which are the basis for understanding the fermentation of soybean foods. Further analyses of the genomic, transcriptomic, and metabolic features of *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* may provide insights into their fermentative features and roles during fermentation of soybean foods and shed light on starter development for the production of fermented soybean foods with safety and high quality.

Acknowledgements

This work was supported by the National Research Foundation (2017M3C1B5019250) of the Ministry of Science and ICT, Republic of Korea.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.fm.2018.09.001.

References

Alazi, E., Niu, J., Kowalczyk, J.E., Peng, M., Aguilar Pontes, M.V., Kan, J.A., Visser, J., Vries, R.P., Ram, A.F., 2016. The transcriptional activator GaaR of *Aspergillus Niger* is required for release and utilization of d-galacturonic acid from pectin. FEBS (Fed. Eur. Biochem. Soc.) Lett. 590, 1804–1815.

Alvindia, D.G., Natsuaki, K.T., 2009. Biocontrol activities of *Bacillus amyloliquefaciens* DGA14 isolated from banana fruit surface against banana crown rot-causing pathogens. Crop Protect. 28, 236–242.

Bezuidt, O.K., Pierneef, R., Gomri, A.M., Adesioye, F., Makhalanyane, T.P., Kharroub, K., Cowan, D.A., 2016. The *geobacillus* Pan-genome: implications for the evolution of the genus. Front. Microbiol. 7, 723.

Borriss, R., Chen, X.H., Rueckert, C., Blom, J., Becker, A., Baumgarth, B., Fan, B., Pukall, R., Schumann, P., Spröer, C., 2011. Relationship of *Bacillus amyloliquefaciens* clades associated with strains DSM 7^T and FZB42^T: a proposal for *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* subsp. nov. and *Bacillus amyloliquefaciens* subsp. *plantarum* subsp. nov. based on complete genome sequence comparisons. Int. J. Syst. Evol. Microbiol. 61, 1786–1801.

Chaudhari, N.M., Gupta, V.K., Dutta, C., 2016. BPGA-an ultra-fast pan-genome analysis pipeline. Sci. Rep. 6, 24373.

Chen, X.H., Scholz, R., Borriss, M., Junge, H., Mögel, G., Kunz, S., Borriss, R., 2009a. Difficidin and bacilysin produced by plant-associated *Bacillus amyloliquefaciens* are efficient in controlling fire blight disease. J. Biotechnol. 140, 38–44.

Chen, X.H., Koumoutsi, A., Scholz, R., Schneider, K., Vater, J., Süssmuth, R., Piel, J., Borriss, R., 2009b. Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. J. Biotechnol. 140, 27–37.

Chen, X.H., Koumoutsi, A., Scholz, R., Eisenreich, A., Schneider, K., Heinemeyer, I., Morgenstern, B., Voss, B., Hess, W.R., Reva, O., 2007. Comparative analysis of the complete genome sequence of the plant growth–promoting bacterium *Bacillus amyloliquefaciens* FZB42. Nat. Biotechnol. 25, 1007–1014.

Chen, Y.-G., Gu, F.-l., Li, J.-h., Xu, F., He, S.-z., Fang, Y.-m., 2015. *Bacillus vanillea* sp. nov., isolated from the cured vanilla bean. Curr. Microbiol. 70, 235–239.

Chun, B.H., Kim, K.H., Jeon, H.H., Lee, S.H., Jeon, C.O., 2017. Pan-genomic and transcriptomic analyses of *Leuconostoc mesenteroides* provide insights into its genomic and metabolic features and roles in kimchi fermentation. Sci. Rep. 7, 11504.

Chung, E.J., Hossain, M.T., Khan, A., Kim, K.H., Jeon, C.O., Chung, Y.R., 2015. *Bacillus oryzicola* sp. nov., an endophytic bacterium isolated from the roots of rice with antimicrobial, plant growth promoting, and systemic resistance inducing activities in rice. Plant Pathol. J. 31, 152.

Clements, L.D., Miller, B.S., Streips, U.N., 2002. Comparative growth analysis of the facultative anaerobes *Bacillus subtilis, Bacillus licheniformis*, and *Escherichia coli*. Syst. Appl. Microbiol. 25, 284–286.

Compaoré, C., Nielsen, D.S., Sawadogo-Lingani, H., Berner, T.S., Nielsen, K.F., Adimpong, D.B., Diawara, B., Ouédraogo, G., Jakobsen, M., Thorsen, L., 2013. *Bacillus amyloliquefaciens* ssp. *plantarum* strains as potential protective starter cultures for the production of Bikalga, an alkaline fermented food. Journal of applied microbiology J. Appl. Microbiol. 115, 133–146.

Deng, X., Phillippy, A.M., Li, Z., Salzberg, S.L., Zhang, W., 2010. Probing the pan-genome of *Listeria monocytogenes*: new insights into intraspecific niche expansion and genomic diversification. BMC Genom. 11, 500.

Douillard, F.P., Ribbera, A., Kant, R., Pietilä, T.E., Järvinen, H.M., Messing, M., Randazzo, C.L., Paulin, L., Laine, P., Ritari, J., 2013. Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. PLoS Genet. 9, e1003683.

Dunlap, C.A., 2015. Phylogenomic analysis shows that '*Bacillus vanillea*'is a later heterotypic synonym of *Bacillus siamensis*. Int. J. Syst. Evol. Microbiol. 65, 3507–3510.

Dunlap, C.A., Bowman, M.J., Schisler, D.A., 2013. Genomic analysis and secondary metabolite production in *Bacillus amyloliquefaciens* AS 43.3: a biocontrol antagonist of Fusarium head blight. Biol. Contr. 64, 166–175.

Dunlap, C.A., Kim, S.-J., Kwon, S.-W., Rooney, A.P., 2016. *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens; Bacillus methylotrophicus, Bacillus amyloliquefaciens* subsp. *plantarum* and '*Bacillus oryzicola*'are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. Int. J. Syst. Evol. Microbiol. 66, 1212–1217.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.

Eom, J.S., Choi, H.S., 2016. Inhibition of *Bacillus cereus* growth and toxin production by *Bacillus amyloliquefaciens* RD7-7 in fermented soybean products. J. Microbiol. Biotechnol. 26, 44–55.

Fan, B., Blom, J., Klenk, H.-P., Borriss, R., 2017. *Bacillus amyloliquefaciens, Bacillus velezensis*, and *Bacillus siamensis* form an "operational group *B. Amyloliquefaciens*" within the *B. Subtilis* species complex. Front. Microbiol. 8, 3843.

Fukumoto, J., 1943. Studies on the production of bacterial amylase. I. Isolation of bacteria secreting potent amylases and their distribution. Nippon. Nogeikagaku Kaishi 19, 487–503.

Gong, W., Wang, J., Chen, Z., Xia, B., Lu, G., 2011. Solution structure of LCI, a novel antimicrobial peptide from *Bacillus subtilis*. Biochem 50, 3621–3627.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M., 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81–91.

Hadwan, M.H., Almashhedy, L.A., Alsalman, A.R.S., 2014. Seminal xanthine oxidase: appropriate fluorometric assay for the examination of spermatozoa disorders. BioTechnol: Indian J. For. 9, 267–272.

He, P., Hao, K., Blom, J., Rückert, C., Vater, J., Mao, Z., Wu, Y., Hou, M., He, P., He, Y., 2013. Genome sequence of the plant growth promoting strain *Bacillus amyloliquefaciens* subsp. *plantarum* B9601-Y2 and expression of mersacidin and other secondary metabolites. J. Biotechnol. 164, 281–291.

Jang, S.-J., Kim, Y.-J., Park, J.-M., Park, Y.-S., 2011. Analysis of microflora in gochujang, Korean traditional fermented food. Food Sci. Biotechnol. 20, 1435–1440.

Jeon, H.H., Jung, J.Y., Chun, B.-H., Kim, M.-D., Baek, S.Y., Moon, J.Y., Yeo, S.-H., Jeon, C.O., 2016. Screening and characterization of potential *Bacillus* starter cultures for fermenting low salt soybean paste (doenjang). J. Microbiol. Biotechnol. 26, 666–674.

Jeong, H., Jeong, D.-E., Kim, S.H., Song, G.C., Park, S.-Y., Ryu, C.-M., Park, S.-H., Choi, S.-K., 2012. Draft Genome sequence of the plant growth-promoting Bacterium *Bacillus siamensis* KCTC 13613^{T}. J. Bacteriol. 194, 4148–4149.
Kanehisa, M., Sato, Y., Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428, 726–731.
Khochamit, N., Siripornadulsil, S., Sukon, P., Siripornadulsil, W., 2015. Antibacterial activity and genotypic–phenotypic characteristics of bacteriocin-producing *Bacillus subtilis* KKU213: potential as a probiotic strain. Microbiol. Res. 170, 36–50.
Kim, Y., Koh, I., Lim, M.Y., Chung, W.-H., Rho, M., 2017. Pan-genome analysis of *Bacillus* for microbiome profiling. Sci. Rep. 7, 10984.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Lee, H.J., Chun, B.H., Jeon, H.H., Kim, Y.B., Lee, S.H., 2017. Complete genome sequence of *Bacillus velezensis* YJ11-1-4, a strain with broad-spectrum antimicrobial activity, isolated from traditional Korean fermented soybean paste. Genome Announc. 5, e01352–e01417.
Lee, I., Kim, Y.O., Park, S.-C., Chun, J., 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int. J. Syst. Evol. Microbiol. 66, 1100–1103.
Madhaiyan, M., Poonguzhali, S., Kwon, S.-W., Sa, T.-M., 2010. *Bacillus methylotrophicus* sp. nov., a methanol-utilizing, plant-growth-promoting bacterium isolated from rice rhizosphere soil. Int. J. Syst. Evol. Microbiol. 60, 2490–2495.
Masuoka, N., Nihei, K.-i., Maeta, A., Yamagiwa, Y., Kubo, I., 2015. Inhibitory effects of cardols and related compounds on superoxide anion generation by xanthine oxidase. Food Chem. 166, 270–274.
Mathews, S.L., Pawlak, J.J., Grunden, A.M., 2014. Isolation of *Paenibacillus glucanolyticus* from pulp mill sources with potential to deconstruct pulping waste. Bioresour. Technol. 164, 100–105.
Meidong, R., Doolgindachbaporn, S., Jamjan, W., Sakai, K., Tashiro, Y., Okugawa, Y., Tongpim, S., 2017. A novel probiotic *Bacillus siamensis* B44v isolated from Thai pickled vegetables (Phak-dong) for potential use as a feed supplement in aquaculture. J. Gen. Appl. Microbiol. 63, 246–253.
Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M., 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinf. 14, 60.
Nakano, M.M., Zuber, P., 1998. Anaerobic growth of a "strict aerobe"(*Bacillus subtilis*). Annu. Rev. Microbiol. 52, 165–190.
Nawrocki, E.P., Eddy, S.R., 2007. Query-dependent banding (QDB) for faster RNA similarity searches. PLoS Comput. Biol. 3, e56.
Niu, L., Xiong, M., Zhang, J., Xiang, Y., Song, L., Hua, Z., Li, W., 2018. *Bacillus camelliae* sp. nov., isolated from Pu'er tea. Int. J. Syst. Evol. Microbiol. 68, 564–569.
Palazzini, J.M., Dunlap, C.A., Bowman, M.J., Chulze, S.N., 2016. *Bacillus velezensis* RC 218 as a biocontrol agent to reduce Fusarium head blight and deoxynivalenol accumulation: genome sequencing and secondary metabolite cluster profiles. Microbiol. Res. 192, 30–36.
Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055.
Passerini, D., Laroute, V., Coddeville, M., Le Bourgeois, P., Loubière, P., Ritzenthaler, P., Cocaign-Bousquet, M., Daveran-Mingot, M.-L., 2013. New insights into *Lactococcus lactis* diacetyl-and acetoin-producing strains isolated from diverse origins. Int. J. Food Microbiol. 160, 329–336.
Peng, Y., Huang, Q., Zhang, R.-h., Zhang, Y.-z., 2003. Purification and characterization of a fibrinolytic enzyme produced by *Bacillus amyloliquefaciens* DC-4 screened from douchi, a traditional Chinese soybean food. Comp. Biochem. Physiol. 134, 45–52.
Periwal, V., Patowary, A., Vellarikkal, S.K., Gupta, A., Singh, M., Mittal, A., Jeyapaul, S., Chauhan, R.K., Singh, A.V., Singh, P.K., 2015. Comparative whole-genome analysis of clinical isolates reveals characteristic architecture of *Mycobacterium tuberculosis* pangenome. PLoS One 10, e0122979.
Perry, D., Abraham, E., 1979. Transport and metabolism of bacilysin and other peptides by suspensions of *Staphylococcus aureus*. Microbiology 115, 213–221.
Priest, F.G., Goodfellow, M., Shute, L.A., Berkeley, R.C.W., 1987. Bacillus amyloliquefaciens sp. nov. nom. rev. Int. J. Syst. Bacteriol. 37, 69¬–71.
Rückert, C., Blom, J., Chen, X., Reva, O., Borriss, R., 2011. Genome sequence of *B. amyloliquefaciens* type strain DSM7^{T} reveals differences to plant-associated *B. amyloliquefaciens* FZB42. J. Biotechnol. 155, 78–85.
Richter, M., Rosselló-Móra, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. Unit. States Am. 106, 19126–19131.
Rosselló-Móra, R., Amann, R., 2015. Past and future species definitions for bacteria and archaea. Syst. Appl. Microbiol. 38, 209–216.
Ruiz-Garcia, C., Bejar, V., Martinez-Checa, F., Llamas, I., Quesada, E., 2005. *Bacillus velezensis* sp. nov., a surfactant-producing bacterium isolated from the river Vélez in Málaga, southern Spain. Int. J. Syst. Evol. Microbiol. 55, 191–195.
Schneider, K., Chen, X.-H., Vater, J., Franke, P., Nicholson, G., Borriss, R., Süssmuth, R.D., 2007. Macrolactin is the polyketide biosynthesis product of the pks2 cluster of *Bacillus amyloliquefaciens* FZB42. J. Nat. Prod. 70, 1417–1423.
Scholz, R., Molohon, K.J., Nachtigall, J., Vater, J., Markley, A.L., Süssmuth, R.D., Mitchell, D.A., Borriss, R., 2011. Plantazolicin, a novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42. J. Bacteriol. 193, 215–224.
Scholz, R., Vater, J., Budiharjo, A., Wang, Z., He, Y., Dietel, K., Schwecke, T., Herfort, S., Lasch, P., Borriss, R., 2014. Amylocyclicin, a novel circular bacteriocin produced by *Bacillus amyloliquefaciens* FZB42. J. Bacteriol. 196, 1842–1852.
Shin, J., Song, Y., Jeong, Y., Cho, B.-K., 2016. Analysis of the core genome and pan-genome of autotrophic acetogenic bacteria. Front. Microbiol. 7, 1531.
Sumpavapol, P., Tongyonk, L., Tanasupawat, S., Chokesajjawatee, N., Luxananil, P., Visessanguan, W., 2010. *Bacillus siamensis* sp. nov., isolated from salted crab (pookhem) in Thailand. Int. J. Syst. Evol. Microbiol. 60, 2364–2370.
Tettelin, H., Riley, D., Cattuto, C., Medini, D., 2008. Comparative genomics: the bacterial pan-genome. Curr. Opin. Microbiol. 11, 472–477.
Willats, W.G., McCartney, L., Mackie, W., Knox, J.P., 2001. Pectin: Cell Biology and Prospects for Functional Analysis, Plant Cell Walls. Springer, pp. 9–27.
Xu, Z., Shao, J., Li, B., Yan, X., Shen, Q., Zhang, R., 2013. Contribution of bacillomycin D in *Bacillus amyloliquefaciens* SQR9 to antifungal activity and biofilm formation. Appl. Environ. Microbiol. 79, 808–815.
Yamada, T., Letunic, I., Okuda, S., Kanehisa, M., Bork, P., 2011. iPath2. 0: interactive pathway explorer. Nucleic Aacids Res 39, W412–W415.
Yang, Y., Deng, Y., Jin, Y., Liu, Y., Xia, B., Sun, Q., 2017. Dynamics of microbial community during the extremely long-term fermentation process of a traditional soy sauce. J. Sci. Food Agric. 97, 3220–3227.
Yao, Z., Liu, X., Shim, J.M., Lee, K.W., Kim, H.-J., Kim, J.H., 2016. Properties of a fibrinolytic enzyme secreted by *Bacillus amyloliquefaciens* RSB34, isolated from doenjang. J. Microbiol. Biotechnol. 27, 9–18.
Yi, H., Chun, J., Cha, C.-J., 2014. Genomic insights into the taxonomic status of the three subspecies of *Bacillus subtilis*. Syst. Appl. Microbiol. 37, 95–99.
Zhang, J., Du, G., Chen, J., Fang, F., 2016a. Characterization of a *Bacillus amyloliquefaciens* strain for reduction of citrulline accumulation during soy sauce fermentation. Biotechnol. Lett. 38, 1723–1731.
Zhang, L., Huang, J., Zhou, R., Wu, C., 2017. Evaluating the feasibility of fermentation starter inoculated with *Bacillus amyloliquefaciens* for improving acetoin and tetramethylpyrazine in Baoning bran vinegar. Int. J. Food Microbiol. 96, 1137–1155.
Zhang, N., Yang, D., Kendall, J.R., Borriss, R., Druzhinina, I.S., Kubicek, C.P., Shen, Q., Zhang, R., 2016b. Comparative genomic analysis of Bacillus amyloliquefaciens and Bacillus subtilis reveals evolutional traits for adaptation to plant-associated habitats. Front. Microbiol. 7, 2039.