BIODIVERSITAS
Volume 23, Number 3, March 2022
Pages: 1293-1300

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d230313

Whole-genome analysis of *Bacillus subtilis* G8 isolated from natto

DIKSON, HANS VICTOR, DELVIN JONG, ASTIA SANJAYA, ARIELA SAMANTHA, JUANDY JO, REINHARD PINONTOAN♥

Department of Biology, Faculty of Science and Technology, Universitas Pelita Harapan. Jl. M.H. Thamrin Boulevard 1100, Tangerang, Banten 15811, Indonesia. Tel.: +62-21-5460901, ♥email: reinhard.pinontoan@uph.edu

Manuscript received: 26 January 2021. Revision accepted: 17 February 2022

Abstract. *Dikson, Victor H, Jong D, Sanjaya A, Samantha A, Jo J, Pinontoan R. 2022. Whole-genome analysis of* Bacillus subtilis *G8 isolated from natto. Biodiversitas 23: 1293-1300. Bacillus subtilis*-fermented soy-based food is associated with multiple health benefits. Various bacterial strains have been isolated from it, including *B. subtilis* G8, recent isolation from Japanese natto commercially available in Indonesia. Both 16S rRNA gene sequencing and fibrinolytic activity characterization have been performed and published in prior studies. After comparison to the genome of a natto-isolated reference strain (i.e., *B. subtilis* BEST195), the *B. subtilis* G8 genome showed a similar guanine-cytosine (GC) content, predicted number of coding sequences (CDS) and predicted number of tRNA genes, but had a shorter sequence length and fewer predicted rRNA genes. Further analysis using multiple genome alignment with Mauve, average nucleotide identity (ANI) matrix calculation, and phylogenetic inference indicated that *B. subtilis* G8 was more related to natto-derived *B. subtilis* than to cheonggukjang-derived *B. subtilis* and *B. subtilis* 168. Finally, sequence analyses of a gene encoding nattokinase as well as two genes regulating poly-gamma-glutamic acid (γ-PGA) production in *B. subtilis* G8, *B. subtilis* BEST195 and *B. subtilis* 168 clearly indicated that *B. subtilis* G8 is able to produce nattokinase and γ-PGA, which both contribute to natto's fermentation process. Therefore, it is proposed that *B. subtilis* G8 should be reclassified as *B. subtilis* subsp. *natto* G8 to reflect that it is a natto-derived *B. subtilis* strain.

Keywords: *Bacillus subtilis* G8, natto, nattokinase, poly-gamma-glutamic acid, whole-genome analysis

Abbreviations: ANI: Average nucleotide identity, bp: basepair, CDS: coding sequence, GC: guanine-cytosine, iTOL: Interactive Tree of Life, LCB: Locally Collinear Block, TYGS: Type Strain Genome Server, γ-PGA: Poly-gamma-glutamic acid

INTRODUCTION

Traditional fermented soy-based food is widely consumed in Asia due to its nutritional value and associated health benefits, which range from the improving the health of gut microbiota to the prevention of cardiovascular diseases (Elshaghabee et al. 2017; Nagata et al. 2017; Pangastuti et al. 2019; Kojima et al. 2020). Fermentation of soy-based food is carried out by either fungi or bacteria (Zhao et al. 2013). *Bacillus subtilis* is one such fermentation agent that is commonly found in various Asian soy-based fermented foods, such as Japanese natto, Korean cheonggukjang, Nepalese kinema, and Burmese pepoke (Kamada et al. 2015). This particular species is also generally recognized as safe, which reinforces its use in various applications in food production (Su et al. 2020).

It is interesting to note that a variety of strains of *B. subtilis* have been isolated from soy-based fermented food, as reported worldwide (Nishito et al. 2010; Kamada et al. 2015; Bang et al. 2018; Heo et al. 2019). Notably, Lucy et al. (2019) isolated *B. subtilis* G8 from a commercially available Japanese natto in Indonesia. Several studies were performed to characterize this strain of *B. subtilis*. First, the fibrinolytic capability of *B. subtilis* G8 was investigated using various *in vitro* fibrinolytic assays. Crude *B. subtilis* G8 extract possessed potent fibrinolytic activity, which was likely to be mediated by various proteases (Pinontoan et al. 2021). Second, the 16S rRNA gene sequencing analysis indicated that *B. subtilis* G8 shared approximately 99% identity with the *B. subtilis* subspecies natto BEST195 reported by Nishito et al. (2010), a fermentation agent for commercialized natto, and *B. subtilis* subspecies *subtilis* 168, used as a laboratory standard. Phylogenetic analysis showed that *B. subtilis* G8 clustered closer to *B. subtilis* BEST195 than to *B. subtilis* subsp. *subtilis* 168 (Lucy et al. 2019).

To better resolve the identity of *B. subtilis* G8, whole-genome analysis is needed. Additionally, whole-genome sequencing would provide useful information regarding other genes that may contribute to natto's previously mentioned health benefits. *B. subtilis* G8 appears to be closely related to other natto-derived *B. subtilis* strains that produce nattokinase and poly-gamma-glutamic acid (γ-PGA), both of which are important in natto fermentation (Kada et al. 2013). Therefore, whole-genome sequencing and analysis of *B. subtilis* G8 were carried out in the present study.

MATERIALS AND METHODS

Extraction of genomic DNA from *B. subtilis* G8

Bacillus subtilis G8 was isolated from Japanese fermented soybean natto sold commercially in Indonesia

(Lucy et al. 2019). *B. subtilis* G8 was inoculated in nutrient broth and incubated at 37°C with agitation for 24 h to produce sufficient biomass for extraction of genomic DNA. After incubation, bacterial cells were harvested from the media by centrifugation at 5,000 × g for 5 min. Genomic DNA was extracted from the harvested cells using the Wizard Genomic DNA Purification Kit (Promega, USA). The extracted genomic DNA was qualified and quantified using BioDrop Duo spectrophotometer which yielded absorbance ratios of 2.30 (A260/A280), 2.14 (A260/A230), and concentration of 446 μg/mL respectively. Moreover, agarose gel electrophoresis yielded one intact band containing the genomic DNA. These indicated that the extracted genomic DNA had a high purity with no fragmentation, which warranted whole-genome sequencing. The genomic DNA was sent for whole-genome sequencing to NovoGene (Hong Kong). Sequencing was performed on the Illumina NovaSeq 6000 using 150-basepair (bp) paired-end reads.

Bacillus subtilis G8 genome assembly

Paired-end *B. subtilis* G8 whole-genome raw reads were first trimmed of any low-quality reads using Sickle (Joshi and Fass 2011) with minimum read length and PHRED score set to 150 bp and 20, respectively. The trimmed reads were then assembled using IDBA-Hybrid guided assembly software (Peng et al. 2012) with *B. subtilis* subsp. natto BEST195 as the reference sequence (accession number AP011541). The kMaxShortSequence variable in the source code was modified to match the 150 bp read length. The assembled *B. subtilis* G8 draft genome was then filtered for contigs flagged as plasmid sequences using PLSDB (Galata et al. 2019) and Nucleotide BLAST (Altschul et al. 1990), as well as contaminant or low-coverage (≤50x) sequences using Bowtie2 (Langmead and Salzberg 2012), SAMtools (Li et al. 2009), and Qualimap2 (García-Alcalde et al. 2012; Okonechnikov et al. 2015). Once cleared of known contaminant and plasmid sequences, the contigs were reordered according to the reference *B. subtilis* subsp. natto BEST195 using Mauve Contig Mover (Rissman et al. 2009) and then further examined using Artemis (Carver et al. 2008). The genome was annotated and then submitted to the DNA Data Bank of Japan using DFAST's genome annotation and submission service (Tanizawa et al. 2016) with the accession number AP025224.

Genomic comparison with natto-derived and cheonggukjang-derived *B. subtilis* strains

The assembled *B. subtilis* G8 genome was compared with the genomes of 14 *B. subtilis* strains isolated from natto and cheonggukjang and the genome of the laboratory standard *B. subtilis* 168, all of which were obtained from various published reports (Table 1). Genome sequences of *B. subtilis* G8, *B. subtilis* BEST195, *B. subtilis* 168, and two representatives of cheonggukjang-derived *B. subtilis* (strains DKU_NT_02 and DKU_NT_03) were initially aligned with Mauve using the progressive Mauve alignment algorithm (Darling et al. 2010). The resulting multiple genome alignment was subsequently visualized by displaying putative homologous segments across different genomes known as Locally Collinear Blocks (LCBs) with vertical lines connecting each of the same LCBs between different genomes. The aforementioned five genomes, along with 11 additional *B. subtilis* genomes isolated from natto and cheonggukjang (16 genomes total) were subsequently submitted to the Kostas Lab Genome Matrix (Rodriguez-R and Konstantinidis 2016) to calculate their average nucleotide identity (ANI) percentage by measuring nucleotide-level genomic similarities between coding regions within genomes. The obtained data was presented in a matrix format of ANI percentages across two different genomes. Lastly, these genomes were submitted to the Type (Strain) Genome Server (Meier-Kolthoff and Göker 2019) for phylogenetic inference. The resulting phylogenetic tree was subsequently visualized using Interactive Tree of Life (Letunic and Bork 2021).

Nattokinase expression and γ-PGA gene regulation analysis

The two hallmark characteristics of natto-derived *B. subtilis* are its ability to produce nattokinase and γ-PGA (Kada et al. 2013). While nattokinase is encoded by a single gene, *apr*N (Nakamura et al. 1992), γ-PGA is produced by a collection of PGA synthase enzymes encoded by the *pgs*A, *pgs*B, and *pgs*C genes, which are themselves regulated by a series of regulatory proteins. These regulatory proteins are encoded by the *deg*Q, *swr*A, *deg*S, *deg*U, *com*P, and *com*A genes (Luo et al. 2016). Notably, γ-PGA production in *B. subtilis* 168 is disrupted due to specific mutations in the *swr*A coding sequence (CDS) and *deg*Q promoter sequence (Stanley and Lazazzera 2005). To further characterize the ability of *B. subtilis* G8 to produce nattokinase and γ-PGA, *B. subtilis* G8 *apr*N CDS, *swr*A CDS, and *deg*Q promoter sequences were compared with those of *B. subtilis* BEST195 and *B. subtilis* subsp. subtilis 168 using MUSCLE multiple sequence alignment software (Edgar 2004; Madeira et al. 2019).

RESULTS AND DISCUSSION

Genome sequence assembly of *B. subtilis* G8

The assembled *B. subtilis* G8 draft genome consisted of 145 contigs with a total length of 4,017,503 bp (Table 2). In comparison to the *B. subtilis* BEST195 genome, the *B. subtilis* G8 genome had a similar GC content, predicted number of CDS, and predicted number of tRNA genes. Intriguingly, *B. subtilis* G8 possessed a much shorter sequence length (87,877 bp shorter) and fewer predicted rRNA genes (20 genes fewer) than *B. subtilis* BEST195. However, this does not imply that the *B. subtilis* G8 genome was truncated or that the strain is dysfunctional. Indeed, this study and another group's study (Kamada et al. 2015) observed that several *B. subtilis* strains had sequence lengths and predicted rRNA gene numbers comparable to those of *B. subtilis* G8 (Table 3). In addition, *B. subtilis* G8 crude extract exhibits strong fibrinolytic activities (Lucy et al. 2019; Pinontoan et al. 2021), supporting the conclusion that *B. subtilis* G8 is a functional strain.

Genomic comparison with natto-derived and cheonggukjang-derived *B. subtilis* strains

Japanese natto and Korean cheonggukjang are *B. subtilis*-fermented soybean foods with known health benefits (Zhao et al. 2013). Genomes of cheonggukjang-derived *B. subtilis* strains were therefore utilized as a control for homology assessment between *B. subtilis* G8 genomes and those of natto-derived *B. subtilis* strains (Table 1). *B. subtilis* 168 was included as another control because it was reported to be incapable of producing γ-PGA, unlike natto-derived *B. subtilis* strains (Nishito et al. 2010). Three methods were utilized: (i) multiple genome alignments with Mauve; (ii) ANI calculation; and (iii) phylogenetic analysis. First, multiple genome alignment with Mauve's progressive Mauve alignment algorithm was performed on the genomes of *B. subtilis* G8, *B. subtilis* BEST195, *B. subtilis* 168, as well as *B. subtilis* DKU_NT_02 and DKU_NT_03 (as representatives of cheonggukjang-derived *B. subtilis*) to identify any shared structure and organization among them, especially in case there had been genomic rearrangements, as LCBs. *B. subtilis* G8 had nine different LCBs (Fig. 1). Six of those LCBs, i.e., five of the largest LCBs and one smaller LCB (in red), were shared across all tested genomes. In contrast, the three smaller LCBs (observed near the end of the *B. subtilis* G8 genome), were only shared by two other *B. subtilis* strains. The first LCB (purple) was shared between *B. subtilis* G8, *B. subtilis* DKU_NT_03, and *B. subtilis* 168. The second LCB (violet) was only shared between *B. subtilis* G8 and *B. subtilis* DKU_NT_02. The third LCB (brown) was shared between *B. subtilis* G8 and *B. subtilis* DKU_NT_03. In summary, multiple genome alignment supported the notion that *B. subtilis* G8's genome was largely similar to the genomes of other tested *B. subtilis* strains.

The *B. subtilis* G8 genome had 100% ANI with natto-derived *B. subtilis* genomes (BEST195, N3-1, N4-2. N2-2, CGMC 2108, and N1_1), 99% with cheonggukjang-derived *B. subtilis* genomes (DKU_NT_02, DKU_NT_03, MH-1, 2RL2-3, 2KL1, GFR-12, PJ-7, and SSJ-1), and 98% ANI with the *B. subtilis* 168 genome (Fig. 2). These results suggest that *B. subtilis* G8 had a higher degree of similarity to Japanese natto-isolated strains than to Korean cheonggukjang-isolated strains or the standard laboratory strain.

Third, evolutionary relationships among the genomes were analyzed and represented as a phylogenetic tree. The genomes of *B. subtilis* G8 and 15 other *B. subtilis* strains were used in this analysis. *B. subtilis* G8 was placed in the same cluster as six natto-derived *B. subtilis* strains (Fig. 3). This finding corroborated the ANI calculation results that showed that *B. subtilis* G8 was more related to the six natto-derived *B. subtilis* strains than to the eight cheonggukjang-derived *B. subtilis* strains and the laboratory strain. Taken together, these findings indicate a high degree of homology between *B. subtilis* G8 and other natto-derived *B. subtilis* strains.

Table 2. Genome annotation statistics for *Bacillus subtilis* strains BEST195 and G8 using the DFAST annotation service

Statistic	*B. subtilis* BEST195	*B. subtilis* G8
Total sequence length (bp)	4,105,380	4,017,503
Predicted GC content (%)	43.5	43.4
Predicted CDS	4,294	4,279
Predicted coding ratio (%)	87.0	87.7
Predicted rRNA	30	10
Predicted tRNA	88	83

Note: CDS, coding sequence; rRNA, ribosomal RNA; tRNA, transfer RNA

Table 1. A list of the *Bacillus subtilis* genomes utilized in the *B. subtilis* G8 comparative analysis

Strain of *B. subtilis*	Source	Accession number	Reference
N1-1	Natto	CP032861.1	Heo et al. (2019)
N2-2	Natto	CP032863.1	Heo et al. (2019)
N3-1	Natto	CP032865.1	Heo et al. (2019)
N4-2	Natto	CP032867.1	Heo et al. (2019)
CGMCC 2108	Natto	CP014471.1	Tan et al. (2016)
BEST195	Natto	AP011541.2	Nishito et al. (2010)
2KL1	Cheonggukjang	CP032872.1	Heo et al. (2019)
2RL2-3	Cheonggukjang	CP032857.1	Heo et al. (2019)
GFR-12	Cheonggukjang	CP032852.1	Heo et al. (2019)
MH-1	Cheonggukjang	CP032853.1	Heo et al. (2019)
PJ-7	Cheonggukjang	CP032855.1	Heo et al. (2019)
SSJ-1	Cheonggukjang	CP032860.1	Heo et al. (2019)
DKU_NT_02	Cheonggukjang	CP022890.1	Bang et al. (2018)
DKU_NT_03	Cheonggukjang	CP022891.1	Jeong et al. (2018)
168	X-ray mutagenesis of *B. subtilis* Marburg	AL009126.3	Borriss et al. (2018); Burkholder and Giles (1947)

Note: Genomes of several *B. subtilis* strains used in the analyses were isolated from either Japanese natto or Korean cheonggukjang. In addition, the genome of the laboratory standard *B. subtilis* 168 was used

Table 3. Genome annotation statistics for several *Bacillus subtilis* strains isolated from Korean cheonggukjang and Japanese natto using the DFAST annotation service

Source	**Japanese Natto**								**Korean Cheonggukjang**							**Lab. standard**
Strain	**BEST195**	**CGMCC2108**	**N1-1**	**N2-2**	**N3-1**	**N4-2**	**DKU_NT_02**	**DKU_NT_03**	**2KL1**	**2RL2-3**	**GFR-12**	**MH-1**	**PJ-7**	**SSJ-1**	**168**
Total sequence length (bp)	4,105,380	4,122,154	4,108,100	4,122,398	4,122,184	4,119,216	4,014,255	4,196,031	4,196,696	4,191,876	4,202,955	4,184,648	4,293,706	4,199,434	4,215,606
Predicted GC content (%)	43.5	43.5	43.5	43.5	43.5	43.5	43.6	43.3	43.3	43.4	43.3	43.3	43.2	43.6	43.5
Predicted CDS	4,294	4,318	4,322	4,333	4,328	4,321	4,187	4,399	4,406	4,421	4,409	4,428	4,606	4,245	4,242
Predicted coding ratio (%)	87	87	86.9	86.9	87	87	87	87	86.8	86.8	86.9	86.8	87.1	87.9	87.8
Predicted rRNA	30	30	30	30	30	30	30	30	33	33	30	30	30	30	30
Predicted tRNA	88	88	88	88	88	88	87	87	93	87	87	87	87	86	87

Note: CDS: coding sequence; rRNA: ribosomal RNA; tRNA: transfer RNA

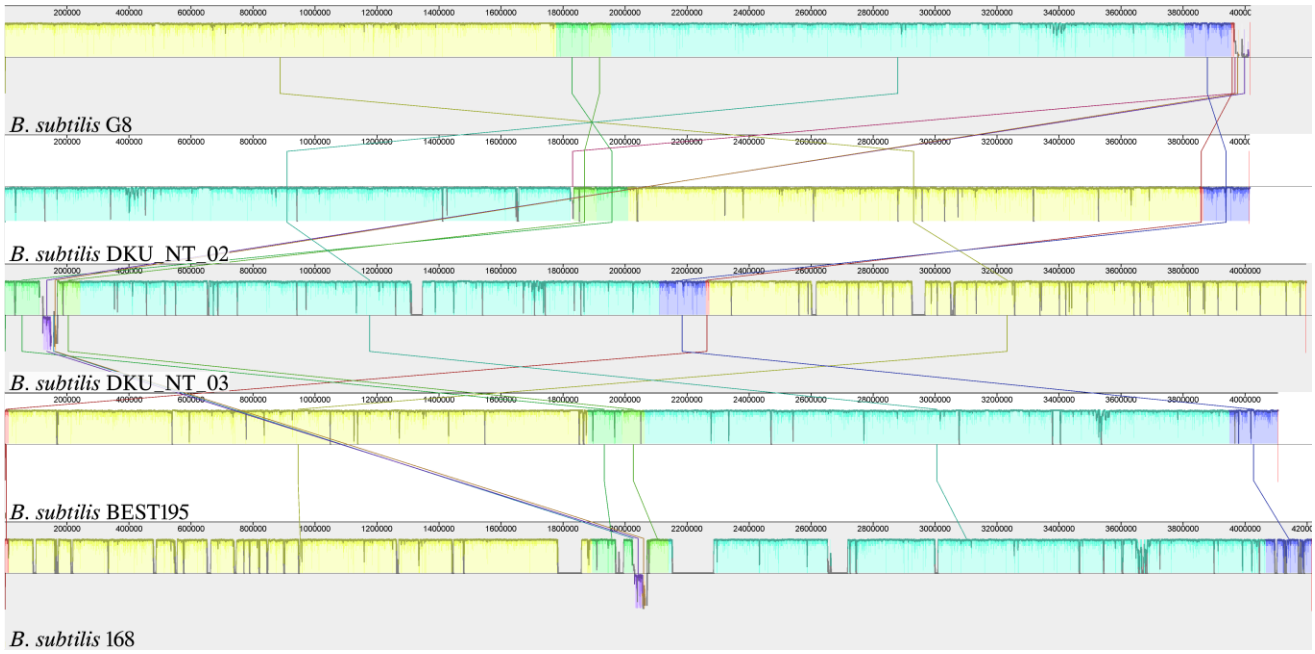

Figure 1. Multiple whole-genome alignments with Mauve on several strains of *Bacillus subtilis*. Note: As depicted in order, genomes of *B. subtilis* G8, *B. subtilis* DKU_NT_02 & DKU_NT_03 (both from cheonggukjang), *B. subtilis* BEST195 (from natto), and the laboratory standard *B. subtilis* 168, were aligned using the progressive Mauve software

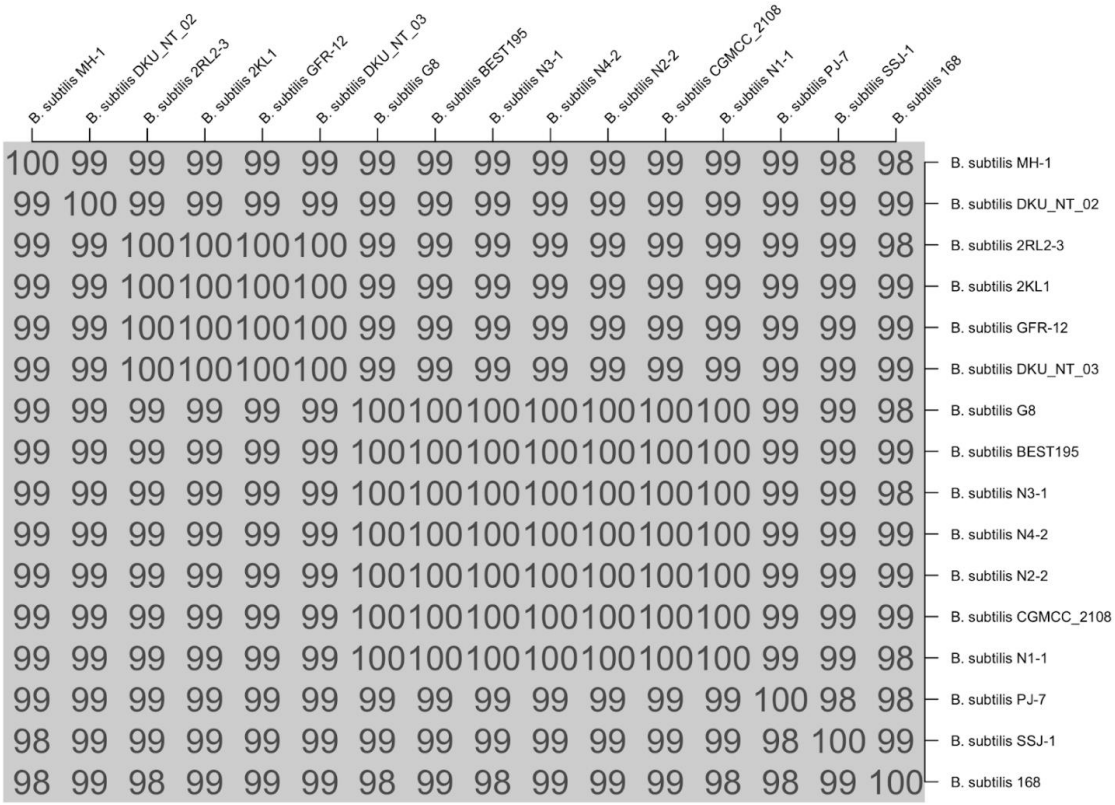

Figure 2. Average nucleotide identity (ANI) calculation among 16 genomes of *Bacillus subtilis* strains. Note: ANI percentage was calculated for *B. subtilis* G8, natto-derived *B. subtilis* (BEST195, N3-1, N4-2. N2-2, CGMC 2108, and N1_1), cheonggukjang-derived *B. subtilis* (DKU_NT_02, DKU_NT_03, MH-1, 2RL2-3, 2KL1, GFR-12, PJ-7, and SSJ-1), and the laboratory standard *B. subtilis* 168

Analyses on nattokinase-encoding and γ-PGA regulating genes

To demonstrate that *B. subtilis* G8 is a natto-derived *B. subtilis* strain that is able to produce nattokinase and γ-PGA, genes responsible for producing nattokinase and regulating γ-PGA production were compared with those of *B. subtilis* BEST195 and *B. subtilis* 168. Nattokinase, also known as subtilisin NAT, is an alkaline serine protease enzyme encoded by the *apr*N gene (Nakamura et al. 1992). As shown in Table 4, The CDS of *B. subtilis* G8 *apr*N was completely identical to that of *B. subtilis* BEST195, but not of *B. subtilis* 168. Since nattokinase is a fibrinolytic enzyme (Weng et al. 2017; Chen et al. 2018), this finding corroborated previous findings (Pinontoan et al. 2021), which showed that *B. subtilis* G8 exhibited strong fibrinolytic activities. Additionally, a 26.7 kDa protein band was observed during zymography analysis of *B. subtilis* G8 crude extract, which was in agreement with the nattokinase molecular weight of 27.7 kDa reported by Li et al. (2021).

Next, the genes regulating γ-PGA production in *B. subtilis* G8 were assessed and compared with those in *B. subtilis* BEST195 and *B. subtilis* 168. Unlike *B. subtilis* BEST195, *B. subtilis* 168 appears incapable of producing γ-PGA due to several mutations in its *swr*A CDS and *deg*Q promoter sequence. A single nucleotide insertion (adenine) in the *swr*A CDS caused a frameshift mutation and introduced a stop codon (TGA) at position 22, which prematurely terminates its translation and results in a small, nonfunctional protein. In addition, a single nucleotide substitution (thymine to cytosine) in its *deg*Q promoter sequence reduced its transcription rate. Both mutations consequently disrupt the regulation of γ-PGA production in *B. subtilis* 168 (Stanley and Lazazzera 2005). Both genes share 100% identity with *B. subtilis* G8 and *B. subtilis* BEST195 (Table 4). This was not the case between *B. subtilis* G8 and *B. subtilis* 168. In addition, the *swr*A CDSs and *deg*Q promoter sequences of the analyzed strains were aligned. Indeed, *B. subtilis* G8 and *B. subtilis* BEST195 did not possess any mutation in their *swr*A CDS and *deg*Q promoter sequences that might disrupt γ-PGA production, as observed in *B. subtilis* 168 (Fig. 4). The lack of such a mutation shows that *B. subtilis* G8 is capable of producing γ-PGA, which contributes to the bulk of natto's mucilage and consequently its sticky consistency (Hsueh et al. 2017; Chan et al. 2021). Collectively, these results strongly suggest that *B. subtilis* G8 is a natto-derived *B. subtilis* strain that is able to produce nattokinase and γ-PGA (Kada et al. 2013).

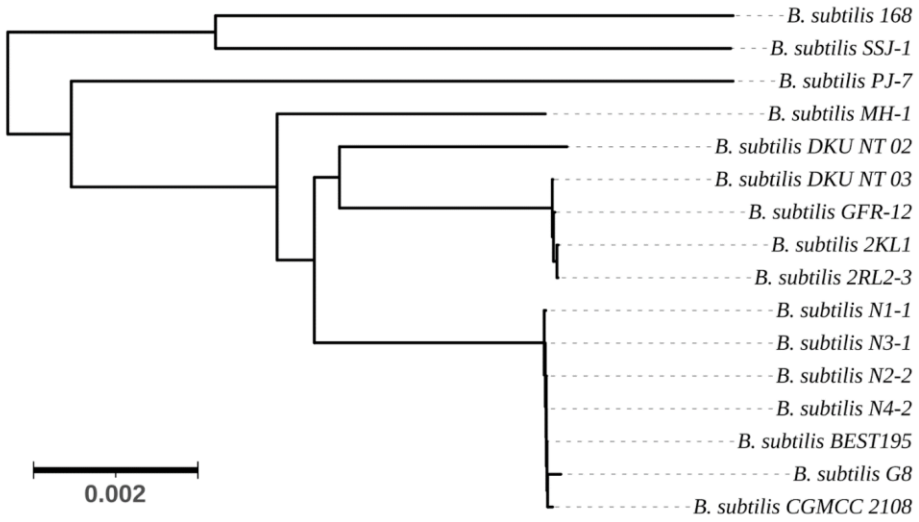

Figure 3. Phylogenetic analysis on the relationship among 16 genomes of *Bacillus subtilis* strains**.** Note: Type (Strain) Genome Server was utilized to infer the phylogenetic relationship among tested strains of *B. subtilis*. The phylogenetic tree was subsequently visualized using the Interactive Tree of Life (iTOL). Scale bar of 0.002

Table 4. Nucleotide BLAST identity percentages of *apr*N, *swr*A, and promoter sequence of *deg*Q sequences between *Bacillus subtilis* G8 and *B. subtilis* 168 or *B. subtilis* BEST195.

Sequences	Length (bp)	G8 to BEST195		G8 to 168	
		Coverage (%)	Identity (%)	Coverage (%)	Identity (%)
*apr*N CDS	1,146	100	100	100	99.21
*swr*A CDS	336	100	100	100	99.70
*deg*Q promoter	118	100	100	100	99.15

Note: CDS, coding sequence

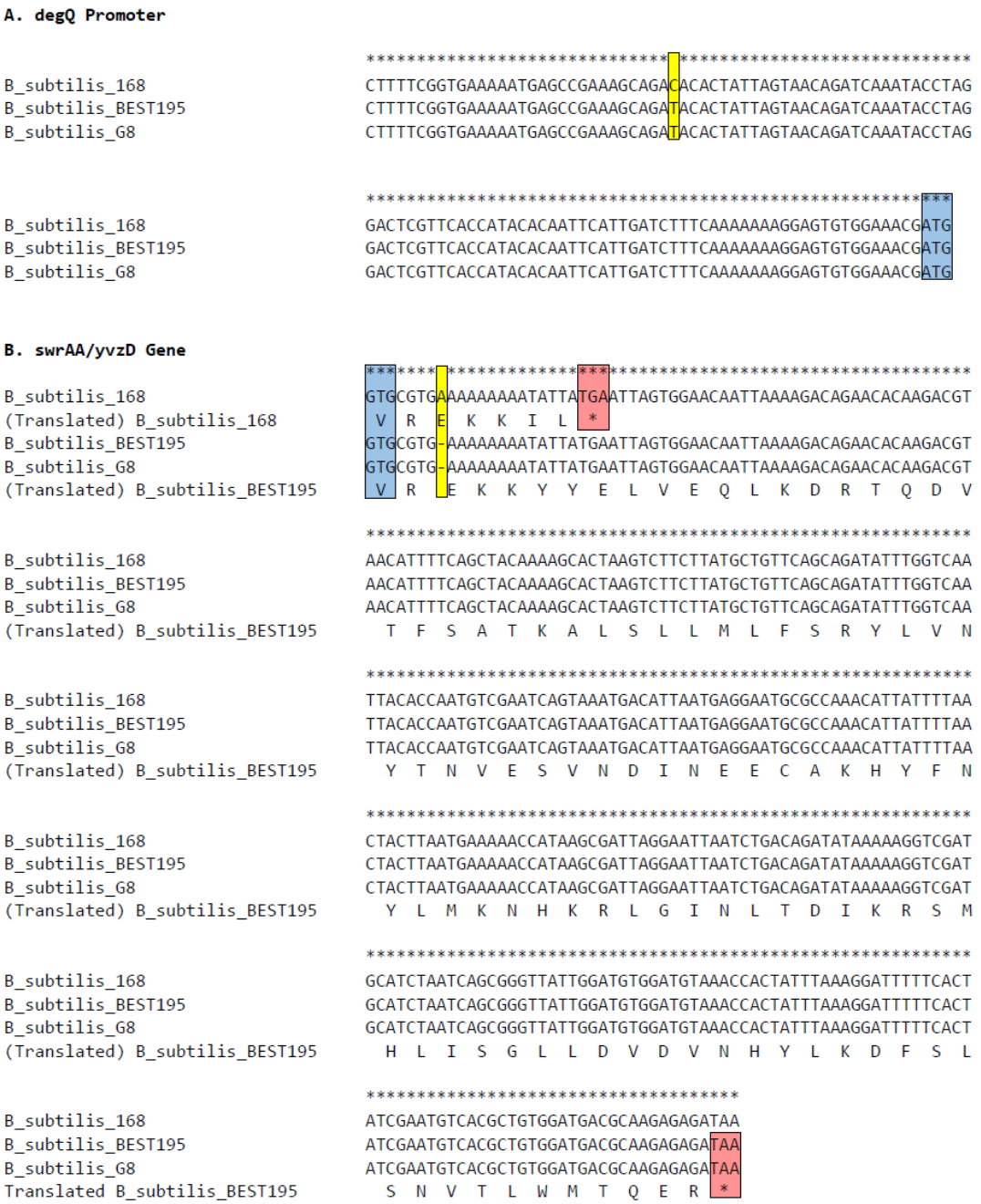

Figure 4. Multiple sequence alignment of genes regulating γ -PGA production. Note: The *deg*Q promoter sequence (A) and *swr*A CDS (B) were compared among genomes of *Bacillus subtilis* 168, *B. subtilis* BEST195, and *B. subtilis* G8. Notable mutations that disrupt the regulation of γ-PGA production in *B. subtilis* 168 are highlighted in yellow. Start codons of both genes are highlighted in blue. *swr*A CDS stop codon highlighted in red

In conclusion, the genomic analysis and comparisons with other *B. subtilis* strains indicate that *B. subtilis* G8 is closely related to natto-derived *B. subtilis* strains that produce nattokinase and γ-PGA, and it is thus proposed that *B. subtilis* G8 be reclassified as *B. subtilis* subspecies *natto* G8. As natto consumption is associated with various health benefits mediated by *B. subtilis* strains, including the prevention of cardiovascular disease, these findings may support a broader utilization of *B. subtilis* subsp. *natto* G8 in the near future.

ACKNOWLEDGEMENTS

This work was supported by the Directorate of Higher Education, Indonesian Ministry of Education, Culture, Research and Technology [No. 1218/LL3/PG/2021, 2021].

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215: 403-410. DOI:10.1016/S0022-2836(05)80360-2.

Bang M-S, Jeong H-W, Lee Y, Lee SJ, Lee S-C, Shin J-I, Oh C-H. 2018. Complete genome sequence of *Bacillus subtilis* strain DKU_NT_02, isolated from traditional Korean food using soybean (chung-gook-jang) for high-quality poly-γ-glutamic acid activity. Genome Announc 6: 25. DOI:10.1128/genomeA.00525-18.

Carver T, Berriman M, Tivey A, Patel C, Böhme U, Barrell BG, Parkhill J, Rajandream M-A. 2008. Artemis and ACT: Viewing, annotating and comparing sequences stored in a relational database. Bioinformatics 24: 2672-2676. DOI:10.1093/bioinformatics/btn529.

Chan EWC, Wong SK, Kezuka M, Oshiro N, Chan HT. 2021. Natto and miso: An overview on their preparation, bioactive components and health-promoting effects. Food Res 5: 446-452. DOI:10.26656/fr.2017.5(3).587.

Chen H, McGowan EM, Ren N, Lal S, Nassif N, Shad-Kaneez F, Qu X, Lin Y. 2018. Nattokinase: A promising alternative in prevention and treatment of cardiovascular diseases. Biomark Insights 13: 1-13. DOI:10.1177/1177271918785130.

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394-1403. DOI:10.1101/gr.2289704.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE 5: e11147. DOI: 10.1371/journal.pone.0011147.

Edgar RC. 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinform 5: 113. DOI: 10.1186/1471-2105-5-113.

Elshaghabee FMF, Rokana N, Gulhane RD, Sharma C, Panwar H, 2017. *Bacillus* as potential probiotics: Status, concerns, and future perspectives. Front Microbiol 8: 1490. DOI: 10.3389/fmicb.2017.01490.

Galata V, Fehlmann T, Backes C, Keller A. 2019. PLSDB: A resource of complete bacterial plasmids. Nucleic Acids Res 47: D195-D202. DOI: 10.1093/nar/gky1050.

García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: Evaluating next-generation sequencing alignment data. Bioinformatics 28: 2678-2679. DOI: 10.1093/bioinformatics/bts503.

Heo J, Kim J-S, Hong S-B, Park B-Y, Kim S-J, Kwon S-W. 2019. Genetic marker gene *rec*Q differentiating *Bacillus subtilis* and the closely related *Bacillus* species. FEMS Microbiol Lett 366: fnz172. DOI: 10.1093/femsle/fnz172.

Hsueh Y-H, Huang K-Y, Kunene SC, Lee T-Y. 2017. Poly-γ-glutamic acid synthesis, gene regulation, phylogenetic relationships, and role in fermentation. Intl J Mol Sci 18: 2644. DOI: 10.3390/ijms18122644.

Joshi NA, Fass JN. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. GitHub. https://github.com/najoshi/sickle.

Kada S, Ishikawa A, Ohshima Y, Yoshida K. 2013. Alkaline serine protease AprE plays an essential role in poly-γ-glutamate production during natto fermentation. Biosci Biotechnol Biochem 77: 802-809. DOI: 10.1271/bbb.120965.

Kamada M, Hase S, Fujii K, Miyake M, Sato K, Kimura K, Sakakibara Y. 2015. Whole-genome sequencing and comparative genome analysis of *Bacillus subtilis* strains isolated from non-salted fermented soybean foods. PLoS One 10: e0141369. DOI: 10.1371/journal.pone.0141369.

Kamada M, Hase S, Sato K, Toyoda A, Fujiyama A, Sakakibara Y. 2014. Whole genome complete resequencing of *Bacillus subtilis* natto by combining long reads with high-quality short reads. PLoS One 9: e0109999. DOI: 10.1371/journal.pone.0109999.

Kojima A, Ikehara S, Kamiya K, Kajita E, Sato Y, Kouda K, Tamaki J, Kagamimori S, Iki M. 2020. Natto intake is inversely associated with osteoporotic fracture risk in postmenopausal Japanese women. J Nutr 150: 599-605. DOI: 10.1093/jn/nxz292.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359. DOI: 10.1038/nmeth.1923.

Letunic I, Bork P. 2021. Interactive Tree of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49: W293-296. DOI: 10.1093/nar/gkab301.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078-2079. DOI: 10.1093/bioinformatics/btp352.

Li M, Zhang Z, Li S, Tian Z, Ma X. 2021. Study on the mechanism of production of γ-PGA and nattokinase in *Bacillus subtilis* natto based on RNA-seq analysis. Microb Cell Fact 20: 1-15. DOI: 10.1186/s12934-021-01570-x.

Lucy J, Raharjo PF, Elvina E, Florencia L, Susanti AI, Pinontoan R. 2019. Clot lysis activity of *Bacillus subtilis* G8 isolated from Japanese fermented natto soybeans. Appl Food Biotechnol 6: 101-109. DOI: 10.22037/afb.v6i2.22479.

Luo Z, Guo Y, Liu J, Qiu H, Zhao M, Zou W, Li S. 2016. Microbial synthesis of poly-γ-glutamic acid: current progress, challenges, and future perspectives. Biotechnol Biofuels 9: 1-12. DOI: 10.1186/s13068-016-0537-7.

Madeira F, Park Y mi, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47: W636-W641. DOI: 10.1093/nar/gkz268.

Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10: 2182. DOI: 10.1038/s41467-019-10210-3.

Nagata C, Wada K, Tamura T, Konishi K, Goto Y, Koda S, Kawachi T, Tsuji M, Nakamura K. 2017. Dietary soy and natto intake and cardiovascular disease mortality in Japanese adults: The Takayama study. Am J Clin Nutr 105: 426-431. DOI: 10.3945/ajcn.116.137281.

Nakamura T, Yamagata Y, Ichishima E. 1992. Nucleotide sequence of the subtilisin NAT gene, aprN, of *Bacillus subtilis* (natto). Biosci Biotechnol Biochem 56: 1869-1871. DOI:10.1271/bbb.56.1869

Nishito Y, Osana Y, Hachiya T, Popendorf K, Toyoda A, Fujiyama A, Itaya M, Sakakibara Y. 2010. Whole genome assembly of a natto production strain *Bacillus subtilis* Natto from very short read data. BMC Genom 11: 243. DOI: 10.1186/1471-2164-11-243.

Okonechnikov K, Conesa A, García-Alcalde F. 2015. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32: 292-294. DOI: 10.1093/bioinformatics/btv566.

Pangastuti A, Alfisah RK, Istiana NI, Sari SLA, Setyaningsih R, Susilowati A, Purwoko T. 2019. Metagenomic analysis of microbial community in over-fermented tempeh. Biodiversitas 20 (4): 1106-1114. DOI: 10.13057/biodiv/d200423.

Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28: 1420-1428. DOI: 10.1093/bioinformatics/bts174.

Pinontoan R, Elvina, Sanjaya A, Jo J. 2021. Fibrinolytic characteristics of *Bacillus subtilis* G8 isolated from natto. Biosci Microbiota Food Health 40: 144-149. DOI: 10.12938/bmfh.2020-071.

Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of draft genomes using the Mauve aligner. Bioinformatics 25: 2071-2073. DOI: 10.1093/bioinformatics/btp356.

Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr 4: 2-10. DOI: 10.7287/peerj.preprints.1900v1.

Stanley NR, Lazazzera BA. 2005. Defining the genetic differences between wild and domestic strains of *Bacillus subtilis* that affect poly-γ-dl-glutamic acid production and biofilm formation. Mol Microbiol 57: 1143-1158. DOI: 10.1111/j.1365-2958.2005.04746.x.

Tanizawa Y, Fujisawa T, Kaminuma E, Nakamura Y, Arita M. 2016. DFAST and DAGA: Web-based integrated genome annotation tools and resources. Biosci Microbiota Food Health 35: 173-184. DOI: 10.12938/bmfh.16-003.

Weng Y, Yao J, Sparks S, Wang KY. 2017. Nattokinase: An oral antithrombotic agent for the prevention of cardiovascular disease. Intl J Mol Sci 18: 523. DOI: 10.3390/ijms18030523.

Zhao X, Song J-L, Wang Q, Qian Y, Li G-J, Pang L. 2013. Comparisons of shuidouchi, natto, and cheonggukjang in their physicochemical properties, and antimutagenic and anticancer effects. Food Sci Biotechnol 22: 1077-1084. DOI: 10.1007/s10068-013-0186-6.