JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

Classification of the Genus *Bacillus* Based on MALDI-TOF MS Analysis of Ribosomal Proteins Coded in *S10* and *spc* Operons

Yudai Hotta,*[,†] Jun Sato,[†] Hiroaki Sato,[‡] Akifumi Hosoda,[†] and Hiroto Tamura[†]

[†]School of Agriculture, Meijo University, 1-501 Shiogamaguchi, Tenpaku-ku, Nagoya, Aichi 468-8502, Japan

[‡]Research Institute for Environmental Management Technology, National Institute of Advanced Industrial Science and Technology (AIST), 16-1 Onogawa, Tsukuba, Ibaraki 305-8569, Japan

Supporting Information

ABSTRACT: A rapid bacterial identification method by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) using ribosomal proteins coded in *S10* and *spc* operons as biomarkers, named the *S10*-GERMS (the *S10-spc-alpha* operon gene encoded ribosomal protein mass spectrum) method, was applied for the genus *Bacillus* a Gram-positive bacterium. The *S10*-GERMS method could successfully distinguish the difference between *B. subtilis* subsp. *subtilis* NBRC 13719^T and *B. subtilis* subsp. *spizizenii* NBRC 101239^T because of the mass difference of 2 ribosomal subunit proteins, despite the difference of only 2 bases in the 16S rRNA gene between them. The 8 selected reliable and reproducible ribosomal subunit proteins without disturbance of S/N level on MALDI-TOF MS analysis, S10, S14, S19, L18, L22, L24, L29, and L30, coded in *S10* and *spc* operons were significantly useful biomarkers for rapid bacterial classification at species and strain levels by the *S10*-GERMS method of genus *Bacillus* strains without purification of ribosomal proteins.

KEYWORDS: ribosomal protein, MALDI-TOF MS, bacterial identification, *S10-spc-alpha* operon, *S10*-GERMS

INTRODUCTION

The classification and identification of bacterial strains at the species and strains levels are important for the fields of clinical and environmental microbiology and the food safety industry. Matrix-assisted laser desorption ionization−time-of-flight mass spectrometry (MALDI-TOF MS) is one of the most widely used mass-based approaches for bacterial identification and classification because of simple sample preparation and extremely rapid analysis within a few minutes without any substantial costs for consumables.[1−3] Because MALDI-TOF MS detection of the bacterial proteome has been shown to be sufficient for bacterial identification and classification at the genus, species, subspecies, and discrimination at strain level,[2,4,5] bacterial identification and classification by MALDI-TOF MS has taken two general approaches toward data analysis, namely, pattern recognition and biomarker assignment based on bacterial genomic databases.

Pattern recognition is based on the identification of an unknown bacterial strain by comparing its mass spectrum to a database of mass spectra of known reference bacterial strains.[4−7] The representative methods are the Biotyper software program (Bruker Daltonics) and SARAMIS database application (AnagnosTec GmbH) comparing the mass spectrum of each strain with the mass spectra of many reference strains.[8,9] These methods demonstrated the reproducibility even if the culture conditions were not identical, although it uses uncertain biomarker peaks. Biomarker assignment methods also have been developed using a theoretical biomarker with a validation procedure for the obtained results. This approach identifies bacteria by a high number of matching between the observed masses of biomarkers by MALDI-TOF MS and theoretical masses of putative proteins in bacterial genomic databases;[10,11] however, the bacterial identification method based on the bioinformatics approach has several caveats: (1) the bacteria being identified or a closely related strain must have been genome-sequenced; (2) the accuracy of the genomic information is critical; (3) because of post-translational modification, there can be a significant difference between the theoretical mass based on a genomic database and the observed mass of the protein by MALDI-TOF MS.[12]

To avoid uncertain biomarker peaks, a bioinformatics-based approach has been proposed using ribosomal subunit proteins as biomarkers for the rapid identification of bacteria.[11] Because ribosomal subunit proteins have fewer post-translational modifications, except for N-terminal methionine loss, it is easy to calculate the theoretical mass on the basis of a genomic database. Therefore, our previous study demonstrated the bacterial identification method at the strain level based on ribosomal protein profile matching by comparison between the observed masses in the MALDI mass spectra of sample strains and theoretical biomarker masses of the genome-sequenced strain.[13] Moreover, 14 ribosomal subunit proteins coded in the *S10-spc-alpha* operon, which encodes half of the ribosomal subunit protein and is highly conserved in eubacterial genomes, were selected as reliable biomarkers for the identification of the genus *Pseudomonas* named the *S10*-GERMS (the *S10-spc-alpha* operon gene encoded ribosomal protein mass spectrum) method.[14] For further development of the *S10*-GERMS method, the genus *Bacillus* was targeted because it is a Gram-positive bacterium and plays an important role in food processing and human pathogens.[15−18] Moreover, *Bacillus* species are difficult to

Received: January 28, 2011
Revised: April 7, 2011
Accepted: April 7, 2011
Published: April 07, 2011

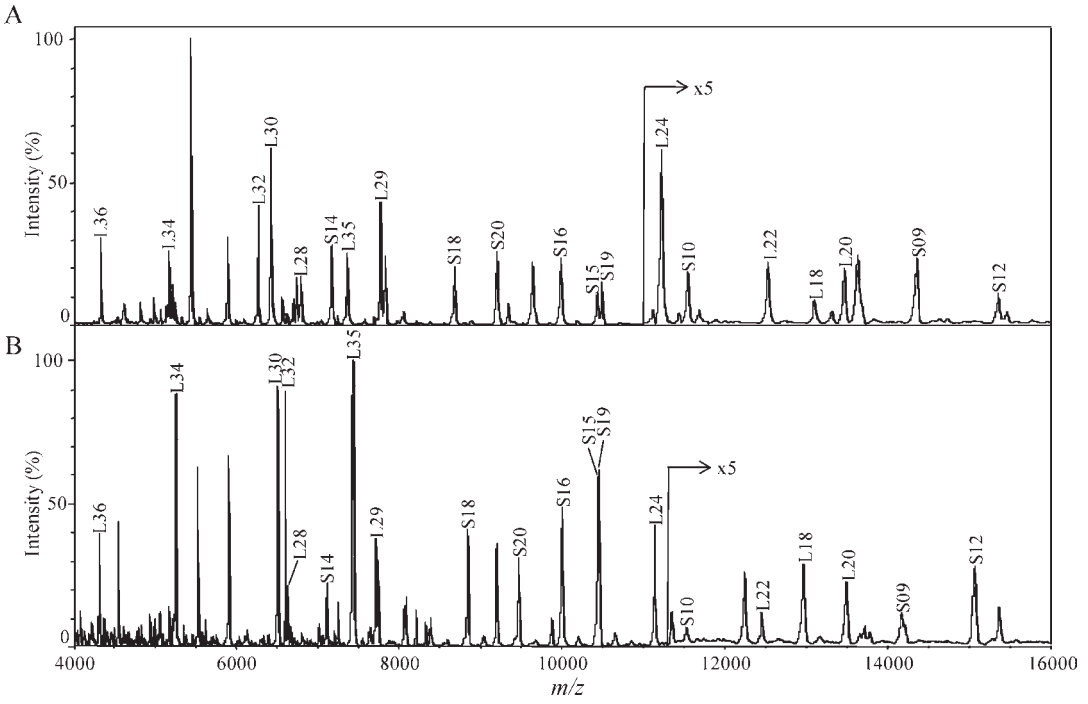

Figure 1. Typical MALDI mass spectra of genome-sequenced type strains: (A) *B. cereus* NBRC 15305^T; (B) *B. subtilis* subsp. *subtilis* NBRC 13719^T.

distinguish from one another because the nucleotide sequence of 16S rRNA gene of the *B. cereus* group exhibited very high levels of sequence similarity.[15]

In this study, therefore, the *S10*-GERMS method was applied for the rapid identification of the genus *Bacillus*. First, the ribosomal protein database was constructed by *S10* and *spc* operons sequencing of the genus *Bacillus*. Second, the reliability of our proposed method was assessed by comparison between the constructed database and MALDI mass spectra. Finally, to demonstrate the significance of ribosomal subunit proteins as biomarkers coded in *S10* and *spc* operons, phylogenetic analysis based on selected ribosomal subunit proteins encoded in *S10* and *spc* operons was compared with that based on selected ribosomal subunit protein from the genome-sequenced type strain.

EXPERIMENTAL PROCEDURES

Bacterial Strains. To compare the results of phylogenetic classification based on the 16S rRNA gene and ribosomal subunit proteins, 22 strains of the genus *Bacillus* were selected as follows: *B. amyloliquefaciens* NBRC 15535^T, *B. atrophaeus* NBRC 15539^T, *B. azotoformans* NBRC 15712^T, *B. badius* NBRC 15713^T, *B. cereus* NBRC 15305^T(=ATCC 14579^T), *B. licheniformis* NBRC 12200^T(=ATCC 14580^T), *B. megaterium* NBRC 15308^T, *B. mycoides* NBRC 101228^T(=DSM 2048^T), *B. niacini* NBRC 15566^T, *B. novalis* NBRC 102450^T, *B. pumilus* NBRC 12092^T(=ATCC 7061^T), *B. simplex* NBRC 15720^T, *B. subtilis* subsp. *subtilis* NBRC 13719^T(=168^T), *B. subtilis* NBRC 13722, *B. subtilis* NBRC 14415, *B. subtilis* NBRC 14474, *B. subtilis* subsp. *subtilis* NBRC 101246, *B. subtilis* subsp. *subtilis* NBRC 101588, *B. subtilis* NBRC 104440, *B. subtilis* subsp. *spizizenii* NBRC 101239^T, *B. thuringiensis* NBRC 101235^T(=ATCC 10792^T), and *B. vietnamensis* NBRC 101237^T. The NBRC strains were purchased from the National Institute of Technology and Evaluation (NITE)—Biological Resource Center (NBRC, Kisarazu, Japan). Each bacterial strain was grown aerobically in the medium and at the temperature recommended by suppliers.

***S10* and *spc* Operon Sequencing.** Chromosomal DNA was extracted from the bacteria as described previously.[14] The procedure included disruption of cells by cell-lysing solution, phenol/chloroform extraction, ethanol precipitation, and RNase treatment. The quantity and quality of the extracted DNA were estimated by measuring the UV absorption spectrum. PCR amplification of *S10* and *alpha* operons was performed using KOD containing dNTP at a concentration of 200 μM, each of the primers at a concentration of 4 μM, 100 ng of template DNA, and 2.5 U KOD polymerase (Toyobo, Tokyo, Japan) in a total volume of 50 μL. PCR amplification conditions of *S10* and *spc* operons were as follows: (1) 2 min at 95 °C, (2) 30 cycles of 30 s at 95 °C, 30 s at 50−55 °C, and 5 min at 72 °C, (3) 5 min at 72 °C. PCR primers and sequencing primers used in this study were designed on the basis of consensus nucleotide sequences of *S10* and *spc* operons from 13 genome-sequenced strains with the Clustal X program for alignment of nucleotide sequences (Table S1 in the Supporting Information). The sequencing reaction was carried out using a BigDye ver. 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) and an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems) according to the manufacturer's instructions. The Basic Local Alignment Search Tool (BLAST) program was used for homology analysis. Ribosomal subunit proteins of the genus *Bacillus* had accession numbers from AB609193 to AB609311, respectively, in the DDBJ/EMBL/GenBank.

MALDI-TOF MS Analysis of Whole Cell of Genus *Bacillus*. Ribosomal subunit protein analysis by MALDI-TOF MS was performed under almost the same conditions as described in our previous study.[13,14] Briefly, bacterial cells were harvested by centrifugation and washed twice in TMA-1 buffer (10 mM Tris-HCl (pH 7.8), 30 mM NH_4Cl, 10 mM $MgCl_2$, and 6 mM 2-mercaptoethanol). Acid treatment by formic acid (FA) and trifluoroacetic acid (TFA) was performed for disruption of the cell walls.[1,2] About 1.5 μL of each sample solution of whole cells adjusted to $OD_{660\ nm}$ = 1.0 was mixed with an equivalent volume of 20−100% (v/v) FA or 1−10% (v/v) TFA solution. Acid-treated sample solution was mixed with 7.0 μL of sinapic acid matrix solution at a concentration of 10 mg/mL in 50% (v/v) acetonitrile with 1% (v/v) trifluoroacetic acid (TFA). About 1.5 μL of sample/matrix

Table 1. Observed and Theoretical Masses (Average Masses, *m/z*) of Ribosomal Subunit Proteins of Genome-Sequenced Genus *Bacillus*[a]

subunit	*B. subtilis* NBRC13719^T	*B. licheniformis* NBRC12200^T	*B. pumilus* NBRC12092^T	*B. cereus* NBRC15305^T	*B. thuringiensis* NBRC101235^T	*B. mycoides* NBRC101228^T
L36	**4306.4**	**4306.4**	**4306.4**	4334.4	4334.4	4334.4
	(4306.5)	(4306.3)	(4305.9)	(4334.3)	(4334.3)	(4333.8)
L34	**5254.3**	**5254.3**	5271.3	*5171.2*	*5171.2*	*5171.2*
	(5254.2)	(5254.1)	(5270.9)	(5171.8)	(5171.1)	(5171.8)
L32	6598.8	6394.6	6411.7	*6263.5*	*6263.5*	*6739.1*
	(6598.7)	(6394.5)	(6411.5)	(6263.2)	(6263.3)	(6741.4)
L30[b]	6507.6	6506.7	6621.7	*6425.6*	*6439.6*	*6425.6*
	(6507.6)	(6506.7)	(6621.8)	(6425.4)	(6439.5)	(6439.5)
L28	6678.9	6722.9	6793.0	*6793.0*	*6793.0*	*6793.0*
	(6678.7)	(6723.1)	(6793.2)	(6793.4)	(6792.8)	(6792.8)
S14[b]	**7115.6**	**7115.6**	7173.6	*7165.6*	*7165.6*	*7165.6*
	(7114.9)	(7115.9)	(7173.8)	(7165.8)	(7165.4)	(7165.4)
L35	7426.9	7423.8	7410.8	*7365.7*	*7365.7*	*7365.7*
	(7426.2)	(7424.6)	(7410.9)	(7366.3)	(7365.9)	(7366.2)
L29[b]	7714.0	7730.0	7729.0	*7769.1*	*7769.1*	*7716.1*
	(7713.8)	(7730.6)	(7729.5)	(7729.5)	(7779.1)	(7715.4)
S18	8839.4	8866.4	8868.5	*8683.3*	*8683.3*	*8699.3*
	(8839.1)	(8866.6)	(8869.0)	(8684.0)	(8683.2)	(8699.5)
S20	9469.0	9391.9	9380.8	*9211.6*	*9227.6*	*9271.7*
	(9468.8)	(9391.6)	(9381.6)	(9213.0)	(9227.6)	(9272.1)
S16	10004.6	10090.8	10059.6	*9987.6*	*9987.6*	*9972.6*
	(10004.5)	(10091.2)	(10060.6)	(9988.5)	(9987.7)	(9973.2)
S15	10443.0	10471.0	10431.0	*10430.0*	*10430.0*	*10430.0*
	(10441.6)	(10471.7)	(10432.5)	(10430.8)	(10430.1)	(10430.6)
S19[b]	10453.1	10497.1	10453.0	*10498.2*	*10498.2*	*10525.2*
	(10451.9)	(10497.5)	(10454.7)	(10499.0)	(10498.9)	(10526.0)
L24[b]	11143.1	11166.2	11106.0	*11229.3*	*11229.3*	*11215.3*
	(11143.5)	(11166.2)	(11107.5)	(11229.6)	(11228.7)	(11216.0)
S10[b]	**11535.5**	**11535.5**	**11535.5**	*11553.5*	*11567.5*	*11567.5*
	(11536.4)	(11535.7)	(11536.5)	(11554.2)	(11568.8)	(11568.7)
L22[b]	12460.6	12599.7	12344.5	*12536.7*	*12536.7*	*12564.8*
	(12461.8)	(12599.6)	(12344.7)	(12536.3)	(12536.7)	(12536.3)
L18[b]	12969.8	13012.8	13071.0	*13107.0*	*13094.9*	*13066.9*
	(12969.6)	(13011.2)	(13071.5)	(13107.0)	(13096.0)	(13067.9)
L20	13507.9	13520.9	13535.9	*13481.9*	*13481.9*	*13454.9*
	(13507.1)	(13519.1)	(13535.2)	(13482.3)	(13481.2)	(13456.6)
S09	14178.3	14212.3	14397.5	*14360.6*	*14360.6*	*14359.6*
	(14178.1)	(14210.9)	(14397.0)	(14360.4)	(14359.8)	(14359.7)
S12	15085.5	15143.6	15302.8	*15370.0*	*15370.0*	*15370.0*
	(15083.5)	(15142.0)	(15302.5)	(15368.6)	(15368.0)	(15369.1)

[a] Bold and italic indicate common theoretical masses among subunit proteins. Numbers given in parentheses indicate observed masses. [b] Eight ribosomal subunit proteins selected for non-genome sequenced genus *Bacillus*.

mixture was spotted onto the MALDI target and dried in air. MALDI mass spectra in the range of m/z 4000−20000 were observed in positive linear mode by averaging 1000 individual laser shots using an AXIMA-Performance time-of-flight mass spectrometer (Shimadzu/Kratos, Kyoto, Japan) equipped with a pulsed N_2 laser. Mass calibration of whole cells of *Pseudomonas putida* NBRC 100650 (=KT2440) was followed by self-calibration using four moderately strong peaks assigned to ribosomal subunit proteins, L36 ($[M + H]^+$, m/z 4435.3), L29 ($[M + H]^+$, m/z 7173.3), S10 ($[M + H]^+$, m/z 10753.6), and L15 ($[M + H]^+$, m/z 15190.4), as internal references after external calibration using two peaks of myoglobin ($[M + H]^+$, m/z 16952.6, and $[M + 2H]^{2+}$, m/z 8476.8). Mass calibration of genus *Bacillus* strains was performed by external calibration using the mass spectra observed for the KT2440 strain. The peak matching of theoretical masses of biomarker proteins was judged from errors within 150 ppm as described previously.[13,14,19,20] The observed masses in the MALDI mass spectra of *B. subtilis* sample strains were compared with the biomarker masses of genome-sequenced type strain *B. subtilis* subsp. *subtilis* NBRC 13719^T as the reference strain for the ribosomal subunit protein profile matching as described in our previous study.[13] The presence or absence of the reference masses was indicated by

1 or 0, respectively, and their data were processed by cluster analysis, generating a phylogenetic tree.

Construction of the Ribosomal Protein Database. The amino acid sequences of all ribosomal subunit proteins of genus *Bacillus* strains were obtained from the NCBInr database. The calculated mass of each subunit protein was predicted using a Compute pI/M_w tool on the ExPASy proteomics server (http://au.expasy.org/tools/pi_tool.html). N-Terminal methionine loss was first considered based on the "N-end rule" as a post-translational modification.[21] In this rule, N-terminal methionine is cleaved from specific penultimate amino acid residues such as glycine, alanine, serine, proline, valine, threonine, and cysteine. To validate the constructed database, theoretical masses of ribosomal subunit proteins of genome-sequenced strains, *B. cereus* NBRC 15305^T, *B. licheniformis* NBRC 12200^T, *B. mycoides* NBRC 101228^T, *B. pumilus* NBRC 12092^T, *B. subtilis* subsp. *subtilis* NBRC 13719^T, and *B. thuringiensis* NBRC 101235^T, were compared with each of the MALDI mass spectra, respectively.

Cluster Analysis. Phylogenetic analysis was performed with the Clustal X program.[22] Phylogenetic tree construction and bootstrap analyses (100 replicates) were performed using the Mega3.1 program.[23] Phylogenetic trees were constructed using three different methods, the neighbor-joining, maximum-likelihood, and maximum parsimony algorithms available in the Mega3.1 program.

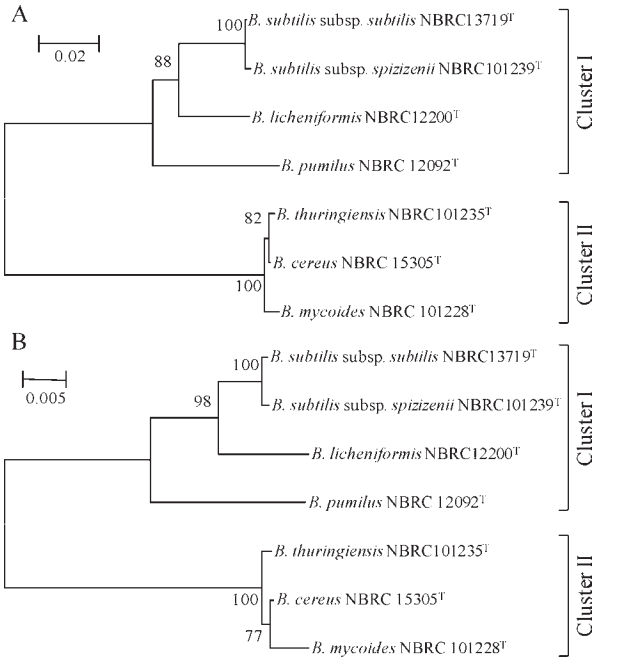

Figure 2. Phylogenetic tree of genome-sequenced type strain of genus *Bacillus* by the NJ method: (A) phylogenetic tree based on amino acid sequences of 20 ribosomal proteins; (B) phylogenetic tree based on 16S rRNA gene sequence.

RESULTS AND DISCUSSION

The genus *Bacillus* includes bacteria that are considered human pathogens. *B. anthracis* is the causative agent of anthrax and is of high relevance to human and animal health, *B. cereus* is occasionally associated with food poisoning, and *B. thuringiensis* is primarily an insect pathogen because of its ability to produce toxins, which have been widely used as bioinsecticides against harmful insects.[15,16] Furthermore, *B. subtilis*, *B. amyloliquefaciens*, *B. pumilus*, and *B. licheniformis* have the capability to produce γ-polyglutamic acid (PGA) and are used for food processing of fermented soybeans foods.[17] On the other hand, *B. subtilis*, *B. licheniformis*, *B. megaterium*, and *B. cereus* are capable of causing food spoilage.[18] Therefore, it is important to establish a rapid and simple identification method for the genus *Bacillus*. According to an announcement by the IMG system, 110 strains of the genus *Bacillus* including 12 type strains, which are important for the identification of bacterial species, had been genome-sequenced by January 1, 2011 (including draft status).

Construction of Ribosomal Protein Database of Genome-Sequenced Genus *Bacillus*. Although the constructed ribosomal protein database based on amino acid sequences of genome-sequenced bacteria has some errors in amino acid sequences, these errors were easily and theoretically corrected by comparison between the masses observed by MALDI mass spectra and theoretical masses of the database based on the *S10-spc-alpha* operon. Therefore, a ribosomal protein database of six genome-sequenced type strains was constructed, and each theoretical mass was compared with MALDI mass spectra, respectively.

Because Gram-positive bacteria such as *Lactobacillus plantarum* have a rigid cell wall and it is difficult to analyze ribosomal subunit proteins,[19] acid extraction treatment was also applied to the genus *Bacillus*. The result using TFA and FA revealed that 30% FA treatment gave the best signal quality on MALDI-TOF MS (Figure S1 in the Supporting Information). In particular, high molecular weights of ribosomal subunit proteins such as L18, L22, and S10 subunits, coded in *S10* and *spc* operons, were detected with high sensitivity; therefore, the acid extraction procedure using 30% FA was used for the *S10*-GERMS method in this study.

Twenty ribosomal subunit proteins were detected in all six genome-sequenced type strains by MALDI mass spectra. Therefore, the ribosomal protein database of genome-sequenced type strains was constructed on the basis of those 20 ribosomal proteins (Figure 1; Table 1). Comparison of the constructed ribosomal protein database with the observed MALDI mass spectra revealed the following errors: misannotation of the start codon of L22 and L32 and sequence error of L29 subunit proteins in *B. cereus* NBRC 15305^T; the presence of unregistered L22, L32, L34, S14, S15, and S18 subunit proteins in *B. pumilus* NBRC 12092^T; misannotation of the start codons of L29 and S20 subunit proteins and the presence of an unregistered L24 subunit protein in *B. thuringiensis* NBRC 101235^T; misannotation of the start codons of L29, S14, and S18 and sequence errors of L22 and S20 subunit proteins and the presence of an unregistered L24 subunit protein in *B. mycoides* NBRC 101228^T. Lauber et al. reported that dozens of ribosomal protein masses of *B. subtilis* subsp. *subtilis* NBRC 13719^T were found to be in error and not easily accounted for by post-translational modifications, and these errors had been corrected.[24] Therefore, 20 selected ribosomal subunit proteins of *B. subtilis* subsp. *subtilis* NBRC 13719^T considering N-terminal methionine loss only as a post-translational modification could be assigned by MALDI-TOF MS analysis. Sequence errors of ribosomal subunit proteins of other type strains were confirmed by sequencing of *S10* and *spc* operons. Furthermore, misannotated and unregistered ribosomal proteins were annotated on the basis of complete genome sequence registered in NCBInr database and their theoretical masses by sequencing of target genes were compared with observed masses by MALDI-TOF MS analysis. Taken together, these results demonstrated that unregistered ribosomal subunit

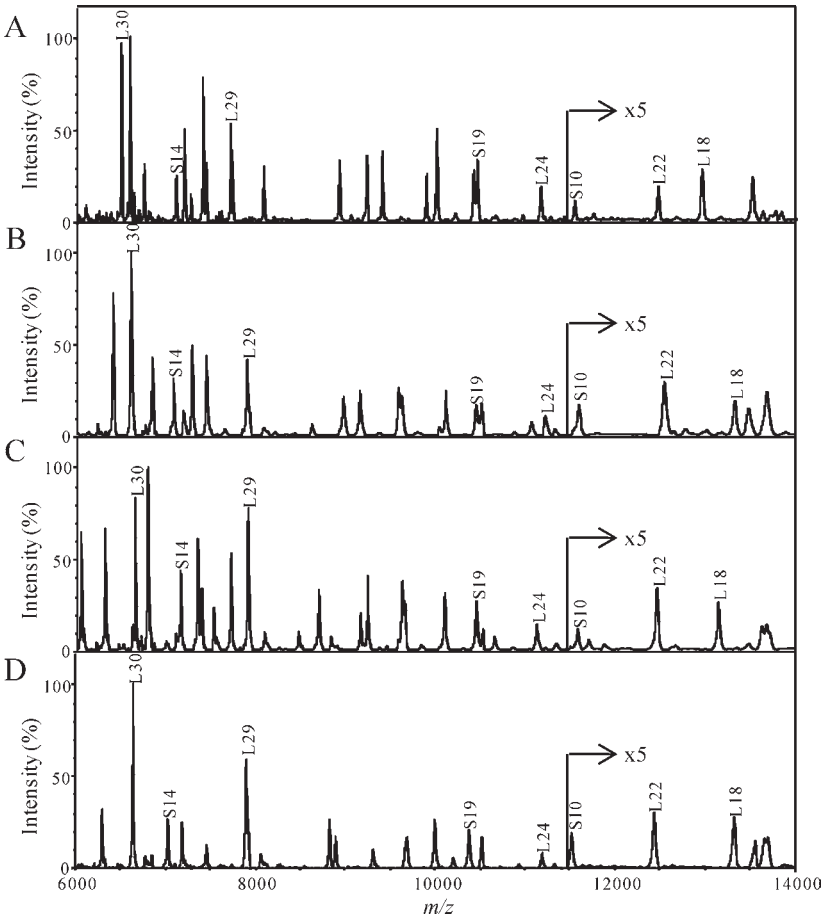

Figure 3. Typical MALDI mass spectra of non-genome-sequenced type strains of genus *Bacillus*: (A) *B. amyloliquefaciens* NBRC 15535^T; (B) *B. azotoformans* NBRC 15712^T; (C) *B. badius* NBRC 15713^T; (D) *B. vietnamensis* NBRC 101237^T.

proteins and registration error should be confirmed on the basis of the nucleotide sequences of genome-sequenced strains and MALDI mass spectra.

To evaluate the classification result, the phylogenetic tree based on amino acid sequences of 20 ribosomal subunit proteins was compared with the phylogenetic tree based on 16S rRNA gene sequence (Figure 2). The basic topologies of the two phylogenetic trees were similar, but were slightly different in their details. The genus *Bacillus* contains two important groups of bacteria named after *B. subtilis* and *B. cereus*, respectively. The *B. subtilis* group contains the closely related taxa *B. amyloliquefaciens*, *B. atrophaeus*, *B. licheniformis*, *B. mojavensis*, *B. sonorensis*, *B. subtilis* subsp. *subtilis*, *B. subtilis* subsp. *spizizenii*, and *B. vallismortis*.[25] The *B. cereus* group contains *B. anthracis*, *B. cereus*, *B. mycoides*, *B. pseudomycoides*, *B. thuringiensis*, and *B. weihenstephanensis*.[25] The phylogenetic tree based on amino acid sequences of 20 ribosomal subunit proteins indicated that *B. licheniformis*, *B. pumilus*, and *B. subtilis* were affiliated in a cluster in the *B. subtilis* group (cluster I) and that *B. cereus*, *B. mycoides*, and *B. thuringiensis* were contained in a cluster in the *B. cereus* group (cluster II). Although the *B. cereus* group taxa were difficult to distinguish from each other by 16S rRNA gene sequences,[15] MALDI mass spectra of ribosomal protein could discriminate them clearly because of mass differences as follows: 4 of 20 ribosomal subunit protein differences between *B. cereus* NBRC 15305^T and *B. thuringiensis* NBRC 101235^T; 12 of 20 ribosomal subunit protein differences between *B. cereus* NBRC 15305^T and *B. mycoides* NBRC 101228^T; and 12 of 20 ribosomal subunit protein differences between *B. mycoides* NBRC 101228^T and *B. thuringiensis* NBRC 101235^T (Table 1). These results demonstrated that bacterial identification by MALDI-TOF MS using ribosomal subunit proteins discriminated the genus *Bacillus* at the species level.

Construction of Ribosomal Protein Database of Non-Genome-Sequenced Genus *Bacillus*. In our previous study, 14 reliable and reproducible ribosomal subunit proteins coded in the *S10-spc-alpha* operon were selected as significantly useful biomarkers for bacterial classification at species and strain levels of genus *Pseudomonas* strains by the *S10*-GERMS method.[14] Although 20 reliable ribosomal subunit proteins on genome-sequenced type strains of the genus *Bacillus* were detected by MALDI-TOF MS, both S11 and S13 subunit proteins coded in the *alpha* operon, which were selected as biomarkers for identification of the genus *Pseudomonas*, were not found; therefore, ribosomal protein databases of non-genome-sequenced type strains of the genus *Bacillus* were constructed by *S10* and *spc* operon sequencing, except for the *alpha* operon.

S10 and *spc* operons of the genus *Bacillus* were sequenced using a specific primer based on consensus nucleotide sequences of *S10* and *spc* operons from 13 genome-sequenced strains (Table S1 of the Supporting Information). The ribosomal protein database of 8 selected ribosomal subunit proteins, L18,

Table 2. Observed and Theoretical Masses (Average Masses, *m*/*z*) of Eight Selected Ribosomal Subunit Proteins of Genus *Bacillus* Type Strains[a]

subunit	*B. amyloliquefaciens* NBRC15535^T	*B. atrophaeus* NBRC15539^T	*B. azotoformans* NBRC15712^7	*B. badius* NBRC15713^T	*B. megaterium* NBRC15308^T	*B. niacini* NBRC15566^T	*B. novalis* NBRC 102450^T	*B. simplex* NBRC15720^7	*B. vietnamensis* NBRC101237^T
L30	6507.6	6507.6	6610.8	6662.7	6577.7	6532.6	6418.5	6679.8	6651.7
	(6507.7)	(6507.7)	(6610.9)	(6662.7)	(6577.6)	(6532.6)	(6418.5)	(6679.9)	(6651.7)
S14	7115.6	7115.6	7087.5	7165.6	7173.6	7124.5	7165.6	7202.7	7033.5
	(7115.6)	(7115.9)	(7087.8)	(7165.4)	(7173.4)	(7124.0)	(7165.5)	(7202.5)	(7033.7)
L29	7714.0	7714.0	7900.2	7912.2	7730.0	7758.1	7744.0	7844.2	7900.2
	(7114.7)	(7714.7)	(7990.0)	(7912.0)	(7729.7)	(7757.2)	(7744.1)	(7844.1)	(7901.4)
S19	10453.1	10439.0	10453.0	10455.0	10452.9	10437.0	10441.0	10449.0	10384.9
	(10454.2)	(10440.4)	(10452.8)	(10455.1)	(10453.6)	(10435.5)	(10440.3)	(10447.5)	(11385.4)
L24	11157.1	11133.0	11225.2	11126.2	11062.9	10930.8	10835.7	11151.1	11191.2
	(11158.0)	(11133.1)	(11224.6)	(11126.3)	(11062.4)	(13929.1)	(10835.9)	(11150.2)	(11192.8)
S10	11535.5	11535.5	11591.6	11579.5	11578.5	11533.5	11544.4	11563.5	11519.5
	(11536.2)	(11536.2)	(11590.9)	(11579.9)	(11577.1)	(11533.2)	(11543.4)	(11563.2)	(11520.5)
L22	12460.6	12429.6	12552.7	12463.6	12407.6	12519.7	12458.6	12384.5	12437.6
	(12460.7)	(12428.3)	(12552.3)	(12463.6)	(12407.8)	(12518.6)	(12458.0)	(12384.0)	(12437.9)
L18	12954.7	13010.9	13332.3	13155.0	13221.2	13152.1	13190.1	13181.1	13325.2
	(12955.2)	(13009.9)	(13331.8)	(13154.9)	(13221.2)	(13154.1)	(13191.7)	(12381.8)	(13324.4)

[a] Numbers given in parentheses indicate observed masses.

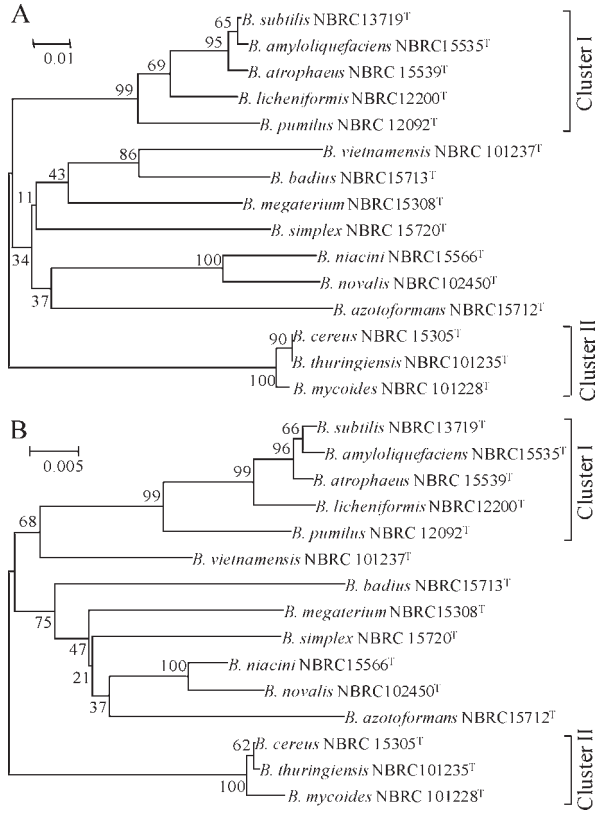

Figure 4. Phylogenetic trees of type strains of genus *Bacillus* strains by the NJ method: (A) phylogenetic tree based on amino acid sequences of 8 selected ribosomal proteins in *S10* and *spc* operons; (B) phylogenetic tree based on 16S rRNA gene sequence.

L22, L24, L29, L30, S10, S14, and S19 in *S10* and *spc* operons, was constructed using the *S10*-GERMS method because the mass spectra of 8 ribosomal subunit proteins of type strains of the genus *Bacillus* had mass reproducibility and sensitivity based on the S/N level (Figure 3; Table 2). To evaluate the availability of the selected 8 ribosomal subunit proteins for bacterial identification and classification, the phylogenetic tree based on amino acid sequences of the 8 ribosomal subunit proteins was compared with the phylogenetic tree based on the 16S rRNA gene sequence (Figure 4). Each phylogenetic tree obtained by the *S10*-GERMS method and 16S rRNA gene sequence gave almost the same results with slight differences in their details. The phylogenetic tree based on amino acid sequences of the 8 ribosomal subunit proteins formed clusters I and II. The most conspicuous difference was the phylogenetic position of *B. vietnamensis* NBRC 101237^T. Noguchi et al. reported that *B. vietnamensis* formed a cluster with *B. aquimaris* and *B. marisflavi*, whose cluster was linked to the *B. subtilis* group.[26] However, the phylogenetic tree constructed by Cerritos et al. clearly separated them.[27] Therefore, the full discussion of phylogenetic analysis based on *S10*-GERMS method may require more information on genus *Bacillus* species. Previously, *B. amyloliquefaciens* NBRC 15535^T and *B. subtilis* subsp. *subtilis* NBRC 13719^T in the *B. subtilis* group and *B. cereus* NBRC 15305^T and *B. thuringiensis* NBRC 101235^T in the *B. cereus* group were recognized as closely related strains, respectively; however, the *S10*-GERMS method could discriminate them clearly because of mass differences as follows: mass differences of L22 and L24 subunit proteins between *B. amyloliquefaciens* NBRC 15535^T and *B. subtilis* subsp. *subtilis* NBRC 13719^T; mass differences of L22, L30, and S10 subunit proteins between *B. cereus* NBRC 15305^T and *B. thuringiensis* NBRC 101235^T (Tables 1 and 2). These results demonstrated that the 8 selected ribosomal subunit proteins were significantly useful biomarkers for bacterial classification at the species level by MALDI-TOF MS analysis of genus *Bacillus* strains, although the phylogenetic position of *B. vietnamensis* NBRC 101237^T remains to be discussed.

Discrimination of *B. subtilis* at the Strain Level. Bacterial identification and classification by MALDI-TOF MS using ribosomal subunit proteins as biomarkers discriminated *P. putida*,

Table 3. Binary Peak Matching Profile of *B. subtilis* Strains[a]

	reference peaks																			
strain NBRC no.	L36	L34	**L30**	L32	L28	**S14**	L35	**L29**	S18	S20	S16	S15	**S19**	**L24**	**S10**	**L22**	**L18**	L20	S09	S12
13719^T	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14415	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14474	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
101588	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
101246	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0	1	1	0
13722	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1
101239	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	0	1	1
104440	1	0	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	0	1	0

[a] Present peaks are indicated as 1 and absent peaks as 0. Boldface indicates 8 ribosomal subunit proteins selected for non-genome-sequenced genus *Bacillus*.

Bifidobacterium longum, and *Rhodococcus erythropolis* at the subspecies and strain levels, respectively.[13,14] Three subspecies in *B. subtilis* are known: *B. subtilis* subsp. *subtilis*, *B. subtilis* subsp. *spizizenii*, and *B. subtilis* subsp. *inaquosorum*.[28,29] The 16S rRNA sequence identity between *B. subtilis* subsp. *subtilis* NBRC 13719^T and *B. subtilis* subsp. *spizizenii* NBRC 101239^T was 99.8% (1473/1475), although the phylogenetic position in the phylogenetic tree of *B. subtilis* based on the 16S rRNA gene sequence was easily changed by only one base difference of the nucleotide sequence.[28] Because it was difficult to discriminate *B. subtilis* at the subspecies and strain levels, *B. subtilis*, which is a type species of the genus *Bacillus*, was used as the model bacteria in this study. To assess bacterial classification at subspecies and strain levels, phylogenetic trees of *B. subtilis* based on 8 and 20 ribosomal subunit protein profile matching using ribosomal protein of *B. subtilis* subsp. *subtilis* NBRC 13719^T as a reference strain were constructed and compared (Table 3; Figures 5 and 6). Bacterial identification using 8 or 20 selected ribosomal subunit proteins as theoretical biomarkers indicated the clear discrimination of *B. subtilis* at subspecies and strain levels rather than 16S rRNA gene analysis. In detail, the basic topologies of their phylogenetic trees were slightly different. The most conspicuous difference was the phylogenetic position between *B. subtilis* NBRC 13722 and *B. subtilis* subsp. *spizizenii* NBRC 101239^T. Despite the difference of only 2 bases in the 16S rRNA gene between *B. subtilis* subsp. *subtilis* NBRC 13719^T and *B. subtilis* subsp. *spizizenii* NBRC 101239^T, they showed 63 and 67% of DNA−DNA relatedness value.[28,30] In the *S10*-GERMS method, 2 of 8 selected ribosomal subunit proteins, L18 and L29, were different (Table 3); therefore, the differences of 2 of 8 and 4 of 20 ribosomal subunit proteins might be standard for discrimination at species and subspecies levels of genus *Bacillus* in bacterial identification by MALDI-TOF MS using ribosomal subunit proteins.

In this study, because 8 selected ribosomal subunit proteins coded in *S10* and *spc* operons for non-genome-sequenced type strains of the genus *Bacillus* were corrected by sequencing using designed specific primers for *S10* and *spc* operons of the genus *Bacillus*, these were suitable biomarkers for construction of the database. Therefore, the *S10*-GERMS method was a significantly useful tool for bacterial classification at species, subspecies, and strain levels of genus *Bacillus* strains based on phylogenetic analysis. Although the *S10*-GERMS method required the development of a ribosomal protein database of type strains, the phylogenetic tree based on this method revealed a clearer discrimination of bacteria than that based on the 16S rRNA gene sequence. Although the bacterial identification method using ribosomal subunit proteins is applicable even to non-genome-sequenced bacterial groups through MALDI-TOF MS analysis of purified ribosomal proteins of the type strain,[20] it is difficult to determine the subunit numbers and theoretical masses of applied biomarkers; therefore, the *S10*-GERMS method can enable ribosomal protein database construction due to the theoretical masses of ribosomal subunit proteins by sequencing the *S10-spc-alpha* operon of target bacteria. Because it is an important for risk assessment in food safety, the *S10*-GERMS method can enable the identifucation of the genus *Bacillus* associated with food processing, food spoilage, and food poisoning. In the future, the rapid identification of bacteria by

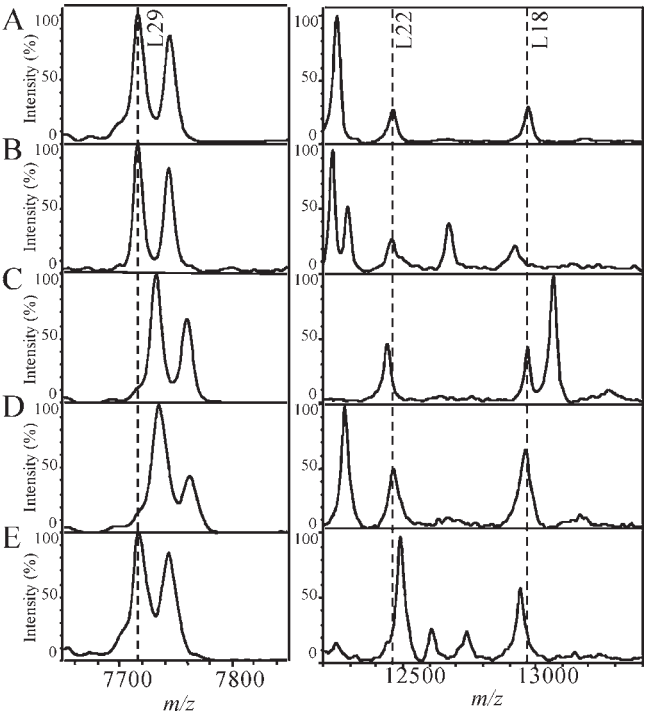

Figure 5. MALDI mass spectra of *B. subtilis* strains: (A) *B. subtilis* subsp. *subtilis* NBRC 13719^T; (B) *B. subtilis* subsp. *subtilis* NBRC 101246; (C) *B. subtilis* NBRC 13722; (D) *B. subtilis* subsp. *spizizenii* NBRC 101239^T; (E) *B. subtilis* NBRC 104440.

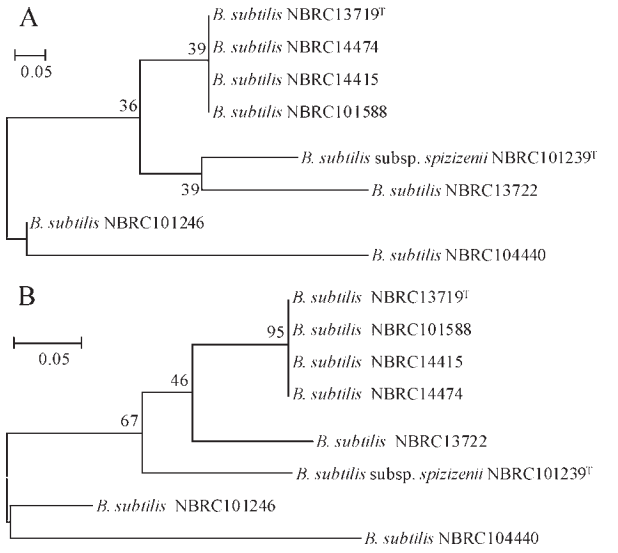

Figure 6. Phylogenetic trees of *B. subtilis* strains by the NJ method: (A) phylogenetic tree based on 8 ribosomal protein profile matching using ribosomal protein of *B. subtilis* subsp. *subtilis* NBRC 13719^T as a reference strain; (B) phylogenetic tree based on 20 ribosomal protein profile matching using ribosomal protein of *B. subtilis* subsp. *subtilis* NBRC 13719^T as a reference strain.

MALDI-TOF MS will develop as an advanced method based on theoretical biomarkers using a constructed ribosomal subunit protein database of type strains.

■ ASSOCIATED CONTENT

Supporting Information. Figure S1 and Table S1. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +81-52-838-2446. Fax: +81-52-833-5524. E-mail: n0861503@ccalumni.meijo-u.ac.jp.

Funding Sources

This study was financially supported by the Agriomics research project of the Ministry of Education, Culture, Sports, Science and Technology (H.T.), and Meijo University (H.T.), the Japan Science and Technology Agency (H.T.), the Steel Industry Foundation for the advancement of environmental protection technology (H.T.), and Research Fellowships from the Japan Society for the Promotion of Science for Young Scientists (Y.H.).

■ REFERENCES

(1) Fenselau, C.; Demirev, P. A. Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom. Rev.* **2001**, *20*, 157–171.

(2) Lay, J. O., Jr. MALDI-TOF mass spectrometry of bacteria. *Mass Spectrom. Rev.* **2001**, *20*, 172–194.

(3) Mellmann, A.; Cloud, J.; Maier, T.; Keckevoet, U.; Ramminger, I.; Iwen, P.; Dunn, J.; Hall, G.; Wilson, D.; Lasala, P.; Kostrzewa, M.; Harmsen, D. A. Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria. *J. Clin. Microbiol.* **2008**, *46*, 1946–1954.

(4) Welham, K. J.; Domin, M. A.; Scannell, D. E.; Cohen, E.; Ashton, D. S. The characterization of micro-organisms by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 176–180.

(5) Arnold, R. J.; Reilly, J. P. Fingerprint matching of *E. coli* strains with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of whole cells using a modified correlation approach. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 630–636.

(6) Jarman, K. H.; Daly, D. S.; Petersen, C. E.; Saenz, A. J.; Valentine, N. B.; Wahl, K. L. Extracting and visualizing matrix-assisted laser desorption/ionization time-of-flight mass spectral fingerprints. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1586–1594.

(7) Fagerquist, C. K.; Miller, W. G.; Harden, L. A.; Bates, A. H.; Vensel, W. H.; Wang, G.; Mandrell, R. E. Genomic and proteomic identification of a DNA-binding protein used in the "fingerprinting" of campylobacter species and strains by MALDI-TOF-MS protein biomarker analysis. *Anal. Chem.* **2005**, *77*, 4897–4907.

(8) Erhard, M.; Hipler, U. C.; Burmester, A.; Brakhage, A. A.; Wostemeyer, J. Identification of dermatophyte species causing onychomycosis and tinea pedis by MALDI-TOF mass spectrometry. *Exp. Dermatol.* **2008**, *17*, 356–361.

(9) Cherkaoui, A.; Hibbs, J.; Emonet, S.; Tangomo, M.; Girard, M.; Francois, P.; Schrenzel, J. Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. *J. Clin. Microbiol.* **2010**, *48*, 1169–1175.

(10) Demirev, P. A.; Ho, Y.-P.; Ryzhov, V.; Fenselau, C. Microorganism identification by mass spectrometry and protein database searches. *Anal. Chem.* **1999**, *71*, 2732–2738.

(11) Pineda, F. J.; Antoine, M. D.; Demirev, P. A.; Feldman, A. B.; Jackman, J.; Longenecker, M.; Lin, J. S. Microorganism identification by matrix-assisted laser/desorption ionization mass spectrometry and model-derived ribosomal protein biomarkers. *Anal. Chem.* **2003**, *75*, 3817–3822.

(12) Fagerquist, C. K.; Garbus, B. R.; Miller, W. G.; Williams, K. E.; Yee, E.; Bates, A. H.; Boyle, S.; Harden, L. A.; Cooley, M. B.; Mandrell, R. E. Rapid identification of protein biomarkers of *Escherichia coli* O157:H7 by matrix-assisted laser desorption ionization-time-of-flight−time-of-flight mass spectrometry and top-down proteomics. *Anal. Chem.* **2010**, *82*, 2717–2725.

(13) Teramoto, K.; Sato, H.; Sun, L.; Torimura, M.; Tao, H.; Yoshikawa, H.; Hotta, Y.; Hosoda, A.; Tamura, H. Phylogenetic classification of *Pseudomonas putida* strains by MALDI-MS using ribosomal subunit proteins as biomarkers. *Anal. Chem.* **2007**, *79*, 8712–8719.

(14) Hotta, Y.; Teramoto, K.; Sato, H.; Yoshikawa, H.; Hosoda, A.; Tamura, H. Classification of genus *Pseudomonas* by MALDI-TOF MS based on ribosomal protein coding in *S10-spc-alpha* operon at strain level. *J. Proteome Res.* **2010**, *9*, 6722–6728.

(15) Yamada, S.; Ohashi, E.; Agata, N.; Venkateswaran, K. Cloning and nucleotide sequence analysis of *gyrB* of *Bacillus cereus*, *B. thuringiensis*, *B. mycoides*, and *B. anthracis* and their application to the detection of *B. cereus* in rice. *Appl. Environ. Microbiol.* **1999**, *65*, 1483–1490.

(16) Schnepf, E.; Crickmore, N.; Van Rie, J.; Lereclus, D.; Baum, J.; Feitelson, J.; Zeigler, D. R.; Dean, D. H. *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 775–806.

(17) Meerak, J.; Yukphan, P.; Miyashita, M.; Sato, H.; Nakagawa, Y.; Tahara, Y. Phylogeny of γ-polyglutamic acid-producing *Bacillus* strains isolated from a fermented locust bean product manufactured in West Africa. *J. Gen. Appl. Microbiol.* **2008**, *54*, 159–166.

(18) Thompson, J. M.; Dodd, C. E. R.; Waites, W. M. Spoilage of bread by *Bacillus*. *Int. Biodeterior. Biodegrad.* **1993**, *32*, 55–66.

(19) Sun, L.; Teramoto, K.; Sato, H.; Torimura, M.; Tao, H.; Shintani, T. Characterization of ribosomal proteins as biomarkers for matrix-assisted laser desorption/ionization mass spectral identification of *Lactobacillus plantarum*. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3789–3798.

(20) Teramoto, K.; Kitagawa, W.; Sato, H.; Torimura, M.; Tamura, T.; Tao, H. Phylogenetic analysis of *Rhodococcus erythropolis* based on the variation of ribosomal proteins as observed by matrix-assisted laser desorption ionization−mass spectrometry without using genome information. *J. Biosci. Bioeng.* **2009**, *108*, 348–353.

(21) Sherman, F.; Stewart, J. W.; Tsunasawa, S. Methionine or not methionine at the beginning of a protein. *Bioessays* **1985**, *3*, 27–31.

(22) Tompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **1997**, *25*, 4876–4882.

(23) Kumar, S.; Tamura, K.; Nei, M. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings Bioinf.* **2004**, *5*, 150–163.

(24) Lauber, M. A.; Running, W. E.; Reilly, J. P. *B. subtilis* ribosomal proteins: structural homology and post-translational modifications. *J. Proteome Res.* **2009**, *8*, 4193–4206.

(25) Fritze, D. Taxonomy of the genus bacillus and related genera: the aerobic endospore-forming bacteria. *Phytopathology* **2004**, *94*, 1245–1248.

(26) Noguchi, H.; Uchino, M.; Shida, O.; Takano, K.; Nakamura, L. K.; Komagata, K. *Bacillus vietnamensis* sp. nov., a moderately halotolerant, aerobic, endospore-forming bacterium isolated from Vietnamese fish sauce. *Int. J. Syst. Evol. Microbiol.* **2004**, *54*, 2117–2120.

(27) Cerritos, R.; Vinuesa, P.; Eguiarte, L. E.; Herrera-Estrella, L.; Alcaraz-Peraza, L. D.; Arvizu-Gómez, J. L.; Olmedo, G.; Ramirez, E.; Siefert, J. L.; Souza, V. *Bacillus coahuilensis* sp. nov., a moderately halophilic species from a desiccation lagoon in the Cuatro Ciénegas Valley in Coahuila, Mexico. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 919–923.

(28) Nakamura, L. K.; Roberts, M. S.; Cohan, F. M. Relationship of *Bacillus subtilis* clades associated with strains 168 and W23: a proposal for *Bacillus subtilis* subsp. *subtilis* subsp. nov, and *Bacillus subtilis* subsp. *spizizenii* subsp. nov. *Int. J. Syst. Bacteriol.* **1999**, *49*, 1211–1215.

(29) Rooney, A. P.; Price, N. P.; Ehrhardt, C.; Swezey, J. L.; Bannan, J. D. Phylogeny and molecular taxonomy of the *Bacillus subtilis* species complex and description of *Bacillus subtilis* subsp. *inaquosorum* subsp. nov. *Int. J. Syst. Evol. Microbiol.* **2009**, *59*, 2429–2436.

(30) Wang, L. T.; Lee, F. L.; Tai, C. J.; Kuo, H. P. *Bacillus velezensis* is a later heterotypic synonym of *Bacillus amyloliquefaciens*. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 671–675.