Classification of Genus Pseudomonas by MALDI-TOF MS Based on Ribosomal Protein Coding in S10-spc-alpha Operon at Strain Level

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · OCTOBER 2010

Impact Factor: 4.25 · DOI: 10.1021/pr100868d · Source: PubMed

6 AUTHORS, INCLUDING:

Hiroaki Sato
National Institute of Advanced Industrial S…
71 PUBLICATIONS **843** CITATIONS
SEE PROFILE

Akifumi Hosoda
Meijo University
25 PUBLICATIONS **254** CITATIONS
SEE PROFILE

research articles Journal of proteome research

Classification of Genus *Pseudomonas* by MALDI-TOF MS Based on Ribosomal Protein Coding in *S10−spc−alpha* Operon at Strain Level

Yudai Hotta,*,† Kanae Teramoto,‡,§ Hiroaki Sato,‡ Hiromichi Yoshikawa,|| 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, Advanced Technology Division, JEOL Ltd., Tokyo 196-8558, Japan, and Department of Life, Environmental and Material Science, FIT, 3-30-1 Wajirohigashi, Higashi-ku, Fukuoka 811-0295, Japan

Received August 25, 2010

We have proposed a rapid phylogenetic classification at the strain level by MALDI-TOF MS using ribosomal protein matching profiling. In this study, the *S10−spc−alpha* operon, encoding half of the ribosomal subunit proteins and highly conserved in eubacterial genomes, was selected for construction of the ribosomal protein database as biomarkers for bacterial identification by MALDI-TOF MS analysis to establish a more reliable phylogenetic classification. Our method revealed that the 14 reliable and reproducible ribosomal subunit proteins with less than *m*/*z* 15 000, except for L14, coded in the *S10−spc−alpha* operon were significantly useful biomarkers for bacterial classification at species and strain levels by MALDI-TOF MS analysis of genus *Pseudomonas* strains. The obtained phylogenetic tree was consisted with that based on genetic sequence (*gyrB*). Since *S10−spc−alpha* operons of genus *Pseudomonas* strains were sequenced using specific primers designed based on nucleotide sequences of genome-sequenced strains, the ribosomal subunit proteins encoded in *S10−spc−alpha* operon were suitable biomarkers for construction and correction of the database. MALDI-TOF MS analysis using these 14 selected ribosomal proteins is a rapid, efficient, and versatile bacterial identification method with the validation procedure for the obtained results.

Keywords: ribosomal protein • MALDI-TOF MS • bacterial identification • *S10-spc-alpha* operon

Introduction

The classification and identification of bacterial strains is significant for the fields of clinical and environmental microbiology, the food and safety industry, molecular epidemiology and counterterrorism. Especially, bacterial identification at the species and strains levels in complex matrices is important for risk assessment. Lifecycle assessment of man-made chemicals in the environment is essential to establish an environmentally conscious material cycle society. Since bacteria play an important role in the biodegradation of man-made chemicals in their lifecycle assessment as well as in the field of medical research, it is important to establish a rapid and simple identification method for bacteria. MALDI-TOF MS analysis of whole cells, cell lysates, or crude bacterial extracts has been developed and has successfully determined the species within a few minutes without any substantial costs for consumables.[1−4] This approach was applied for aerobic and anaerobic bacteria, and fungi.[5−12] Recently, species identification by MALDI-TOF MS has been carried out with the Biotyper software program (Bruker Daltonics) or SARAMIS database application (AnagnosTec GmbH) by comparing the mass spectrum of each strain with the mass spectra of the many reference strains.[13−15] This method demonstrated the reproducibility even if the culture conditions were not identical, although it uses uncertain biomarker peaks. However, it is important to develop a bacterial identification method based on theoretical biomarker with a validation procedure for the obtained results.

To avoid uncertain biomarker peaks, a bioinformatics-based approach has been proposed using ribosomal subunit proteins as biomarkers for rapid identification of bacteria.[1,16−19] Ribosomal subunit proteins have a highly conserved amino acid sequence with a slight difference between bacteria and archaea, as the 16S rRNA sequence is commonly used for taxonomic classification.[20−22] Most of the ribosomal subunit proteins could be observed by MALDI-TOF MS analysis of the cell lysate sample by following their characteristics: significant amount of protein expression, appropriate molecular weight range, higher p*I*, moderate sequence conservation, fewer post-translational modifications, except for N-terminal methionine loss, and database accessibility.[19,23−25]

The differences between the calculated and observed masses are caused by post-translational modifications and errors in

* To whom correspondence should be addressed. Phone, (+81)52-838-2446; fax, +81-52-833-5524; e-mail, n0861503@ccalumni.meijo-u.ac.jp.
† Meijo University.
‡ Research Institute for Environmental Management Technology.
§ Advanced Technology Division, JEOL Ltd.
|| Department of Life, Environmental and Material Science.

10.1021/pr100868d

Published on Web 10/14/2010

amino acid sequences in the databases.[24−27] Since post-translational modifications for ribosomal subunit proteins have been confirmed[26−29] and few sequence errors in ribosomal subunit proteins have been identified by mass spectral characterization of whole ribosomal subunit proteins, where the errors were due to misannotation of N-termini, bacterial identification using ribosomal subunit proteins has been developed as a reliable bacterial identification technique.[23−25]

Although the presence of ribosomal protein databases and the characteristics of ribosomal proteins allowed a robust identification method of bacteria coupled with MALDI-TOF MS measurements and a protein database search, the development of this approach depended on the advancement of complete microbial genome sequencing projects. According to an announcement by the Integrated Microbial Genomes (IMG) system, 969 strains of bacteria had been completely sequenced by June 1, 2010, and it is certain that the numbers of sequenced bacteria will rapidly increase. However, only 17 strains of genus *Pseudomonas* have been genome-sequenced: 4 strains of *P. aeruginosa*, 4 strains of *P. putida*, 3 strains of *P. syringae*, 3 strains of *P. fluorescens*, 1 strain of *P. entomophila*, 1 strain of *P. mendocina* and 1 strain of *P. stutzeri*. Therefore, the database shortage of ribosomal proteins prevented advanced discrimination at strain level, although a rapid identification method of bacteria by MALDI-TOF MS based on a bioinformatics-based approach using ribosomal subunit proteins as biomarkers could be developed for the highly reliable bacterial classification at strain level; therefore, it is essential to establish database construction method of ribosomal subunit proteins.

The Gram-negative, rod-shaped saprotrophic soil bacteria *Pseudomonas* used in this study are ubiquitous bacteria in nature and possess variable metabolic abilities to utilize a wide range of organic compounds. Our studies showed that *P. putida* strains which can degrade alkylphenol polyethoxylates to estrogenic and antiandrogenic metabolites exist fairly universally in paddy fields in Japan.[30,31] Therefore, genus *Pseudomonas* occupies an important ecological position in the lifecycle assessment of the chemicals. Furthermore, some species or strains of genus *Pseudomonas* are also important pathogens of animals and plants.

Although our previous study demonstrated the bacterial identification method at strain level based on ribosomal protein matching profiling by comparison between the observed masses in the MALDI mass spectra of sample strains and the biomarker masses of the reference strain,[25] there are some disadvantages in making phylogenetic trees as follows: there are some errors in the ribosomal protein database and there is no ribosomal protein database of type strains which are important for the identification of bacterial species. Therefore, our effort focused on developing a method for phylogenetic analysis at species and strain levels with theoretical evidence.

In this study, our research suggested a rapid bacterial classification method by MALDI-TOF MS using ribosomal subunit proteins coded in *S10−spc−alpha* operon as biomarkers, because the operon encodes half of the ribosomal subunit and is highly conserved in eubacterial genomes,[32−35] and the phylogenetic trees based on *S10−spc−alpha* operon reflects the phylogenetic tree based on 16S rRNA sequence.[36] Our proposed method will be developed as a highly reliable advanced method for phylogenetic analysis at strain level with theoretical biomarkers using ribosomal subunit proteins encoded in the *S10−spc−alpha* operon with a validation procedure.

Experimental Section

Bacterial Strains. To compare the results of phylogenetic classification based on *gyrB* and ribosomal subunit proteins, the 21 strains of genus *Pseudomonas* were selected as follows: *P. putida* NBRC 100650 (=KT2440), *P. putida* NBRC 14164^T, *P. putida* NBRC 3738, *P. putida* NBRC 14671, *P. putida* NBRC 15366. *P. putida* NBRC 100988, *P. putida* NBRC 101019, *P. putida* JCM 6156, *P. putida* JCM 13061, *P. putida* ATCC 700007 (=F1), *P. fluorescens* NBRC 14160^T, *P. fluorescens* NBRC 15831, *P. fluorescens* ATCC BAA-477 (=Pf-5), *P. alcaligenes* NBRC 14159^T, *P. aeruginosa* NBRC 12689^T, *P. azotoformans* NBRC 12693^T, *P. chlororaphis* NBRC 3904^T, *P. fulva* NBRC 16637^T, *P. mendocina* NBRC 14162^T, *P. straminea* NBRC 16665^T, and *P. stutzeri* NBRC 14165^T. Since *P. putida* NBRC 100650, *P. putida* ATCC 700007, and *P. fluorescens* ATCC BAA-477 were originally the same strain as the genome-sequenced KT2440, F1, and Pf-5 strains, respectively, the strain names of KT2440, F1, and Pf-5 are used in this paper. The NBRC, JCM, and ATCC strains were purchased from the National Institute of Technology and Evaluation (NITE)-Biological Resource Center (NBRC, Kisarazu, Japan), the RIKEN BRC (JCM, Wako, Japan) through the National Bio-Resource Project of MEXT, Japan, and the American Type Culture Collection (ATCC, Rockville, MD), respectively. Each bacterial strain was grown aerobically in the recommended medium and temperature by suppliers.

***S10−spc−alpha* Operon Sequencing.** Chromosomal DNA was extracted from the bacteria as described previously.[37] 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−spc−alpha* operon was performed using KOD containing dNTP at a concentration of 200 μM, each of the primers at a concentration of 400 μM, 100 ng template DNA, and 2.5 U KOD polymerase (TOYOBO) in a total volume of 50 μL. PCR amplification conditions of *S10*, *spc*, and *alpha* operons were as follows: (1) 2 min at 95 °C; (2) 30 cycles of 30 s at 95 °C, 30 s at 57 °C, and 5 min at 72 °C; (3) 5 min at 72 °C. PCR primers and sequencing primers used in this study were designed based on consensus nucleotide sequences of *S10−spc−alpha* operon from 17 genome-sequenced strains with the Clustal X program for alignment of nucleotide sequences (Table 1). The sequencing reaction was carried out using a BigDye ver.3.1 Cycle Sequencing Kit (Applied Biosystems) and an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems) according to the manufacturer's instructions. The Basic Local Alignment Search Tool (BLAST) (http://www.ncbi.nlm.nih.gov/blast/) program was used for homology analysis. Ribosomal subunit proteins of genus *Pseudomonas* had accession numbers from AB560216 to AB560485, respectively, in the DDBJ/EMBL/GenBank.

Bacterial Identification by MALDI-TOF MS. Ribosomal subunit proteins analysis by MALDI-TOF MS was performed under almost the same conditions as described in our previous study.[25] 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). Each sample solution of whole cells adjusted to $OD_{660\ nm} = 1.0$ was mixed with a sinapinic acid matrix solution at a concentration of 10 mg/mL in 50% acetonitrile with 1% trifluoroacetic acid. About 1.5 μL of sample/matrix mixture (1/5 vol %) was spotted onto the MALDI target and dried in air. MALDI mass

Table 1. PCR and Sequencing Primers Used in This Study[a]

primer	sequence (5′-3′)	primer	sequence (5′-3′)
−220f	TTCTTCAARGGCTACCGTCC	6010r	GCGGAACTTCCATCACGTT
119f	CTTTTARCGGGCGTAKTSCG	6490f	CATGAAAAACCGYGAGCTGA
615f	TTCACCGAAGAAGGTGTCTC	6510r	TCAGCTCRCGGTTTTTCATG
635r	GAGACACCTTCTTCGGTGAA	7400f	TGTGTCGCGTCGCTAAGAAC
1505f	AACAAGAAGATGTAYCGCGC	7420r	TCTTAGCGACGCGACACATG
1525r	GCGCGRTACATCTTCTTGTT	8290f	GGGCTGGAGTTCTAAGTTATG
2310f	AAGCAGCATTACCGTCTGGT	8310r	CATAACTTAGAACTCCAGCCC
2330r	ACCAGACGGTAATGCTGCTT	8955f	CAACAAGGCTTACTACATGC
3265f	GAAGTAGCCGCTAAGTTGTC	8975r	GCATGTAGTAAGCCTTGTTG
3285r	GACAACTTAGCGGCTACTTC	9430f	AATGGCTAAGCAAGGTGCTC
4105f	TGGTATCGCGAAGGTCGTGT	9450r	GAGAGAGCACCTTGCTTAGC
4125r	ACACGACCTTCGCGATACCA	9920f	GCGATGTTCATGATGTGGC
4710f	GACGGTGATGTGATGAAAGC	9940r	AGCCACATCATGAACATCGC
4730r	GCTTTCATCACATCACCGTC	10710f	CCCTGATGAAGAAAGCCAAC
5205f	TCAGACTCAATCCATGCTCG	11390f	AAGCCGATCCGCAAGTAATC
5225r	TCGAGCATGGATTGAGTCTG	11495r	GACGCCATCAACCACTGTCT
5990f	AACGTGATGGAAGTTCCGC	12510r	ACTCATTTACCGAAMTCTGC

[a] Primers were designed from the conserved regions of the nucleotide sequences of *S10−spc−alpn* operon of 17 genome-sequenced genus *Pseudomonas*. R = A or G; Y = C or T; K = G or T; M = A or C; S = C or G; W = A or T.

spectra in the range of m/z 4000−25 000 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 *P. putida* KT2440 was carried out by internal calibration using two peaks of myoglobin ($[M + H]^+$, m/z 16952.6, and $[M + H]^{2+}$, m/z 8476.8), 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. Mass calibration of the other samples was performed by external calibration using the mass spectra observed for the whole cell of the KT2440 strain. The peak matching of the biomarker proteins was judged from errors within 150 ppm. The observed masses in the MALDI mass spectra of sample strains were compared with the biomarker masses of genome-sequenced *P. putida* KT2440 as the reference strain for the ribosomal subunit protein profile matching as described in our previous study.[25] The presence or absence of the reference masses 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 *Pseudomonas* strains are obtained from the NCBInr database (http://www.ncbi.nlm.nih.gov/). The calculated mass of each subunit protein was predicted using a Compute pI/Mw tool on the ExPASy proteomics server (http://www.expasy.org/tools/pi_tool.html). N-terminal methionine loss was first considered based on the "N-end rule" as a post-translational modification.[38,39] In this rule, N-terminal methionine is cleaved from specific penultimate amino acid residues such as glycine, alanine, serine, proline, valine, threonine, and cysteine. Since N-terminal methionine loss of the ribosomal subunit S16, whose penultimate amino acid residue is valine, was not observed in our previous study, it was not considered in this study. Other post-translational modifications were further considered referring to previous reports on ribosomal proteins of Gram-negative bacteria: two methylation and oxidation of L16 in *S10* operon and methylation of S11 in *alpha* operon.[26−29] In a previous study, 43 ribosomal subunits of *P. putida* KT2440 considering these post-translational modifications were indicated as the reliable biomarkers. Since MALDI-TOF MS analysis indicated that the same post-translational modification was confirmed on ribosomal subunit proteins of genus *Pseudomonas*, the database considered the post-translational modification constructed in this study.

Cluster Analysis. Phylogenetic analysis was constructed with the Clustal X program.[40] Phylogenetic tree construction and bootstrap analyses (100 replicates) were performed using the Mega3.1 program.[41] Phylogenetic trees were constructed using three different methods, the neighbor-joining, maximum-likelihood, and maximum parsimony algorithms available in the Mega3.1 program.

Results and Discussion

Our strategy of MALDI-TOF MS analysis is as follows: (1) construction of the ribosomal protein database based on *S10−spc−alpha* operon sequencing of genus *Pseudomonas* as model bacteria, (2) confirmation of the assessment of the reliability of our proposed method by comparison with the constructed database and MALDI mass spectra, (3) establishment of phylogenetic analysis based on ribosomal subunit proteins encoded in *S10−spc−alpha* operon.

Ribosomal Protein Database Construction of Genome-Sequenced Genus *Pseudomonas*. To advance a rapid identification method of bacteria by MALDI-TOF MS based on ribosomal subunit proteins as biomarkers, the ribosomal protein database of genome-sequenced genus *Pseudomonas* was constructed. In the process of database construction of ribosomal subunit proteins, at least 17 ribosomal subunit proteins which were not registered or were missing annotations of the start codon were confirmed in the ribosomal protein database of 17 genome-sequenced genus *Pseudomonas* strains. MALDI mass spectra of genome-sequenced *P. putida* F1 and *P. fluorescens* Pf-5 revealed the misannotation of the start codon on L36 subunit protein and the presence of an unregistered L34 subunit protein in *P. putida* F1, and the misannotation of the start codon on L22 subunit protein and the presence of unregistered L36 subunit protein in *P. fluorescens* Pf-5, respectively (data not shown). The unregistered ribosomal subunit proteins were confirmed based on nucleotide sequences of genome-sequenced strains. Therefore, some errors in the

Table 2. Theoretical Masses of Ribosomal Subunit Proteins in *S10−spc−alpha* Operon in Type Strains of Genus *Pseudomonas* ($\leq m/z$ 15 000)[a]

	P. putida 14164	*P. fluorescens* 14160	*P. alcaligenes* 14159	*P. aeruginosa* 12689	*P. azotoformans* 12693	*P. chlororaphis* 3904	*P. fulva* 16637	*P. mendocina* 14162	*P. straminea* 16665	*P. stutzeri* 14165
***S10* operon**										
L22	**11912.0**	**11912.0**	11893.9	**11912.0**	**11912.0**	**11912.0**	**11912.0**	**11912.0**	**11912.0**	11897.9
L23	*10900.7*	**10945.7**	10955.6	10950.7	**10945.7**	**10945.7**	*10900.7*	11015.7	11085.8	10920.6
L29	**7173.3**	**7173.3**	**7215.4**	7202.4	**7173.3**	**7173.3**	**7173.3**	7205.4	7215.4	7274.4
S10	**11753.6**	**11753.6**	*11783.6*	11767.6	**11753.6**	**11753.6**	**11753.6**	*11783.6*	11755.6	**11753.6**
S17	**9902.5**	*9966.6*	9957.6	9955.6	*9966.6*	9984.6	**9902.5**	9974.6	10014.7	9973.6
S19	**10218.1**	10246.1	10186.0	10227.1	10189.1	10204.0	**10218.1**	10176.0	10190.0	10163.0
***spc* operon**										
L14	**13410.9**	**13410.9**	*13396.8*	13412.9	**13410.9**	**13410.9**	**13410.9**	*13396.8*	**13410.9**	13436.9
L18	12497.4	12556.4	12561.4	12531.4	**12512.4**	**12512.4**	12485.4	12413.3	12457.3	12477.3
L24	**11330.2**	*11336.3*	11340.3	11471.5	*11336.3*	11345.3	**11330.2**	11344.3	11344.3	11413.4
L30	6334.5	**6395.6**	6278.3	6347.4	**6395.6**	**6395.6**	6292.5	6363.5	6448.5	6463.6
L36	**4435.4**	**4435.4**	4407.3	**4435.4**	**4435.4**	**4435.4**	**4435.4**	**4435.4**	**4435.4**	4421.4
S08	13845.1	13962.3	13951.3	14040.4	13920.2	13973.3	13861.1	13928.2	13914.2	13869.1
S14	11259.3	**11304.3**	11394.3	11435.3	**11304.3**	11274.2	11288.3	11359.2	11385.3	11326.2
***alpha* operon**										
S11	13529.5	13485.4	13517.5	**13513.5**	13499.4	**13513.5**	13543.5	13531.5	13493.4	13527.5
S13	13126.3	13210.4	13177.4	13135.2	13164.5	13239.4	13140.3	13118.3	13058.3	13176.4

[a] Bold, italic and underline: common theoretical masses among subunit proteins.

ribosomal protein database were easily and theoretically corrected by comparison between the observed masses by MALDI mass spectra and theoretical masses of the database based on *S10−spc−alpha* operon.

Selection of Biomarker Protein Coded in *S10−spc−alpha* Operon by Using Type Strains. Although type strains are important for the identification of bacterial species, there is no ribosomal protein database of type strains. To verify the utilization of *S10−spc−alpha* operon, the ribosomal protein database of type strains of genus *Pseudomonas* was constructed by *S10−spc−alpha* operon sequencing. To sequence *S10−spc−alpha* operon of 10 type strains of genus *Pseudomonas* strains, specific primers were designed based on consensus nucleotide sequences of *S10−spc−alpha* operon from 17 genome-sequenced strains (Table 1). The fact that these designed primers were demonstrated as useful sequencing primers for *S10−spc−alpha* operon of genus *Pseudomonas* demonstrated that ribosomal subunit proteins coded in *S10−spc−alpha* operon were suitable biomarkers for database construction (Table 2).

Since it is difficult to carry out accurate phylogenetic analysis using the bacterial identification method with MALDI mass spectral fingerprints, phylogenetic analysis based on the amino acid sequences of ribosomal subunit proteins coded in *S10−spc−alpha* operon was applied for accurate phylogenetic distance at the species and strain levels by our proposed method in this study.

The results of comparing the observed masses by MALDI-TOF MS analysis and theoretical masses based on nucleotide sequences revealed that 14 ribosomal subunit proteins from 26 ribosomal subunit proteins coded in the *S10−spc−alpha* operon, whose masses are less than m/z 15 000, except for L14, could be selected as reliable biomarkers for the bacterial identification method using MALDI-TOF MS because mass spectra of 14 ribosomal subunit proteins of 10 type strains of genus *Pseudomonas* had the reproducibility and the sensitivity of masses based on S/N level (Figure 1, Table 2), while L14 was eliminated because of low sensitivity and reproducibility of MALDI-TOF MS analysis.

Phylogenetic analysis based on amino acid sequence of 26 and 14 selected ribosomal protein was compared with that

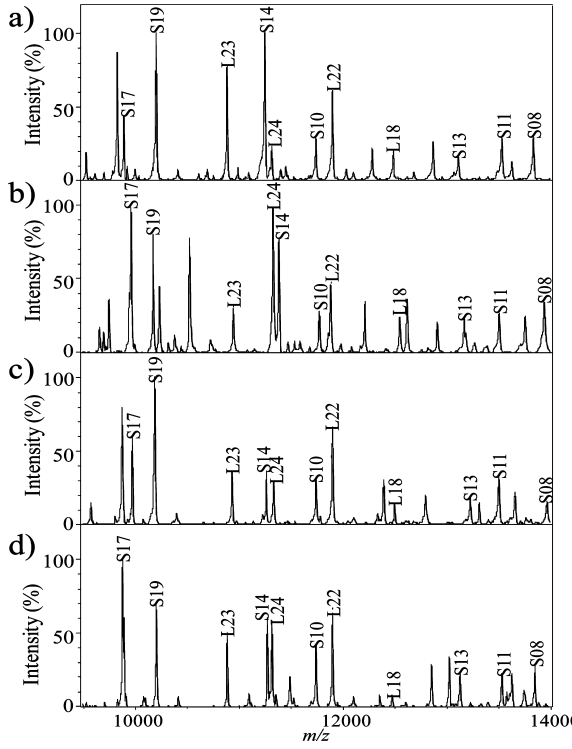

Figure 1. Representative MALDI mass spectra of type strains of genus *Pseudomonas*. (a) *P. putida* NBRC 14164^T, (b) *P. alcaligenes* NBRC 14159^T, (c) *P. chlororaphis* NBRC 3904^T, (d) *P. fulva* NBRC 16637^T.

based on *gyrB* sequences instead of 16S rRNA gene sequences because results of *gyrB* sequence analysis were correlated very well with DNA−DNA reassociation values determined by DNA hybridization experiments.[42] Phylogenetic trees of 10 type strains of genus *Pseudomonas* based on *gyrB* sequences, and on amino acid sequences of 26 and 14 selected ribosomal subunit proteins, were constructed using individual data sets (Figure 2). Since the basic topologies of phylogenetic trees based on 26 and 14 selected ribosomal subunit proteins were mostly identical, 14 ribosomal subunit proteins were sufficient for bacterial identification by MALDI-TOF MS analysis. The

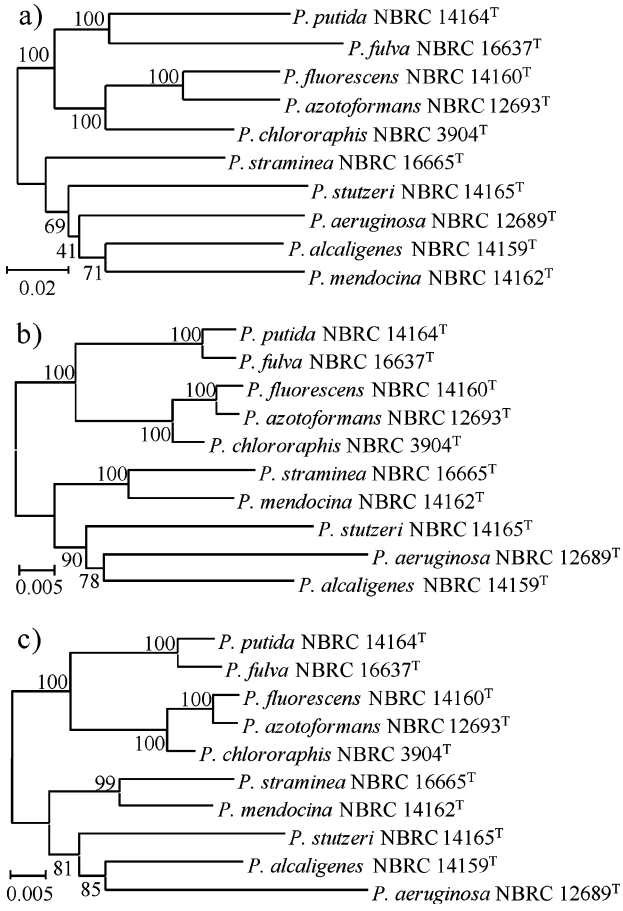

Figure 2. Phylogenetic tree of 10 type strain of genus *Pseudomonas*. (a) Phylogenetic tree based on *gyrB* sequences (872 bp). (b) Phylogenetic tree based on amino acid sequences of 26 ribosomal proteins in *S10−spc−alpha* operon. (c) Phylogenetic tree based on amino acid sequences of 14 selected ribosomal proteins in *S10−spc−alpha* operon.

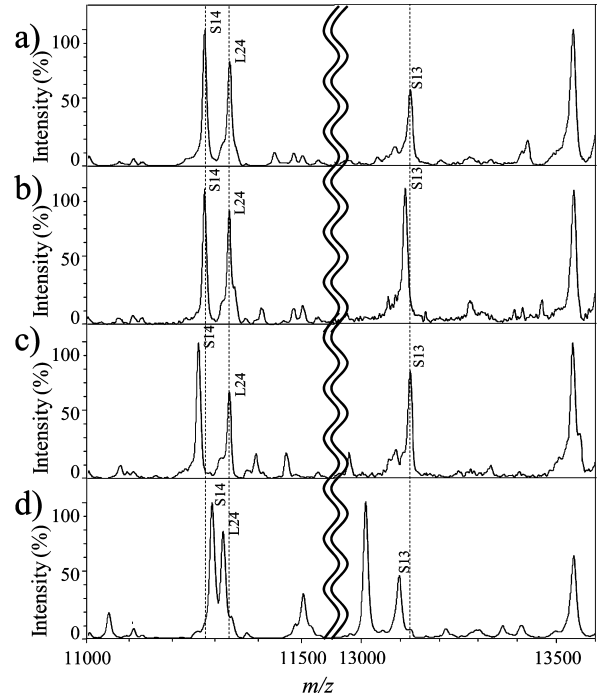

Figure 3. Typical MALDI mass spectra of whole cell of *P. putida*. (a) *P. putida* JCM6165, (b) *P. putida* NBRC 101019, (c) *P. putida* JCM13061, (d) *P. putida* NBRC 15366.

basic topologies of phylogenetic trees based on *gyrB* sequences and amino acid sequences of 14 selected ribosomal subunit proteins were similar, except for the shift of the phylogenetic position of *P. mendocina* NBRC 14162T.

Yamamoto proposed the major clusters and subclusters of genus *Pseudomonas* based on combined *gyrB* and *rpoD* phylogenetic analysis.[43] Phylogenetic analysis of the 10 type strains of genus *Pseudomonas* based on amino acid sequences of 14 selected ribosomal subunit proteins revealed that *P. aeruginosa* NBRC 12689T and *P. alcaligenes* NBRC 14159T formed '*P. aeruginosa* complex' and *P. stutzeri* NBRC 14165T which concluded in '*P. stutzeri* complex' neighboring '*P. aeruginosa* complex' (Figure 3). Although *P. straminea* NBRC 16665T, which was included in the suggested independent third subcluster in IGC I, and *P. mendocina* NBRC 14162T, which was included in '*P. aeruginosa* complex', formed another group in the phylogenetic tree based on the amino acid sequences of 14 selected ribosomal subunit proteins, *P. mendocina* NBRC 14162T formed '*P. aeruginosa* complex' with *P. aeruginosa* NBRC 12689T and *P. alcaligenes* NBRC 14159T in the phylogenetic tree based on *gyrB* sequences. In the phylogenetic tree of genome-sequenced strains based on amino acid sequences of 14 selected ribosomal subunit proteins and 16S rRNA gene sequences, *P. mendocina* ymp was also not included in '*P. aeruginosa* complex' (data not shown). Since both ribosomal proteins and 16S rRNA are components of the bacterial ribosome, their phylogenetic analysis was based on closely involved rather than based on *gyrB* sequences. These results suggest that the reclassification of *P. mendocina* species was required and bacterial identification using MALDI-TOF MS may be an advanced method of phylogenetic analysis. *P. putida* NBRC 14164T and *P. fulva* NBRC 16637T belonged to subcluster '*P. putida* complex' in IGC II, *P. fluorescens* NBRC 14160T and *P.azotoformans* NBRC 12693T formed subpopulation '*P. fluorescens* lineage', which belonged to subcluster '*P. putida* complex' in IGC II, and the phylogenetic position of *P. chlororaphis* NBRC 3904T, belonged to '*P. chlororaphis* lineage', neighboring to '*P. fluorescens* lineage' in all phylogenetic trees (Figure 2). These results demonstrated that bacterial identification by MALDI-TOF MS using 14 selected ribosomal subunit proteins as biomarkers was a useful advanced method for phylogenetic classification at the species level.

Discrimination at the Strain Level of *P. putida*. In our previous study, 43 ribosomal subunit proteins were selected as reliable biomarkers for the high-throughput classification of *P. putida* strains at the strain level.[25] However, MALDI mass spectra of the whole cell have some ribosomal subunit proteins with low sensitivity of detection, and it is difficult to construct a database of ribosomal subunit proteins which are not present in *S10−spc−alpha* operon. In this study, since our proposed method using 14 selected ribosomal subunit proteins coded in *S10−spc−alpha* operon as biomarkers in MALDI-TOF MS analysis was developed as high reliable and reproducible bacterial classification at species level, phylogenetic classification of *P. putida* at the strain level was carried out. As the first step, a ribosomal protein database of 7 strains of *P. putida* was constructed by *S10−spc−alpha* operon sequencing and calculation of theoretical masses based on the amino acid sequences of each subunit protein. As the second step, these strains were analyzed by MALDI-TOF MS, and then their mass spectra were assigned to theoretical masses of 14 selected ribosomal proteins.

Although the mass differences of S3 and S5 among KT2440, JCM 6156, NBRC 100988, and NBRC 101019 were proposed as key biomarkers to distinguish the strains,[25] MALDI mass spectra of S3 and S5 were often misassigned because of their low sensitivity of detection. Our results revealed that the theoretical masses of S3 and S5 calculated based on *S10−spc−alpha* operon sequences of these 4 strains have identical masses, with *m*/*z* 25594.7 and *m*/*z* 17578.3, respectively. Therefore, the selection of 14 ribosomal subunit proteins was significantly important to prevent misidentification by MALDI-TOF MS analysis. In particular, four typical strains of *P. putida* revealed different masses of ribosomal subunit proteins, respectively: the masses of S14, 11259.3, 11273.3 and 11289.3, the masses of L24, 11315.2, and 11330.2, and the masses of S13, 13096.3, 13112.3, and 13126.3 (Figure 3).

The basic topologies of these phylogenetic trees of *P. putida* based on *gyrB* sequences (872 bp), amino acid sequences of 14 selected ribosomal subunit proteins in *S10−spc−alpha* operon, and 14 ribosomal subunit protein profile matching using ribosomal protein of *P. putida* KT2440 as a reference strain were similar, but were slightly different in their details (Figure 4).

The most conspicuous difference was the phylogenetic position between *P. putida* NBRC14671 and *P. putida* GB-1, and between *P. putida* NBRC 15366 and *P. putida* W619. In these results, phylogenetic trees based on the 14 selected ribosomal subunit proteins profile matching using ribosomal protein of *P. putida* KT2440 as a reference strain gave almost the same results as the phylogenetic trees based on *gyrB* sequences and on the amino acid sequences of 14 selected ribosomal subunit proteins. Moreover, the ribosomal protein database for bacterial identification doubled by sequencing the *S10−spc−alpha* operon of 18 strains including 10 type strains of genus *Pseudomonas* during this research since it is easy to construct an in-house database using *S10−spc−alpha* operon.

Taken together, the phylogenetic tree based on the 14 ribosomal subunit proteins profile matching using ribosomal protein of a reference strain are easy to construct using only the database of type strains and phylogenetic analysis based on amino acid sequences of 14 selected ribosomal subunit proteins in *S10−spc−alpha* operon will be applied for accurate phylogenetic distance at the species and strain levels with accumulation of advanced database of ribosomal subunit proteins coded in *S10−spc−alpha* operon.

Recently, the Biotyper software program or SARAMIS database application has been used for species identification by MALDI-TOF MS. Since those biomarker peaks are uncertain, it is hard to correct the observed results in the phylogenetic analysis by those methods. In contrast, our proposed method using 14 selected ribosomal subunit proteins as theoretical biomarkers in MALDI-TOF MS analysis was developed as a highly reliable and reproducible bacterial classification at the strain level capable of validating the obtained results based on theoretical masses.

Conclusions

Our proposed bacterial identification method is performed by three processes: (1) Bacteria are identified by MALDI-TOF MS spectral fingerprints based on the ribosomal protein database. (2) MALDI-TOF MS spectrum is converted to amino acid sequences based on the ribosomal protein database. (3) The conversion was verified by independently constructed database based on the results obtained by sequence analysis

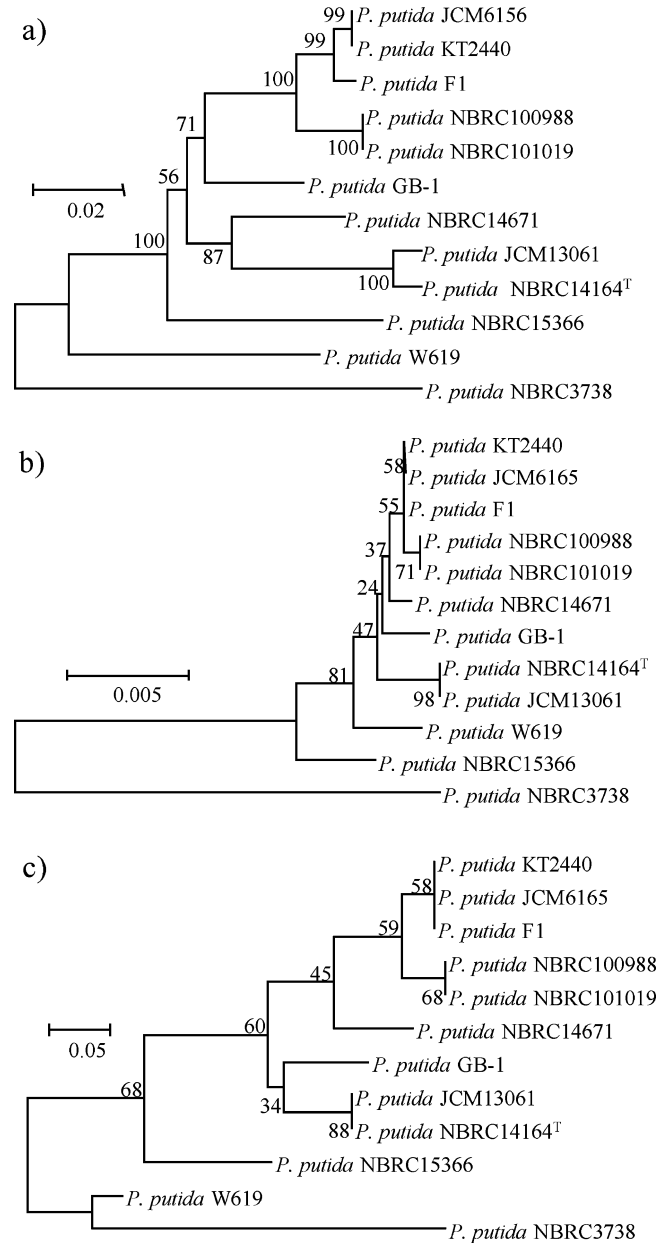

Figure 4. Phylogenetic trees of 12 *P. putida* strains. (a) Phylogenetic tree based on *gyrB* sequences (872 bp). (b) Phylogenetic tree based on amino acid sequences of 14 selected ribosomal proteins in *S10−spc−alpha* operon. (c) Phylogenetic tree based on 14 ribosomal protein profile matching using ribosomal protein of *P. putida* KT2440 as a reference strain.

of *S10−spc−alpha* operon. Bacteria are classified by phylogenetic analysis based on amino acid sequences assigned reliable and reproducible ribosomal subunit proteins encoded in *S10−spc−alpha* operon (Figure 5). Therefore, our proposed method could avoid the accidental coincidence of MALDI mass spectra. Moreover, analysts can construct their own database based on the sequences of *S10−spc−alpha* operon for this improved high-throughput method. This is the first report on detailed phylogenetic analysis at the strain level using MALDI-TOF MS with the validation procedure of the obtained results. Our proposed method has a merit to establish more reliable phylogenetic classification by application for various bacteria. In the future, the rapid identification of bacteria by MALDI-TOF MS will develop as an advanced

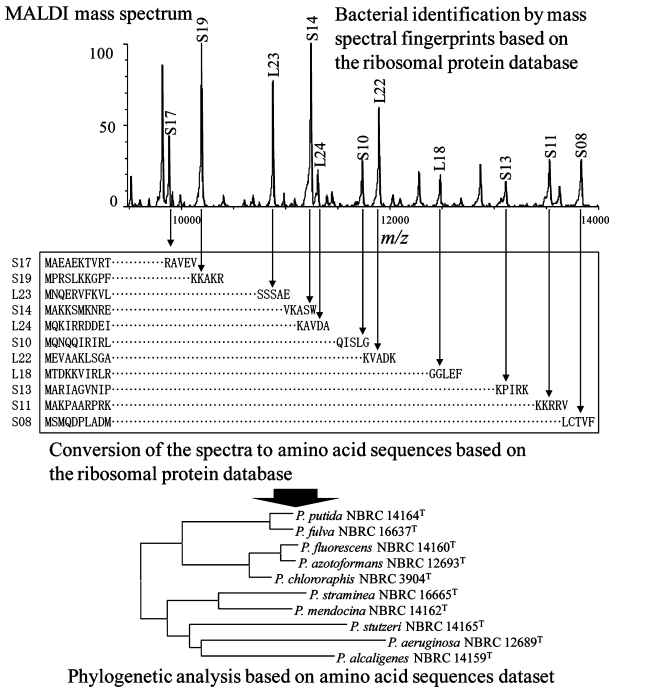

Figure 5. Procedure of phylogenetic analysis by the MALDI-TOF MS method using ribosomal subunit protein coded in *S10−spc−alpha* operon.

method of phylogenetic analysis based on theoretical biomarkers using an advanced database of ribosomal subunit proteins coded in *S10−spc−alpha* operon.

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

References

(1) Fenselau, C.; Demirev, P. A. *Mass Spectrom. Rev.* **2001**, *20*, 157–171.
(2) Holland, R. D.; Wilkes, J. G.; Rafii, F.; Sutherland, J. B.; Persons, C. C.; Voorhees, K. J.; Lay, J. O. Jr. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1227–1232.
(3) Krishnamurthy, T.; Ross, P. L.; Rajamani, U. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 883–888.
(4) 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. *J. Clin. Microbiol.* **2008**, *46*, 1946–1954.
(5) Stingu, C. S.; Rodloff, A. C.; Jentsch, H.; Schaumann, R.; Eschrich, K. *Oral Microbiol. Immunol.* **2008**, *23*, 372–376.
(6) Grosse-Herrenthey, A.; Maier, T.; Gessler, F.; Schaumann, R.; Böhnel, H.; Kostrzewa, M.; Krüger, M. *Anaerobe* **2008**, *14*, 242–249.
(7) Marinach-Patrice, C.; Lethuillier, A.; Marly, A.; Brossas, J. Y.; Gené, J.; Symoens, F.; Datry, A.; Guarro, J.; Mazier, D.; Hennequin, C. *Clin Microbiol Infect.* **2009**, *15*, 634–642.
(8) Lay, J. O. Jr. *Mass Spectrom Rev* **2001**, *20*, 172–194.
(9) Chen, H. Y.; Chen, Y. C. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3564–3568.
(10) Li, T. Y.; Liu, B. H.; Chen, Y. C. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2393–2400.
(11) Valentine, N. B.; Wahl, J. H.; Kingsley, M. T.; Wahl, K. L. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1352–1357.
(12) Welham, K. J.; Domin, M. A.; Johnson, K.; Jones, L.; Ashton, D. S. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 307–310.
(13) Erhard, M.; Hipler, U. C.; Burmester, A.; Brakhage, A. A.; Wöstemeyer, J. *Exp Dermatol.* **2008**, *17*, 356–361.
(14) Nagy, E.; Maier, T.; Urban, E.; Terhes, G.; Kostrzewa, M. *Clin Microbiol Infect.* **2009**, *15*, 796–802.
(15) Cherkaoui, A.; Hibbs, J.; Emonet, S.; Tangomo, M.; Girard, M.; Francois, P.; Schrenzel, J. *J Clin Microbiol.* **2010**, *48*, 1169–1175.
(16) Demirev, P. A.; Ho, Y. P.; Ryzhov, V.; Fenselau, C. *Anal. Chem.* **1999**, *71*, 2732–2738.
(17) Pineda, F. J.; Lin, J. S.; Fenselau, C.; Demirev, P. A. *Anal. Chem.* **2000**, *72*, 3739–3744.
(18) Demirev, P. A.; Lin, J. S.; Pineda, F. J.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 4566–4573.
(19) Pineda, F. J.; Antonie, M. D.; Demirev, P. A.; Feldman, A. B.; Jackman, J.; Longenecker, M.; Lin, J. S. *Anal. Chem.* **2003**, *75*, 3817–3822.
(20) Ochi, K. Heterogeneity of ribosomal proteins among Streptomyces species and its application to identification. *J. Gen. Microbiol.* **1989**, *135*, 2635–2642.
(21) Ochiai, K. *I. Int. J. Syst. Bacteriol.* **1993**, *43*, 69–76.
(22) Matte-Tailliez, O.; Brochier, C.; Forterre, P.; Philippe, H. *Mol. Biol. Evol.* **2002**, *19*, 631–639.
(23) Sun, L.; Teramoto, K.; Sato, H.; Torimura, M.; Tao, H.; Shintani, T. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3789–3798.
(24) Teramoto, K.; Sato, H.; Sun, L.; Torimura, M.; Tao, H. *J. Proteome Res.* **2007**, *6*, 3899–3907.
(25) Teramoto, K.; Sato, H.; Sun, L.; Torimura, M.; Tao, H.; Yoshikawa, H.; Hotta, Y.; Hosoda, A.; Tamura, H. *Anal. Chem.* **2007**, *79*, 8712–8719.
(26) Suh, M. J.; Hamburg, D. M.; Gregory, S. T.; Dahlberg, A. E.; Limbach, P. A. *Proteomics* **2005**, *5*, 4818–4831.
(27) Running, W. E.; Ravipaty, S.; Karty, J. A.; Reilly, J. P. *J.Proteome Res.* **2007**, *6*, 337–347.
(28) Arnold, R. J.; Reilly, J. P. *Anal. Biochem.* **1999**, *269*, 105–112.
(29) Strader, M. B.; VerBerkmoes, N. C.; Tabb, D. L.; Connelly, H. M.; Barton, J. W.; Bruce, B. D.; Pelletier, D. A.; Davison, B. H.; Hettich, R. L.; Larimer, F. W.; Hurst, G. B. *J. Proteome Res.* **2004**, *3*, 965–978.
(30) Sato, H.; Shibata, A.; Wang, Y.; Yoshikawa, H.; Tamura, H. *Polym. Degrad. Stab.* **2001**, *74*, 69–75.
(31) Nishino, E.; Ichiki, Y.; Tamura, H.; Morita, S.; Watanabe, K.; Yoshikawa, H. *Biosci. Biotechnol. Biochem.* **2002**, *66*, 1792–1798.
(32) Watanabe, H.; Mori, H.; Ithoh, T.; Gojobori, T. *J. Mol. Evol.* **1997**, *44*, S57–S64.
(33) Itoh, T.; Takemoto, K.; Mori, H.; Gojobori, T. *Mol. Biol. Evol.* **1999**, *16*, 332–346.
(34) Coenye, T.; Vandamme, P. *FEMS Microbiol Lett.* **2005**, *242*, 117–126.
(35) Victoria, B.; Ahmed, A.; Zuerner, R. L.; Ahmed, N.; Bulach, D. M.; Quinteiro, J.; Hartskeerl, R. A. *PLoS One* **2008**, *3*, e2752.
(36) Barloy-Hubler, F.; Lelaure, V.; Galibert, F. *Nucleic Acids Res.* **2001**, *29*, 2747–2756.
(37) Sambrook, J.; Russell, D. W. *Molecular Cloning: A Laboratory Manual*, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.
(38) Sherman, F.; Stewart, J. W.; Tsunasawa, S. *BioEssays* **1985**, *3*, 27–31.
(39) Moerschell, R. P.; Hosokawa, Y.; Tsunasawa, S.; Sherman, F. *J. Biol. Chem.* **1990**, *265*, 19638–19643.
(40) Tompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. *Nucleic Acids Res.* **1997**, *25*, 4876–4882.
(41) Kumar, S.; Tamura, K.; Nei, M. *Briefings Bioinf.* **2004**, *5*, 150–163.
(42) Yamamoto, S.; Bouvet, P. J. M.; Harayama, S. *Int. J. Syst. Bacteriol.* **1999**, *49*, 87–95.
(43) Yamamoto, S.; Kasai, H.; Arnold, D. L.; Jackson, R. W.; Vivian, A.; Harayama, S. *Microbiology* **2000**, *146*, 2385–2394.

PR100868D