Phylogenetic Classification of Pseudomonas putida Strains by MALDI−MS Using Ribosomal Subunit Proteins as Biomarkers

ARTICLE *in* ANALYTICAL CHEMISTRY · DECEMBER 2007

Impact Factor: 5.64 · DOI: 10.1021/ac701905r · Source: PubMed

9 AUTHORS, INCLUDING:

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

Liwei Sun
National Institute of Advanced Industrial Sci…
10 PUBLICATIONS **161** CITATIONS
SEE PROFILE

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

Anal. Chem. **2007,** *79,* 8712−8719

Phylogenetic Classification of *Pseudomonas putida* Strains by MALDI-MS Using Ribosomal Subunit Proteins as Biomarkers

Kanae Teramoto,[†] Hiroaki Sato,*[,†] Liwei Sun,[†] Masaki Torimura,[†] Hiroaki Tao,[†] Hiromichi Yoshikawa,[‡] Yudai Hotta,[§] Akifumi Hosoda,[§] and Hiroto Tamura[§]

Research Institute for Environmental Management Technology, National Institute of Advanced Industrial Science and Technology (AIST), 16-1 Onogawa, Tsukuba, Ibaraki 305-8569, Japan, Fukuoka Institute of Technology, 3-30-1 Wajirohigashi, Higashi-ku, Fukuoka 811-0295, Japan, and Department of Agriculture, Meijo University, 1-501 Shiogamaguchi, Tenpaku-ku, Nagoya, Aichi 468-8502, Japan

A new method for phylogenetic classification of bacterial strains using matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS) is proposed. This method was developed using a bioinformatics-based approach to the rapid identification of bacteria as previously proposed by Demirev and co-workers, which uses ribosomal proteins composed of ~50 subunit proteins as biomarkers. Although the amino acid sequences of ribosomal proteins are highly conserved, slight sequence variations can occur at the strain level. Since ribosomal subunit proteins are a complex of housekeeping proteins that have different phylogenetic evolution rates, sequence variation detected as mass differences by MALDI-MS may be useful for the phylogenetic classification of bacteria at strain level. In our proposed method, the first step is the selection of reliable biomarkers through characterization of the expressed ribosomal subunit proteins of a reference strain (usually a genome-sequenced strain) by MALDI-MS. The observed masses in the MALDI mass spectra of cell lysates of sample strains are then compared with the biomarker masses of the reference strain. The biomarkers for each sample strain were designated as present or absent at the reference masses, indicated by 1 or 0, respectively, which were summarized in a table. This table is processed by cluster analysis, generating a phylogenetic tree. In this study, the success of this approach was confirmed by classification of *Pseudomonas putida* strains because its classification is much more complicated than that of other bacterial strains. Forty-three reliable biomarkers were selected from ribosomal subunit proteins of a genome-sequenced strain, *P. putida* KT2440. The numbers and kinds of biomarkers observed for 16 strains of *P. putida*, including different biovars, were markedly different, reflecting the variety of the strains. The classification results by the proposed method were highly comparable to those based on the DNA gyrase subunit B gene (*gyrB*) sequence analysis, suggesting our proposed method would be a useful high-throughput method for phylogenetic classification of newly isolated bacteria.

The classification or typing of bacterial strains is of key importance in the fields of environmental microbiology, molecular epidemiology, and the fermentation industry. Two DNA sequence-based methods, DNA fragment analyses by electrophoresis and DNA sequence analyses, are currently employed for the genotypic classification of bacteria. Pulsed-field gel electrophoresis,[1] randomly amplified polymorphic DNA,[2] amplified fragment length polymorphism,[3] and ribotyping[4] are the chief forms of DNA fragment analysis used for characterization of bacterial strains. Since fragment analyses rely on gel electrophoretic patterns of DNA fragments by restriction enzyme treatment, band broadening and poor reproducibility of electrophoretic mobility risk impeding the acquisition of reproducible results.

Automated DNA sequence analysis is, nowadays, a powerful method for classifying bacteria. The 16S rRNA gene sequence analysis is generally used for classifying bacteria at the genus and species levels. To classify bacteria at the subspecies and strain levels, the use of housekeeping genes with a higher evolution rate, such as the DNA gyrase subunit B gene (*gyrB*)[5,6] and the RNA polymerase σ^{70} factor gene (*rpoD*)[7] have also been proposed. The limitation of a single gene sequence analysis has, however, been pointed out for the genotypic classification of nonclonal population whose genes are modified by horizontal transmission and homologous recombination.[8−11]

* To whom correspondence should be addressed. Fax: +81 (29) 861-8308. E-mail: sato-hiroaki@aist.go.jp.

[†] AIST.

[‡] Fukuoka Institute of Technology.

[§] Meijo University.

(1) Olson, M. V. *J. Chromatogr.* **1989**, *470*, 377−383.
(2) Bingen, E.; Barc, M. C.; Brahimi, N.; Vilmer, E.; Beaufils, F. *J. Clin. Microbiol.* **1995**, *33*, 1657−1659.
(3) Keim, P.; Kalif, A.; Schupp, J.; Hill, K.; Travis, S. E.; Richmond, K.; Adair, D. M.; HughJones, M.; Kuske, C. R.; Jackson, P. *J. Bacteriol.* **1997**, *179*, 818−824.
(4) Faruque, S. M.; Roy, S. K.; Alim, A. R. M. A.; Siddique, A. K.; Albert, M. J. *J. Clin. Microbiol.* **1995**, *33*, 2833−2838.
(5) Yamamoto, S.; Harayama, S. *Appl. Environ. Microbiol.* **1995**, *61*, 1104−1109.
(6) Yamamoto, S.; Harayama, S. *Int. J. Syst. Bacteriol.* **1998**, *48*, 813−819.
(7) Lonetto, M.; Gribskov, M.; Gross, C. A. *J. Bacteriol.* **1992**, *174*, 3843−3849.
(8) Urwin, R.; Maiden, M. C. J. *Trends Microbiol.* **2003**, *11*, 479−487.

10.1021/ac701905r CCC: $37.00
Published on Web 10/16/2007

To avoid this problem, multilocus sequence typing (MLST) has been proposed for the characterization of bacterial strains.[12] In this method, partial sequences of many conserved housekeeping genes (typically seven genes) are classified into sequence types. The genotypic classification of bacterial strains is then performed based on combinations of these sequence types. The advantages of MLST are the lack of influence on the results of horizontal transmission and homologous recombination and portability of the results between different laboratories.

Since the information of gene sequences is translated to their corresponding proteins, another approach to the classification of bacteria based on protein profiling has been developed to classify the bacteria. Instead of the gel electrophoretic fingerprinting of whole bacteria proteins, mass spectral fingerprinting has been proposed in the past decade, in which matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS) plays an important role in protein analysis.[13−17] Although this method can be used to classify bacterial strains simply and rapidly by a combination with chemometrics data analysis,[18−20] it has serious disadvantages: it shows low reproducibility of the mass spectra, which are strongly influenced by both culture and measurement conditions,[21−25] and, moreover, less phylogenetic reliability caused by the use of unassigned peaks.

To avoid the uncertainty of biomarker peaks, a bioinformatics-based approach for rapid identification of bacteria has recently been proposed by Demirev and co-workers, in which bacterial species could be identified based on matching protein masses in the mass spectra with calculated protein masses predicted from amino acid sequences.[26−30] Ribosomal subunit proteins have been proposed as the suitable biomarkers for this approach.[30]

The ribosome is an organelle that coordinates protein synthesis in all cells. The bacterial ribosome consists of more than 50 ribosomal subunit proteins and 3 rRNAs. Since bacterial cells contain vast amounts of ribosomes, most ribosomal subunit proteins can be observed as main peaks by MALDI-MS of cell lysate.[30−32] Since major post-translational modifications of ribosomal subunit proteins are N-terminal methionine loss, depending on the penultimate amino acid residue, the masses of expressed proteins can be predicted if their amino acid sequences are known.[31−36] With the progress of complete microbial genome sequencing projects, many amino acid sequences of ribosomal subunit proteins, have been registered in public protein sequence databases such as Swiss-Prot, TrEMBL, and NCBInr. Therefore, an Internet search engine for identifying bacterial species using the observed masses of ribosomal subunit proteins is now available.[37] Recently, we have characterized the expressed ribosomal proteins of typical lactic acid bacteria strains of *Lactobacillus plantarum*,[31] *Lactobacillus delbrueckii* subsp. *bulgaricus,* and *Streptococcus thermophilus*[32] by comparison of observed MALDI-MS data with an available public protein sequence databases. Although the amino acid sequences of ribosomal proteins are highly conserved, the observed masses of some ribosomal subunit proteins varied within the species. This fact indicates that mass differences of this type reflecting variation of amino acid sequence may be of use in classifying bacteria at the strain level.

In this study, we propose a new method for phylogenetic classification of bacterial strains based on the profiling of ribosomal subunit proteins by MALDI-MS. The masses of biomarker ribosomal subunit proteins of sample strains are compared with those of a reference strain (usually a genome-sequenced strain). The biomarkers commonly observed at the same mass for sample strains are summarized to compile a biomarker matching table in a binary format. This table can be processed using cluster analysis, generating a dendrogram tree. The important thing is that the results obtained using this method should guarantee genotypic classification, because ribosomal subunit proteins used as biomarkers are a complex of typical housekeeping proteins. Furthermore, since the sequence variation of ribosomal subunit proteins results from molecular evolution, our proposed method would enable "phylogenetic" classification, whereas prior simple fingerprint approaches using MALDI-MS have only performed

(9) Smith, J. M.; Dowson, C. G.; Spratt, B. G. *Nature* **1991**, *349*, 29−31.
(10) Smith, J. M.; Smith, N. H.; O'Rourke, M.; Spratt, B. G. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4384−4388.
(11) Urwin, R.; Maiden, M. C. J. *Trends Microbiol.* **2003**, *11*, 479−487.
(12) Maiden, M. C. J.; Bygraves, J. A.; Feil, E.; Morelli, G.; Russell, J. E.; Urwin, R.; Zhang, Q.; Zhou, J. J.; Zurth, K.; Caugant, D. A.; Feavers, I. M.; Achtman, M.; Spratt, B. G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3140−3145.
(13) Claydon, M. A.; Davey, S. N.; Edward-Jones, V.; Gordon, D. B. *Nat. Biotechnol.* **1996**, *14*, 1584−1586.
(14) 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.
(15) Welham, K. J.; Domin, M. A.; Scannell, D. E.; Cohen, E.; Ashton, D. S. *Rapid Commun. Mass Spectrom*. **1998**, *12*, 176−180.
(16) Arnold, R. J.; Reilly, J. P. *Rapid Commun. Mass Spectrom*. **1998**, *12*, 630−636.
(17) Jarman, K. H.; Daly, D. S.; Peterson, C. E.; Saenz, A. J.; Valentine, N. B.; Wahl, K. L. *Rapid Commun. Mass Spectrom*. **1999**, *13*, 1586−1594.
(18) Dickinson, D. N.; La Duc, M. T.; Haskins, W. E.; Gornushkin, I.; Winefordner, J. D.; Powell, D. H.; Venkateswaran, K. *Appl. Environ. Microbiol.* **2004**, *70*, 475−482.
(19) Stackebrandt, E.; Pauker, O.; Erhard, M. *Curr. Microbiol.* **2005**, *50*, 71−77.
(20) Donohue, M. J.; Smallwood, A. W.; Pfaller, S.; Rodgers, M.; Shoemaker, J. A. *J. Microbiol. Methods* **2006**, *65*, 380−389.
(21) Wang, Z. P.; Russon, L.; Li, L.; Roser, D. C.; Long, S. R. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 456−464.
(22) Arnold, R. J.; Karty, J. A.; Ellington, A. D.; Reilly, J. P. *Anal. Chem.* **1999**, *71*, 1990−1996.
(23) Saenz, A. J.; Petersen, C. E.; Valentine, N. B.; Gantt, S. L.; Jarman, K. H.; Kingsley, M. T.; Wahl, K. L. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1580−1585.
(24) Williams, T. L.; Andrzejewski, D.; Lay, J. O.; Musser, S. M. Jr. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 342−351.
(25) Wunschel, D. S.; Hill, E. A.; McLean, J. S.; Jarman, K.; Gorby, Y. A.; Valentine, N.; Wahl, K. *J. Microbiol. Methods* **2005**, *62*, 259−271.
(26) Demirev, P. A.; Ho, Y. P.; Ryzhov, V.; Fenselau, C. *Anal. Chem.* **1999**, *71*, 2732−2738.
(27) Pineda, F. J.; Lin, J. S.; Fenselau, C.; Demirev, P. A. *Anal. Chem.* **2000**, *72*, 3739−3744.
(28) Demirev, P. A.; Lin, J. S.; Pineda, F. J.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 4566−4573.
(29) Fenselau, C.; Demirev, P. A. *Mass Spectrom. Rev.* **2001**, *20*, 157−171.
(30) 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.
(31) Sun, L.; Teramoto, K.; Sato, H.; Torimura, M.; Tao, H. *Rapid Commun. Mass Spectrom*. **2006**, *20*, 3789−3798.0
(32) Teramoto, K.; Sato, H.; Sun, L.; Torimura, M.; Tao, H. *J. Proteome Res.* **2007**, *6*, 3899−3907.
(33) Arnold, R. J.; Reilly, J. P. *Anal. Biochem.* **1999**, *269*, 105−112.
(34) Suh, M. J.; Hamburg, D. M.; Gregory, S. T.; Dahlberg, A. E.; Limbach, P. A. *Proteomics* **2005**, *5*, 4818−4831.
(35) 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.
(36) Running, W. E.; Ravipaty, S.; Karty, J. A.; Reilly, J. P. *J. Proteome Res.* **2007**, *6*, 337−347.
(37) http://www.rmidb.org/cgi-bin/index.pl.

"chemotaxonomic" classification. The fundamental concept is similar to that of MLST,[12] which is based on a combination of several sequence types of plural housekeeping genes. As with MLST, but unlike previous mass spectral fingerprint approaches, the data are in principle portable and comparable between laboratories, irrespective of culture or measurement conditions. Moreover, this method is much simpler and faster than DNA sequence-based methods that require gel electrophoresis or DNA sequencing.

Phylogenetic classification of a total of 16 strains of *Pseudomonas putida* as model bacteria was demonstrated in this study. Our investigation on the environmental fate of alkylphenol polyethoxylates[38−40] suggested that *P. putida*, as well as other species belonging to genus Pseudomonas, are widely distributed in soil, sediments, and freshwater as responsible bacteria to produce endocrine disruptor-like alkylphenols. Therefore, to monitor or remediate the alkylphenol contamination caused by bacteria, a high-throughput classification method of bacterial should be developed. However, they are highly variable, some of which are further classified into different biotypes or biovars. Currently, phylogenetic classification of *Pseudomonas* strains by sequencing of a single gene such as 16S rRNA and *gyrB* needs to be verified by a combination of other genetic and biochemical information. Phylogenetic classification of *P. putida* strains proved to be a good model for demonstrating the performance of our proposed method using numerous housekeeping ribosomal proteins as biomarkers.

EXPERIMENTAL SECTION

Bacterial Strains. To compare the result of phylogenetic classification based on *gyrB* sequences,[6] we selected the following 16 *P. putida* strains whose *gyrB* sequences are available from public databases: NBRC 100650 (= KT2440),[41] NBRC 3738, NBRC 14164^T, NBRC 14671, NBRC 100986, NBRC 100988, NBRC 101019, JCM 6156, JCM 6158, JCM 13061, MBIC 3930, MBIC 5315, ATCC 17484, ATCC 17522, ATCC 23973, and NCIMB 9816. Since NBRC 100650 is essentially the same strain as the genome-sequenced KT2440 strain,[41] the strain name of KT2440 is used in this paper. The NBRC, JCM, MBIC, ATCC, and NCIMB strains were purchased from the National Institute of Technology and Evaluation (NITE)-Biological Resource Center (NBRC, Kisarazu, Japan), the Japan Collection of Microorganisms at Riken (JCM, Wako, Japan), the Marine Biotechnology Institute Co. (MBI, Kamaishi, Japan), the American Type Culture Collection (ATCC, Rockville, MD), and the National Collections of Industrial and Marine Bacteria (NCIMB, Aberdeen, UK), respectively. Some *P. putida* strains are subclassified into biovars A and B, based on their ability to grow on L-tryptophan and L-kynurenine.[42] Among the strains used in this study, NBRC 14164^T (= IFO 14164^T), JCM 6158 (= PpG7), and JCM 13061 (= ATCC 11172) are biovar A, whereas ATCC 17484 and ATCC 17522 are biovar B, according to a previous paper.[6] Although these biovers are based on phenotypic classification, it has been confirmed that biovers A and B are also genotypically distant based on the gyrB sequences.[6] A partial nucleotide sequence (872 bp) of the *gyrB* gene was obtained from the *gyrB* database of MBI.[43] Each bacterial strain was grown aerobically in the recommended medium and temperature by suppliers. Accession numbers of *gyrB* sequences and growth conditions are summarized in the Supporting Information Table SI-1.

Cell Lysates. Cell lysates were prepared as described in our previous papers.[31,32] Briefly, 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). The cells were suspended in the buffer and ground (two times for 30 s each at 3000 rpm) with zirconia silica beads in a Mini Bead-Beater 8 (Biospec, Bartlesville, OK). After removing the beads and cell debris by centrifugation, a portion of the cell lysates was subjected to MALDI-MS measurements.

Purification of Ribosome Fraction. Ribosomal subunit proteins of *P. putida* KT2440 were isolated and purified using the reported method.[44−46] The lysate of KT2440 was subjected to ultracentrifugation at 80000*g* for 6 h, followed by ultracentrifugation through a 30% sucrose cushion at 45000*g* for 16 h to obtain 70S ribosome particles. After removing the ribosomal rRNA by precipitation using 1 M magnesium acetate and glacial acetic acid, precipitation using acetone at −20 °C, centrifugation, and dialysis against 2% acetic acid, the solutions of purified ribosomal protein were finally obtained.

MALDI-MS. Sample preparation, apparatus, and data acquisition of MALDI-MS were almost the same as described in our previous papers.[31,32] Each sample solution of cell lysates 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 was spotted onto the MALDI target and dried in air. MALDI mass spectra in the range of m/z 2000−40 000 were observed in positive linear mode by averaging 500 individual laser shots using an Axima CFR plus time-of-flight mass spectrometer (Shimadzu/Kratos, Kyoto, Japan) equipped with a pulsed N_2 laser (λ = 337 nm, pulse width 3 ns, frequency 10 Hz). Mass calibration of purified ribosomal protein and cell lysate of *P. putida* KT2440 was carried out by internal calibration using three peaks of ACTH clip 18−39 ($[M + H]^+$, m/z 2465.7) and myoglobin ($[M + H]^+$, m/z 16952.6, and $[M + H]^{2+}$, m/z 8476.8), followed by self-calibration using moderately strong peaks assigned to ribosomal subunit proteins as internal references. Mass calibration of the other samples was performed by external calibration using the mass spectra observed for the cell lysate of the KT2440 strain. The peak matching of the biomarker proteins was judged from errors within 150 ppm.

(38) Sato, H.; Shibata, A.; Wang, Y.; Yoshikawa, H.; Tamura, H. *Polym. Degrad. Stab.* **2001**, *74*, 69−75.

(39) Nishio, E.; Ichiki, Y.; Tamura, H.; Morita, S.; Watanabe, K.; Yoshikawa, H. *Biosci. Biotechnol. Biochem.* **2002**, *66*, 1792−1798.

(40) Sato, H.; Shibata, A.; Wang, Y.; Yoshikawa, H.; Tamura, H. *Biomacromolecules* **2003**, *4*, 46−51.

(41) Nelson, K. E.; Weinel, C.; Paulsen, I. T.; Dodson, R. J.; Hilbert, H.; dos Santos, V. A. P. M.; Fouts, D. E.; Gill, S. R.; Pop, M.; Holmes, M.; Brinkac, L.; Beanan, M.; DeBoy, R. T.; Daugherty, S.; Kolonay, J.; Madupu, R.; Nelson, W.; White, O.; Peterson, J.; Khouri, H.; Hance, I.; Lee, C. P.; Holtzapple, E.; Scanlan, D.; Tran, K.; Moazzez, A.; Utterback, T.; Rizzo, M.; Lee, K.; Kosack, D.; Moestl, D.; Wedler, H.; Lauber, J.; Stjepandic, D.; Hoheisel, J.; Straetz, M.; Heim, S.; Kiewitz, C.; Eisen, J. A.; Timmis, K. N.; Dusterhoft, A.; Tummler, B.; Fraser, C. M. *Environ. Microbiol.* **2002**, *4*, 799−808.

(42) Stanier, R. Y.; Palleroni, N. J.; Doudoroff, M. *J. Gen. Microbiol.* **1966**, *43*, 159−271.

(43) http://www.mbio.jp/icb/.

(44) Kurland, C. G. *Methods Enzymol.* **1971**, *20*, 379−381.

(45) Traub, P.; Mizushima, S.; Lowry, C. V.; Nomura, M. *Methods Enzymol.* **1971**, *20*, 391−407.

(46) Hardy, S. J. S.; Kurland, C. G.; Voynow, P.; Mora, G. *Biochem.* **1969**, *8*, 2897.

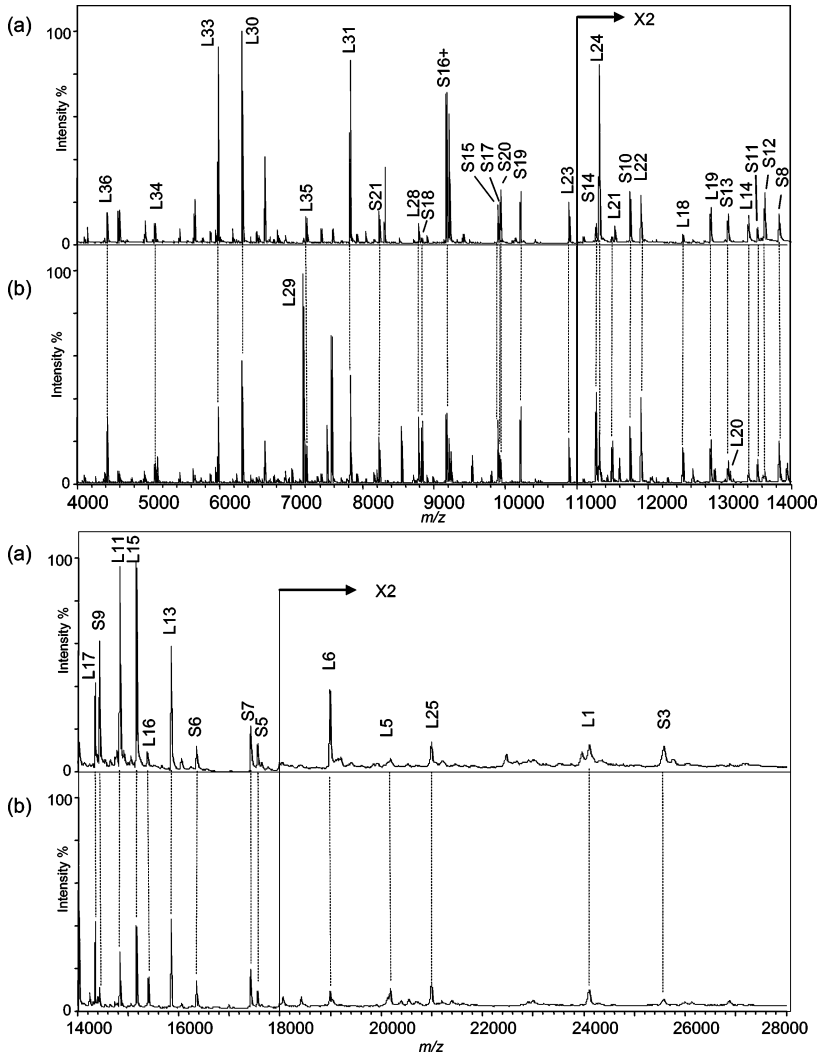

Figure 1. MALDI mass spectra of purified ribosomal protein (a) and cell lysate (b) of *P. putida* KT2440.

Cluster Analysis. Phylogenetic analysis based on the MALDI mass spectra was conducted by the BioNumerics software (version 3.5; Applied Maths, Kortrijk, Belgium). A binary biomarker matching table constructed from MALDI mass spectral analysis was processed to build a phylogenetic tree by the unweighted pair group method with arithmetic mean (UPGMA) cluster analysis using the simple matching similarity coefficient. A phylogenetic tree based on the *gyrB* gene sequence analysis was also constructed by UPGMA using Kimura two-parameter model in MeGA 3.1 program.[47] The details of the procedures are described in the Results and Discussion.

RESULTS AND DISCUSSION

Characterization of Ribosomal Subunit Proteins of *P. putida* KT2440 as Biomarkers. The first step in this method is the selection of reliable biomarker proteins from more than 50 ribosomal subunit proteins of the reference strain KT2440, because the actual masses of some expressed ribosomal subunit proteins would be different from the calculated masses based on their amino acid sequences due to post-translational modifications and sequence errors.[32,34,36] Figure 1 compares the MALDI mass spectra of purified ribosomal protein and cell lysate of *P. putida* KT2440 in the range of m/z 4000−28 000. Table 1 summarizes the assigned ribosomal proteins of *P. putida* KT2440, together with calculated masses and possible post-translational modifications. Detailed characterization results are supplied in Supporting Information Table SI-2. As a post-translational modification, N-Terminal methionine loss was first considered based on the "N-end rule". In this rule, N-terminal methionine is cleaved from specific penultimate amino acid residues such as glycine, alanine, serine, proline, valine, threonine, and cysteine.[48,49] In the case of S16, whose penultimate amino acid residue is valine, however, no N-terminal methionine loss was observed. This type of exception has been reported for S16 in *Escherichia coli*[33] and S27 and L42 in *Saccharomyces cerevisiae*.[50] In the case of L29, whose N-terminal amino acid sequence is MMKAN-, the observed mass was matched to the calculated mass when N-terminal methionine loss was considered. This in fact appears to be misannotation of the start codon rather than N-terminal methionine loss.

(47) Kumar, S.; Tamura, K.; Nei, M. *Brief. Bioinform.* **2004**, *5*, 150−163.
(48) Sherman, F.; Stewart, J. W.; Tsunasawa, S. *BioEssays* **1985**, *3*, 27−31.
(49) Moerschell, R. P.; Hosokawa, Y.; Tsunasawa, S.; Sherman, F. *J. Biol. Chem.* **1990**, *265*, 19638−19643.
(50) Arnold, R. J.; Polevoda, B.; Reilly, J. P.; Sherman, F. *J. Biol. Chem.* **1999**, *274*, 37035−37040.

Table 1. Assigned Ribosomal Subunit Proteins of *P. putida* KT2440

protein name	mass of $[M + H]^+$	post-translational modifications[a]	second amino acid residue
		Large Subunit Proteins	
L1	24120.8	−Met	A
L5	20198.6	−Met	A
L6	19012.8	−Met	S
L11	14861.8	−Met, methylation (9 times)	A
L13	15863.3		K
L14	13410.9		I
L15	15190.4		K
L16	15374.2	methylation (2 times) and oxidation	L
L17	14363.7		R
L18	12497.4	−Met	T
L19	12881.9	−Met	T
L20	13149.7	−Met	A
L21	11501.4	−Met	S
L22	11911.9		E
L23	10900.6		N
L24	11330.2		Q
L25	21008.9	−Met	T
L28	8792.3	−Met	S
L29	7173.3	−Met[b]	M
L30	6321.5	−Met	A
L31	7845.1		K
L33	5991.0		R
L34	5139.2		K
L35	7214.7	−Met	P
L36	4435.4		K
		Small Subunit Proteins	
S3	25594.7	−Met	G
S5	17578.7	−Met, acetylation	A
S6	16373.9		R
S7	17448.3	−Met	P
S8	13845.1	−Met	S
S9	14461.5	−Met	S
S10	11753.6		Q
S11	13543.4	methylation	A
S12	13642.0	−Met, β-methylthiolation	A
S13	13126.3	−Met	A
S14	11273.3	−Met	A
S15	9900.3	−Met	A
S16	9189.5		V
S17	9927.6	−Met	A
S18	8838.9	−Met, acetylation	A
S19	10218.1	−Met	P
S20	9938.6	−Met	A
S21	8239.6	−Met	P

[a] −Met means N-terminal methionine loss. [b] Misannotation of the start codon?

The other post-translational modifications were further considered referring to the previous reports for ribosomal proteins of Gram-negative bacteria.[33−36] Six subunit proteins could be assigned in the light of the following post-translational modifications: methylation (nine times) of L11, methylation (two times), and oxidation of L16, acetylation of S5, methylation of S11, and β-methylthiolation of S12, and acetylation of S18, respectively. These appear to be conserved post-translational modifications.

Two subunit proteins (L20 and L29) were observed only for cell lysate. Those in the purified ribosomal protein sample might be lost by coprecipitation with rRNA during the purification procedure, since these subunit proteins would directly bind to 23S rRNA in the 70S ribosome particle.[51] A similar phenomenon was observed in our previous study.[31,32]

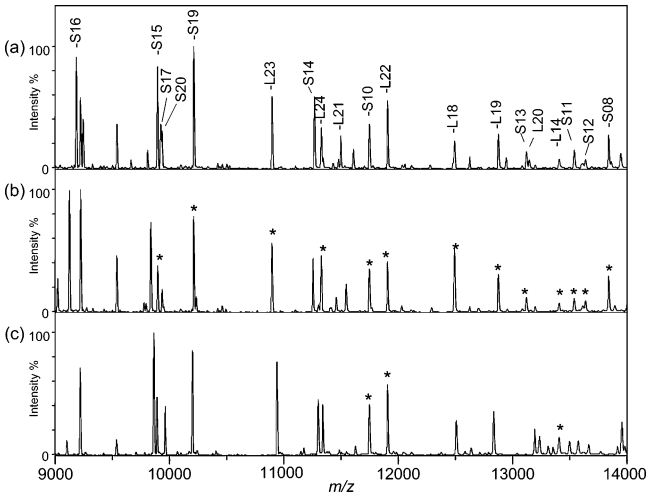

Figure 2. MALDI mass spectra of cell lysate of *P. putida* strains. (a) KT2440 (reference strain), (b) NBRC 14164^T (biover A), and (c) ATCC 17522 (biover B). Asterisks indicate the biomarkers with the same masses with those of KT2440.

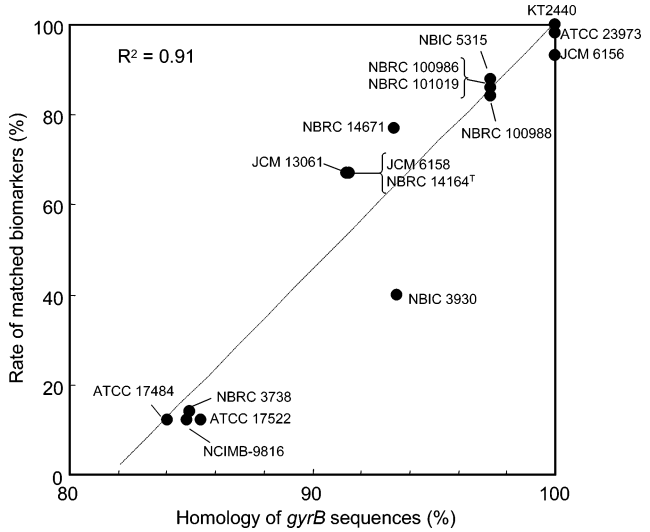

Figure 3. Relationship between homology of *gyrB* sequences and rate of matched biomarkers for *P. putida* strains.

As described above, 43 ribosomal subunit proteins listed in Table 1 were finally selected as reliable biomarkers for the classification of *P. putida* strains.

Sequence Variations of Ribosomal Subunit Proteins. Figure 2 compares expanded MALDI mass spectra in the range of m/z 9000−14 000 for cell lysates of *P. putida* KT2440 (reference strain), NBRC 14164^T (biover A), and ATCC 17522 (biover B). The peaks marked with asterisks indicate the matched biomarkers, which are observed at the same mass of the biomarker subunit proteins within the error of 150 ppm. The MALDI mass spectrum of NBRC 14164^T (biover A) is similar to that of KT2440. For the case of ATCC 17522 (biover B), however, only three biomarkers (S10, L22, and L14) are matched with the reference masses in the shown m/z range. These results suggested that KT2440 would be genetically close to NBRC 14164^T, rather than ATCC 17522. In fact, the homology of *gyrB* sequences between KT2440 and NBRC 14164^T (~91%) is higher than that between KT2440 and ATCC 17522 (~85%).

(51) Rohl, R.; Nierhaus, K. H. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 729−733.

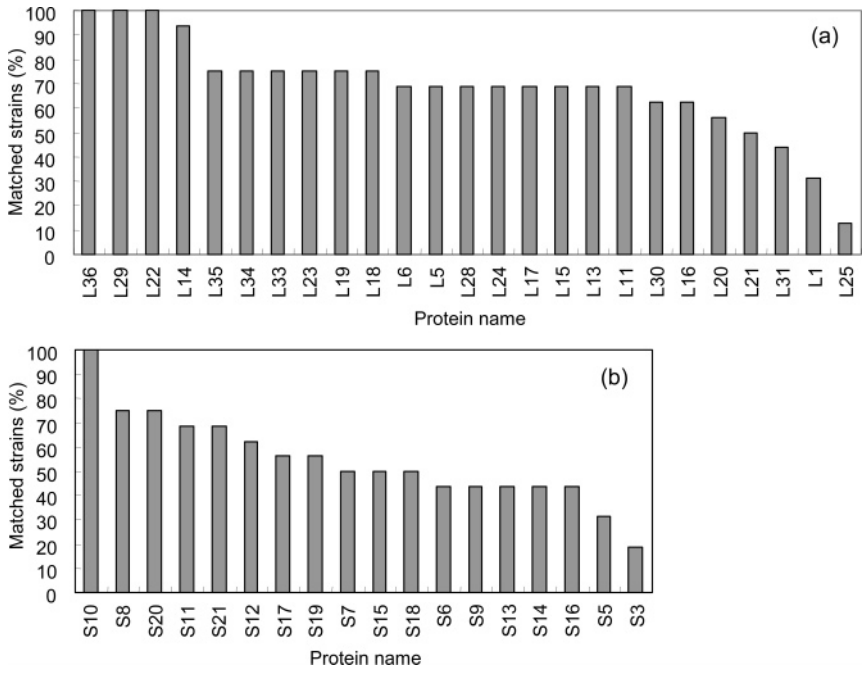

Figure 4. Distribution of the percentage of the matched strains for each biomarker. (a) Large subunit proteins and (b) small subunit proteins.

Figure 3 shows the relationship between the rate of matched biomarkers and the homology values of *gyrB* sequences against to KT2440. Detailed result of the matched biomarkers is summarized in Supporting Information Table SI-3. Interestingly, a good linear relationship was obtained with a correlation coefficient of ~0.91. Genetically close or distant strains based on *gyrB* can be clearly distinguished based on the matching rate of biomarkers. For example, the strains with the homology value larger than 85% (ATCC 23973, JCM 6156, NBIC 5315, NBRC 100986, NBRC 10988, and NBRC 101019) are distributed more than 80% in the matching rate. On the other hand, four strains containing biover B (ATCC 17484, ATCC 17522, NBRC 3738, and NCIMB 9816) have lower homology values less than 85%, also giving considerably lower rate of matched biomarkers (~12−14%). Such results are not surprising, because both ribosomal proteins and DNA gyrase subunit B are housekeeping proteins whose amino acid sequences (and coding DNA sequences) should vary reflecting molecular evolution. Consequently, the sequence variation of ribosomal subunit proteins detected as mass differences by MALDI-MS would be a good criterion for phylogenetic analysis.

The degrees of the sequence variation of each subunit protein are then investigated. Figure 4 shows the distribution of the percentage of the matched strains for each biomarker. This value would reflect the conservation of amino acid sequence of each biomarker. The values differ widely, ranging from 13 to 100%. Among 43 biomarkers, L22, L29, L36, and S10 were matched for all *P. putida* strains, giving a score of 100%. These highly conserved subunit proteins have important functions within ribosome particle. L22 and L29, together with L23 (score 75%) and L24 (score 69%), form the polypeptide exit tunnel in the 50S ribosome subunit.[52] L36 plays a significant role in organizing 23S rRNA structure.[53] S10 is involved in tRNA binding to the ribosomes.[54] In addition, among the other biomarkers with high scores, L14 (score 94%), L5 (score 75%) and L19 (score 75%) are located on the inter-subunit surface, forming bridges with the 30S ribosome subunit.[52,55] On the other hand, 10 subunit proteins (L1, L25, L31, S3, S5, S6, S9, S13, S14, and S16) were matched in fewer than half of the strains, suggesting low conservation of their amino acid sequences even within the same species.

To further investigate the conservation of amino acid sequences of ribosomal subunit proteins, a homology search was performed using a BLASTp search against the nonredundant NCBI database. As an example, homologies of L29 and L31 of *P. putida* KT2440 are discussed here. These subunit proteins have a similar length of amino acid sequences (63 residues for L29 and 71 residues for L31) but different percentages of the matched strains (score 100 and 43% for L29 and L31, respectively). Table 2 shows bacterial strains that have the same amino acid sequence for L29 of *P. putida* KT2440 and identities (%) of amino acid sequences of L31 for these strains. The L29 amino acid sequence of *P. putida* KT2440 was the same, not only for other *P. putida* strains (F1 and GB-1) but also for other species in the *Pseudomonas* genus, such as *Pseudomonas entomophila, Pseudomonas fluorescens*, and *Pseudomonas syringae*. In the case of L31, however, only *P. putida* F1 has the same amino acid sequence; the others show lower homology, notably 69% for *P. syringae* strains. Thus, the amino acid sequence of L31 mutates within the same species. Other subunit proteins also exhibited a different degree of conservation of amino acid sequences. These results indicate that the molecular evolution rates are different depending on ribosomal subunit proteins. Therefore, profiling of matched ribosomal

(52) Gao, H. X.; Sengupta, J.; Valle, M.; Korostelev, A.; Eswar, N.; Stagg, S. M.; Van Roey, P.; Agrawal, R. K.; Harvey, S. C.; Sali, A.; Chapman, M. S.; Frank, J. *Cell* **2003**, *113*, 789−801.
(53) Maeder, C.; Draper, D. E. *J. Mol. Biol.* **2005**, *354*, 436−446.
(54) Nomura, M.; Held, W. A. In *Ribosomes*; Nomura, M., et al. Eds.; Cold Spring Harbor Laboratory Press: Long Island, NY, 1974; pp 193−223.
(55) Maisnier-Patin, S.; Paulander, W.; Pennhag, A.; Andersson, D. I. *J. Mol. Biol.* **2007**, *366*, 207−215.

Table 2. Bacteria Strains Having the Same Amino Acid Sequence for L29 of *P. putida* KT2440 and Identities (%) of Amino Acid Sequences of L31

strain names	identities (%) L29 (length 63)	identities (%) L31 (length 71)
P. putida F1	100	100
P. putida GB-1	100	98
P. entomophila L48	100	92
P. putida W619	100	88
P. fluorescens PfO-1	100	84
P. fluorescens Pf-5	100	70
P. syringae pv. tomato str. DC3000	100	69
P. syringae pv. phaseolicola 1448A	100	69
P. syringae pv. syringae B728a	100	69

subunit proteins would make it possible to perform phylogenetic classification of *P. putida* strains.

Phylogenetic Classification of *P. putida* Strains. The profile of matched ribosomal proteins (see Supporting Information Table SI-3) was processed by UPGMA cluster analysis using the simple matching similarity coefficient to generate a phylogenetic tree. Other clustering methods such as Ward, single-linkage, and complete-linkage also generated similar topology. Figure 5 shows the phylogenetic trees of *P. putida* strains based on the partial sequence of the *gyrB* gene (a) and on the ribosomal profiling using MALDI-MS (b). Interestingly, the topologies of both trees are highly comparable. Based on *gyrB* sequences, *P. putida* strains can be split into two major clusters: cluster I containing KT2440 and biovar A, and cluster II containing biovar B. Two major clusters and six further classified groups based on *gyrB* sequences are produced by our proposed method. The comparable result provides strong evidence that phylogenetic classification can be achieved by the ribosomal protein profiling using MALDI-MS.

Here, it should be noted that two major clusters are separated at ~25% similarity by our proposed method, while those are separated at ~90% similarity by the *gyrB* sequence analysis. Six groups could be separated at ~85% similarity, which were separated at ~97% similarity in the case of the *gyrB* sequence analysis. These facts indicate that resolving power of bacteria classification at strain level by our proposed method is higher than that of the *gyrB* sequence method. Furthermore, the strains belonging to group I-1, which were difficult to distinguish by the *gyrB* sequence method, could be successfully separated. In this case, The mass differences due to sequence mutation of L25, S3, and S5 were the keys to distinguishing KT2440, ATCC 23973, and JCM 6156 strains. For the discrimination of NBRC 101019, NBRC 100986, and NBRC 100988 strains, six subunit proteins (L1, S3, S5, S12, S18, and S19) were key biomarkers. On the other hand, the strains within groups I-4 and II-2 could not be discriminated by our method. Originally, *gyrB* sequencing revealed very little difference in the strains, so the slight difference between the two methods is unlikely to be important.

Our proposed method features several advantages. It is potentially applicable to the phylogenetic classification of other bacteria, since all bacterial cells contain vast amounts of ribosomal proteins. Moreover, ribosomal proteins can easily be observed by MALDI-MS as major peaks in the cell lysate sample, or

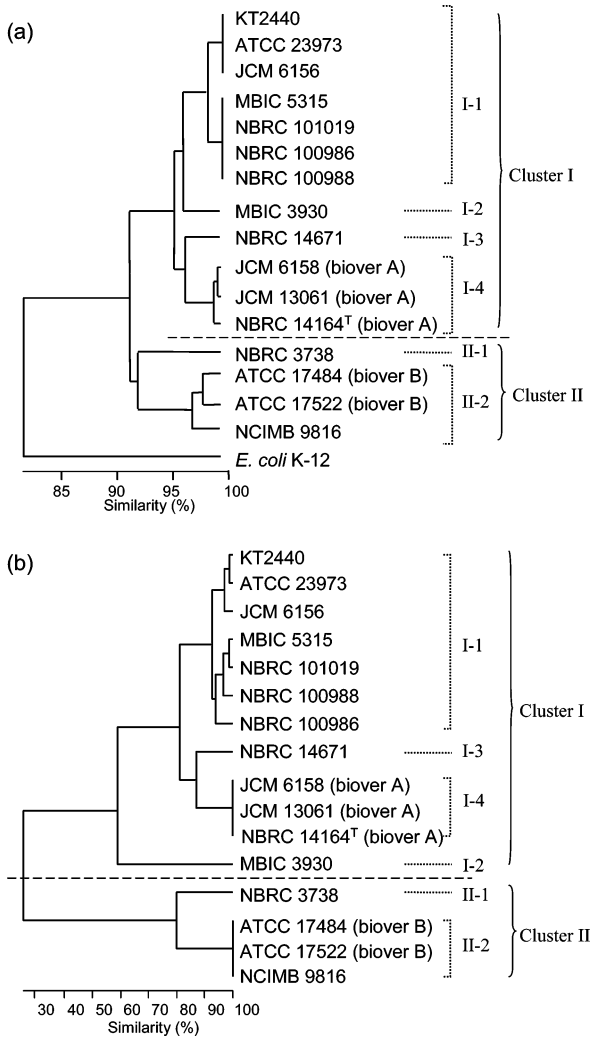

Figure 5. Phylogenetic trees of *P. putida* strains based on *gyrB* sequences (872 bp) (a) and ribosomal protein matching profiling by MALDI-MS (b).

sometimes even for intact cells, without purification of ribosomal proteins under optimized experimental conditions.[31,32] Since sample preparation is very simple and MALDI-MS measurements can be sequentially performed for many samples, classification results can be obtained within 30 min, even for several dozen samples, if an automatic diagnosis system can be realized. The most important advantage of this method is that our method targets several dozen ribosomal subunit proteins expressed from corresponding housekeeping genes with different evolution rates, while general DNA sequence-based methods use only single or few genes. The classification results using the integrated information of the target ribosomal subunit proteins appear to give a more accurate view of the phylogenetic evolution of bacterial strains.

CONCLUSIONS

We have proposed a new phylogenetic classification based on ribosomal protein profiling by MALDI-MS, which was developed using the bioinformatics-based method for rapid identification of bacteria proposed by Demirev and co-workers.[26−30] The classification result of *P. putida* strains including different biovars using our proposed method was in very good agreement with that based on *gyrB* sequences.

The masses of ribosomal subunit proteins can be simply and accurately determined by MALDI-MS without painstaking procedures or specialized reagents. Slight variations of the amino acid sequence can be detected as mass differences in the observed mass spectra. Since the phylogenetic tree is constructed by integrating the sequence information of multiple housekeeping proteins with different evolution rates, the classification results are expected to be more reliable than those obtained by single-gene sequence analyses.

Our proposed method is still at the preliminary stage. To establish more reliable phylogenetic classification, the data processing method should be further investigated, especially the minimum number of biomarkers needed and which ribosomal subunit proteins should be selected. Assuming progress in our investigations, this method has the potential to be developed into a practical and promising technique for the phylogenetic classification of bacterial strains in addition to the DNA sequence-based methods.

SUPPORTING INFORMATION AVAILABLE

Culture conditions and accession numbers of *gyrB* sequences of *P. putida* strains, detailed characterization result of assigned ribosomal subunit proteins of *P. putida* KT2440, and biomarker matching table in a binary format. Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review June 21, 2007. Accepted September 12, 2007.

AC701905R