PO-CON1313E

SHIMADZU
Excellence in Science

Novel bacterial classification method by MALDI-TOF MS based on ribosomal protein coding in *S10-spc-alpha* Operon at Strain level

ASMS 2013 ThP25-493

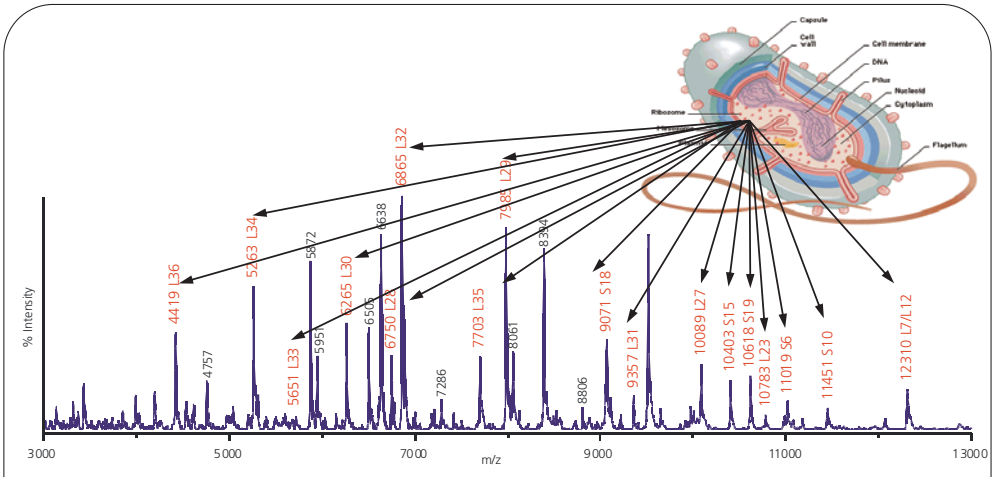
Hiroto Tamura[1,6]; Yudai Hotta[1,2];
Naomi Yamamoto[1,6]; Hiroaki Sato[3]; Keisuke Shima[4];
Shinji Funatsu[4]; Yuzo Yamazaki[4];
Helen Montgomery[5]; Akifumi Hosoda[1];
Noriyuki Ojima[4]
[1]Meijo University, Nagoya, Japan;
[2]Kumiai Chemical Industry, Tokyo, Japan;
[3]AIST, Tsukuba, Japan;
[4]SHIMADZU Corporation, Kyoto, Japan;
[5]SHIMADZU UK, Manchester, UK;
[6]"The Knowledge Hub" of AICHI, Nagoya, Japan

Novel bacterial classification method by MALDI-TOF MS based on ribosomal protein coding in *S10-spc-alpha* Operon at Strain level

1. Introduction

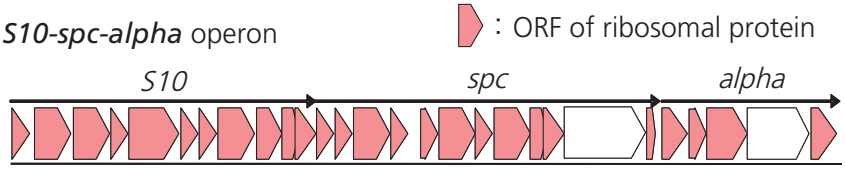
Whole-Cell MALDI-TOF MS (WC-MS) is the most widely used mass-based approaches for bacterial identification and classification. Commercialized bacterial identification systems on the basis of protein mass fingerprinting are rapidly expanding. However, the mass fingerprinting methodology can not assure to discriminate bacteria at strain level, partly because strain specific biomarker based on bacterial genomic database has not been assigned with the method. Until now we have proposed a reliable phylogenetic classification method at the strain level by MALDI-TOF MS using ribosomal subunit proteins coded in the *S10-spc-alpha* operon, which encodes half of the ribosomal proteins in eubacterial genomes, as proteomic biomarkers.

In this study, the method, named *S10*-GERMS (*S10-spc-alpha* operon Gene Encoded Ribosomal Protein Mass Spectrum) was applied to discriminate the pathovar of *P. syringae*, a plant-pathogenic bacterium.



- ~50% of detected peaks in WC -MS (> *m/z* 4000) are derived from ribosomal proteins.
- Assigned ribosomal subunit proteins based on DNA sequences and WC-MS can be specific markers for bacterial classification at species and strain level [1), 2), 3)].
- However, available DNA sequence information of ribosomal proteins is limited, because most of bacterial strains are not genome -sequenced.

Fig. 1 Typical example of WC-MS spectrum



- *S10-spc-alpha* operon contains more than ribosomal subunits, which are commonly conserved in bacteria.
- At the moment, DNA sequencing of this operon using conventional Sanger DNA sequencer is more accessible and accurate than whole genome sequencing.
- Using gene sequence of this operon, a considerable number of WC-MS peaks of bacterial strain can be assigned even if the strain is not genome sequenced.

Fig. 2 Diagram of *S10-spc-alpha* operon

Novel bacterial classification method by MALDI-TOF MS based on ribosomal protein coding in *S10-spc-alpha* Operon at Strain level

2. Methods and Materials

Workflow of the phylogenetic biomarker finding and constructing database

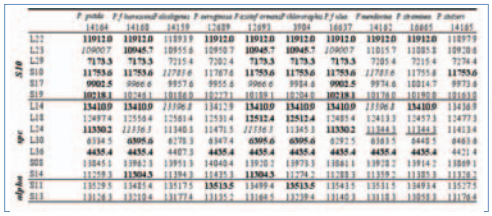
Step 1- Analyzing bacterial strains by WC-MS (observed mass).

Step 2- DNA sequencing and translating of *S10-spc-alpha* operon of the strains.

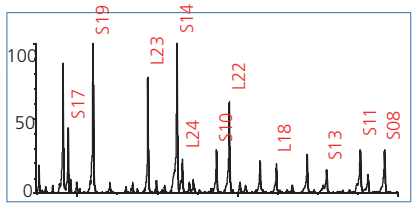
Step 3- Construction of theoretical ribosomal proteins database of the strains.
Theoretical mass of each subunit protein was calculated taking account of N-terminal methionine loss.

Step 4- Construction of accurate database by comparison of theoretical vs. observed mass.
To validate the constructed database, the theoretical masses of MS analyzed bacterial strains at step 1 were compared with each of the MALDI mass spectra.

Theoretical mass from **Step 3**



Observed mass from **Step 1**



Accurate working database

	P. putida NBRC 14164	P. fulva NBRC 16637	P. fluorescens NBRC 14160	P.azotoformans NBRC 12693	P.chlororaphis NBRC 3904	P. aeruginosa NBRC 12689	P.mendocina NBRC 14162	P. straminea NBRC 16665	P. stutzeri NBRC 14165	P.alcaligenes NBRC 14159
L22	11911.95	11911.95	11911.95	11911.95	11911.95	11911.95	11911.95	11911.95	11897.92	11893.91
L23	10900.65	10900.65	10945.74	10945.74	10945.74	10950.71	11015.74	11085.83	10920.64	10955.60
L29	7173.31	7173.31	7173.31	7173.31	7173.31	7202.35	7205.35	7215.39	7274.42	7215.44
S10	11753.58	11753.58	11753.58	11753.58	11753.58	11767.61	11783.61	11755.55	11753.58	11783.61
S17	9902.53	9902.53	9966.58	9966.58	9984.61	9955.60	9974.55	10014.66	9973.57	9957.61
S19	10218.07	10218.07	10246.12	10189.07	10204.04	10227.08	10175.98	10190.01	10162.99	10186.01
L18	12497.41	12485.36	12556.43	12512.38	12512.38	12531.43	12413.29	12457.30	12477.34	12561.35
L24	11330.24	11330.24	11336.25	11336.25	11345.26	11471.46	11344.27	11344.27	11413.37	11340.32
L30	6334.54	6292.46	6395.60	6395.60	6395.60	6347.44	6363.48	6448.54	6463.62	6278.33
L36	4435.39	4435.39	4435.39	4435.39	4435.39	4435.39	4435.39	4435.39	4421.36	4407.34
S08	13845.12	13861.12	13962.26	13920.23	13973.29	14040.42	13928.16	13914.18	13869.10	13951.25
S14	11259.27	11288.27	11304.32	11304.32	11274.24	11435.25	11359.15	11385.32	11326.23	11394.33
S11	13529.50	13529.50	13485.44	13485.44	13499.47	13499.47	13517.48	13479.39	13513.50	13503.46
S13	13126.31	13140.34	13210.43	13164.45	13239.43	13135.23	13118.33	13058.28	13176.42	13177.40

Using this database, discrimination of sample bacteria at the strain level was performed.

Bacterial Strains

Commercially available strains of *Pseudomonas syringae* were purchased from the National Institute of Technology and Evaluation (NITE)-Biological Resource Center (NBRC, Kisarazu, Japan)

P. syringae NBRC 3310, NBRC 3508, NBRC 12655, NBRC 12656, NBRC 14053, NBRC 14083 and NBRC 14084. Each bacterial strain was grown aerobically in the medium and at the temperature recommended by its suppliers.

Construction of the ribosomal protein database

The amino acid sequences of all ribosomal subunit proteins of the *P. syringae* strains were obtained from the NCBInr database. The theoretical mass of each subunit protein as $[M + H]^+$ was calculated from translated amino acid sequences using a Compute pI/Mw tool on the ExPASy proteomics server. N-Terminal methionine loss was only considered based on the 'N-end rule' as a possible post-translational modification. When there were some errors in their theoretical masses they were corrected by comparison with the observed masses by WC-MS. The selected biomarker proteins based on the corrected ribosomal protein database were used for bacterial discrimination at the strain level of *P. syringae*.

Novel bacterial classification method by MALDI-TOF MS based on ribosomal protein coding in *S10-spc-alpha* Operon at Strain level

MALDI-TOFMS

Instrument : AXIMA Performance (Shimadzu Biotech/Kratos)
Tuning mode : Linear positive
Matrix solution : Sinapinic acid 10 mg/mL (50% acetonitrile 1% TFA)
Sample preparation : Bacterial cells were harvested by centrifugation and washed twice in TMA-I buffer (10 mM Tris-HCl (pH 7.8), 30 mM NH_4Cl, 10 mM $MgCl_2$, and 6 mM 2-mercaptoethanol). Each harvested solution of whole cells was mixed with the matrix solution. Then, the mixed solution was spotted onto the MALDI target and dried in air.

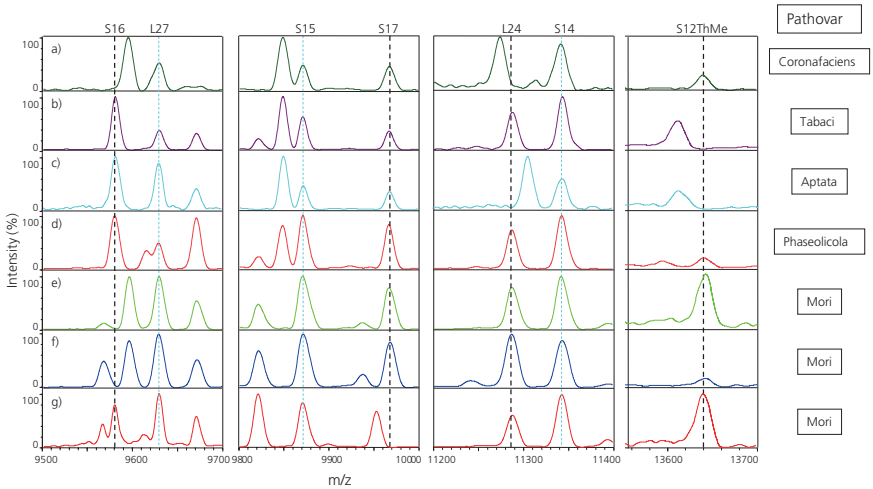
3. Results

In this study, the *S10*-GERMS method was applied for discrimination of the pathovar of *P. syringae* using genome-sequenced and commercially available strains as model samples. Theoretical masses of ribosomal proteins were calculated by *S10-spc-alpha* operon sequence information of the 12 genome-sequenced strains of *P. syringae*. 14 ribosomal subunit proteins with less than *m/z* 15,000, coded in the *S10-spc-alpha* operon (except for L14) and two additional ribosomal proteins (S16, S12) were selected as reliable and reproducible biomarkers for typing the pathovar of *P. syringae* (Table 1). Several peaks seem to be pathovar-specific. And then, 7 sample strains of *P. syringae* were analyzed by MALDI-TOF MS (Fig. 3).

Table 1 Theoretical mass of genome-sequenced strains of *P. syringae*

Protein name	Coded operon	Theoretical masses as [M + H]+											
		Aesculi 2250	Aesculi NCPPB 3681	Phaseolicola 1448A ATCC 1	Tabaci 1528	Syringae B728a	Syringae 642	Oryzae 1_6	Tomato K40	Tomato Max13NCPPB 1	Tomato 108	Tomato DC3000	Tomato T1
L36	spc	4435.4	4435.4	4435.4	4435.4	4435.4	4435.4	4435.4	4435.4	4435.4	4435.4	4435.4	4435.4
L30	spc	6422.7	6422.7	6422.7	6422.7	6422.7	6408.6	6422.7	6422.7	6422.7	6422.7	6422.7	6426.7
L29	S10	7173.3	7173.3	7173.3	7173.3	7173.3	7173.3	7173.3	7173.3	7173.3	7173.3	7173.3	7173.3
S17	S10	9966.6	9966.6	9966.6	9966.6	9966.6	9966.6	9966.6	9966.6	9966.6	9966.6	9966.6	9966.6
S19	S10	10204.0	10204.0	10204.0	10204.0	10204.0	10204.0	10204.0	10177.0	10177.0	10177.0	10177.0	10177.0
L23	S10	10945.7	10945.7	10945.7	10945.7	10945.7	10945.7	10945.7	10945.7	10945.7	10945.7	10945.7	10945.7
L24	spc	11288.2	11288.2	11288.2	11288.2	11306.2	11306.2	11274.2	11288.2	11288.2	11288.2	11288.2	11288.2
S14	spc	11362.4	11362.4	11343.3	11343.3	11346.4	11343.3	11343.3	11343.3	11343.3	11343.3	11343.3	11343.3
S10	S10	11796.6	11796.6	14733.2	11767.6	11767.6	11767.6	11767.6	11767.6	11767.6	11767.6	11767.6	11767.6
L22	S10	11893.9	11893.9	11893.9	11893.9	11893.9	11893.9	11893.9	11893.9	11893.9	11893.9	11893.9	11893.9
L18	spc	12473.4	12473.4	12473.4	12473.4	12473.4	12473.4	12473.4	12473.4	12473.4	12473.4	12473.4	- ※
S13	alpha	13239.5	13239.5	13239.5	13239.5	13239.5	13239.5	13239.5	13239.5	13239.5	13239.5	13239.5	13239.5
S11	alpha	13485.4	13485.4	13485.4	13485.4	13485.4	13485.4	13485.4	13485.4	13485.4	13485.4	13485.4	13485.4
S08	spc	13892.1	13892.1	13892.1	13892.1	13892.1	13892.1	13892.1	13892.1	13892.1	13892.1	13892.1	13892.1
S16		9580.9	9580.9	9580.9	9580.9	9580.9	9580.9	- ※	9580.8	9594.9	9594.9	9594.9	9594.9
S12		13594.7	13594.7	13594.7	13594.7	13594.7	13594.7	13594.7	13628.7	13594.7	13594.7	13594.7	13594.7

※ There was a mis-sequence in the genome sequence.



P. syringae : NBRC 3310 (a), NBRC 3508 (b), NBRC 12655 (c), NBRC 12656 (d), NBRC 14053 (e), NBRC 14083 (f), and NBRC 14084 (g).

Fig. 3 MALDI mass spectra of *P. syringae*: NBRC 3310 (a), NBRC 3508 (b), NBRC 12655 (c), NBRC 12656 (d), NBRC 14053 (e), NBRC 14083 (f), and NBRC 14084 (g).

Novel bacterial classification method by MALDI-TOF MS based on ribosomal protein coding in *S10-spc-alpha* Operon at Strain level

In this analysis, 2 ribosomal subunits, L24 and S17 which are encoded in the *S10-spc-alpha* operon, were biomarkers for discrimination of these strains. Furthermore, MALDI-TOF MS analysis revealed that 2 ribosomal subunits, S12 and S16, play a useful role in discrimination of the strain of *P. syringae*.

Using these ribosormal protein biomarkers, the 7 sample strains were classified into 5 clusters as color-coded (Table 2). Each cluster seems to reflect their pathovar. Further analysis of other *P. syringae* strains is needed to confirm this point.

Table 2 Ribosomal protein profiling table of *P. syringae* strains.

Pathovar	Strain	Ribosomal protein types															
		L36	L30	L29	S17	S19	L23	L24	S14	S10	L22	L18	S13	S11	S08	S16	S12
Genomesequenced strains																	
Aesculi	2250	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
Aesculi	NCPPB 3681	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
Phaseolicola	1448A	I	I	I	I	I	I	I	II	III	I	I	I	I	I	I	I
Tabaci	ATCC 11528	I	I	I	I	I	I	I	II	II	I	I	I	I	I	I	I
Syringae	B728a	I	I	I	I	I	I	II	III	II	I	I	I	I	I	I	I
Syringae	642	I	II	I	I	I	I	II	II	II	I	I	I	I	I	I	I
Oryzae	1_6	I	I	I	I	I	I	III	II	II	I	I	I	I	I	-	I
Tomato	K40	I	I	I	I	II	I	I	II	II	I	I	I	I	I	I	II
Tomato	Max13	I	I	I	I	II	I	I	II	II	I	I	I	I	I	II	I
Tomato	NCPPB 1108	I	I	I	I	II	I	I	II	II	I	I	I	I	I	II	I
Tomato	DC3000	I	I	I	I	II	I	I	II	II	I	I	I	I	I	II	I
Tomato	T1	I	III	I	I	II	I	I	II	II	I	-	I	I	I	II	I
Sample strains																	
Coronafaciens	NBRC3310	I	I	I	I	I	I	III	II	I	I	I	I	-	I	II	I
Tabaci	NBRC3508	I	I	I	I	I	I	I	II	I	I	I	I	I	I	I	I
Aptata	NBRC12655	I	I	I	I	I	I	II	II	I	I	I	I	I	I	I	I
Phaseolicola	NBRC12656	I	I	I	I	I	I	I	II	I	I	I	I	I	I	I	I
Mori	NBRC14053	I	I	I	I	I	I	I	II	I	I	I	I	I	I	II	III
Mori	NBRC14083	I	I	I	I	I	I	I	II	I	I	I	I	I	I	II	III
Mori	NBRC14084	I	I	I	II	I	I	I	II	I	I	I	I	I	I	II	I

Ⅰ：Type of the ribosomal proteins of *P. syringae* 2250 strain.
Ⅱ and Ⅲ：the mass of the biomarkers of other *P. syringae* strains which are differed from type Ⅰ.

4. Conclusions

- The *S10*-GERMS (*S10-spc-alpha* operon gene encoded ribosomal protein mass spectrum) method is a significantly useful tool for bacterial discrimination at strain level because WC-MS is significantly simpler and faster than DNA-based classification methods.
- Using gene sequence of *S10-spc-alpha* operon, a considerable number of WC-MS peaks of bacterial strain can be assigned even if the strain is not genome sequenced.
- This *S10*-GERMS and relevant method may have potential for the rapid discrimination of *P. syringae* at the pathovar as well as the strain level.
- The *S10*-GERMS method can expand the potential of the WC-MS into strain typing, which is important for bacterial outbreak analysis and tracking sources of bacterial contamination.

Novel bacterial classification method by MALDI-TOF MS based on ribosomal protein coding in *S10-spc-alpha* Operon at Strain level

5. Acknowledgement

This study was financially supported by the Ministry of Education, Culture, Sports, Science and Technology (HT), the Japan Science and Technology Agency (HT), Aichi Science and Technology Foundation (HT), and Research Fellowships from the Japan Society for the Promotion of Science for Young Scientists (YH).

6. Reference

1). Pineda, F. J. *et al.*, Anal. Chem., (2003), **75**, 3817-22
2). Teramoto, K. *et al.*, Anal. Chem., (2007), **79**, 8712-8719
3). Hotta, Y. *et al.*, *J. Proteome Res.*, (2011), **9**, 6722–6728
4). Tamura, H. *et al.*, *J. Am. Soc. Mass Spectrom.*, (2013) doi:10.1007/s13361-013-0627-8

First Edition: June, 2013

SHIMADZU
Shimadzu Corporation
www.shimadzu.com/an/