The Journal of Microbiology, October 2009, p. 646-650 Vol. 47, No. 5
DOI 10.1007/s12275-009-0092-z

Development of a Method Based on Surface Enhanced Laser Desorption and Ionization Time of Flight Mass Spectrometry for Rapid Identification of *Klebsiella pneumoniae*

Daiwen Xiao[1], Yongchang Yang[1], Hua Liu[1], Hua Yu[1], Yingjun Yan[1], Wenfang Huang[1]*, Wei Jiang[1], Weijin Liao[1], Qi Hu[2], and Bo Huang[2]

[1]Clinical Laboratory Department, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Chengdu 610072, P. R. China
[2]Laboratory Medicine Department, Chongqing Medical University, Chongqing 400016, P. R. China

(Received March 24, 2009 / Accepted June 14, 2009)

A method based on surface enhanced laser desorption and ionization time of flight mass spectrometry (SELDI-TOF MS) was developed for the rapid identification of *Klebsiella pneumoniae* by directly applying bacterial colonies without further protein extraction. A total of 40 *K. pneumoniae* and 114 other related microorganisms isolated clinically were analyzed by SELDI-TOF MS. An identification model for *K. pneumoniae* was established by artificial neural networks (ANNs) with classification accuracy of 100%. The model was blindly tested with 43 *K. pneumoniae* and 53 control bacteria again. The results showed that the model was successful with accuracy of 96.9%, sensitivity of 100% and specificity of 94.3%. This strategy is potential for rapid identification of *K. pneumoniae*.

***Keywords*:** identification, *K. pneumoniae*, SELDI-TOF MS

Rapid and accurate classification of microorganisms is essential for both disease treatment and bacterial taxonomy. Now, there are several different ways for classifying microorganisms based on the analysis of chemical signatures or by using molecular microbiology methods such as sequencing of 16S rDNA and selected protein coding genes. But currently used identification methods are labor intense, time consuming and inadequate on sensitivity and specificity. For example, PCR-based methods are available but not appropriate for classification, especially in the case of unknown bacterial samples. So there is an urgent need to develop direct, sensitive, and specific methods for bacterial identification.

Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) has been used widely for identification and typing of microorganisms. It can directly examine the profile of proteins from the intact bacterial cell surface, which provides a reproducible spectrum within minutes (Lynn *et al.*, 1999; Carbonnelle *et al.*, 2007; Degand *et al.*, 2008; Hsieh *et al*., 2008). Besides MALDI-TOF MS, surface enhanced laser desorption ionization time of flight mass spectrometry (SELDI-TOF MS) is another effective tool for proteomics research. A number of published articles have reported using this technology in disease diagnosis such as cancer, inflammation, and pulmonary arterial hypertension (Mendrinos *et al.*, 2005; Semmes *et al.*, 2005; Abdul-Salam *et al.*, 2006; De Torre *et al.*, 2006). In recent years, this technology has been successfully applied to classify species and strains with bacterial lysates (Schmid *et al.*, 2005; Seibold *et al.*, 2007).

K. pneumoniae are opportunistic pathogens that cause septicaemia, pneumonia, urinary tract infection and soft tissue infection. Pneumonia caused by *K. pneumoniae* is difficult to treat and leads to mortality rates up to 50%. Successful treatment of *K. pneumoniae* infection is dependent on the early diagnosis and prompt administration of the appropriate antibiotic. In this work, we proposed to directly identify *K. pneumoniae* by SELDI-TOF MS on AU (golden) proteinchip arrays. Before SELDI-TOF MS analysis, *K. pneumoniae* was confirmed by conventional procedures for microorganism identification and DNA sequencing. Then identification model for *K. pneumoniae* was established and validated by artificial neural networks (ANNs).

Materials and Methods

Materials

The 250 strains isolated from clinical microbiology laboratory in our hospital were characterized by BD fully automatic microbiological identification system and stored at -80°C before culturing, including 83 *K. pneumoniae*, 50 *S. aureus*, and 117 other *Enterobacteriaceae* bacteria. Protein calibration standards (insulin, cytochrome C, and myoglobin), ACN (acetonitrile), TFA (trifluoroacetic acid), and SPA (sinapic acid) were purchased from Sigma-Aldrich Corporation (USA), AU proteinchip arrays and SELDI-TOF MS, Biomarker Wizard Software were obtained from Ciphergen Corporation (USA).

Identification of *K. pneumoniae* by *dnaJ* gene sequencing

K. pneumoniae strains were further confirmed by molecular

* To whom correspondence should be addressed.
(Tel) 86-28-8739-4056; (Fax) 86-28-8739-4056
(E-mail) Wenfang.huang@163.com

method. Briefly, genomic DNA of pure culture strains was obtained by heating cells for 10 min at 100°C with DNA extraction reagents. The suspension was centrifuged for 5 min at 10,000×g. One microliter supernatant was used for the PCR. The primers used (Pham *et al.*, 2007) were DN1-1F; 5'-GATYTRCGHTAYAACATGGA-3' and DN1-2R; 5'-TTCACRCCRTYDAAGAARC-3' (product size, 758 bp). PCR cycle conditions were as follows: 94°C for 3 min, followed by 30 cycles of 94°C for 1 min, 58°C for 1 min and 72°C for 1 min. Final extension was carried out at 72°C for 10 min. PCR products were sequenced by ABI3100. The results were compared with sequences of GenBank database for confirmation of these microorganisms.

Sample preparation for SELDI-TOF MS

Two independently prepared groups of bacteria were used. Group1 (training group) composed of 40 *K. pneumoniae*, 30 *S. aureus*, 31 *E. coli*, 14 *E. cloacae*, 8 *K. oxytoca*, 9 *M. morganiii*, 8 *P. mirabilis*, 10 *S. marcescens*, and 4 *C. freundii* was used for construction of diagnosis model. Group2 (blind test

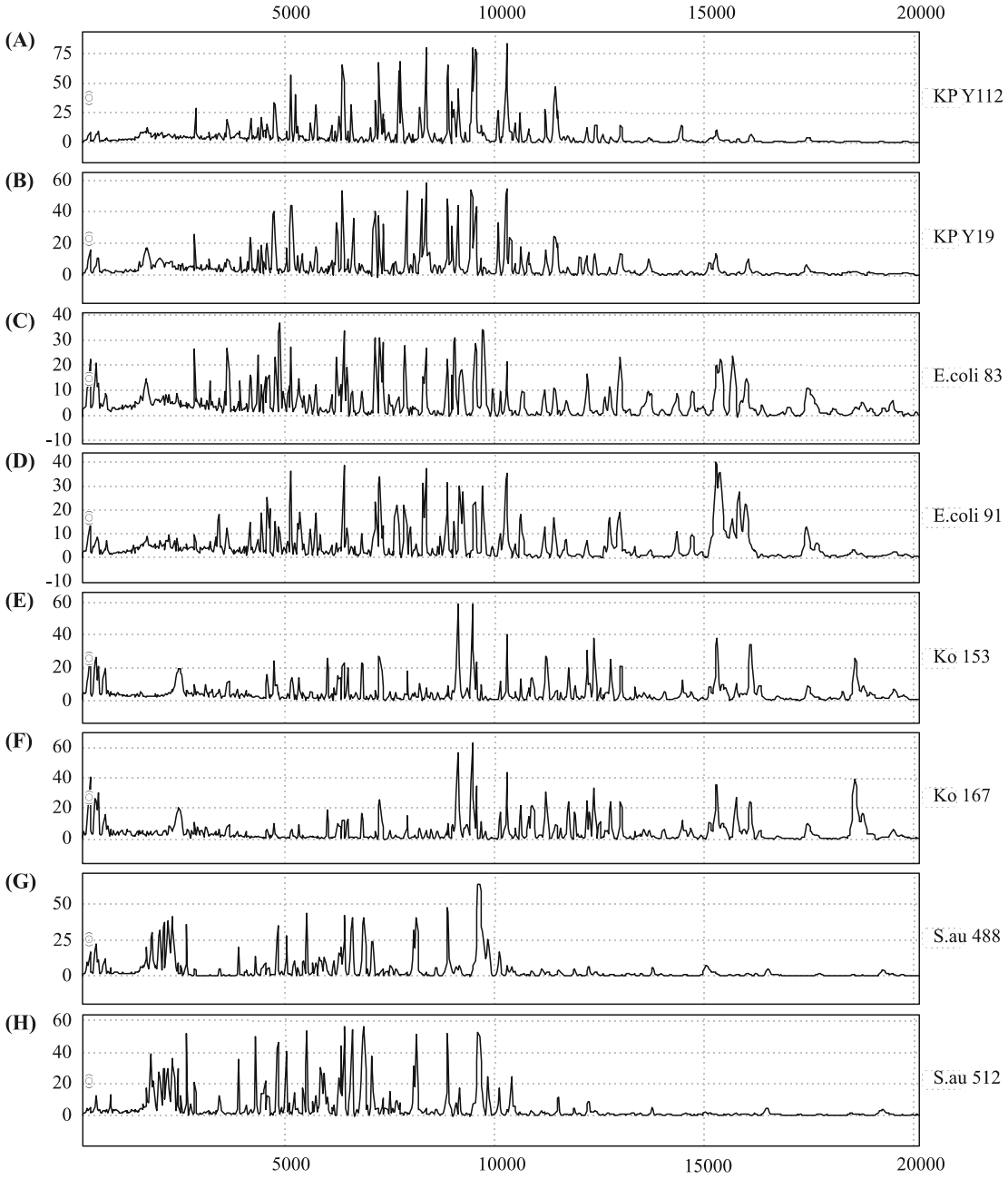

Fig. 1. The representative SELDI-TOF MS spectra for intact bacteria. The mass spectrometric profiles of intact bacteria were obtained and analyzed with Ciphergen Proteinchip Software. Different species had their characteristic protein profiles, while the same species had similar protein profiles. (A), (B) *K. pneumoniae*; (C), (D) *E. coli*; (E), (F) *K. oxytoca*; (G), (H) *S. aureus*

Table 1. Six protein peaks used for construction of ANNs diagnosis model

Mass (kDa)	p	Mean 1	SD 1	Mean 2	SD 2
6.38	0	34.58	13.45	4.39	6.00
7.23	0	35.92	12.33	1.44	2.64
7.73	0	25.70	15.00	0.38	1.63
9.12	0	31.49	10.64	5.98	8.25
9.46	0	45.70	15.05	7.80	14.31
12.57	0	4.633	3.36	0.68	1.60

The spectra obtained were analyzed by Biomarker Wizard Software, the 6 protein peaks were differently expressed and were considered as biomarkers by ANNs to discriminate *K. pneumoniae* from other bacteria. Mean1: mean of *K. pneumoniae*; SD1: standard deviation of *K. pneumoniae*; Mean2: mean of control bacteria; SD2: standard deviation of control bacteria

group) composed of 43 *K. pneumoniae*, 24 *E. coli*, 20 *S. aureus*, 5 *E. cloacae*, 4 *S. marcescens* was used to test the diagnosis model blindly.

Bacteria stored at -80°C were cultured on MacConkey or blood plates for one night. Then single colony was selected and incubated on another plate for 24 h at 35°C. Bacteria collected in 1.5 ml Eppendorf tube were centrifuged at 10,000×g for 5 min. The sediment was washed twice and re-suspended with deionized water. Three microliter bacteria suspension was mixed with the same volume of half-saturation SPA solution containing 50% ACN and 0.5% TFA. One microliter mixture was pipetted onto AU proteinchip arrays. After the spots dried, 1 μl half-saturation SPA was applied and left to dry for 10 min.

SELDI-TOF MS data collection and analysis

The AU proteinchip arrays were analyzed by SELDI-TOF MS, applying an optimization mass range from 3 kDa to 30 kDa and a mean laser intensity and detector sensitivity of 190 and 8 respectively. Spectra were collected and analyzed with ProteinChip Software (Version 3.1.0, Ciphergen Biosystems, USA). Mass charge was calibrated with a mixture of standards including insulin (5,733.49 Da), cytochrome C (12,230 Da), and myoglobin (16,950 Da). Biomarker Wizard (Ciphergen Biosystems, USA) was then used to identify corresponding peaks in each spectrum within 0.3% of the mass. The signal-to-noise ratio was set to 5 for the first pass and to 2 for the second pass. Intensity values for each peak were averaged for each sample.

Development of ANNs model

After analysis with Biomarker Wizard Software, a three-layer multi-layer perceptron (MLP) ANNs with a feed forward back propagation (BP) algorithm was used to construct classification model. The nodes of input layer, hidden layer and output layer were set at 6, 5, 1, respectively. The intensities of 6 significant difference peaks ($P<0.0000001$) selected were input to the nodes. The objective output value of 40 *K. pneumoniae* in the training group was set at 1, and the objective output value of 114 control bacteria in the same group was set at 0. The training frequency was set at 1000, and the learning rate was set at 0.01. A random initialization of the network was performed. The artificial neural network model was trained on data from the training group and stored.

Blind test of ANNs model

The model was blindly tested with the data from 98 bacteria. When the output value less than 0.5 was labeled as control, conversely, the output value not less than 0.5 was labeled as *K. pneumoniae*.

Results

Identification of *K. pneumoniae* by *DnaJ* gene sequencing

83 *K. pneumoniae* strains confirmed by BD automatic microbial identification system were further identified by *DnaJ* gene sequencing. *DnaJ* gene fragments of expected size about 750 bp were obtained. After compared with GenBank database, the sequences of all strains matched perfectly with corresponding type strain (matching rate ≥98.5%).

Protein profiles results

All *K. pneumoniae* strains and other control bacteria were measured by SELDI-TOF MS. The representative SELDI-TOF MS spectra for intact *K. pneumoniae*, *E. coli*, *K. oxytoca*, and *S. aureus* cells were shown in Fig. 1, where a substantial number of peaks were observed in the mass range 5 to 20 kDa. By visual inspection, *K. pneumoniae* strains showed similar spectra but different protein peaks compared with other bacteria. Data were analyzed using Biomarker Wizard Software. After automatic baseline noise calibration, the spectra were normalized by the "total ion current" method as described by the manufacturer. About 380 protein peaks were detected by SELDI-TOF-MS between 2,000 and 30,000 m/z in bacteria of the training group, among which 30 peaks were significantly different between *K. pneumoniae* and controls ($P<1\times10^{-10}$).

Establishment of ANNs model

30 protein peaks from the training group with significant difference ($P<1\times10^{-10}$) were used to construct diagnosis model by ANNs. The model was trained and tested by using 154 training samples. The results showed that the model constructed by 6 protein peaks at 6.38 kDa, 7.23 kDa, 7.73 kDa, 9.12 kDa, 9.46 kDa, and 12.57 kDa could distinguish *K. pneumoniae* from control bacteria perfectly with classification accuracy of 100% (Table 1).

Blind test of model

The model was tested again by blind test group, consisting of 43 *K. pneumoniae* and 53 control bacteria. The results

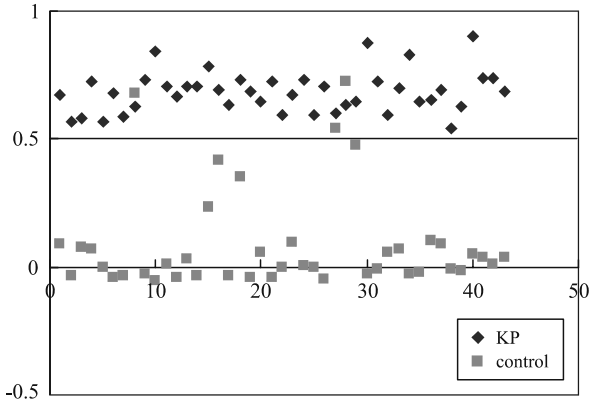

Fig. 2. Distribution of the cases of blind test group. When the output value < 0.5 was judged as control, conversely, the output value ≥0.5 was labeled as *K. pneumoniae*. 43 *K. pneumoniae* isolates had been correctly classified, while only 3 bacteria from 53 control bacteria were wrongly identified.

showed that the accuracy of the model was 96.9%. The sensitivity was determined to be 100%, which means that 43 *K. pneumoniae* samples could be correctly classified as positives. Conversely, the specificity of the test was determined to be 94.3%, which corresponds to an identification of 50 true negatives in a population of 53 *K. pneumoniae* negative samples (Fig. 2).

Discussion

In recent years, the incidence of nosocomial bacteremia caused by Gram-negative aerobic bacilli has increased markedly. After *E. coli*, *K. pneumoniae* is the second most frequent agents of Gram-negative bacteria caused severe morbidity and mortality among newborn, the elderly and immunocompromised patients. Now, there is some controversy about which is the best way of classifying bacteria as definitive identification by conventional methods based on microbiologic procedures, antibody recognition and PCR amplification may be cumbersome (Hansen *et al.*, 2004).

High sensitivity, specificity, and simple preparation procedures are indispensable for ideal protocol in classifying and identifying human pathogenic microorganisms. With the development of proteomics, MALDI-TOF MS with high efficiency for proteome profiling was used to microbial identification. Some articles have showed the methods based on extraction of bacterial proteins employed for bacterial identification by MALDI-TOF MS could be successful identification and characterization of microorganisms. But they have not been used widely because of relatively time-consuming, technique demanding, or have low reproducibility. In recent years, many researchers have proved that directly applying bacterial colonies without further protein extraction to MALDI-TOF MS analysis is a simple and reliable method for rapid protein profiling (Hettick *et al.*, 2006; Hsieh *et al.*, 2008).

SELDI-TOF MS is a mass spectrometric method which allows selective absorption of proteins on a chromatographic array surface, so it can directly analyze bacteria or other protein-containing samples with minimizing sample losses and accelerating sample preparation. In this study, we used SELDI-TOF MS to detect the intact bacteria without special protein extraction. Before the examination, *K. pneumoniae* strains were confirmed by BD automatic microbial identification system and *DnaJ* gene sequencing. The *DnaJ* gene encoding heat shock protein 40 that acts as a functional cohort of *DnaK* in various aspects of protein dynamics was reported to be a suitable target for phylogenetic study and identification of bacteria such as Legionella (Liu *et al.*, 2003), Streptococcus (Itoh *et al.*, 2006), and Enterobacteriaceae species (Pham *et al.*, 2007). Compared with 16S rDNA, *tuf*, and *atpK* genes, *DnaJ* gene showed better resolution for differentiating closely related enterobacteria at the species level. Our results also proved the matching rate of *K. pneumoniae* comparing with GenBank database was satisfactory.

In order to analyze vast amounts of proteomic data gained by SELDI-TOF MS, we employed ANNs software to develop and validate identification model. ANNs is a form of machine learning capable of accurately modelling proteomic data and identifying biomarkers which are capable of discriminating between one class and another. It has been used to the diagnosis of cancers (Ball *et al.*, 2002; Anagnostou *et al.*, 2003; Gamito and Grawford, 2004) and cardiovascular disease (Baldassarre *et al.*, 2004). Moreover, several models for identification of bacteria have been successfully built and trained (Lancashire *et al.*, 2005; Schmid *et al.*, 2005).

In this study, a model had been established by ANNs and showed astonished performance in discrimination *K. pneumoniae* from control bacteria on the species level. Our results demonstrated that this strategy based on SELDI-TOF MS was feasible for bacterial identification in clinical laboratories. Furthermore, abundant mass spectral information was obtained directly from whole bacterial cells without preliminary cellular protein extraction. Compared to other techniques, the method is inexpensive and hundreds of samples can be analyzed per day. Recently, a report (Hsieh *et al.*, 2008) has proved that 5×10^3 and 3×10^4 was the minimum number of cells for bacterial identification in a pure strain and in bacterial mixtures by MS techniques respectively, which suggested that a rapid and accurate identification of pathogenic microorganisms without culture will be possible in the future.

In summary, directly applying a bacterial colony to SELDI-TOF MS provides a highly accurate and rapid method for *K. pneumoniae* identification. And this strategy may play an important role to identification of pathogenic bacteria in clinical microbiology laboratory.

References

Abdul-Salam, V.B., G.A. Paul, J.O. Ali, S.R. Gibbs, D. Rahman, G.W. Taylor, M.R. Wilkins, and R.J. Edwards. 2006. Identification of plasma protein biomarkers associated with idiopathic pulmonary arterial hypertension. *Proteomics* 6, 2286-2294.

Anagnostou, T., M. Remzi, M. Lykourinas, and B. Djavan. 2003. Artificial neural networks for decision-making in urologic oncology. *Eur. Urol*. 43, 596-603.

Baldassarre, D., E. Grossi, M. Buscema, M. Intraligi, M. Amato, E. Tremoli, L. Pustina, S. Castelnuovo, S. Sanvito, L. Gerosa, and C.R. Sirtori. 2004. Recognition of patients with cardiovas-

cular disease by artificial neural networks. *Ann. Med.* 36, 630-640.

Ball, G., S. Mian, F. Holding, R.O. Allibone, J. Lowe, S. Ali, G. Li, S. McCardle, I.O. Ellis, C. Creaser, and R.C. Rees. 2002. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. *Bioinformatics* 18, 395-404.

Carbonnelle, E., J.L. Beretti, S. Cottyn, G. Quesne, P. Berche, X. Nassif, and A. Ferroni. 2007. Rapid identification of Staphylococci isolated in clinical microbiology laboratories by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol*. 45, 2156-2161.

Degand, N., E. Carbonnelle, B. Dauphin, J.L. Beretti, M. Le-Bourgeois, I. Sermet-Gaudelus, C. Segonds, P. Berche, X. Nassif, and A. Ferroni. 2008. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of non-fermenting gram-negative bacilli isolated from cystic fibrosis patients. *J. Clin. Microbiol*. 46, 3361-3367.

De Torre, C., S.X. Ying, P.J. Munson, G.U. Meduri, and A.F. Suffredini. 2006. Proteomic analysis of inflammatory biomarkers in bronchoalveolar lavage. *Proteomics* 6, 3949-3957.

Gamito, E.J. and E.D. Crawford. 2004. Artificial neural networks for predictive modeling in prostate cancer. *Curr. Oncol. Rep*. 6, 216-221.

Hansen, D.S., H.M. Aucken, T. Abiola, and R. Podschun. 2004. Recommended test panel for differentiation of *Klebsiella* species on the basis of a trilateral interlaboratory evaluation of 18 biochemical tests. *J. Clin. Microbiol*. 42, 3665-3669.

Hettick, J.M., M.L. Kashon, J.E. Slaven, Y. Ma, J.P. Simpson, P.D. Siegel, G.N. Mazurek, and D.N. Weissman. 2006. Discrimination of intact mycobacteria at the strain level: a combined MALDI-TOF MS and biostatistical analysis. *Proteomics* 6, 6416-6425.

Hsieh, S.Y., C.L. Tseng, Y.S. Lee, A.J. Kuo, C.F. Sun, Y.H. Lin, and J.K. Chen. 2008. Highly efficient classification and identification of human pathogenic bacteria by MALDI-TOF MS. *Mol. Cell. Proteoms* 7, 448-456.

Itoh, Y., Y. Kawamura, H. Kasai, M.M. Shah, P.H. Nhung, M. Yamada, X. Sun, T. Koyana, M. Hayashi, K. Ohkusu, and T. Ezaki. 2006. *DnaJ* and gyrB gene sequence relationship among species and strains of genus *Streptococcus*. *Syst. Appl. Microbiol*. 29, 368-374.

Lancashire, L., O. Schmid, H. Shah, and G. Ball. 2005. Classification of bacterial species from proteomic data using combinatorial approaches incorporating artificial neural networks, cluster analysis and principal components analysis. *Bioinformatics* 21, 2191-2199.

Liu, H., Y. Li, X. Huang, Y. Kawamura, and T. Ezaki. 2003. Use of the *dnaJ* gene for the detection and identification of all *Legionella pneumophila* serogroups and description of the primers used to detect 16S rDNA gene sequences of major members of the genus *Legionella*. *Microbiol. Immunol*. 47, 859-869.

Lynn, E.C., M.C. Chung, W.C. Tsai, and C.C. Han. 1999. Identification of Enterobacteriaceae bacteria by direct matrix-assisted laser desorptiom/ionization mass spectrometric analysis of whole cells. *Rapid. Commun. Mass. Spectrom.* 13, 2022-2027.

Mendrinos, S., J.D. Nolen, T. Styblo, G. Carlson, J. Pohl, M. Lewis, and J. Ritchie. 2005. Cytologic findings and protein expression profiles associated with ductal carcinoma of the breast in ductal lavage specimens using surface-enhanced laser desorption and ionization-time of flight mass spectrometry. *Cancer* 105, 178-183.

Pham, H.N., K. Ohkusu, N. Mishima, M. Noda, M. Monir-Shah, X. Sun, M. Hayashi, and T. Ezaki. 2007. Phylogeny and species identification of the family Enterobacteriaceae based on *dnaJ* sequences. *Diagn. Microbiol. Infect. Dis*. 58, 153-161.

Schmid, O., G. Ball, L. Lancashire, R. Culak, and H. Shah. 2005. New approaches to identification of bacterial pathogens by surface enhanced laser desorption/ionization time of flight mass spectrometry in concert with artificial neural networks, with special reference to *Neisseria gonorrhoeae*. *J. Med. Microbiol*. 54, 1205-1211.

Seibold, E., R. Bogumil, S. Vorderwülbecke, S. Al-Dahouk, A. Buckendahl, H. Tomaso, and W. Splettstoesser. 2007. Optimized application of surface-enhanced laser desorption/ionization time-of-flight MS to differentiate *Francisella tularensis* at the level of subspecies and individual strains. *FEMS Immunol. Med. Microbiol*. 49, 364-373.

Semmes, O.J., Z. Feng, B.L. Adam, L.L. Banez, W.L. Bigbee, D. Campos, L.H. Cazares, D.W. Chan, W.E. Grizzle, E. Izbicka, J. Kagan, G. Malik, D. McLerran, J.W. Moul, A. Partin, P. Prasanna, J. Rosenzweig, L.J. Sokoll, S. Srivastava, S. Srivastava, I. Thompson, M.J. Welsh, N. White, M. Winget, Y. Yasui, Z. Zhang, and L. Zhu. 2005. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin. Chem.* 51, 102-112.