Original Article
Application of MALDI-TOF MS for rapid identification and virulence analysis of *Helicobacter pylori*

Fei Chen[1*], Danian Ji[2*], Jinghao Zhang[1], Feng Yang[1], Jun Zhou[2], Shiwen Wang[1], Yi Fang[1], Michal A Olszewski[4], Zhijun Bao[3], Ping Xiang[2], Yanmei Zhang[1,5,6], Hu Zhao[1,5,6]

Departments of [1]Laboratory Medicine, [2]Endoscopy, [3]Gastroenterology, Huadong Hospital Affiliated to Fudan University, Shanghai, China; [4]Department of Internal Medicine, Division of Pulmonary and Critical Care Medicine, University of Michigan Medical School and Department of Veterans Affairs Ann Arbor Health System, Ann Arbor, MI 48105, USA; [5]Shanghai Key Laboratory of Clinical Geriatric Medicine, Shanghai, China; [6]Research Center on Aging and Medicine, Fudan University, Shanghai, China. *Equal contributors.

Abstract: Purpose: The aim of this study was to extend the reference database of MALDI-TOF MS and enhance its efficiency correctly identifying clinical *H. pylori* isolates, while assessing their virulence potential. Methods: Clinical *H. pylori* isolates, obtained from fresh upper GI biopsy samples, were prepared for analysis using several pretreatment methods to optimize MALDI-TOF MS detection. Major spectra projections (MSP), generated through analysis of the isolates, were compared against the original reference database in MALDI Biotyper CA System. Dendrogram and virulence analysis models were generated and evaluated to determine the virulence of *H. pylori.* Results: 70% formic acid was found to be the optimal pretreatment reagent for MALDI-TOF MS analysis of *H. pylori*. For 31 validated isolates, the identification capacity of the database was improved from 7 (22.6%) to 29 (93.5%), upon expanding the original reference database with data from the 37 clinical *H. pylori* isolates. Created MSP dendrogram and virulence analysis models achieved 100% accuracy, 60% specificity, 83.3% positive predictive value, and 100% negative predictive value for determining the virulence of gene *dupA.* Conclusion: Application of formic acid for sample pretreatment increased efficiency of detection and virulence analysis of *H. pylori* for MALDI-TOF MS.

Keywords: MALDI-TOF MS, *Helicobacter pylori*, identification, virulence analysis

Introduction

Helicobacter pylori (*H. pylori*) has been recognized as an etiological agent closely related to occurrence and development of acute and chronic gastritis and gastroduodenal ulcers. It has been recognized as an important factor for occurrence of gastric malignancies with a worldwide infection rate approaching 50% [1, 2]. Association of *H. pylori* pathogenicity with the presence of virulence genes, such as *dupA,* has been shown by several studies [3-5]. Consequently, virulence gene detection/analysis has become increasingly explored as a potential predictor of overall risk and potential severity of GI complications that may arise from infection with specific strains of *H. pylori* [6]. Thus, development and broader application of rapid *H. pylori* detection methods that simultaneously assess genetic predictors of pathogenicity have become important priorities in laboratory diagnostics.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is a new soft ionization mass spectrometry system, successfully applied for effective detection of some bacteria, yeast, and other microorganisms [7-10]. A few studies [11-13] have reported on MALDI-TOF MS detection capability of *H. pylori*, however, sample sizes were not sufficient and operations were complicated.

The aim of this present study was to test MALDI-TOF MS efficacy for *H. pylori* detection and to explore potential for its further enhancement via application of modified specimen pre-treatments and through incorporation of simultaneous analysis of virulence genes. This was achieved by: 1) Optimization of the pretreat-

ment procedure to obtain high-quality spectra; 2) Construction of new reference database for improved spectra analysis; 3) Testing MALDI Biotyper capability for distinguishing the virulence gene of *H. pylori* strains expanding bioinformatics tools integrated in MALDI-TOF MS; and 4) Validating the reliability of the new database and virulence analysis models developed using *H. pylori* clinical isolates.

Materials and methods

Bacterial isolation and phenotypic identification

This study was approved by the Ethics Committee for Human Studies of Huadong Hospital Affiliated to Fudan University (Ethics approval number: [2013]-077) and conducted with strict adherence to approved protocols. At total of 351 gastric mucosal tissue specimens were collected in Huadong Hospital, from June 2015 to July 2016, including 186 males and 165 females, with an average age of 50.3±13.8 years old. Informed consent was obtained from all individual participants included in the study. All gastric biopsy samples were collected by designated GI physicians at Huadong Hospital after receiving patient consent. All collected biopsies were placed in specimen caps with normal saline and stored at 4°C for less than 4 hours before further processing. Following thorough homogenization with tissue lyser (Jingxin Co., Ltd., Shanghai, China) under sterile conditions, samples were plated on *Columbia agar* medium (OXOID, Thermo Fisher Scientific Inc., USA), containing 8% sterile sheep blood and 0.5% selective antibiotics supplement. Agar was kept under microaerophilic conditions (5% O_2, 10% CO_2, 85% N_2) at 35°C for 5-7 days. Suspected colonies (colorless, translucent, 0.5~1 mm in diameter) grown on *H. pylori* selective plates were given Gram staining urease, oxidase, and catalase tests. Colonies cultured for 48-72 hours were collected, transferred to 30% glycerol Bruce broth (OXOID, Thermo Fisher Scientific Inc., USA), and preserved at -80°C until analyzed [4].

H. pylori isolates recovery

Containers with frozen stocks of *H. pylori* isolates were rapidly recovered in 37°C water baths. Stocks were replated on Columbia blood agar and cultured under conditions of 5% O_2 and 10% CO_2 at 35°C for 3-5 days.

16S rRNA and dupA gene sequencing

H. pylori colonies grown on the plates were transferred to the Eppendorf tube containing sterile normal saline and centrifuged at 13000 rpm for 60 seconds to collect bacteria. Total bacterial genome DNA of *H. pylori* isolates were extracted, referring to steps in the instructions of the bacterial DNA extraction kit (Tiangen Biotech, Beijing, China). Identification gene *16S rRNA* and virulence gene *dupA* were amplified individually from *H. pylori* isolates with Taq PCR MasterMix (Tiangen Biotech Co., Ltd.) and specific primers [4]. Total DNA of *H. pylori* clinical isolates were used as a template to carry out PCR amplification reaction with an automatic thermal cycler (Applied Biosystems, MD, USA). Reaction settings were as follows: 3 minutes at 95°C; 30 cycles of 30 seconds at 95°C, 30 seconds at 52-60°C, and 2 minutes at 72°C, with a final extension of 10 minutes at 72°C. PCR products were then purified and sequenced by Shanghai Sunny Biotechnology Company [4].

Sample preparation and MALDI-TOF MS analysis of H. pylori

Bacterial Test Standard (BTS) was used for external calibration of MALDI-TOF MS. For sample analysis, individual *H. pylori* colonies of clinical isolates were individually transferred from agar plate cultures to a MALDI target plate (Bruker Daltonics, Germany), using 1 µL sterile inoculation loop supplemented with 1 µL of pretreatment reagent, then air dried. Samples were then treated with 1 µL of a-cyano-4-hydroxycinnamic acid matrix solution (a-CHCA, Bruker Daltonics, Germany), air dried, and analyzed in Microflex LT MALDI-TOF mass spectrometer (Bruker Daltonics, Germany). Linear positive-ion mode, N_2 laser, λ = 355 nm, pulse duration: 150 ns, laser frequency: 200 Hz. Spectra were recorded over the range m/z 2000-20,000. Duplicate spots were prepared for each sample and triplicate spectra containing 200 shots each were generated from each spot.

Optimization of pretreatment conditions

Clinical *H. pylori* isolate *HP* 151205-1 with known genetic background was pretreated with

Table 1. Evaluation of the different reference databases[1]

	The cross validation				The external validation				Total			
	Numbers of strains (spectra)		Percentage of strains (spectra)		Numbers of strains (spectra)		Percentage of strains (spectra)		Numbers of strains (spectra)		Percentage of strains (spectra)	
Database	ORD	NRD	ORD	NRD	ORD	NRD	ORD	NRD	ORD	NRD	ORD	NRD
Species	2 (8)	37 (222)	5.4 (3.6)	100 (100)	2 (7)	17 (56)	6.5 (3.8)	54.8 (30.1)	4 (15)	54 (278)	5.9 (3.6)	79.5 (68.1)
Genus	1 (2)	0 (0)	2.7 (0.9)	0 (0)	5 (12)	12 (82)	16.1 (8.6)	38.7 (44.1)	6 (14)	12 (82)	8.8 (3.4)	17.6 (20.1)
Total	37 (222)	37 (222)	100 (100)	100 (100)	31 (186)	31 (186)	100 (100)	100 (100)	68 (408)	68 (408)	100 (100)	100 (100)

[1]ORD: The original reference database of Biotyper 3.1, containing seven H. pylori MSP included by the manufacturer (5989). NRD: The new reference database, enriched with 37 new MSPs added to the original reference database of Biotyper 3.1 (6026).

Table 2. Background of clinical *H. pylori* isolates used in this study[2]

Group	Sample no.	Diseases	*dupA* sequencing	Group	Sample no.	Diseases	*dupA* sequencing
G1-1	*HP* 151204-27	CAG	+	G2-1	*HP* 151224-14	CAG+DU	+
	HP 151204-14	CAG	-		*HP* 160128-10	N	+
	HP 151204-23	CAG+DU	-		*HP* 151008-12	DU	+
	HP 151204-11	CAG	-		*HP* 160114-6	CAG	+
G1-2	*HP* 150710-32	CAG+DU	+		*HP* 160114-11	CAG	-
	HP 151022-5	N	+	G2-2	*HP* 160716-31	CAG	+
	HP 151202-31	CAG+DU	+		*HP* 161210-5	CAG	-
	HP 151105-5	CAG+DU	+		*HP* 150724-38	CAG	-
	HP 150723-11	CAG	+		*HP* 160121-4	N	-
	HP 160123-11	CAG+DU	+		*HP* 160121-9	CAG	-
G2-1	*HP* 151119-3	N	+		*HP* 160121-7	N	-
	HP 151224-7	N	-	Others	*HP* 151224-13	N	-
	HP 151029-10	N	+		*HP* 160121-5	CAG+DU	-
	HP 150703-10	CAG+DU	-		*HP* 151202-42	CAG	-
	HP 151015-15	CAG	+		*HP* 151214-10	CAG	+
	HP 150827-8	CAG+DU	-		*HP* 151202-32	CAG	+
	HP 150724-21	CAG	+		*HP* 160526-43	CAG	+
	HP 151203-2	DU	+		*HP* 151202-12	CAG+DU+GU	+
	HP 150723-10	CAG	+				

[2]CAG: Chronic atrophy gastritis; N: Non-atrophic gastritis; DU: Duodenal ulcer; GU: Gastric ulcer; HP: *Helicobacter pylori*.

the following four different pretreatment reagents: 70% formic acid, 70% formic acid and acetonitrile (1:1), 50% isopropanol (containing 0.1% trifluoroacetic acid), and double distilled water. Each sample was treated with 1 µL of pretreatment reagents, respectively. After drying at room temperature, a-CHCA matrix solution was added to samples and dried with the air before MALDI-TOF MS analysis. Pretreatment efficiency of these reagents was evaluated by comparing the intensity and number of *H. pylori* specific peaks.

Construction of MSPs

MSPs were constructed based on 6 raw spectra of each isolate using MALDI Biotyper 3.1 software (Bruker Daltonics, Germany). All newly-created MSPs were added to the original reference database for cross validation and external validation. For subsequent cluster analysis, MSP dendrogram containing 44 MSPs (7 MSPs in the original database and 37 MSPs from clinical isolates) was generated. Based on outcomes of cluster analysis, isolates with common features were sorted into different subgroups. All new MSPs were added to the original database to further enhance *H. pylori* strain differentiation and the capability of updated MSP to distinguish different isolates based on clinical and genetic backgrounds of virulence.

Validation of the new reference database

Thirty-seven *H. pylori* isolates, including 222 spectra, were used for cross validation and 31 *H. pylori* isolates, including 186 spectra, were used for external validation. The highest score value of each isolate was selected as the identification result.

Construction of virulence analysis models using ClinProTools

ClinProTools (version 3.0.22; Bruker Daltonics) was used to generate classification models to differentiate among *dupA*-positive and *dupA*-negative *H. pylori* isolates. ClinProTools software was operated with the following settings: threshold for signal-to-noise (S/N) ratio was 5, relative intensity threshold was 1.5%, and maximal number of best peaks was 10.

Validation of MSP dendrogram and virulence analysis models of ClinProTools

Thirty-one *H. pylori* isolates were used to validate the identification ability of MSP dendro-

Table 3. Validation of MSP dendrogram and ClinProTools models[3]

Sample no.	Diseases	*dupA* sequencing	MDM	CM	Sample no.	Diseases	*dupA* sequencing	MDM	CM
HP 150716-10	CAG	-	G1-1	-	*HP* 160604-25	CAG	-	Others	+
HP 151202-14	CAG	-	G1-1	-	*HP* 151105-8	CAG	-	Others	+
HP 160105-5	CAG	-	G1-1	-	*HP* 150707-26	CAG	+	Others	+
HP 151119-10	CAG	+	G1-2	+	*HP* 151105-4	CAG	-	Others	+
HP 151015-11	CAG+GU	+	G1-2	+	*HP* 150709-22	CAG+DU	+	Others	+
HP 151204-14	CAG	-	G1-2	+	*HP* 151202-53	N	-	Others	+
HP 151205-51	CAG	+	G1-2	+	*HP* 151204-45	CAG	-	Others	+
HP 150910-8	DU	+	G1-2	+	*HP* 150703-22	CAG	-	Others	+
HP 160526-47	CAG	+	G1-2	+	*HP* 151105-10	N	-	Others	+
HP 151202-59	CAG	+	G2-1	+	*HP* 150605-24	CAG+DU	-	Others	-
HP 151224-8	CAG	+	G2-1	+	*HP* 151008-17	CAG+DU	-	Others	-
HP 151203-5	N	+	G2-1	+	*HP* 151205-52	CAG	-	Others	-
HP 150714-11	CAG	-	G2-1	+	*HP* 151202-52	CAG	-	Others	+
HP 151119-4	CAG	+	G2-1	+	*HP* 150917-14	CAG	-	Others	-
HP 150726-38	N	+	G2-1	+	*HP* 151205-14	CAG	-	Others	+
					HP 150917-16	N	+	Others	+

[3]CAG: Chronic atrophy gastritis; N: Non-atrophic gastritis; DU: Duodenal ulcer; GU: Gastric ulcer; HP: *Helicobacter pylori*; MDM: MSP Dendrogram Model; CM: ClinProTools Model.

gram and virulence analysis models of ClinProTools. Six raw spectra of each isolate were used to create one MSP, which was subsequently used to create an MSP dendrogram with the 37 newly constructed and 7 original MSPs for validation. One hundred and eighty-six spectra from 31 isolates were imported to ClinProTools to validate the virulence profiles of *H. pylori* isolates.

Results

Isolation rate and detection of dupA gene of clinical H. pylori isolates

To construct and validate the new reference database for MALDI-TOF, collection of confirmed clinical *H. pylori* isolates needed to be secured. From 351 gastric biopsy specimens, 68 cultures with *H. pylori* growth characteristics were cultured (positive rate of 19.4%). These isolates were confirmed to be *H. pylori* by gene sequencing of *16S rRNA*. Out of 68 clinical *H. pylori* isolates, data from the initial 37 were selected to be used to enrich the original database and establish virulence analysis models. The remaining 31 isolates were selected for validation of the new reference database and virulence analysis models. In the first set, 24 isolates were *dupA*-positive and 13 were *dupA*-negative. In the second set, 13 were

dupA-positive and 18 were *dupA*-negative. Genetic and clinical backgrounds of individual isolate samples are listed in **Tables 2** and **3**.

Pretreatment of specimens with formic acid allows for optimization of MALDI-TOF H. pylori detection

Selected clinical *H. pylori* isolates with known genetic backgrounds were used for optimization of pretreatment for MALDI-TOF MS detection. **Figure 1** shows the mass spectrum of clinical isolates from different pretreatment reagents: no pretreatment, ddH$_2$O, 50% isopropanol (containing 0.1% TFA), 70% formic acid, 70% formic acid, and acetonitrile (1:1), respectively. Compared with the other four pretreatment reagents, 70% formic acid showed the highest intensity and largest number of *H. pylori* specific peaks. Furthermore, ion peaks could be observed at the position after 16000 Da. Thus, among tested pretreatment reagents for MALDI-TOF MS, 70% formic acid was found to be optimal for detection of *H. pylori*.

Establishment and validation of the new reference database

To test the capability for cross-validation of ORD (original reference database), 2 isolates (5.4%) and 1 isolate (2.7%) were identified at
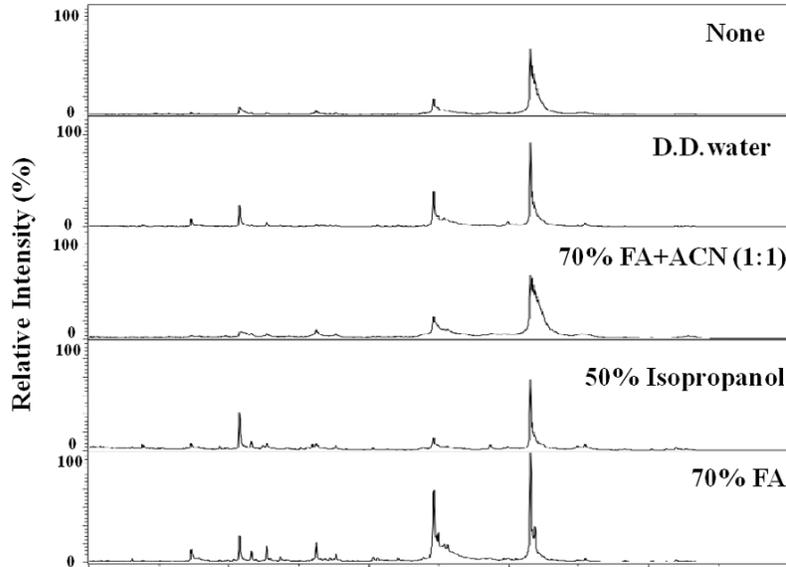
Figure 1. Optimization of the pretreatment method. Clinical isolates were recovered using various pretreatment reagents prior to analysis by MALDI-TOF analyzer. Images show individual mass spectra profiles (MSP) of the same representative clinical *H. pylori* isolate processed after recovery witch each reagent. Note optimal amplification of MSP in sample recovered in 70% formic acid. m/z: mass-to-charge ratio; MALDI matrix: α-CHCA.
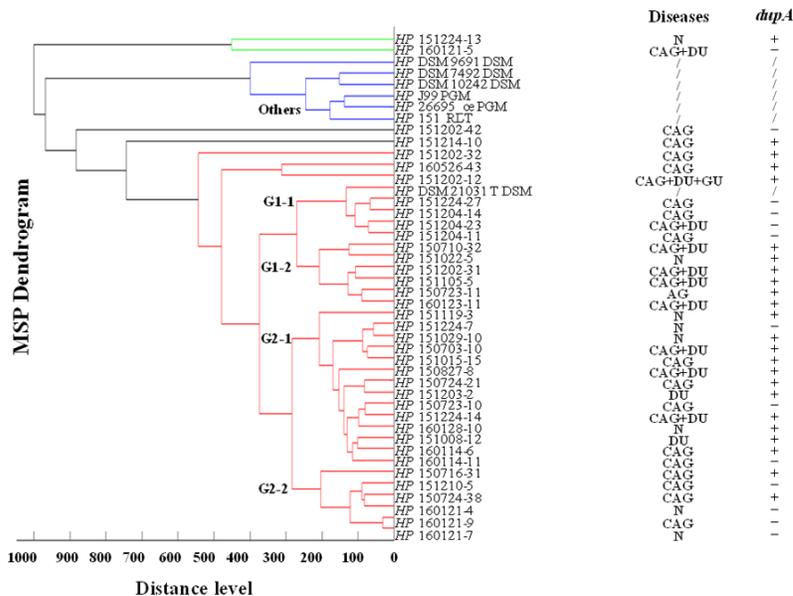
Figure 2. Cluster analysis of 37 clinical *H. pylori* isolates and 7 standard *H. pylori* strains. MSP dendrogram of the mass spectra was generated by MALDI Biotyper 3.1. The dendrogram was generated with default settings. Thirty out of 37 clinical isolates were divided into four groups. G1-1: Group 1-1; G1-2: Group 1-2; G2-1: Group 2-1; G2-2: Group 2-2. The remaining seven clinical isolates and six reference spectra were designated as "Others"; *HP*: *Helicobacter pylori*.

isolates (100%) were identified at the species level by the NRD (new reference database) (**Table 1**). Furthermore, 31 *H. pylori* isolates were used for external-validation of ORD and NRD. For external validation, 3 isolates (8.1%) could be correctly identified at the species level and genus level by the ORD. In contrast, 17 (54.8%) and 12 isolates (38.7%) were identified at the species level and genus level by the NRD, respectively (**Table 1**). In total, for 68 clinical isolates, 66 isolates (97.1%) were accurately identified, with 54 (79.5%) identified at the species level. The remaining 12 isolates (17.6%) were identified only at genus level. This was sufficient for positive diagnosis. Noticeably, the number of identified spectra vastly improved in the NRD relative to ORD (88.2% vs 7.0%) (**Table 1**).

The dupA profile associated with clinical characteristics by cluster analysis of MSP dendrogram

An MSP dendrogram was generated for seven standard *H. pylori* strains from the original database and 37 clinical *H. pylori* isolates. Six out of seven standard *H. pylori* strains MSPs, generated by the original manufacturer, were sorted to one separated group (**Figure 2**). Thirty out of 37 clinical isolates were divided into two groups: Group 1 (10 isolates) and Group 2 (20 isolates). Each group was divided into two subgroups: Group 1-1 (G1-1), Group 1-2 (G1-2), Group 2-1

the species level and genus level by comparison against the ORD, respectively. Thirty-seven

(G2-1), and Group 2-2 (G2-2). The proportion of G1-2 and G2-1 with duodenal/gastric ulcers was higher than that of G1-1 and G2-2 (90% vs 10%). Correspondingly, the number of *dupA*-positive isolates in G1-2 and G2-1 was also significantly higher than that in G1-1 and G2-2 (85% vs 20%). This phenomenon indicates that the profile of *dupA*-positive and *dupA*-negative may relate with clinical characteristics which could be reflected by different spectrum patterns.

Construction of virulence analysis models using ClinProTools

Apart from MSP dendrogram analysis, ClinProTools was further used to generate classification models to differentiate *dupA*-positive and *dupA*-negative *H. pylori* isolates. Twenty-four *dupA*-positive isolates and 13 *dupA-negative* isolates were used to create the classification models. Ten specific peaks at mass to charge ratios (m/z) of 7767, 8391, 7286, 6069, 2789, 6272, 7668, 13880, 6554, and 14019, generated by ClinProTools with the genetic algorithm, could reflect the profiles of the two genotypes. Using FlexAnalysis software, the signals of m/z 8391, 7286, and 7684 were further confirmed as specific for *dupA*-negative *H. pylori* isolates while m/z 7767 and 14019 were confirmed for *dupA*-positive isolates. The genetic algorithm model generated the highest recognition capability (99.29%) and cross validation (94.32%), indicating that *dupA*-positive/*dupA*-negative *H. pylori* isolates could be distinguished well.

Validation of MSP dendrogram and virulence analysis models

Thirty-one *H. pylori* isolates (186 spectra) were used to validate the capacities of virulence analysis models based on clustering analysis of MSP dendrogram and ClinProTools. Each analysis model, alone, could not give ideal results, but combination of these two virulence analysis models significantly improved the capability of virulence analysis. Resultant accuracy, specificity, positive predictive value (PPV), and negative predictive value (NPV) for determining virulence gene *dupA* of *H. pylori* isolates reached 100%, 60%, 83.3%, and 100%, respectively (**Table 3**).

Discussion

Epidemiological studies have shown that nearly half of the world's population is infected with *H.*

pylori [14, 15]. Virulence factor DupA has been closely linked with pathogenicity [5]. Identification and virulence analysis may provide useful information for suitable treatment. Conventional methods such as cultures, ^{13}C-urea breath test (UBT), rapid urease test (RUT), and histopathology cannot simultaneously detect and assess the virulence of *H. pylori*. This has triggered exploration and adaptation of other diagnostic methods for this purpose [4, 16].

In recent years, with advantages of quickness and low costs in identification of microorganisms, MALDI-TOF MS has been popularized in clinical application as an effective microbiological identification tool in clinical laboratories [7]. However, due to the low amount of reference spectra of *H. pylori*, the Biotyper database has poor capability for *H. pylori* identification [11, 12]. In this study, clinical *H. pylori* isolates confirmed by gene sequencing were added to the original database of Bruker MS to increase the identification and clinical diagnosis capability MALDI-TOF MS for *H. pylori* detection.

Pretreatment procedures were critical for MALDI-TOF MS in obtaining high-quality mass spectra [13]. Clinical *H. pylori* isolates with known genetic backgrounds were used to evaluate pretreatment efficiency. The two most common reagents for pretreatment, 70% formic acid and 70% formic acid/acetonitrile (1:1), were compared in the *H. pylori* pretreatment procedure [7]. The 70% formic acid treatment provided optimal spectra in terms of peak number, relative peak intensity, and resolution. Furthermore, ion peaks could still be observed at the position after 16000 Da (data not shown). In Cho's report, the greatest number of ions were detected when using 50% isopropanol compared with other reagents including ddH$_2$O, PBS, 50% ACN, and 50% EtOH, but formic acid was not compared [13]. Furthermore, the sample preparation procedures in the two studies were different. Cho used the supernatant of *H. pylori* cell extract solution while this study used the intact bacteria specimen, which was simple and quicker. This present study verified analysis using VITEK MS (BioMérieux, France) platform and obtained consistent results (data not shown), demonstrating that for bacterial isolate analysis, formic acid pretreatment is optimal for enhancing efficiency of MS-based analysis.

An integral database is critical for the identification capability of MS [11, 14]. This study found that only 10 (14.7%) of the *H. pylori* isolates could be correctly identified (score value ≥1.70) using the original Biotyper database. The main reason for this could be the limited number of reference spectra in the ORD (only 7 reference spectra). Furthermore, poor ability to identify clinical isolates from China could be dictated by fact that standard strains in the ORD were from different sources (e.g. *H. pylori* J99 isolated from America and *H. pylori* 26695 from England). In MSP dendrogram analysis, researchers noticed that only one of 7 MSPs in the original database had a close protein pattern with clinical isolates. However, identification rates were largely improved (14.7% vs 97.1%) after adding 37 clinical *H. pylori* isolates to the ORD. Additionally, these results indicated that increasing the identification spectra of each isolate could also improve overall identification rates. Additionally, this study extended another database (SARAMIS database, only containing 2 *H. pylori* standard strains) in VITEK MS which significantly increased the identification rate of *H. pylori* from 0% to 90.2% (data not shown).

Expression of certain virulence factors has been reported to be closely related with pathogenicity of *H. pylori* [3]. In recent years, correlation between infection with *dupA*-positive *H. pylori* and risk for duodenal ulcers has been confirmed by many studies [4, 5, 17]. Therefore, simultaneous evaluation of key virulence factors during identification of *H. pylori* isolates could serve as useful predictor for development of gastrointestinal diseases and provide useful information for clinical treatment. One previous study used hierarchical cluster analysis and principle component analysis (PCA) to evaluate the ability of MALDI-TOF MS for differentiation of *cagA/vacA* genes of *H. pylori* [13]. In the present study, MSP dendrogram was a grouping method to differentiate the microorganisms according to distance levels. ClinProTools is an easy-to-use data post-processing software capable of generating pattern recognition models for classification and prediction of diseases from mass spectrometry-based profiling data. This study used both MSP dendrogram and ClinProTools to rapidly classify MALDI-TOF MS mass spectra of *dupA*-positive/*dupA*-negative isolates. Only 48.4% of *H. pylori* isolates could be identified by the MSP dendrogram. ClinProTools could identify all isolates, making it a good complement to MSP dendrogram for virulence analysis. Combination of these two models significantly improved the capability of virulence analysis. Furthermore, external validation using 31 *H. pylori* isolates further confirmed the reliability of virulence analysis models. Therefore, the combined models could also differentiate clinical complications, to some extent, indicating that this approach is applicable for rapid determination of the virulence of clinical *H. pylori* isolates.

In conclusion, this study optimized the pretreatment procedure and generated a new reference database that can be shared among clinical laboratories for identification of *H. pylori* by MALDI-TOF MS*,* providing a new method for clinical diagnosis. Additionally, virulence analysis models were established for simultaneous analysis of virulence profiles of *H. pylori*. These may help to assess risks associated with development of clinical complications of *H. pylori* infection and further help prioritize treatment of *H. pylori* infection. Continual improvement of efficacy parameters can be achieved by broadening the electronic database of MSP in analytical programs.

Disclosure of conflict of interest

None.

Address correspondence to: Hu Zhao, Yanmei Zhang and Ping Xiang, Department of Laboratory Medicine, Huadong Hospital Affiliated to Fudan University, 221 West Yan'an Road, Shanghai 200040, China. Tel: +86-21-62483180-20878; E-mail: hubertzhao@163.com (HZ); Tel: +86-21-62483180-60215; E-mail: 15618653286@163.com (YMZ); Tel: +86-21-62483180-80601; E-mail: xiangping813@126.com (PX)

References

[1] Makola D, Peura DA, Crowe SE. Helicobacter pylori infection and related gastrointestinal diseases. J Clin Gastroenterol 2007; 41: 548-558.

[2] Cammarota G, Sanguinetti M, Gallo A, Posteraro B. Review article: biofilm formation by Helicobacter pylori as a target for eradication of resistant infection. Aliment Pharmacol Ther 2012; 36: 222-230.

[3] Yamaoka Y. Mechanisms of disease: helicobacter pylori virulence factors. Nat Rev Gastroenterol Hepatol 2010; 7: 629-641.

[4] Hu B, Zhao F, Wang S, Olszewski MA, Bian H, Wu Y, Kong M, Xu L, Miao Y, Fang Y, Yang C, Zhao H, Zhang Y. A high-throughput multiplex genetic detection system for Helicobacter pylori identification, virulence and resistance analysis. Future Microbiol 2016; 11: 1261-1278.

[5] Zhang Z, Zheng Q, Chen X, Xiao S, Liu W, Lu A. The Helicobacter pylori duodenal ulcer promoting gene, dupA in China. BMC Gastroenterol 2008; 8: 1-6.

[6] Tarkhashvili N, Beriashvili R, Chakvetadze N, Moistsrapishvili M, Chokheli M, Sikharulidze M, Malania L, Abazashvili N, Jhorjholiani E, Chubinidze M, Ninashvili N, Zardiashvili T, Gabunia U, Kordzaia D, Imnadze P, Sobel J, Guarner J. Helicobacter pylori Infection in patients undergoing upper endoscopy, republic of Georgia. Emerg Infect Dis 2009; 15: 504-505.

[7] Patel R. MALDI-TOF MS for the diagnosis of infectious diseases. Clin Chem 2015; 61: 100-111.

[8] Angeletti S. Matrix assisted laser desorption time of flight mass spectrometry (MALDI-TOF MS) in clinical microbiology. J Microbiol Methods 2017; 138: 20-29.

[9] Cassagne C, Normand A, L'Ollivier C, Ranque S, Piarroux R. Performance of MALDI-TOF MS platforms for fungal identification. Mycoses 2016; 59: 678-690.

[10] Cheng K, Chui H, Domish L, Hernandez D, Wang G. Recent development of mass spectrometry and proteomics applications in identification and typing of bacteria. Proteomics Clin Appl 2016; 10: 346-357.

[11] Ilina EN, Borovskaya AD, Serebryakova MV, Chelysheva VV, Momynaliev KT, Maier T, Kostrzewa M, Govorun VM. Application of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for the study of Helicobacter pylori. Rapid Commun Mass Sp 2010; 24: 328-334.

[12] Xiao D, Zhang H, He L, Peng X, Wang Y, Xue G, Su P, Zhang J. High natural variability bacteria identification and typing: helicobacter pylori analysis based on peptide mass fingerprinting. J Proteomics 2014; 98: 112-122.

[13] Cho YT, Kuo CH, Wang SS, Chen YS, Weng BC, Lee YC, Cheng CN, Shiea J, Wu DC. Differentiation of virulence of Helicobacter pylori by matrix-assisted laser desorption/ionization mass spectrometry and multivariate analyses. Clin Chim Acta 2013; 424: 123-130.

[14] Lacy BE, Rosemore J. Helicobacter pylori: ulcers and more: the beginning of an era. J Nutr 2011; 131: 2789S-2793S.

[15] Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, Malfertheiner P, Graham DY, Wong VWS, Wu JCY, Chan FKL, Sung JJY, Kaplan GG, Ng SC. Global prevalence of helicobacter pylori infection: systematic review and meta-analysis. Gastroenterology 2017; 153: 420-429.

[16] Wang YK, Kuo FC, Liu CJ, Wu MC, Shih HY, Wang SS, Wu JY, Kuo CH, Huang YK, Wu DC. Diagnosis of Helicobacter pylori infection: current options and developments. World J Gastroenterol 2015; 21: 11221-11235.

[17] Arachchi HS, Kalra V, Lal B, Bhatia V, Baba CS, Chakravarthy S, Rohatgi S, Sarma PM, Mishra V, Das B, Ahuja V. Prevalence of duodenal ulcer-promoting gene (dupA) of Helicobacter pylori in patients with duodenal ulcer in North Indian population. Helicobacter 2007; 12: 591-597.