http://dx.doi.org/10.5012/bkcs.2014.35.3.933

SWATH-based Comparative Proteomic Analysis of the *Mycobacterium bovis* BCG-Korea Strain†

Won-Kyu Lee,[a] Je-Hyun Baek,[‡,a] Sung Weon Ryoo,[§,*] and Yeon Gyu Yu[*]

Department of Chemistry, College of Natural Sciences, Kookmin University, Seoul 136-702, Korea
[*]E-mail: ygyu@kookmin.ac.kr
[‡]Center of Biomedical Mass Spectrometry (CBMS), Diatech Korea Co., Ltd., Seoul 138-826, Korea
[§]Department of Research and Development, Korean Institute of Tuberculosis, Chungbuk 363-954, Korea
[*]E-mail: scientist1@empal.com
Received December 16, 2013, Accepted January 13, 2014

A derivative of *Mycobacterium bovis* Bacillus Calmette-Guérin (BCG) has been used for the preparation of tuberculosis vaccines. To establish a Korean tuberculosis vaccine derived from BCG-Pasteur $1173P_2$, genome sequencing of a BCG-Korea strain was completed by Joung and coworkers.[1] A comparison analysis of the genome sequences of the BCG-Pasteur $1173P_2$ and BCG-Korea strains showed marginal increases in the total genome length (~0.05%) and the number of genes (~4%) in the BCG-Korea genome. However, how the genomic changes affect the BCG-Korea protein expression levels remains unknown. Here, we provide evidence of the proteomic alterations in the BCG-Korea strain by using a SWATH-based mass spectrometric approach (Sequential Window Acquisition of all THeoretical mass spectra). Twenty BCG proteins were selected by top-rank identification in the BCG proteome analysis and the proteins were quantified by the SWATH method. Thirteen of 20 proteins showing significant changes were enough to discriminate between the two BCG proteomes. The SWATH method is very straightforward and provides a promising approach owing to its strong reliability and reproducibility during the proteomic analysis.

Key Words : *Mycobacterium bovis*, BCG-Korea, SWATH, Proteomics

Introduction

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis* (*M. tuberculosis*). It typically affects the respiratory system as well as the skin, vertebra, lymph nodes, and brain. Recent estimates showed a prevalence of 8.6 million new TB cases and 1.3 million TB deaths in 2012.[2] *Mycobacterium bovis* Bacillus Calmette-Guérin (*M. bovis* BCG) is the only strain that has been used in the prevention of TB. This attenuated strain was derived from an *M. bovis* strain after 230 serial passages on glycerol-potato-bile medium from 1908 to 1921 by Albert Calmette and Camille Guérin at the Institut Pasteur in Lille, France.[3-5] This strain (Pasteur-BCG) was distributed to laboratories worldwide and introduced into the World Health Organization (WHO) Expanded Program on Immunization (EPI) in 1974.[6]

To characterize the distributed BCG strains, many proteomic studies were performed using mostly classical proteomics methods, such as two-dimensional electrophoresis (2DE): the culture supernatants and whole proteomes of *M. tuberculosis* and *M. bovis* BCG were compared;[7,8] the supernatants of virulent *M. tuberculosis* H37Rv and attenuated *M. bovis* BCG-Copenhagen were compared[9] and the culture filtrate of *M. bovis* BCG Moreau and *M. bovis* BCG Pasteur were compared.[10] In 2011, there was also a report on the whole proteome of *M. bovis* BCG-Mexico 1931.[11]

Among the many distributed BCG strains are sub-strains from BCG-Pasteur, $1173P_2$ Lot.D (BCG-Pasteur) seed strain that have been maintained, and one of these sub-strains, BCG-Korea, was prepared for the production of the Korean TB vaccine (Fig. 1).[12] The entire BCG-Korea stain genome sequence has been determined,[1] and it consists of 4,376,711 bp (GenBank accession number CP003900) and 4,189 genes. Interestingly, an increased genome length (2,189 bp) and

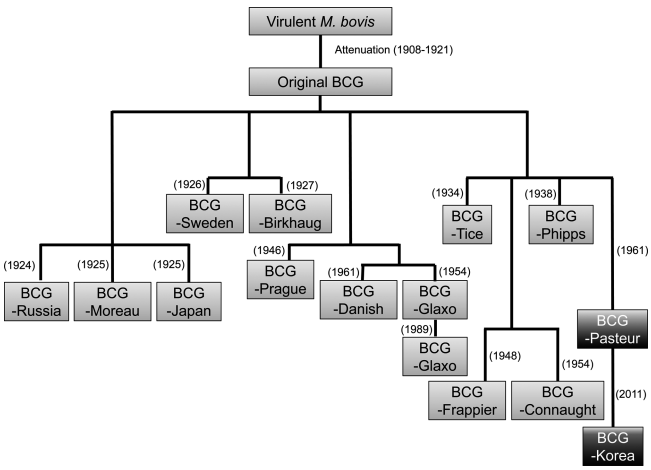

Figure 1. Refined genealogy of BCG vaccines, modified from Liu *et al.*[4] and Joung and Ryoo.[7]

†This paper is to commemorate Professor Myung Soo Kim's honourable retirement.
[a]These authors equally contributed to this work.

number of genes (156 genes) were observed in the BCG-Korea strain compared to the genome of the BCG-Pasteur strain (http://www.ncbi.nlm.nih.gov/genome/161?project_id =58781), the mother strain of BCG-Korea. However, no study has been reported showing the effects of genomic changes in BCG-Korea on the expression profile of its proteins.

In this study, we identified from 923 proteins in BCG-Pasteur and BCG-Korea strains and extracted quantitative information on 20 kind of most abundantly expressed BCG proteins using a SWATH-based approach (Sequential Window Acquisition of all THeoretical mass spectra), which is straightforward, highly reproducible, reliable and applicable to a large number of samples and characteristic proteomic analyses.[13] Ten proteins out of 20 showed different expression levels in the two BCG strains within a 15-35% range. The SWATH method provided evidence of genomic effects on protein levels using the selected 20 BCG proteins.

Materials and Methods

Chemicals. Iodoacetamide was purchased from Sigma-Aldrich (USA). Tris-HCl, urea and dithiothreitol (DTT) were obtained from Merck (Germany). All other chemicals were acquired from standard sources and were of molecular biology grade or higher.

BCG Culture. BCG-Korea (*M. bovis* BCG-Korea 1168P_2) was grown in 5 round beakers with 250 mL of Sauton's liquid medium for 3 weeks at 37 °C under stationary conditions. The freeze-dried BCG-Pasteur 1173 P_2 secondary seed LotD was reconstituted in Sauton's medium (1 mL) and cultured in Sauton's potato medium in Roux tubes.

Preparation of Proteome Samples. For the preparation of the protein extracts from the BCG strains, cells were collected by centrifugation at 12,000 × *g* for 10 min and washed 3 times with PBS buffer (pH 7.4). The collected cells were lysed by ultra-sonication in ice-cold lysis buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 0.02% sodium azide, pH 7.4) that included protease and phosphatase inhibitors (Roche, USA). The lysates were incubated for 3 h in denaturing buffer (8 M urea, 50 mM Tris-HCl, pH 8.0), treated with 10 mM DTT at 37 °C for 1 h, and alkylated with 55 mM iodoacetamide at room temperature for 30 min. The proteins in the sample were further digested with modified trypsin (Promega, USA) at 37 °C overnight. The peptides in the mixture were recovered using a desalting C18 cartridge (Sep-Pak, Waters, USA) and completely dried in a speed-vacuum before mass analysis.

LC-MS/MS Analysis. Triple-TOF-TM 5600+ (AB Sciex, Canada) was used for the BCG strain proteomes. The instrument was coupled with an Eksigent NanoLC-2D+ with Nanoflex cHiPLC system for reproducible peptide sample separation. Solvent A contained 0.1% formic acid (v/v) in water and solvent B contained of 0.1% formic acid (v/v) in 100% acetonitrile. The samples were loaded onto a trap column (0.5 mm × 200 μm) at a flow rate of 1 μL/min, and separated on an analytical column (15 cm × 75 μm) with a linear gradient of 2% to 35% solvent B over 30 min at a flow rate of 400 nL/min on the Nanoflex cHiPLC system. The Chip nanoLC column was regenerated by washing woth 60% solvent B for 50 min and equilibrating with 2% solvent B for 10 min. For data-dependent acquisition experiment, a Triple-TOF-TM 5600+ mass spectrometer was used with a 50-ms survey scan (TOF-MS) and 50-ms automated MS/MS scan for the 15 ions with the highest intensity. The MS/MS triggering criteria for parent ions were as follows: precursor intensity (> 150 counts) and charge state (> 1) with dynamic exclusion option (exclusion time: 6 s). Ions were isolated using a quadruple UNIT resolution (0.7 Da) and fragmented in the collision cell using collision energy ramped from 15 to 45 eV within the 50-ms accumulation time. For SWATH MS-based experiments, the mass spectrometer was operated as previously described by Gillet *et al.*[13] with minor changes**.** In detail, Triple-TOF 5600+ was used in the looped product ion mode with 20 Da/mass windows (each SWATH window has a 1 Da overlap) in the range of 400 to 1000 Da (Experiment 1: MS1 scan (400-1,600 Da); Experiment 2: 400-420 Da; Experiment 3: 419-444 Da ~ Experiment 24: 979-1,000 Da). The MS2 scan range was set to 100-1,600 *m/z*. The collision energy for each window was determined based on the appropriate collision energy for a two-charged ion centered upon the window with a spread of 15 eV. An accumulation time of 80 ms was used for each fragment ion scan and for the survey scans (total duty cycle, 2.5 s; number of total cycles, 1,406) in high sensitivity mode.

Data Analysis. All spectra generated from data-dependent acquisitions were searched against the BCG database using ProteinPilot™ (mycobacterium_bovis_bcg_1_proteins.fasta: total 3,953 protein entries) with the following search parameters: fully tryptic digestion, precursor ion and fragment ion mass tolerance for high-resolution Triple-TOF 5600, fixed modifications for cysteine (+57 Da: carbamidomethylation) and biological modifications/artifact such as methionine oxidation (+16 Da). For SWATH MS data, all SWATH MS runs were processed using the PeakView™ software (Version 1.2). The peak extraction mass window was 50 ppm, and the top 20 proteins were selected for SWATH data analysis. Five peptides per protein were selected and 5 transitions per peptide were also considered. All quantitative analyses (*e.g.* normalization, analysis of variance (ANOVA) and clustering) were processed using the Perseus software v1.2.0.16 (Max Planck Institute of Biochemistry).

Results and Discussion

Identification of Top Abundant Proteins by LC-MS/MS Analysis. The protein extract from the BCG strains was treated with trypsin, and the peptide fragments were analyzed by LC-MS/MS (Fig. 2). Among the several hundred proteins identified, the top 20 most abundant were further analyzed using the SWATH method. The top 20 identified proteins are listed in Table 1 along with accession number, protein/gene name, function, localization, and number of peptides. Many of these proteins are involved in the protein

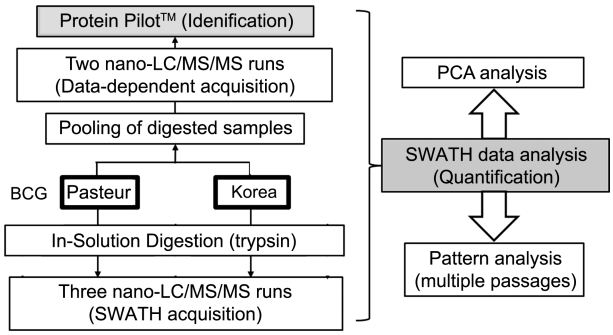

Figure 2. A workflow of SWATH-based quantification for multiple BCG-Pasteur and BCG-Korea strains.

folding and lipid biosynthesis pathway, such as GroEL1, GroLE2, DnaK, ClpC, HtpG, Fas, Mas, and Pks13. The abundance of these proteins indicates that they are critical for the survival of the *Mycobacterium* species. These proteins were known to be located in the cell wall, cytoplasm, periplasmic space, and plasma membrane. Among the identified proteins, preprotein translocase subunit 1 (SecA1) is a transmembrane protein involved in *Mycobacterium* virulence.[14]

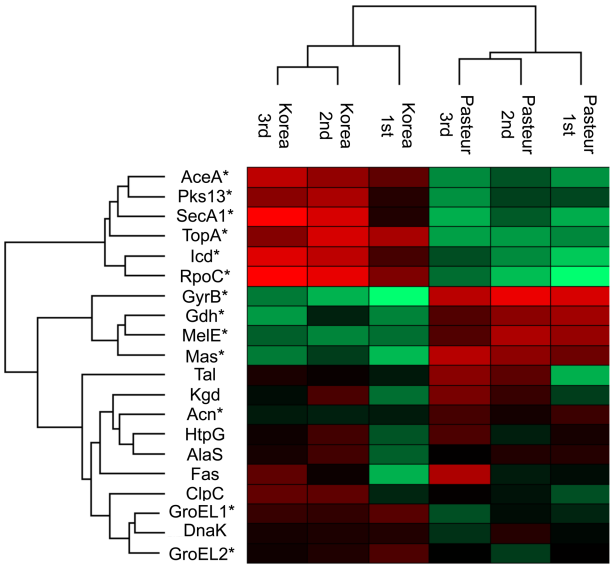

Figure 3. Hierarchical clustering (by average distance correlation) of 20 representative BCG proteins. Proteins were selected by top-rank identification in the BCG proteome by SWATH. Green represents low levels of protein expression, black represents intermediate levels, and red represents high levels of expression. Asterisk-marked proteins indicate a statistically significant change by ANOVA.

Table 1. Summary of 20 BCG proteins quantified by a SWATH-based approach

#	Accession #	Gene name	Description	Amino Acid #	Function	Localization[a]	# Peptides
1	BCG_0479	GroEL2	60 kDa chaperonin 2	541	Chaperone	Cp	99
2	BCG_2545c	Fas	Putative fatty acid synthase	3,070	Lipid synthesis	Cw, Cs, Pm, Fsc	33
3	BCG_0389	DnaK	Putative chaperone protein	626	Chaperone	NA	31
4	BCG_2962c	Mas	Putative multifunctional mycocerosic acid synthase	2,112	Lipid synthesis	Cw, Pm	23
5	BCG_0717	RpoC	DNA-directed RNA polymerase	1,317	Transcription	DRPc, Cw, Pm	22
6	BCG_1194c	MetE	Putative 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	760	Amino acid synthesis	Cw, Cs, Pm	23
7	BCG_3269c	SecA1	Putative preprotein translocase	950	Transport	Cp, Pm	15
8	BCG_2578c	AlaS	Putative alanyl-tRNA synthetase	905	Protein synthesis	Cp	12
9	BCG_3487c	GroEL1	60 kDa chaperonin 1	540	Chaperone	Bn, Cw, Cs, Pm	18
10	BCG_3862c	Pks13	Polyketide synthase	1,734	Lipid synthesis	Cw, Cs, Pm	15
11	BCG_1537c	Can	Putative aconitate hydratase	944	Metabolic reaction, cellular respiration	NA	12
12	BCG_3661c	ClpC	Putative ATP-dependent clp proteasE	849	Chaperone	Cw, Cs, Pm	13
13	BCG_1954	AceA	Putative isocitrate lyase	767	Carboxylic acid metabolism	Cp	12
14	BCG_3704c	TopA	DNA topoisomerase I	935	Intron homing	Cw, Cs, Pm	11
15	BCG_1308c	Kgd	Multifunctional 2-oxoglutarate metabolism enzyme	1,215	Metabolic reaction, cellular respiration	Cw, Cs, Pm, PDc	11
16	BCG_2315c	HtpG	Putative chaperone protein	648	Chaperone	Cp	15
17	BCG_0097c	Idh	Putative isocitrate dehydrogenase	746	Metabolic reaction, cellular respiration	Pm	13
18	BCG_1509c	Tal	Putative transaldolase	374	Pentose shunt	Cp	10
19	BCG_2496c	Gdh	Putative NAD-dependent glutamate dehydrogenase	1,624	Oxidoreductase	Cw, Cs, Pm	11
20	BCG_0035	GyrB	Dna gyrase	715	DNA topological change	Ch, Cp	9

[a]Subcellular location was determined using UniProtKB-SubCell (http://www.uniprot.org/locations), Abbreviation key: **Bn**, bacterial nucleoid; **Ch**, chromosome; **Cp**, cytoplasm; **Cs**, cytosol; **Cw**, cell wall; **DRPc**, DNA-directed RNA polymerase complex; **Fsc**, fatty acid synthase complex; **PDc**, pyruvate dehydrogenase complex; **Pm**, plasma membrane; **NA**, information is not available.

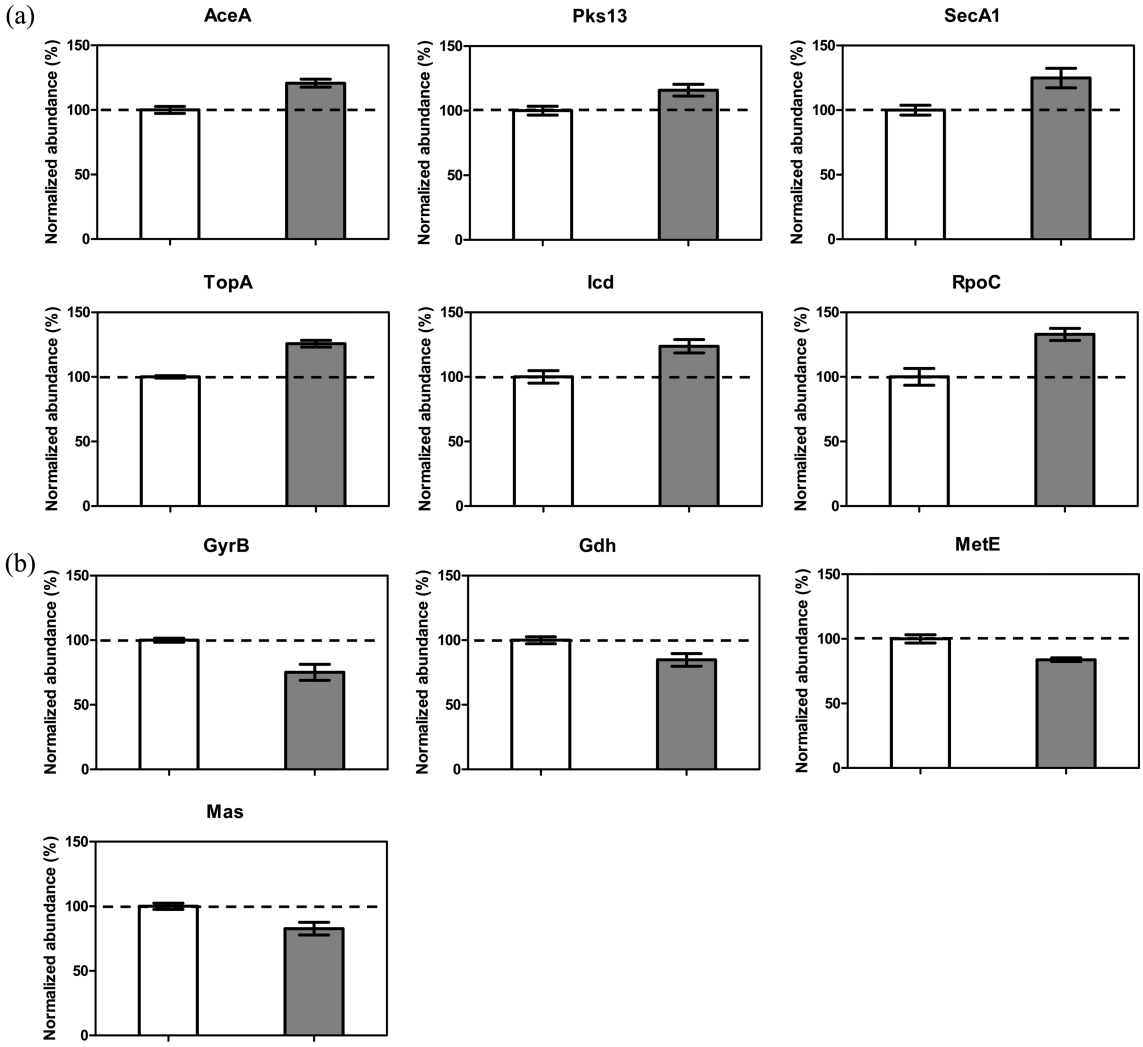

Figure 4. Groups based on responsive expression pattern. All expression levels were normalized to total protein abundances. Errors were calculated from three SWATH run replicates. The expression levels of proteins in BCG-Korea (grey bar) were compared to those of BCG-Pasteur (white bar).

Quantitative Analysis of the BCG top 20 Proteins. The expression levels of the selected 20 proteins in the BCG-Pasteur and BCG-Korea strains were compared using the SWATH method and are color coded (Fig. 3). They were divided into three groups according to their expression level changes. The expression levels of the first group (AceA, Pks13, SecA1, TopA, Icd, and RpoC) were slightly increased in BCG-Korea. The expression levels of these proteins in BCG Korea were 15-40% higher than those in BCG Pasteur (Fig. 4(a)). It is notable that the proteins in the first group are involved in nucleotide metabolism (RpoC and TopA), lipid metabolism (Pks13), carboxylic acid metabolism (AceA), cellular respiration (Icd) and protein transport (SecA1). The expression levels of the second group (GyrB, Gdh, MelE and Mas) in BCG Korea were 10-20% lower than those in BCG Pasteur (Fig. 4(b)). Proteins in the second group are involved in lipid metabolism (Mas), amino acid metabolism (MelE), DNA modification, or are members of the oxido-reductase family (Gdh). The remainder (GroEL2, GroEL1, Can, Tal, Kgd, Fas, DnaK, AlaS, ClpC, and HtpG) showed similar expression levels. Interestingly, the expression levels of the chaperone proteins (GloEL1, GloEL2, DnaK, ClpC, and HtpG) were similar in the BCG Pasteur and BCG Korea strains (Fig. 3). The constant expression levels of these chaperones suggest that these proteins could serve as internal standards when comparing the proteomes of different BCG strains.

The sequences of the 20 genes investigated in this study were identical in the BCG Pasteur and BCG Korea strains (data not shown). In addition, all proteins, except GyrB, had the same copy number in the BCG Pasteur and BCG Korea strains. There are two copies of the *gyrB* gene in BCG-Pasteur and a single copy in the BCG Korea strain. Interestingly, the expression level of the GyrB protein was slightly reduced in BCG Korea (Fig. 4(b)), suggesting that the higher copy number of *gyrB* was not correlated to the expression level.

During the establishment of various BCG strains, several regions of the genome sequence were deleted or duplicated.[5] In this study, we comprehensively analyzed the proteomes from two BCG strains with different passage numbers. From

the identified 923 proteins, the relative expression levels of 20 most abundantly expressed proteins were compared using the SWATH method. Ten proteins out of 20 showed different expression levels in the two BCG strains. The expression level of RpoC, which encode a beta subunit of RNA polymerase, showed the highest change (32.9% increase in BCG Korea strain). The differences of the expression levels in the two BCG strains were within a 15-35% range, and it is premature to correlate the difference in the protein expression levels to any biochemical or physiological differences between the two strains. However, our study showed that the protein expression levels in different laboratory BCG strains may vary within a limited range.

Acknowledgments. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0025229 and 2010-359-C00022). We also thank the Research Program 2014 of Kookmin University, Korea.

References

1. Joung, S. M.; Jeon, S. J.; Lim, Y. J.; Lim, J. S.; Choi, B. S.; Choi, I. Y.; Yu, J. H.; Na, K. I.; Cho, E. H.; Shin, S. S.; Pakr, Y. K.; Kim, C. K.; Kim, H. J.; Ryoo, S. W. *Genome Anounc*. **2013**, *1*, 1-2.
2. World Health Organization 2013, Global tuberculosis report 2013, (WHO/HTM/TB/2013.11)
3. Calmette, A.; Guérin, C.; Weill-Hallé, B. *Rev. Tuberc.* **1924**, *5*, 481-491.
4. Oettinger, T.; Jørgensen, M.; Ladefoged, A.; Hasløv, K.; Andersen, P. *Tuber. Lung Dis.* **1999**, *79*, 243-250.
5. Liu, J.; Tran, V.; Leung, A. S.; Alexander, D. C.; Zhu, B. *Hum. Vaccin.* **2009**, *5*, 70-78.
6. World Health Assembly, 27th, 1974; *Assignment Child*. **1985**, 69-72; 87-88.
7. Jungblut, P. R.; Schaible, U. E.; Mollenkopf, H. J.; Zimny-Arndt, U.; Raupach, B.; Mattow, J.; Halada, P.; Lamer, S.; Kaufmann, S. H. E. *Mol. Microbiol.* **1999**, *33*, 1103-1117.
8. Rodríguez-Alvarez, M.; Mendoza-Hernández, G.; Encarnación, S.; Calva, J. J.; López-Vidal, Y. *Tuberculosis* **2009**, *89*, 126-135.
9. Mattow, J.; Schaible, U. E.; Schmidt, F.; Hagens, K.; Siejak, F.; Brestrich, G.; Haeselbarth, G.; Muller, E. C.; Jungblut, P. R.; Kaufmann, S. H. E. *Electrophoresis* **2003**, *24*, 3405-3420.
10. Berrêdo-Pinho, M.; Kalume, D. E.; Correa, P. R.; Gomes, L. H. F.; Pereira, M. P.; da Silva, R. F.; Castello-Branco, L. R. R.; Degrave, W. M.; Mendonça-Lima, L. *BMC Microbiology* **2011**, *11*, 80.
11. Orduña, P.; Cevallos, M. A.; de León, S. P.; Arvizu, A.; Hernández-González, I. L.; Mendoza-Hernández, G.; López-Vidal1, Y. *BMC Genomics* **2011**, *12*, 493.
12. Joung, S. M.; Ryoo, S. W. *Clin. Exp. Vaccine Res*. **2013**, *2*, 83-91.
13. Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. *Mol. Cell Proteomics* **2012**, *11*, 1-17.
14. Hou, J. M.; D'Lima, N. G.; Rigel, N. W.; Gibbons, H. S.; McCann, J. R.; Braunstein, M.; Teschke, C. M. *J. Bacteriol*. **2008**, *190*, 4880-4887.