EDGEWOOD CHEMICAL BIOLOGICAL CENTER
U.S. ARMY RESEARCH, DEVELOPMENT AND ENGINEERING COMMAND
Aberdeen Proving Ground, MD 21010-5424

ECBC-TR-1154

EFFECT OF GENETIC DATABASE COMPREHENSIVENESS ON FRACTIONAL PROTEOMICS OF *ESCHERICHIA COLI* O157:H7

Vicky L. H. Bevilacqua
Rabih E. Jabbour
Mary M. Wade

RESEARCH AND TECHNOLOGY DIRECTORATE

Samir V. Deshpande

SCIENCE AND TECHNOLOGY CORPORATION
Edgewood, MD 21040-2734

Patrick E. McCubbin

OPTIMETRICS, INC.
Abingdon, MD 21009-1283

January 2014

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
XX-01-2014	Final	Oct 2010–Sep 2012

4. TITLE AND SUBTITLE
Effect of Genetic Database Comprehensiveness on Fractional Proteomics of *Escherichia* coli O157:H7

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Bevilacqua, Vicky L.H.; Jabbour, Rabih E.; Wade, Mary M. (ECBC); Deshpande, Samir V. (STC); and McCubbin, Patrick E. (OptiMetrics)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Director, ECBC, ATTN: RDCB-DRD-D, APG, MD 21010-5424
STC, 500 Edgewood Road, Suite 205, Edgewood, MD 21040-2734
OptiMetrics, Inc., 100 Walter Ward Boulevard Suite 100, Abingdon, MD 21009-1283

8. PERFORMING ORGANIZATION REPORT NUMBER
ECBC-TR-1154

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Edgewood Chemical Biological Center, In-House Laboratory Independent Research Program, 5183 Blackhawk Road, APG, MD 21010-5424

10. SPONSOR/MONITOR'S ACRONYM(S)
ECBC ILIR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report is required for the U.S. Army Edgewood Chemical Biological Center (ECBC) In-House Laboratory Independent Research project "Membrane Vesicles and Extracellular Proteins in Antibiotic Resistance and Virulence." The binding of extracellular proteins [membrane vesicle (MV) or secreted] to antibiotics could contribute toward antibiotic resistance mechanisms. We are characterizing the extracellular, fimbriae, and whole cell proteins produced by the pathogenic Gram-negative bacterium *Escherichia coli* (*E. coli*) O157:H7 in terms of proteomics and antibiotic-binding using mass spectrometry. Here we report a study on the effect of the comprehensive nature of the database used for proteomics on the analysis of various protein fractions. In this case, analysis using a more restricted database (chosen based on sample knowledge) resulted in identification of a higher percentage of proteins than a more comprehensive database. For *E. coli* O157:H7, we identified proteins that were specific to certain cellular fractions. Of these, four penicillin binding proteins were identified solely in the secretome fraction. Although the three penicillin binding proteins (PBPs) whose functions have been determined are antibiotic targets for penicillin and they therefore do not play an antibiotic resistance role, the identification of PBPs solely in the secretome does agree with our hypothesis that antibiotic proteins would be observed in the extracellular fraction.

15. SUBJECT TERMS

Escherichia coli	O157:H7	Liquid chromatography	Mass spectrometry
Virulence	Penicillin binding	Proteomics	Cellular fractions
Antibiotic resistance	LC-MS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Renu B. Rastogi
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	20	(410) 436-7545.

Blank

PREFACE

The work described in this report was authorized under the U.S. Army Edgewood Chemical Biological Center (ECBC) In-House Laboratory Independent Research Program (ILIR). This work was started in October 2010 and completed in September 2012.

The use of either trade or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

This report has been approved for public release.

Acknowledgments

The authors acknowledge the ILIR program for funding and Cindy Swim, A. Way Fountain, and Alan Zulich for administrative support.

Blank

CONTENTS

FIGURE

TABLE

EFFECT OF GENETIC DATABASE COMPREHENSIVENESS ON FRACTIONAL PROTEOMICS OF *ESCHERICHIA COLI* O157:H7

1. INTRODUCTION

The objective of this project is to investigate the role of membrane vesicles (MVs) and extracellular proteins in defining the mechanism(s) of antibiotic resistance and virulence. Certain extracellular proteins of pathogenic bacteria have been shown to function in survival mechanisms such as host immune system modulation (*1*) and biofilm formation (*2*). In addition, Gram-negative bacteria release a subset of extracellular proteins as MV components. Gram-negative bacteria form MVs by pinching off of the outer membrane to form liposomes. Bacteria form MVs at an energy loss and MVs contain periplasmic space components including enzymes, which suggest a functional role for MVs. Pathogenic Gram-negative bacteria produce more MVs than their non-pathogenic counterparts. Not surprisingly, several studies have provided evidence for a number of roles for MVs (*3*), such as transfer of antibiotic-resistance enzymes to other bacteria (*4*) and directed intercellular transport of virulence factors (*5*). In addition, work by Schooling and Beveridge (*6*) indicates that MVs shed by Gram-negative bacteria are a ubiquitous component in the biofilms of these bacteria. Levin and Rozen cite biofilm formation as one of three means by which bacterial populations can attain non-inherited antibiotic resistance (*7*).

Our hypothesis is that the binding of extracellular proteins (MV or secreted) could contribute toward mechanisms of antibiotic resistance. To address this possibility, we are characterizing the extracellular, fimbriae, and whole cell proteins produced by the pathogenic Gram-negative bacterium *Escherichia coli* (*E. coli*) O157:H7 in terms of proteomics and binding of antibiotics. We are using a mass spectrometry (MS) based proteomics approach to classify the proteins. MS proteomics experiments generate a vast amount of information in the form of spectra. The interpretation of the spectra depends on peptide mass fingerprinting (PMF) algorithms such as SEQUEST (*8*) and MASCOT (*9*). The PMF algorithm compares the experimental spectra with theoretical spectra of the protein sequences stored in FASTA format. The number of available protein sequences has increased dramatically since 2003. It is thus expected that the comprehensive nature of the database used for analysis will affect the outcome of the results. Here we report on a study of the effect of the comprehensive nature of the database used for proteomics on the fractional analysis of secreted, whole cell lysate, and fimbriae protein fractions.

2. METHODS

2.1 *Escherichia coli* O157:H7 Growth and Protein Fraction Preparation

E. coli O157:H7 (substrain Sakai) was grown in trypticase soy broth (TSB) to the late exponential phase (~10^8 cfu/mL) in an orbital shaker (125 rpm) at 37 °C. The cell culture was stored at 4 °C until fractionation. For isolation of the whole cell lysate and secreted protein fractions, 30 mL of culture was centrifuged at 11,300 x g/h using a Beckman Coulter (Brea, CA) J2-MC centrifuge. The supernatant was decanted to separate it from the pellet. This supernatant

containing the secreted proteins is referred to as the secreted fraction. The pellet was re-suspended in ~3.5 mL of 100 mM ammonium bicarbonate (ABC). This suspension was divided into three aliquots of approximately equal volume. The cell pellet suspension samples were thawed and lysed by ultrasonication (25 s on, 5 s off, 4 min total) using a Branson Digital Sonifier (Danbury, CT). The lysate was centrifuged at 14,000 rpm for 20 min at 10 °C using a Beckman GS-15R centrifuge. This fraction is referred to as the whole cell lysate fraction. A microwave lysis procedure was also attempted in which the sample was subject to microwaves using a Discover System (CEM Corporation, Matthews, NC) was performed at 55 °C for time periods of 5, 10, and 15 min rather than being subjected to sonication.

For isolation of fimbriae, cell culture aliquots (3 x 30 mL) were centrifuged at 15,000 x *g*/30 using a Beckman J2-MC centrifuge. Each pellet was re-suspended in 7 mL of ABC. These solutions were sheared through a 2 in., 22 gauge needle, 10 times each. Samples were divided into 1.5 mL centrifuge tubes and centrifuged at 15,000 x g/15 min using a Beckman GS-15R centrifuge. The supernatants were combined and filtered through a 0.45 µm acetate syringe filter. The filtrate was then heated at 60 °C/h in a block heater. These samples are referred to as the fimbriae fraction. Samples were frozen at -25 °C for up to four days.

2.2 Liquid Chromatography/Mass Spectrometry Sample Preparation.

Samples were prepared for liquid chromatography tandem mass spectrometry (LC-MS/MS) in a similar manner to that previously reported (*10*). Briefly, proteins were extracted from the whole cell lysate and secreted fractions by transferring each sample to a separate Microcon YM-3 filter unit (Millipore, Billerica, MA) and centrifuging at 14,100 x g/20-30 min. The filter membrane was washed with ABC and centrifuged at 14,100 x g/20 min. For the fimbriae fraction, the frozen samples were thawed and pipetted into Microcon YM-3 filter units (Millipore, Billerica, MA) for purification. The filters were each centrifuged at 14,000 x g/25 min three times with a 200 µL ABC wash in between centrifugations.

Generally, the proteins in the retentate were denatured at 40 °C for 1 h with 300 µL of 7.2 M urea and 3 µg/mL dithiothreitol in ABC. The urea was removed by centrifugation (14,100 x g/30-40 min) and the retentate was washed three times with ABC (150 µL ABC followed by centrifugation at 14,100 x g/30-40 min using an Eppendorf North America (Westbury, NY) centrifuge 5415C or 5415D. The filter unit was then transferred to a new receptor tube, and the proteins in the retentate were digested overnight at 37 °C with 5 µL sequencing grade trypsin (Product # 511A, Promega, Madison, WI) in 10 µL acetonitrile and 240 µL ABC. The tryptic peptides were isolated by centrifuging at 14,100 x g/20-30 min. Alternative digestion protocols involved adjusting trypsin concentration, incubation time, and temperature.

2.3 Liquid Chromatography/Mass Spectrometry Experiments

The tryptic peptides were separated in a similar manner to that previously described (*10*) on a capillary column using the Dionex (Sunnyvale, CA) UltiMate 3000 and the resolved peptides were electrosprayed into a Thermo Scientific (San Jose, CA) LTQ XL linear ion trap mass spectrometer. Product ion mass spectra were obtained in the data-dependent

acquisition mode, with a survey scan followed by tandem mass spectrometry (MS/MS) of the top five most intense precursor ions.

2.4 Proteomics Analysis

A protein database was constructed as previously described (*10*) in a FASTA format using the annotated proteome sequences derived from the genomes in the National Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov, accessed November 16, 2010. For this task, an in-house Practical Extraction and Programming Language (PERL) (http://www.activatestate.com/ActivePerl, accessed November 16, 2010) program was used to automatically download proteome sequences from the NCBI. The database was constructed by translating putative protein-encoding genes and contains amino acid sequences of potential tryptic peptides obtained by the in silico digestion of all proteins, assuming up to two missed cleavages. The acquired mass spectra were searched against this database with the SEQUEST algorithm (Thermo Scientific, Sunnyvale, CA). The SEQUEST thresholds for searching the product ion mass spectra were Xcorr, deltaCn, Sp, RSp, and deltaMpep. These parameters provide a uniform matching score for all candidate peptides. The files containing candidate peptides generated by SEQUEST were validated using the PeptideProphet algorithm (*11*). Peptide sequences with probability scores of 95% and higher were retained and used to generate a binary matrix of sequence-to-bacterium (STB) assignments. The binary matrix was populated by matching the peptides with corresponding proteins in the database and assigning a score of one. A score of zero was assigned for a non-match. The column in the binary matrix represents the proteome of a given bacterium, and each row represents a tryptic peptide sequence from an LC product ion mass spectral analysis. A sample microorganism was matched with a database bacterium by the number of unique peptides that remained after filtering of degenerate peptides from the binary matrix. Verification of the classification and identification of candidate microorganisms is performed through hierarchical clustering analysis and taxonomic classification using the in-house developed software package ABOid (*12*).

2.5 Biochemical Pathway Mapping

An algorithm was developed in-house for automated comparison of proteins observed in samples from a given fraction (whole cell, secreted, fimbriae). This algorithm was used to compare the proteins observed by LC-MS/MS in the three fractions and to determine which proteins were common between two or three fractions and which proteins were specific to a given fraction (fraction-specific proteins). Fraction-specific proteins were mapped to *E. coli* metabolic pathways using the Kyoto Encyclopedia of Genes and Genomes database (KEGG, www.genome.jp/kegg/, accessed November 16, 2011, Copyright 1995-2011 Kanehisa Laboratories).

3. RESULTS AND DISCUSSION

Three whole cell, three fimbriae, and five secreted fraction peptide samples were prepared for LC-MS/MS experiments and proteomics analysis. For PMF, we constructed three different databases, named EC_Sakai, Escherichia, and WholeDB with protein sequences from

$5x10^1$, $4x10^2$ and $2x10^3$ microorganisms, respectively. The WholeDB, Escherichia, and EC_Sakai databases were constructed from the genomes of all sequenced bacteria, all bacteria genomes of the *E. coli* genus, and only the *E. coli* O157:H7 substrain Sakai genome, respectively. In addition, a decoy database was constructed in which the theoretical peptide sequences were determined by reversing the protein sequences. The table below provides the number of proteins and peptides for each database.

Table. Number of Protein and Peptide Sequences in Databases Used

Database	Number of Microorganisms	Number of Proteins	Number of Peptides	Number of Unique Peptides
WholeDB	2×10^3	6,376,733	419,145,721	2.21E+08
Escherichia	4×10^2	298,264	17,716,320	1.26E+06
EC_Sakai	5×10^1	5,433	323,872	3.04E+05
Decoy	5×10^1	5,433	325,303	3.05E+05

After database searching using SEQUEST, followed by ABOid analysis with the standard PeptideProphet cutoff of 95%, different numbers of proteins were observed for a given cellular fraction for each database as shown in the Figure below. These results show that, if information is known about the sample, a higher percentage of proteins will be identified using a database based on prior sample knowledge rather than a more comprehensive database.

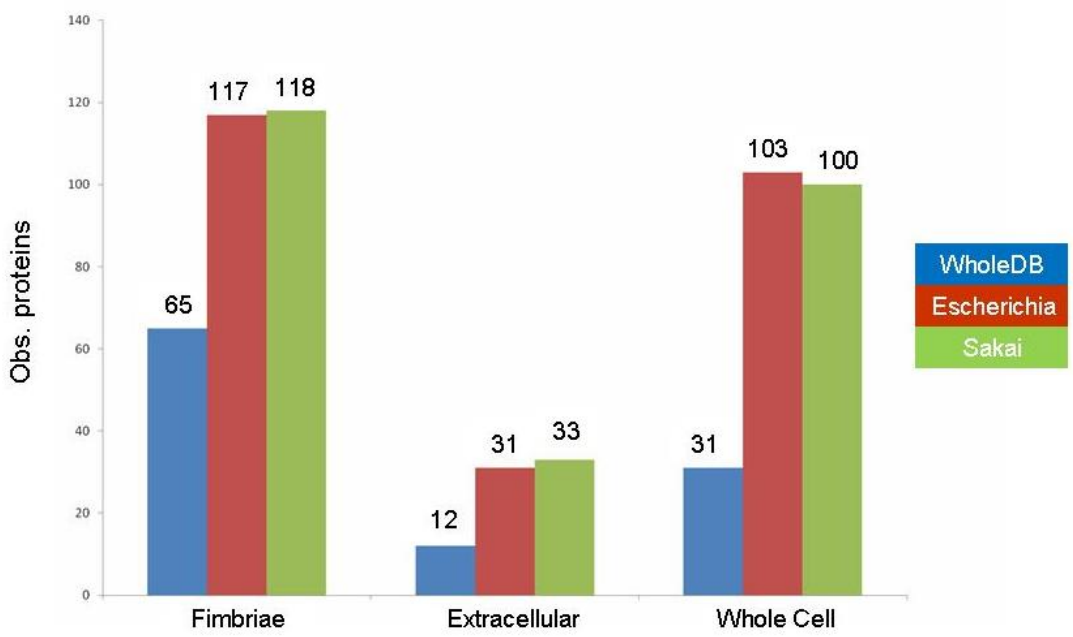

Figure. Proteins identified per cellular fraction for each database.

There may be loss of protein information resulting from the strict probability cutoff of 95%. We analyzed the data for the cellular fractions by preparing Receiver Operating Characteristic (ROC) curves. ROC curves were plotted for each replicate sample for a fraction. We used a binary classifier to determine the optimum cutoff by calculating the areas under the ROC curves (AUC) for that fraction. The statistical software R (www.r-project.org, accessed October 2012) and the package ROCR (www.cran.R-project.org, accessed October 2012) were used for computing the optimum cutoff values. The optimum cutoff values were not identical for the different cellular fractions (~95% for whole cell fraction, ~90% for fimbriae fraction, and 90–95% for secretome), indicating that samples from different fractions and/or bacteria require separate ROC analysis to determine the peptide confidence cutoff for optimum results. However, final choice of cutoff involves a compromise between use of the optimum cutoff and the increasing analysis time required for ROC analysis with increasing number of samples.

From the samples analyzed using the 95% cutoff, 200 *E. coli* proteins were identified. Of these proteins, 15% were common to all fractions. In addition, proteins specific to the secreted (3%), fimbriae (5%), and whole cell (29%) fractions were observed. Biochemical pathway mapping using Kyoto Encyclopedia of Genes and Genomes (KEGG) was carried out and the fimbriae-specific subset included a glucose-specific phosphotransferase system (G-PTS) component protein, inositol monophosphatase (IMP), and a DNA-binding transcription dual regulator. IMP has a role in streptomycin synthesis, and G-PTS is involved in environmental processing. Sequence alignment of the DNA-binding transcription dual regulator protein showed that it is homologous with a hydrogen peroxide-inducible genes activator. Furthermore, a putative stress protein and four penicillin binding proteins (PBPs) were identified solely in the secretome. Three of the PBPs are part of the peptidoglycan biosythesis complex involved in bacteria cell wall synthesis and targeted by penicillin in its antibiotic role. We have yet to determine the function of the fourth PBP. Each of these proteins (PBPs, IMP, G-PTS, putative stress protein) was also identified in the analysis of samples from smaller initial culture volumes, less than 10 mL as compared to 30 mL for the previous samples, from a new batch of *E. coli* O157:H7. Although a limit of detection was not determined, each of these proteins was identified in all replicate samples from the larger culture volume, but some were not identified in all replicate samples prepared from the smaller culture volumes, which implies that the concentrations of some of the proteins may be near the limit of detection in the smaller-volume samples. Only penicillin binding proteins were identifiable by searching the identified protein names for the word "penicillin". It is possible that there are extracellular or other proteins that bind to antibiotics other than penicillin. To address this consideration, we attempted coupling of ampicillin to magnetic beads having three different functional groups linked to the beads through differing chain lengths. Two types of groups resulted in successful coupling, one with a tosyl activation group having a 6-carbon chain (coupled through an ampicillin amine), the other with an amine terminal group having an 18-carbon chain (coupled through an ampicillin carboxyl). We incubated secretome proteins with these two ampicillin-bead complexes and with control beads (no ampicillin) and are currently carrying out the LC-MS/MS analysis on any proteins that may have bound to the beads to ascertain if any secretome proteins selectively bind to ampicillin. Details of the magnetic bead work will be reported separately upon completion of the LC-MS/MS analysis.

4. CONCLUSIONS

We analyzed *E. coli* O157:H7 whole cell, fimbriae, and secreted protein fractions by LC-MS/MS using protein databases of increasing comprehensiveness. We found that a more restricted database chosen based on sample knowledge will result in the identification of a higher percentage of sample proteins than a database that is more comprehensive. However, if strain-unique proteins are of interest, care must be taken to ensure that a protein that has been identified with a more restrictive database is truly unique when compared to the proteomes of organisms that were not included in the database. For *E. coli* O157:H7, we identified proteins that were specific to certain cellular fractions. Based on the functions noted above, the fimbriae-associated proteins IMP, G-PTS and the fimbriae-associated putative stress protein would be expected to be part of survival mechanisms. The DNA binding dual regulator also found in the fimbriae functions as a hydrogen peroxide-inducible genes activator, which has a positive regulatory effect on production of surface proteins that control colony morphology and auto-aggregation, indicating that this protein is a virulence factor. Finally, although the PBPs whose functions were determined are antibiotic targets for penicillin and they therefore do not play an antibiotic resistance role, the identification of PBPs solely in the secretome does agree with our hypothesis that antibiotic proteins would be observed in the extracellular fraction.

LITERATURE CITED

1. Vranakis, I.; De Bock, P.-J.; Papadioti, A.; Samoilis, G.; Tselentis, Y.; Gevaert, K.; Tsiotis, G.; Psaroulaki, A. Unraveling Persistent Host Cell Infection with *Coxiella burnetii* by Quantitative Proteomics. *J. Proteome Res.* **2011**, *10* (9), 4241–4251, DOI: 10.1021/pr200422f. Publication date (Web): July 27, 2011.

2. Ostrowski, A.; Mehert, A.; Prescott, A.; Kiley, T.B.; Stanley-Wall, N.R. Microbial Communities and Interactions: YuaB Functions Synergistically with the Exopolysaccharide and TasA Amyloid Fibers to Allow Biofilm Formation by *Bacillus subtilis*. *J. Bacteriol.* **2011**, *193*, 4821–4831; published ahead of print 8 July 2011, DOI:10.1128/JB.00223-11.

3. Kuehn, M.J.; Kesty, N.C. Bacterial Outer Membrane Vesicles and the Host-Pathogen Interaction. *Gene. Dev.* **2005**, *19*, 2645.

4. Ciofu, O.; Beveridge, T.J.; Kadurugamuwa, J.; Walther-Rasmussen, J., Høiby, N. Chromosomal *β*-lactamase is Packaged into Membrane Vesicles and Secreted from *Pseudomonas aeruginosa*. *J. Antimicrob. Chemother.* **2000**, *45*, 9–13.

5. Sidhu, V.K.; Vorhölter, F.-J.; Niehause, K.; Watt, S.A. Analysis of Outer Membrane Vesicle Associated Proteins Isolated from the Plant Pathogenic Bacterium *Xanthomonas campastris* pv. *campastris*. *BMC Microbiol.* **2008**, *8*, 87–92.

6. Schooling, S.R.; Beveridge, T.J. Membrane Vesicles: An Overlooked Component of the Matrices of Biofilms. *J. Bacteriol.* **2006**, *188*(16), 5945–5957.

7. Levin, B.R.; Rozen, D.E. Non-Inherited Antibiotic Resistance. *Nat. Rev. Microbiol.* **2006**, *4*, 556–562.

8. Eng, J.K.; McCormack, A.L.; Yates, J.R. III, An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Amer. Soc. Mass. Spec.* **1994**, *5*(11), 976.

9. Perkins, D.N.; Pappin, J.C.; Creasy, D.M.; Cottrell, J.S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *ELECTROPHORESIS* **1999**, *20*(18), 3551.

10. Jabbour, R.E.; Deshpande, S.V.; Wade, M.M.; Stanford, M.F.; Wick, C.H.; Zulich, A.W.; Skowronski, E.W.; Snyder, A.P. Double-Blind Characterization of Non-Genome-Sequenced Bacteria by Mass Spectrometry-Based Proteomics. *Appl. Environ. Microbiol.* **2010**, *76*(11) 3637–3644.

11. Keller, A.; Nesvizhskii, A.L.; Kolker, E.; Aebersold, R. Empirical Statistical Model to Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal. Chem.* **2002**, *74*, 5383–5392.

12. Deshpande, S.E.; Rabih E. Jabbour; Peter A. Snyder; Stanford, M.; Wick, C.H.; Zulich, A.W. ABOid: A Software for Automated Identification and Phyloproteomics Classification of Tandem Mass Spectrometric Data. *J. Chromatograph. Separat. Techniq.* **2011**, S5:001. DOI:10.4172/2157-7064.

ACRONYMS AND ABBREVIATIONS

ABC	100 mM ammonium bicarbonate
ABOid	Agents of Biological Origin Identifier
AUC	area under the curve
DNA	deoxyribonucleic acid
ECBC	U.S. Army Edgewood Chemical Biological Center
G-PTS	glucose-specific phosphotransferase system
IMP	inositol monophosphatase
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	liquid chromatography
LC-MS	liquid chromatography-mass spectrometry
LC-MS/MS	liquid chromatography tandem mass spectrometry
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MV	membrane vesicle
NCBI	National Center for Biotechnology Information
PBP	penicillin binding proteins
PERL	Practical Extraction and Programming Languate
PMF	peptide mass fingerprinting
ROC	Receiver Operating Characteristic
STB	sequence-to-bacterium
TSB	trypticase soy broth


EDGEWOOD CHEMICAL BIOLOGICAL CENTER
A U.S. Army RDECOM Laboratory