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

ECBC-TR-1124

SECRETOME BIOMARKERS FOR THE IDENTIFICATION AND DIFFERENTIATION OF ENTEROHEMORRHAGIC AND ENTEROPATHOGENIC *ESCHERICHIA COLI* STRAINS

Rabih E. Jabbour
James D. Wright
Mary Margaret Wade
Vicky L.H. Bevilacqua

RESEARCH AND TECHNOLOGY DIRECTORATE

Samir V. Deshpande

SCIENCE AND TECHNOLOGY CORPORATION
Edgewood, MD 21040-2734

Patrick E. McCubbin

OPTIMETRICS, INC.
Abingdon, MD 21009-1283

September 2013

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-09-2013	Final	Oct 2011 - Sep 2012

4. TITLE AND SUBTITLE
Secretome Biomarkers for the Identification and Differentiation of Enterohemorrhagic and Enteropathogenic *Escherichia coli* Strains

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jabbour, Rabih E.; Wright, James D.; Wade, Mary Margaret; Bevilacqua, Vicky L.H. (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
Science and Technology Corporation, 500 Edgewood Road, Suite 205, Edgewood, MD 21040-2734
OptiMetrics, Inc., 100 Walter Ward Blvd. Suite 100, Abingdon, MD 21009-1283

8. PERFORMING ORGANIZATION REPORT NUMBER
ECBC-TR-1124

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

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

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

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

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The secreted proteins of the enterohemorrhagic and enteropathogenic *Escherichia coli* (EHEC and EPEC) are the most common cause of hemorrhagic colitis, which is a bloody diarrhea with EHEC infection that can often lead to life-threatening hemolytic-uremic syndrome. We are employing a metaproteomic approach as an effective and complementary technique to the current genomic-based approaches. This metaproteomic approach will evaluate the secreted proteins associated with pathogenicity and utilize their signatures as differentiation biomarkers between EHEC and EPEC strains. Analysis of extract from EHEC O104:H4 resulted in the identification of a multidrug efflux protein that belongs to the family of fusion proteins, which are responsible for cell transportation. The experimental peptides identified lie in the region of the HlyD hemolysin secretion protein-D, which is responsible for transporting the hemolysin A toxin. Moreover, the taxonomic classification of EHEC O104:H4 showed the closest match with *E. coli* E55989, which is in agreement with genomic-sequencing studies that were done extensively on the aforementioned strain. Comparative proteomic calculations showed separation between EHEC O157:H7 and O104:H4 in replicate samples using cluster analysis. There were no reported studies that addressed the characterization of secreted proteins in various enhanced-growth media and utilized them as biomarkers for strain differentiation.

15. SUBJECT TERMS

Enterohemorrhagic	Enteropathogenic	Enteroaggregative
Mass spectrometry	Data analysis	Bioinformatics
Identification	Detection	*Escherichia coli*

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	26	(410) 436-7545

Blank

PREFACE

The work described in this report was authorized under the U.S. Army Edgewood Chemical Biological Center (ECBC; Aberdeen Proving Ground, MD) In-House Laboratory Independent Research Program. This work was started in October 2011 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 wish to thank Cynthia Swim for her administrative assistance of this research project and Augustus Fountain for his support and management of the in-house Laboratory Innovation Research Program at ECBC.

Blank

CONTENTS

FIGURES

TABLES

SECRETOME BIOMARKERS FOR THE IDENTIFICATION AND DIFFERENTIATION OF ENTEROHEMORRHAGIC AND ENTEROPATHOGENIC *ESCHERICHIA COLI* STRAINS

1. INTRODUCTION

The U.S. Government has initiated extensive efforts in the detection and identification of biological threat species in their Defense Advanced Research Projects Agency programs that explore the "detect-to-protect" and "detect-to-treat" paradigms (*1*,*2*). Those initiatives cover areas of general health risk, bioterrorism utility, homeland security, agricultural monitoring, food safety, environmental monitoring, and biological warfare agents in battlefield situations (*3*). Some of the health concerns include food contamination outbreaks that affect the military and civilian populations and can be transmitted from abroad to the U.S. soil. One such event was the fatal *Escherichia coli* strain O104:H4 outbreak that occurred in Germany in 2011, which infected citizens from 16 different industrial nations including the USA (*4*–*7*). The recent use of mass spectrometry (MS)-based proteomic analysis has proven useful for characterizing and identifying biological agents without prior knowledge of the sample contents (*8*). Therefore, the present study sought to determine whether MS proteomics could be used to distinguish between enterohemorrhagic and enteropathogenic *E. coli* (EHEC and EPEC) strains. Specifically, MS was used to discriminate between EHEC and EPEC strains on the basis of their secreted protein composition.

Through their presence in food and water matrices, EHEC and EPEC are major causes of disease in humans. Their infection to host cells is through an attaching and effacing mechanism in which the pathogen secretes various proteins that compromise the integrity of the cytoskeleton of the host cell (*9*). EHEC and EPEC pathogens exhibited different responses to antibiotics and, at times, their pathogenicity in humans was enhanced by an antibiotic regimen, as was the case with EHEC strains. In addition, studies have reported differences in the number and nature of the secreted proteins when comparing EHEC and EPEC (*10*). Therefore, the development of techniques that are capable of distinguishing between EHEC and EPEC is imperative to provide effective medical countermeasures in case of an outbreak in food or water supplies.

High-throughput, tandem MS-based proteomics was applied to characterize cellular proteins and produce amino acid sequence information for peptides that are derived from these proteins for *Burkholderia* and *Yersinia* species and strains. Whole-cell and secreted proteins from various bacterial strains were compared and contrasted using the U.S. Army Edgewood Chemical Biological Center (ECBC) in-house ABOid algorithm (software for the classification and identification of agents of biological origin) for species- and strain-level discrimination (*11*).

Therefore, the objective was to establish the sequence-based identity of secreted proteins that were isolated from the aforementioned *E. coli* strains. To achieve this goal, we utilized a high-throughput proteomic analytical system to rapidly characterize virulence proteins and produce amino acid sequence information to be used as differentiation biomarkers of EHEC

and EPEC strains in various biological matrices. This biological identification is essential to enhance the effectiveness of food and water supply safety for U.S. soldiers and to provide health personnel with reliable strain-level discrimination for effective medical countermeasures when needed.

2. METHODS

2.1 Preparation of the EHEC and EPEC Strains

In the present study, the pathogenic *E. coli* strains were O157:H7, O104:H4, and O11:H2 working cultures, which were prepared by streaking cells from cryopreserved stocks onto tryptic soy broth (TSB) and incubating at 37 °C until the cells reached the stationary growth phase. After incubation, the cells were harvested, and colony counts were performed using optical density measurements.

2.2 Isolation of the Secreted Proteins

The harvested cells were pelleted by centrifugation at a relative centrifugal force (RCF) of 2300 for 30 min, and the supernatant was immediately separated into 30 mL aliquots. The supernatants were then filtered using 0.22 µm hollow-fiber dialysis filters to ensure no large particulates or cellular debris were present in the samples. Pelleted and supernatant samples were frozen at –70 °C until further processing.

2.3 Processing of Secreted and Whole-Cell Proteins

The whole-cell samples were lysed using a bead-beating technique (30 s on then 10 s off for a 3 min duration). The lysates were centrifuged at 14,100×*g* for 30 min to remove cellular debris and large particulates. The supernatant from the whole-cell lysates and the filtered secretome samples were loaded separately on Pall molecular weight cutoff (MWCO) 3 kDa filter units (Pall Corporation, Ann Arbor, MI) and centrifuged at 14,100×*g* for 30 min. The effluents were discarded, and the filter membranes were washed with 100 mM of ammonium bicarbonate (ABC) then centrifuged for 20 min at 14,100×*g*. Proteins from the whole-cell and secretome fractions were denatured by adding 8 M of urea and 30 mg/mL of dithiothreitol to the filter and incubating for 1 h at 40 °C. The tubes were then centrifuged at 14,100×*g* for 40 min and washed three times using 150 mL of 100 mM ABC solution. On the last wash, ABC was allowed to sit on the membrane for 20 min it was shaken, followed by centrifugation at 14,100×*g* for 40 min. The filter units were then transferred to new receptor tubes, and the proteins were digested with 5 µL of trypsin in 240 µL of ABC solution plus 5 µL of acetonitrile (ACN). Proteins were digested overnight at 37 °C on an orbital shaker set to 90 rpm. To quench the trypsin digestion, 60 µL of 5% ACN/0.5% formic acid (FA) was added to each filter followed by 2 min of vortexing to mix the sample. The tubes were centrifuged for 10 min at 14,100×*g*. An additional 60 mL of 5% ACN/0.5% FA mixture was added to the filter and centrifuged. The effluents were then analyzed using liquid chromatography (LC)–electrospray ionization–tandem MS.

2.4 Protein Database and Database Search Engine

A protein database was constructed in a FASTA format using the annotated bacterial proteome sequences derived from fully sequenced chromosomes of all available *E. coli* strains, which consisted of 54 strains (as of September 2012). A Perl program (ActiveState Software Inc.; Vancouver, BC; http://www.activestate.com/Products/ActivePerl; accessed April 2011) was written to download these sequences automatically from the National Institutes of Health National Center for Biotechnology Information (NCBI) site (http://www.ncbi.nlm.nih.gov; accessed September 2012). Each database protein sequence was supplemented with information about a source organism and a genomic position for the respective open reading frame (ORF) embedded into a header line. The database for the *E. coli* bacterial proteome, which was constructed by translating putative protein-coding genes, consists of millions of amino acid sequences of potential tryptic peptides obtained by the in-silico digestion of all proteins (allowing up to two missed cleavages).

The experimental MS/MS spectral data of bacterial peptides were searched using a SEQUEST algorithm (Yates Laboratory, The Scripps Research Institute; La Jolla, CA) against a constructed proteome database of microorganisms. The SEQUEST thresholds for searching the product ion mass spectra of peptides were Xcorr, deltaCn, Sp, RSp, and deltaMpep. These parameters provided a uniform matching score for all candidate peptides. The generated out files of these candidate peptides were then validated using a peptide-prophet algorithm. Peptide sequences with a probability score of 95% and higher were retained in the dataset and used to generate a binary matrix of sequence-to-bacterium assignments. The binary matrix assignment was populated by matching the peptides with corresponding proteins in the database and assigning a score of 1. A score of 0 was assigned when there was no match. The column in the binary matrix represents the proteome of a given *E. coli* strain, and each row represents a tryptic peptide sequence resulting from the LC–MS/MS analysis. Analyzed samples were matched with the *E. coli* strains on the basis of the number of unique peptides that remained after further filtering of degenerate peptides from the binary matrix. Verification of the classification and identification of candidate microorganisms was performed through hierarchical clustering analysis and taxonomic classification.

The use of ABOid transformed the results from searching the MS/MS spectra of peptide ions against a custom protein database into a taxonomically meaningful and easy-to-interpret output. It was used to calculate the probability that the peptide sequence assignment to an MS/MS spectrum was correct and that it used accepted spectrum-to-sequence matches to generate a sequence-to-bacterium (STB) binary matrix of assignments. Validated peptide sequences, differentially present or absent in various strains (STB matrices), were visualized as assignment bitmaps and analyzed using an ABOid module, which used phylogenetic relationships among *E. coli* strains as a part of the decision tree process. The bacterial classification and identification algorithm used assignments of organisms to taxonomic groups (phylogenetic classification) on the basis of an organized scheme that begins at the phylum level and follows through classes, orders, families, and genus then down to the strain level. ABOid was developed using Perl, MATLAB (MathWorks, Natick, MA), and Microsoft Visual Basic (Microsoft Corporation; Redmond, WA).

3. RESULTS AND DISCUSSION

3.1 ABOid Algorithm Output

The ABOid algorithm provides results in different formats and can be tailored to address the appropriate factors. For example, Figure 1 provides a typical output that was generated for the LC–MS/MS analyses of bacterial proteins digestion. Bioinformatics tools were used to process the peptide sequence information for the bacterial differentiation and classification. The top window in the software lists the unique proteins that were identified and their corresponding bacterium matches. The program's middle window shows the binary matrix resulting from the STB search-matching process. The total row in the middle window represents the total number of unique proteins that were identified for a given bacterium. The lower section of the program window represents the histogram output of bacterial identification.

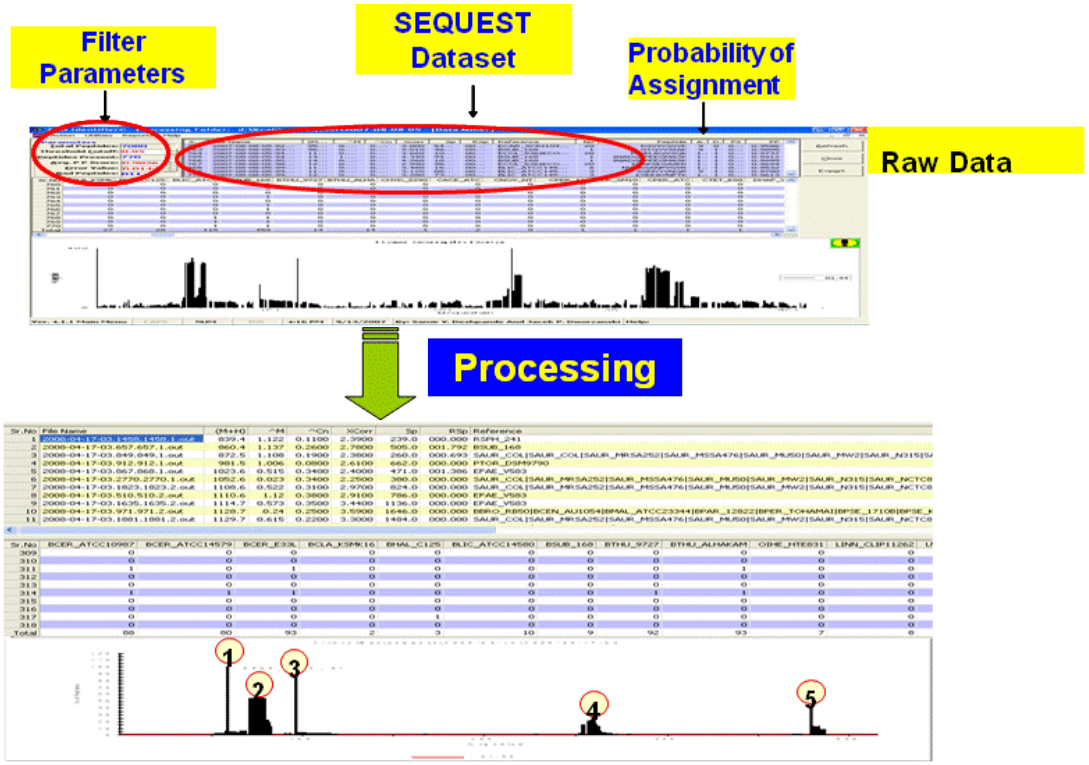

Figure 1. MS-based proteomic approach output.

Figure 2 shows another set of results from the ABOid program that presents an identification output in histogram format. This graph was generated by plotting the number of unique proteins versus the *E. coli* strain match found in the database. The *y*-axis represents the percentage of unique peptides matched with a 95% confidence level for all of the strains on the *x*-axis. In this figure, the identified *E. coli* strain O157:H7 was matched with the analyzed bacterial sample. Common degenerate peptides among various bacteria within the constructed proteome database are shown below the threshold cutoff, which is represented by the horizontal

red line. These degenerate peptides are removed from the total number of unique peptides for the identified species.

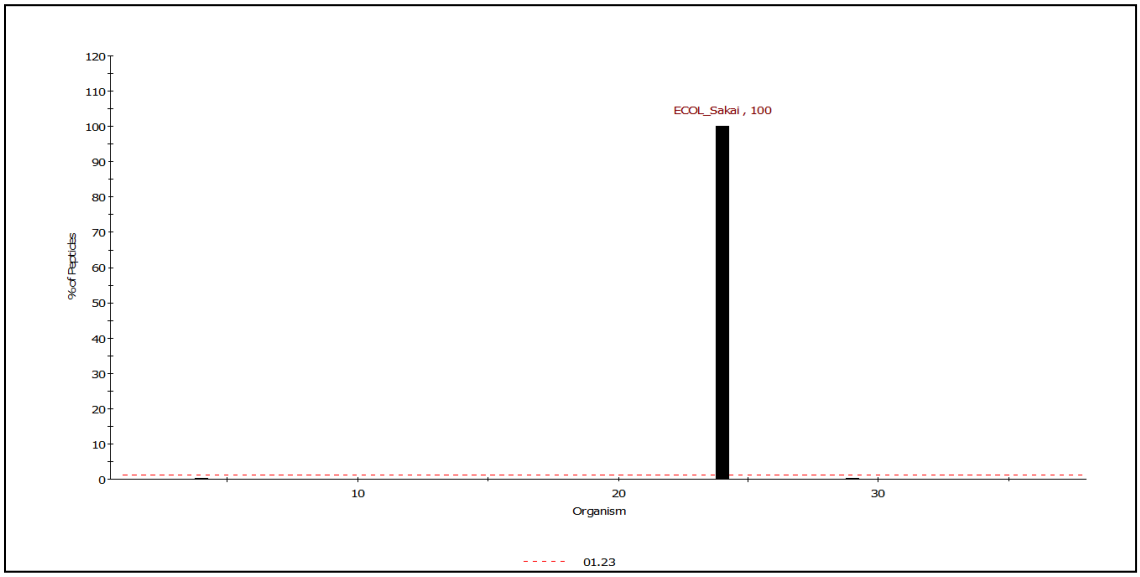

Figure 2. Histogram representing the output of the binary matrix of the unique peptides identified for the *E. coli* strain O157:H7 sample that was analyzed and processed using the ABOid program. All identified peptides were extracted at a 95% confidence level.

3.2 Determination of Common Proteins Using Secretome Lysates for EHEC and EPEC Strains

Strains O157:H7 (EHEC), O104:H4 (EHEC), and O111:H2 (EPEC) were analyzed by proteomic MS to determine the common proteins from replicate analyses generated from their secretome lysates. Tables 1 and 2 show the list of common proteins obtained from three analyses of *E. coli* strains O157:H7 and O104:H4, respectively. The matching of most common proteins was done using UniProtKB database (*12*). The UniProtKB is a nonredundant database that includes all sequenced microbes and provides biological ontologies, classifications and cross-references, cellular processes, and biochemical functions for each protein. In Table 1, the data showed that most of the common proteins identified had the highest match and identification with strain O157:H7 and cellular functionality related to a flagellar type. The dominant flagellar functions are often observed with EHEC bacteria as the responsible pathogenic factors in the attaching and effacing mechanism (*9*). This agreement between the genomics and proteomics studies showed that this approach could be used as an effective complement to the genomic-based techniques.

On the other hand, the data showed that the commonly identified proteins were strain-unique, regardless of the database used. For example, when we utilized our database that included only *E. coli* strains, the identification was the same as that from UniProtKB database, which included all sequenced bacteria. Table 2 represents the output of UniProtKB analyses for the common proteins identified in the secretome fraction of the *E. coli* O104:H4 strain. The

common proteins were first identified using the ABOid algorithm, and then UniProtKB was utilized to determine nonredundant matching and cellular functions and processes. At the time of this study, the *E. coli* strain O104:H4 was not fully sequenced and was not included in either database. The third column in Table 2 represents the closest matches between the studied strains and the bacterial strains in the UniProtKB database. Most of the matches were with *E. coli* strains that were considered to have more of enteroaggregative *E. coli* (EAEC) and/or EPEC strains. None of the matches were with the *E. coli* strain O157:H7, which indicates that the O104:H7 strain is not closely related to EHEC strains. In addition, the common proteins identified for *E. coli* strain O104:H4 were diverse in their cellular functions, unlike those of O157:H7, which had mainly flagellar functions.

Using the UniProtKB utilities for further examination of the cellular functions of the common proteins for the O104:H4 strain revealed that the potential cellular functionality of the tryptic peptides could be identified from the LC–MS/MS analyses. The UniProtKB cellular function tools use various solid, thick, colored lines to represent the different cellular functions for each active site in a given protein. For example, the tryptic peptides that correspond to the identified secreted autotransporter serine protease were located in the region of the protein that indicates a virulence function, as shown in Figure 3. The dotted circle represents the region of the identified peptides for the secreted autotransporter serine protease proteins that were common among the replicate LC–MS/MS analyses of the secreted fraction of the O104:H4 strain.

Table 1. Common Strain-Unique Proteins from Replicate Analyses of the Secretome Fraction of *E. coli* Strain O157:H7

Accession Number	Protein Name	Closest Match	Process	Function	Component
AP_002538.1	Flagellar filament structural protein	EC O157:H7/ EC K12	Ciliary or flagellar motility	ND	Bacterial-type flagellum hook
AP_003849.1	DNA-binding transcriptional dual regulator	EC O157:H7	Binding	Transcription	ND
NP_288384.1	Flagellin	EC O157:H7	Ciliary or flagellar motility	Structural molecule activity	Bacterial-type flagellum filament
YP_001882351.1	Hypothetical protein SbBS512_E4084	*Shigella byodii* /EC NC101	ND	ND	ND

EC: *E. coli*
ND: not determined

Table 2. Common Strain-Unique Proteins from Replicate Analyses of the Secretome Fraction of *E. coli* Strain O104:H4

Accession Number	Protein Name	Closest Match	Process	Function	Component
YP_003223560.1	Secreted autotransporter serine protease	EC O103:H2	Proteolysis	Serine-type endopeptidase activity	Peptidase activity
YP_001463426.1	Multidrug efflux system subunit MdtA	EC O139:H28	Transport	Transporter activity	Plasma membrane
YP_002292692.1	Conserved hypothetical protein	EC SE11	ND	ND	ND
YP_003229309.1	Putative DNA primase	EC O26:H11	ND	ND	ND
YP_541664.1	DNA-binding protein	EC UTI89_C2667	Nitrogen utilization	DNA binding	ND
NP_286019.1	Hypothetical protein	EC O157:H7	Lipoprotein metabolic process	Lipase/hydrolase activities	lipid particle

EC: *E. coli*
ND: not determined

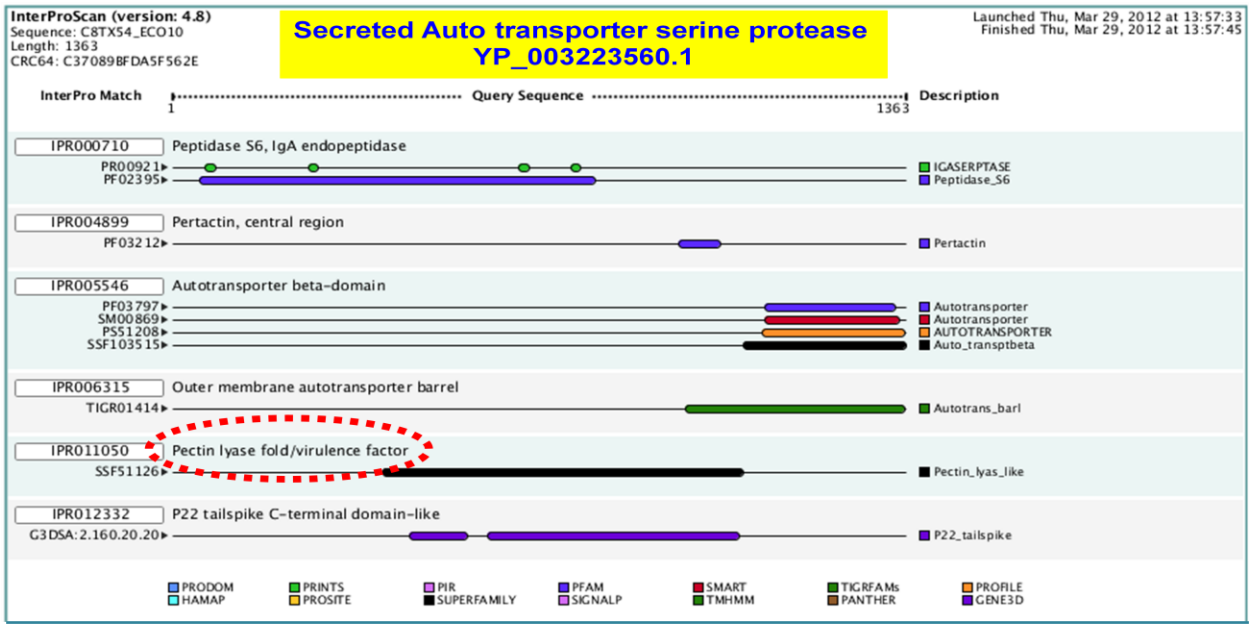

Figure 3. Results from the UniProtKB cellular functions identification tool, InterProScan, for a common protein identified in the secreted fractions of *E. coli* strain O104:H4. The dotted oval shape represents the cellular function of the peptides identified using LC–MS/MS analyses.

3.3 Effect of Cellular Fraction on the Differentiation of EHEC O157:H7 Strain

Whole-cell and secreted fractions from the *E. coli* O157:H7 strain were analyzed using LC–MS/MS followed by data processing using the ABOid algorithm. Identification of the samples was correctly established to be the *E. coli* O157:H7 strain but with more ambiguity using the whole-cell fraction rather than a secreted fraction. The results of the near-neighbor analysis using the Euclidean distance-linkage approach for these cellular fractions showed that the unique set of proteins identified from the secreted fraction (Figure 4a) matched with the *E. coli* strain O157:H7 more closely than that of the whole-cell fraction (Figure 4b). The similarity between the analyzed secretome and the closest neighbor in the database exhibited 100% matching with *E. coli* strain O157:H7 (Figure 4a), but there was only around 35% similarity between the whole-cell fraction and the *E. coli* strain O157:H7 from the database. This difference in matching between the whole-cell and secretome fractions could be attributed to the presence of more strain-unique proteins from the secretome fraction. The whole-cell fractions exhibited common proteins present across all *E. coli* strains that were found in higher concentration than those of the secreted fractions. The proteins identified in the whole-cell fraction showed large numbers of ribosomal proteins, which are commonly found in other strains and species of *E. coli* and other bacteria. Such types of proteins would result in less differentiation than those of the secretome proteins, which did not have ribosomal or other highly expressed and conserved proteins. This difference in the types of proteins from the two studied fractions was reflected in the taxonomic classification as shown in Figure 4.

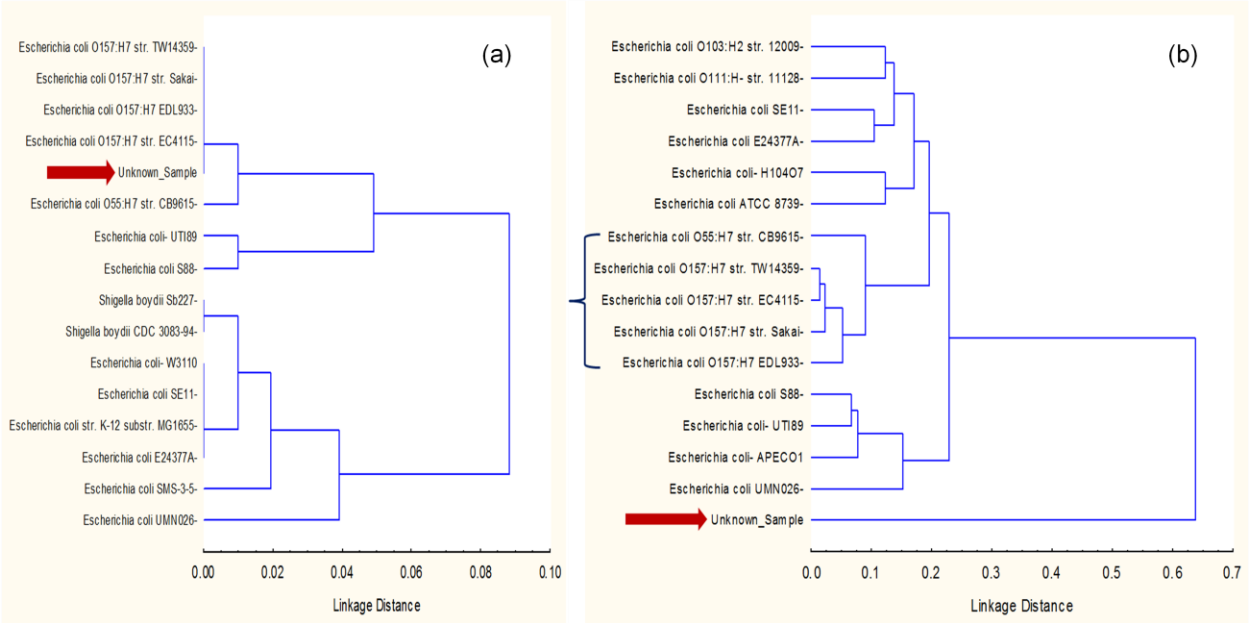

Figure 4. Single-linkage Euclidean distancing for the near-neighbor classification of EHEC *E. coli* O157:H7 strains from (a) secretome and (b) whole-cell fractions.

3.4 Differentiation of *E. coli* O157:H7 and O104:H4 Strains Using Secretome Lysates

Pathogenic *E. coli* strains O157:H7 and O104:H4 were analyzed by proteomic MS for strain identification and differentiation using the secretome fractions of each strain. The identification of the samples was correctly established, and those results were observed in the output of the STB binary matrix with the number of unique peptides on the *y*-axis and bacterium proteome on the *x*-axis. The near-neighbor analysis using the Euclidean distance linkage approach for these *E. coli* strains showed that the unique set of proteins that was identified had the closest match with the *E. coli* O157:H7 and O104:H4 strains. However, the database did not contain the O104:H4 strain because it was absent from the list of fully sequenced *E. coli* strains in the public repository. Using the Euclidean distance linkage approach for the near-neighbor analysis of the *E. coli* strain O104:H4 showed the closest match with the *E. coli* 55989 strain (Figure 5). The *E. coli* strain 55989 is an EAEC strain that was originally isolated in 2002 from the diarrheagenic stools of an HIV-positive adult suffering from persistent watery diarrhea in the Central African Republic. The EAEC strains form aggregates, as their name suggests, and are an emerging cause of gastroenteritis (*13*). This taxonomic classification of *E. coli* strain O104:H4 agrees with the genomic-sequencing efforts that were extensively done on the O104:H4 strain due to its implication in the deadly outbreak of *E. coli* in Germany in 2011 (*14*). The genomic-sequencing of *E. coli* strain O104:H4 showed that this strain is 95% genomically similar to EAEC 55989, which implies that this strain is more of a hybrid clone between the *E. coli* 55989 and ancestor *E. coli* O104:H4 strains. On the basis of genomic classification, this new strain was distant from EHEC strains including O157:H7, a common culprit in food contamination outbreaks (*7*). Such genomic studies provide strong support to our findings in terms of proteomic identification of the strains and in agreement with the phylogenetic classification. The utilization

of proteomics-based identification and phylogenetic classification of the *E. coli* strains from their secretome fractions showed that this approach is an effective and reliable complementary approach to those of the whole-genome sequencing and optical genetic-mapping techniques. Moreover, a recent study on the pathogenicity mechanism of the *E. coli* O104:H4 strain showed that this *E. coli* strain behaves as an EAEC in its characteristic verotoxicity to the host cells (*15*).

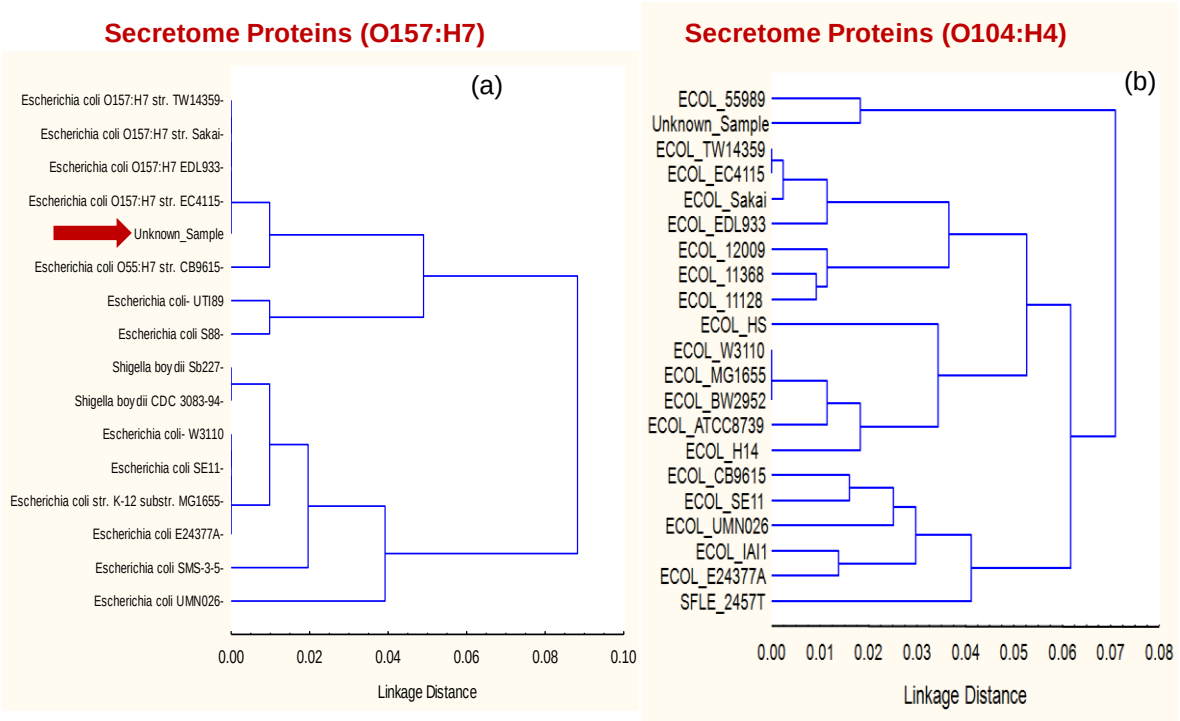

Figure 5. Euclidean distance single linkage of the near-neighbor classification of pathogenic *E. coli* strains (a) O157:H7 and (b) O104:H4 using secretome proteins.

Although the proteomics classification showed strain-level classification for the studied *E. coli* strain, each strain did not show any close relationship to the others. This observation is important to support the findings reported in genomic studies that those strains are different in their protein expression, which was the conclusion of several pathogenesis and sequencing studies (*14,15*).

4. CONCLUSIONS

The results of this study revealed that using secretome proteins as biomarkers for the differentiation of EHEC and EPEC strains is useful when employing metaproteomic analyses. The strain-level differentiation among the EHEC strains studied was improved by the use of secreted proteins as biomarkers. Secretome proteins provide a unique source of cellular variability that was not observed when compared with whole-cell lysates. The extensive genomic studies on the studied strains showed strong agreement with the classification of a strain that was not in the database (i.e., *E. coli* strain O104:H4), which was determined using an MS-based

proteomics approach. Such agreement needs to be further examined with a larger set of samples and under various environmental conditions to verify the effectiveness of the utilized approach. In addition, once such studies are validated, this could increase our confidence in the identification of microbes during the early stages of outbreaks at the strain level using protein biomarkers. This, in turn, would enhance medical countermeasures and diagnostics.

Overall, tandem MS-based proteomics and bioinformatics were useful in the comparative proteomics study for the differentiation of EHEC strains. This resulted in different degrees of separation between the correctly determined database organism and the next nearest-neighbor organism(s). Moreover, this approach relies on taxonomic correlation within the constructed proteome database. Therefore, inferring the identification of a sample organism that is not present in the genome database is possible, as was the case with *E. coli* strain O104:H4. This capability is corroborated because prokaryotic organisms are arranged in hierarchical order; their common proteins increase as we move from strain to phyla and vice versa. Such properties allow the use of an MS-based proteomic approach to infer taxonomic classification based on the depth of available genomic-sequencing information for such microbes.

LITERATURE CITED

1. National Research Council. *Sensor Systems for Biological Agent Attacks*, National Academy Press: Washington, DC, 2005.

2. Demirev, P.A.; Feldman, A.B.; Lin, J.S. Chemical and Biological Weapons: Current Concepts for Future Defenses. *Johns Hopkins APL Tech. Dig.* **2005**, *26*, pp 321–333.

3. Demirev, P.A.; Fenselau, C. Mass Spectrometry for Rapid Characterization of Microorganisms. *Annu. Rev. Anal. Chem.* **2008**, *1*, pp 71–93.

4. World Health Organization. *Outbreaks of E. Coli O104:H4 Infection: Update 30*. [Online] **2011**, http://www.euro.who.int/en/what-we-do/health-topics/emergencies/international-health-regulations/news/news/2011/07/outbreaks-of-e.-coli-o104h4-infection-update-30 (accessed October 2012).

5. Perna, N.T.; Plunkett, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Pósfai, G.; Hackett, J.; Klink; S.; Boutin; A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.T.; Potamousis, K.D.; Apodaca, J.; Anantharaman, T.S.; Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R. Genome Sequence of Enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **2001**, *409*, pp 529–533.

6. European Food Safety Authority (EFSA). *Shiga Toxin/Verotoxin-Producing* Escherichia coli *in Humans, Food and Animals in the EU/EEA, with Special Reference to the German Outbreak Strain STEC O104*, [Online] **2011,** Joint EFSA/ECDC Technical Report, http://www.efsa.europa.eu/en/supporting/pub/166e.htm (accessed October 2012).

7. Mellmann, A.; Harmsen, D.; Cummings, C.A.; Zentz, E.B.; Leopold, S.R.; Rico, A.; Prior, K.; Szczepanowski, R.; Ji, Y.; Zhang, W.; McLaughlin, S.F.; Henkhaus, J.K.; Leopold, B.; Karch, H. Prospective Genomic Characterization of the German Enterohemorrhagic *Escherichia coli* O104:H4 Outbreak by Rapid Next Generation Sequencing Technology. *PLoS ONE* **2011**, *6*(7).

8. 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 with Non-Genome Sequenced Bacteria by Mass Spectrometry-Based Proteomics. *Appl. Environ. Microbiol.* **2010**, *76*(11), pp 3637–3644.

9. Frankel, G.; Phillips, A.D.; Rosenshine, I.; Dougan, G.; Kaper, J.B.; Knutton, S. Enteropathogenic and Enterohaemorrhagic *Escherichia coli*: More Subversive Elements. *Mol. Microbiol.* **1998**, *30***,** pp 911–921.

10. Deng, W.; Yu, H.B.; de Hoog, C.L.; Stoynov, N.; Li, Y.; Foster, L.J.; Finlay, B.B. Quantitative Proteomic Analysis of Type III Secretome of Enteropathogenic *Escherichia coli* Reveals an Expanded Effector Repertoire for Attaching/Effacing Bacterial Pathogens. *Mol. Cell. Proteomics* **2012**, *11*(9), pp 692–709.

11. Jabbour, R.E.; Wade, M.M.; Deshpande, S.V.; Stanford, M.F.; Wick, C.H.; Zulich, A.L.; Snyder, A.P. Identification of *Yersinia pestis* and *Escherichia coli* Strains by Whole Cell and Outer Membrane Protein Extracts with Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2010**, *9*, pp 3647–3655.

12. UniProtKB Protein Knowledgebase [Online]; http://www.uniprot.org/help/uniprotKB (accessed October 2012).

13. *HAMAP: Escherichia coli (Strain 55989 / EAEC) Complete Proteome*. ExPASy Bioinformatics Resource Portal [Online]; http://hamap.expasy.org/proteomes/ECO55.html (accessed October 2012).

14. Kupferschmidt, K. Scientists Rush to Study Genome of Lethal *E. coli*. *Science* **2011**, *332*(6035), pp 1249–1250.

15. Al-Safadi, R.; Abu-Ali, G.S.; Sloup, R.E.; Rudrik, J.T.; Waters, C.M.; Eaton, K.A.; Manning, S.D. Correlation between *In Vivo* Biofilm Formation and Virulence Gene Expression in *Escherichia coli* O104:H4. *PLoS ONE* **2012**, *7*(7).

ACRONYMS AND ABBREVIATIONS

ABC	ammonium bicarbonate
ABOid	(software for the classification and identification of agents of biological origin)
ACN	acetonitrile
EAEC	enteroaggregative *E. coli*
ECBC	U.S. Army Edgewood Chemical Biological Center
EHEC	enterohemorrhagic *E. coli*
EPEC	enteropathogenic *E. coli*
FA	formic acid
LC	liquid chromatography
MWCO	molecular weight cutoff
MS	mass spectrometry
NCBI	National Center for Biotechnology Information
ND	not determined
ORF	open reading frame
RCF	relative centrifugal force
STB	sequence-to-bacterium (binary matrix)
TSB	tryptic soy broth


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