JCM Journals.ASM.org

Comparative Study of Traditional Flagellum Serotyping and Liquid Chromatography–Tandem Mass Spectrometry-Based Flagellum Typing with Clinical *Escherichia coli* Isolates

Keding Cheng,[a,b] Angela Sloan,[a] Lorea Peterson,[a] Stuart McCorrister,[a] Alyssia Robinson,[a] Matthew Walker,[a] Tracy Drew,[a] Joanne McCrea,[a] Linda Chui,[c,d] John Wylie,[e] Sadjia Bekal,[f] Aleisha Reimer,[a] Garrett Westmacott,[a] Mike Drebot,[a] Celine Nadon,[a] J. David Knox,[a,g] Gehua Wang[a]

National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada[a]; Department of Human Anatomy and Cell Sciences, Faculty of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada[b]; Alberta Provincial Laboratory for Public Health, Calgary, Alberta, Canada[c]; Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, Alberta, Canada[d]; Cadham Provincial Laboratory, Manitoba Health, Winnipeg, Manitoba, Canada[e]; Laboratoire de Santé Publique du Quebéc, Saine-Anne-de-Bellevue, Quebéc, Canada[f]; Department of Medical Microbiology, Faculty of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada[g]

Traditional methods of phenotyping *Escherichia coli* bacteria include the serotyping of surface O antigens (lipopolysaccharides), capsule K antigens, and H antigens found on the bacteria's flagellar filaments (1). Despite their usefulness, these conventional antibody-based assays can be costly and laborious to perform due to the wide-ranging quality of antibodies (serum) and the number of antibody agglutination reactions needed to assign a final classification (2, 3). H serotyping is further lengthened by the motility induction required before typing among many distinctive flagellar antigens (H1 to H56; designations H13, H22, and H50 are no longer in use [2, 3]).

Molecular typing methods using PCR-based amplification and genetic sequencing are gaining popularity for serotype classifications of *E. coli* due to their potential for high throughput and accuracy (4, 5). Problems with this approach, however, arise because genetics do not necessarily indicate phenotypes and because multiple primers need to be used for amplifying the sequences of unknown antigens. Matrix-associated laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry (MS) usage for whole-bacterial-protein profiling to classify and type bacteria has also shown some promising results due to its ease of use and high-throughput potential (6, 7). This platform can currently produce data for some successful bacterial subtyping at the species level but hardly the strain level (i.e., the H and O antigen levels [8]).

Recently, we reported a new approach called MS-H for H typing *E. coli* by purifying and digesting *E. coli* flagella of reference strains of all 53 known H types on a 0.22-μm-pore-size filter membrane, followed by online liquid chromatography-tandem mass spectrometry (LC-MS/MS) of the resulting flagellin peptides. The H-antigen serogroups from flagellin peptide data were identified using a minimum of two sequence-specific peptides (9), and MS-H types were assigned as the top-scoring hit in the identified protein list possessing the highest confidence score (10). An example of data output is shown in Fig. S1 in the supplemental material. This novel approach is thoroughly evaluated here by using 127 clinical isolates, including both motile and nonmotile strains collected from three provinces in Canada over the 1-year period of 2012, and by employing a standardized high-throughput method. The evaluation process comprises two steps: preliminary tests and confirmation assays. Preliminary tests were conducted using both the previously described MS-H typing method and the ISO-certified serotyping method in parallel (10). Rather than testing samples of known H types one by one, however, MS-H was conducted in a blind, high-throughput, batch-by-batch mode in order to replicate the clinical-sample scenario. In the absence of motility induction, flagella of 8 isolates of unknown motilities per batch were extracted and digested, together with 2 reference strains of known H types as control samples. Flagellar peptide sequence data were searched against an updated custom database containing all *E. coli* flagellar protein sequences extracted from the publicly available database NCBI nr (see Text S1 in the supplemental material [FASTA file type]). MS-H types were designated the top hit based on the number of H-type-specific peptide sequences identified with confidence. The serotyping procedure involved motility induction and antigen-antibody agglutination reactions to designate an H type (see Text S2). Confirmation assays were conducted with different techniques 1 year after preliminary tests were completed in order to determine correct types if the H types assigned to any isolates by the two methods did not agree.

The results of this clinical-sample evaluation are summarized in Table 1. Detailed isolate information is presented in Tables S1 through S6 in the supplemental material. For 75 of the 127 clinical isolates preliminarily tested (59%; 61 motile and 14 nonmotile by serotyping) (Tables S1 and S2 combined), the results of MS-H and traditional serotyping were in agreement with regard to the H type or absence of flagella. For an additional 5 samples (4%) (Table S3), there was agreement on the isolate H type between the two methods when MS-H typing was performed on motility-induced isolates during confirmation assays. For the remaining 47 samples (37%) (Tables S4, S5, and S6), for which traditional serotyping results were contradicted by MS-H results obtained from preliminary tests, motility induction and/or genetic sequencing of the *fliC* gene (11) was used to conclusively establish the H types of the isolates during confirmation assays. For samples whose genetic

Published ahead of print 26 March 2014

Editor: E. Munson

Address correspondence to Keding Cheng, keding.cheng@phac-aspc.gc.ca, or Gehua Wang, gehua.wang@phac-aspc.gc.ca.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.00174-14.

doi:10.1128/JCM.00174-14

The authors have paid a fee to allow immediate free access to this article.

TABLE 1 Summary of MS-H's performance on clinical *E. coli* isolates in comparison with traditional serotyping[a]

Preliminary testing[b]		Results of indicated confirmation assay(s)[c]			
No. of isolates (n = 127)	Categories	Repeated MS-H and/or motility assay	PCR and genetic sequencing of *fliC*	Whole-genome sequencing	Conclusions
75 (61 motile, 14 NM)	Serotyping and MS-H (results were in agreement)	NA	NA	NA	NA
20	Serotyping (all isolates were undetermined) and MS-H (all isolates were identified with H typing)	16 preliminary MS-H types were confirmed; 4 isolates were confirmed to be NM	14 preliminary MS-H types were confirmed	2 preliminary MS-H types were confirmed	16 of 20 isolates not determined by serotyping were clearly identified by MS-H; sample carryover during LC-MS/MS caused 1 unmatched H type during preliminary testing
17	Serotyping and MS-H (results were different)	3 preliminary serotyping results were confirmed after motility induction	10 preliminary MS-H types were confirmed	4 preliminary MS-H types were confirmed	14 of 17 isolates were incorrectly identified by serotyping but were correctly identified by MS-H; sample carryover during LC-MS/MS caused 1 unmatched MS-H type during preliminary testing
10	Serotyping (all isolates were deemed NM) and MS-H (all isolates were identified with H typing)	6 isolates were confirmed to be NM; 2 isolates were found to be motile, with clear MS-H types; 2 isolates were found to be sluggish, with clear MS-H types after motility induction	2 preliminary MS-H types were confirmed	2 preliminary MS-H types were confirmed	4 of 10 isolates were mislabeled as NM by serotyping; sample carryover during LC-MS/MS caused 5 false MS-H types during preliminary testing
5	Serotyping (all isolates were identified with H typing) and MS-H (no isolates were identified)	All 5 isolates were identified after motility induction	4 preliminary serotyping results were confirmed	1 preliminary serotyping result was confirmed	Identification by MS-H was aided by motility induction

[a] NA, not applicable; NM, nonmotile.
[b] Using the traditional standard of protein identification by mass spectrometry (9).
[c] Using the refined MS-H standard (see Text S2 in the supplemental material).

sequence data could not be obtained following traditional PCR (9 isolates), whole-genome sequencing (WGS) was used to obtain *fliC* sequences (12). In 17 instances (Table S6), different H types were assigned to an isolate by preliminary serotyping and MS-H testing. In these situations, *fliC* analyses demonstrated that the preliminary MS-H method had accurately typed 14 of the isolates (3 were mistyped due to the need for motility induction, 1 of which was also mistyped due to sample carryover), while serotyping correctly identified only 3. After a refinement of the MS-H protocol was established (described in Text S2) in confirmation tests, all 17 isolates were successfully identified by MS-H. The remaining 30 isolates were classified as either undetermined (Table S5) or nonmotile (Table S4) by traditional serotyping. Using MS-H, 23 were assigned an H type during preliminary testing (Tables S4 and S5 combined). Retesting motility during confirmation assays indicated that 4 of the 20 isolates classified as undetermined by serotyping and only 6 of the 10 isolates classified as nonmotile were nonmotile. MS-H typing results for the remaining 20 motile isolates were confirmed as correct by genetic sequencing of the *fliC* gene, including 2 "sluggish" (13) isolates classified as nonmotile in preliminary serotyping tests. Thus, MS-H correctly H typed 20 samples that could not be typed by conventional means but at the same time incorrectly assigned an H type to 1 confirmed motile and 5 confirmed nonmotile samples during preliminary tests based on the traditional protein identification standard of a minimum of two sequence-specific peptides (9). We hypothesized that these misclassifications were due to sample carryovers from previous motile sample runs. Sample carryover is a common phenomenon in high-throughput LC-MS/MS (14, 15), and this problem was identified and solved during confirmation assays using repeated blank washes and a higher sequence coverage requirement (rather than the requirement of two sequence-specific peptides used in conventional protein identification [9]), for the assignment of MS-H types (Text S2 and Table 1). In brief, blank washes consisted of one rigorous wash comprising several organic solvent cycles, followed by a second wash applying a solvent gradient like that of a sample run. Regarding higher sequence coverage, the need for exponentially modified protein abundance index (emPAI) values of 1 or greater was enforced, indicating that at least 30% of detectable peptides were indeed detected (16). With this refined method, MS-H correctly typed 93 of the 127 isolates (73.2%) without motility induction or 103 of 127 (81.1%) if the 10 H types correctly identified after motility induction were included. This is in contrast to the performance of traditional serotyping, which correctly assigned 69 of the 127 isolates (54.3%) with routine motility induction. Even without motility induction, MS-H identified 19% more isolates than did conventional serotyping. One possible reason is that antibody serum quality affected the agglutination reaction (17, 18) and final serotype assignment. It is also possible that increasing the length of motility induction serves to identify certain isolates during serotyping, albeit in a prolonged manner not suitable for emergency situations. Providing an extra wash step between samples and raising the threshold of protein identification eliminated the false positives produced by MS-H due to sample carryover that occurred during the high-throughput LC-MS/MS step of the procedure. Notably, the increased threshold did not affect the H typing results of motile strains due to the high protein sequence coverage (often 30% or above) already provided by the quality of the MS-H data (10). It is conceivable that had this modification been made prior to the start of this study, only 2 isolates would have been wrongly identified by MS-H and its accuracy could have been 115 of 127 isolates (90.5%) without motility induction (10 isolates), compared to that of serotyping with motility induction at 89 of 127 isolates (70.1%, with 20 undetermined and 18 wrongly identified). Since O-antigen serotyping is more stable than H-antigen typing and does not require motility induction (1, 13), its combination with MS-H will certainly render the process of *E. coli* typing more efficient.

In conclusion, this 1-year study of 127 clinical isolates showed that MS-H is more diagnostically sensitive for *E. coli* H typing than is serotyping. MS-H is also faster and easier with regard to sample preparation and detection, resulting in more accurate data output at the protein sequence level. Of utmost importance, this is a molecular-level detection platform with phenotypic properties considered. Using a refined standard operating protocol with flagellar preparation, high-throughput LC-MS/MS with extra wash steps, and a higher threshold for protein identification, MS-H should prove a valuable supplementary or independent method for *E. coli* H typing, particularly in outbreak situations.

ACKNOWLEDGMENTS

We thank Helen Tabor and Matthew Gilmour for supporting this project at its early stage.

K.C. designed the method and confirmation assays, summarized data, and wrote the manuscript. A.S. refined the flagellum database, performed sequence alignments of *fliC* sequences, and provided critical writing. L.P., A. Robinson, and T.D. performed serotyping, preliminary flagellum extraction, and MS-H sample preparation. S.M. executed mass spectrometry runs. M.W. and A. Reimer performed whole-genome sequencing and sequence assembly. L.C., J.W., and S.B. provided clinical isolates and critical writing. M.D., C.N., and J.D.K. contributed project ideas and critical writing. G. Westmacott contributed to database creation and search results for preliminary tests. G. Wang contributed project ideas, method design, flagellum extraction, data summary, and manuscript writing.

We declare that we have no competing financial interests.

REFERENCES

1. **Ewing WH.** 1986. Edwards and Ewing's identification of Enterobacteriaceae. Elsevier Science Publishing Co, Inc, New York, NY.
2. **Ratiner YA, Sihvonen LM, Liu Y, Wang L, Siitonen A.** 2010. Alteration of flagellar phenotype of *Escherichia coli* strain P12b, the standard type strain for flagellar antigen H17, possessing a new non-*fliC* flagellin gene *flnA*, and possible loss of original flagellar phenotype and genotype in the course of subculturing through semisolid media. Arch. Microbiol. **192:** 267–278. http://dx.doi.org/10.1007/s00203-010-0556-x.
3. **Murray PR, Baron EJ, Pfaller MA, Tenover FC, Yolken RH (ed).** 1999. Manual of clinical microbiology, 7th ed. ASM Press, Washington, DC.
4. **Machado J, Grimont F, Grimont PA.** 2000. Identification of *Escherichia coli* flagellar types by restriction of the amplified *fliC* gene. Res. Microbiol. **151:**535–546. http://dx.doi.org/10.1016/S0923-2508(00)00223-0.
5. **Fields PI, Blom K, Hughes HJ, Helsel LO, Feng P, Swaminathan B.** 1997. Molecular characterization of the gene encoding H antigen in *Escherichia coli* and development of a PCR-restriction fragment length polymorphism test for identification of *E. coli* O157:H7 and O157:NM. J. Clin. Microbiol. **35:**1066–1070.
6. **Welker M, Moore ER.** 2011. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. Syst. Appl. Microbiol. **34:**2–11. http://dx.doi.org/10.1016/j.syapm.2010.11.013.
7. **Karger A, Ziller M, Bettin B, Mintel B, Schares S, Geue L.** 2011. Determination of serotypes of Shiga toxin-producing *Escherichia coli* isolates by intact cell matrix-assisted laser desorption ionization–time of flight mass spectrometry. Appl. Environ. Microbiol. **77:**896–905. http://dx.doi.org/10.1128/AEM.01686-10.
8. **Clark CG, Kruczkiewicz P, Guan C, McCorrister SJ, Chong P, Wylie J, van Caeseele P, Tabor HA, Snarr P, Gilmour MW, Taboada EN,**

Westmacott GR. 2013. Evaluation of MALDI-TOF mass spectroscopy methods for determination of *Escherichia coli* pathotypes. J. Microbiol. Methods **94:**180–191. http://dx.doi.org/10.1016/j.mimet.2013.06.020.

9. **Taylor GK, Goodlett DR.** 2005. Rules governing protein identification by mass spectrometry. Rapid Commun. Mass Spectrom. **19:**3420. http://dx.doi.org/10.1002/rcm.2225.
10. **Cheng K, Drebot M, McCrea J, Peterson L, Lee D, McCorrister S, Nickel R, Gerbasi A, Sloan A, Janella D, Van Domselaar G, Beniac D, Booth T, Chui L, Tabor H, Westmacott G, Gilmour M, Wang G.** 2013. MS-H: a novel proteomic approach to isolate and type the *E. coli* H antigen using membrane filtration and liquid chromatography-tandem mass spectrometry (LC-MS/MS). PLoS One **8:**e57339. http://dx.doi.org/10.1371/journal.pone.0057339.
11. **Wang L, Rothemund D, Curd H, Reeves PR.** 2003. Species-wide variation in the *Escherichia coli* flagellin (H-antigen) gene. J. Bacteriol. **185:**2936–2943. http://dx.doi.org/10.1128/JB.185.9.2936-2943.2003.
12. **Gilmour MW, Graham M, Van Domselaar G, Tyler S, Kent H, Trout-Yakel KM, Larios O, Allen V, Lee B, Nadon C.** 2010. High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. BMC Genomics **11:**120. http://dx.doi.org/10.1186/1471-2164-11-120.
13. **Orskov F, Orskov I.** 1992. *Escherichia coli* serotyping and disease in man and animals. Can. J. Microbiol. **38:**699–704. http://dx.doi.org/10.1139/m92-115.
14. **Hughes NC, Wong EY, Fan J, Bajaj N.** 2007. Determination of carryover and contamination for mass spectrometry-based chromatographic assays. AAPS J. **9:**E353–E360. http://dx.doi.org/10.1208/aapsj0903042.
15. **Clouser-Roche A, Johnson K, Fast D, Tang D.** 2008. Beyond pass/fail: a procedure for evaluating the effect of carryover in bioanalytical LC/MS/MS methods. J. Pharm. Biomed. Anal. **47:**146–155. http://dx.doi.org/10.1016/j.jpba.2007.12.019.
16. **Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M.** 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol. Cell. Proteomics **4:**1265–1272. http://dx.doi.org/10.1074/mcp.M500061-MCP200.
17. **Pacheco AB, Ferreira LC, Pichel MG, Almeida DF, Binsztein N, Viboud GI.** 2001. Beyond serotypes and virulence-associated factors: detection of genetic diversity among O153:H45 CFA/I heat-stable enterotoxigenic *Escherichia coli* strains. J. Clin. Microbiol. **39:**4500–4505. http://dx.doi.org/10.1128/JCM.39.12.4500-4505.2001.
18. **Prager R, Strutz U, Fruth A, Tschape H.** 2003. Subtyping of pathogenic *Escherichia coli* strains using flagellar (H)-antigens: serotyping versus *fliC* polymorphisms. Int. J. Med. Microbiol. **292:**477–486. http://dx.doi.org/10.1078/1438-4221-00226.