Phenotypic H-Antigen Typing by Mass Spectrometry Combined with Genetic Typing of H Antigens, O Antigens, and Toxins by Whole-Genome Sequencing Enhances Identification of *Escherichia coli* Isolates

Keding Cheng,a,b Huixia Chui, Carissa Domish,a Angela Sloan,a Drexler Hernandez,a Stuart McCorrister,a Alyssa Robinson,a Matthew Walker,a Lorea A. M. Peterson,a Miles Majcher,a Sam Ratnam,a,b David J. M. Haldane,a Sadjia Bekal,c John Wylie,b Linda Chui,a,b Shaun Tyner,a Bianli Xu,a,c Aleisha Reimer,a Celine Naden,a,b,c,d J. David Knox,a,c,d Gehua Wanga

National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada; Department of Human Anatomy and Cell Sciences, College of Medicine, University of Manitoba, Winnipeg, MB, Canada; Henan Center of Disease Control and Prevention, Zhengzhou, Henan Province, China; Newfoundland and Labrador Public Health Laboratory, St. John’s, NL, Canada; Nova Scotia Provincial Public Health Laboratory Network, Halifax, NS, Canada; Laboratoire de Santé Publique du Québec, Sainte-Anne-de-Bellevue, QC, Canada; Cadham Provincial Laboratory, Winnipeg, MB, Canada; Provincial Laboratory for Public Health and Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, AB, Canada; Department of Medical Microbiology, College of Medicine, University of Manitoba, Winnipeg, MB, Canada;

Mass spectrometry-based phenotypic H-antigen typing (MS-H) combined with whole-genome-sequencing-based genetic identification of H antigens, O antigens, and toxins (WGS-HOT) was used to type 60 clinical *Escherichia coli* isolates, 43 of which were previously identified as nonmotile, H type undetermined, or O rough by serotyping or having shown discordant MS-H and serotyping results. Whole-genome sequencing confirmed that MS-H was able to provide more accurate data regarding H antigen expression than serotyping. Further, enhanced and more confident O antigen identification resulted from gene cluster based typing in combination with conventional typing based on the gene pair comprising *wzx* and *wzy* and that comprising *wzm* and *wzt*. The O antigen was identified in 94.6% of the isolates when the two genetic O typing approaches (gene pair and gene cluster) were used in conjunction, in comparison to 78.6% when the gene pair database was used alone. In addition, 98.2% of the isolates showed the existence of genes for various toxins and/or virulence factors, among which verotoxins (Shiga toxin 1 and/or Shiga toxin 2) were 100% concordant with conventional PCR based testing results. With more applications of mass spectrometry and whole-genome sequencing in clinical microbiology laboratories, this combined phenotypic and genetic typing platform (MS-H plus WGS-HOT) should be ideal for pathogenic *E. coli* typing.

*Escherichia coli* is a common bacterium, and pathogenic *E. coli*-contaminated food and/or water can cause severe health problems, such as hemolytic-uremic syndrome (HUS), in humans (1). Consequently, rapid and accurate identification and typing of *E. coli* are very important to those who have been affected by such contamination, particularly during *E. coli* outbreaks. *E. coli* is conventionally identified via biochemical tests (2) and typed by the serotyping of two major surface antigens, lipopolysaccharides (LPS; O antigens) and flagellar proteins (H antigens) (3). The pathogenicity of these bacteria is examined through cytotoxicity assays (4) or by checking for the presence of suspected toxins (or their related genes) by methods such as enzyme-linked immunosorbent assay or PCR, especially on common toxins such as Shiga toxin 1 (*Stx*1) or Shiga toxin 2 (*Stx*2) (4, 5). Currently, serotyping methods are not sufficiently rapid to quickly and accurately identify an unexpected *E. coli* strain because of the large number of possible O (*O1–O188; O31, O47, O67, O72, O94, and O122 have been withdrawn*) and H (*H1 to H56; H13, H22, and H50 have been withdrawn*) antigens (6–8). In addition, during O-antigen serotyping, rough strains often arise, likely because of unfavorable growth conditions or genetic mutations (9–11) that cause a lack of antigen-specific sugar/sugar chains on the LPS molecule. For H-antigen serotyping, clinical isolates routinely require flagellum growth induction to optimize the H-antigen–antiseraum agglutination reactions, a very time-consuming process (8). Even with motility induction, many isolates are still designated H nonmotile (NM) or undetermined after repeated serotyping (12). Over the past few years, we have focused our efforts on developing a new method for more accurate and rapid typing of *E. coli* H antigens by using mass spectrometry (8, 12). This H typing method was named MS-H (8), and validation using clinical isolates proved that it displayed higher speed, better sensitivity, and better specificity than conventional serotyping (12–14). In this study, we formally applied whole-genome sequencing (WGS) to the MS-H...
typing process, with a focus on resolving clinical isolates that could not be assigned an H type by conventional serotyping and on those isolates whose H serotypes were in disagreement with MS-H-assigned types. The H antigens, O antigens, and toxin genes of these isolates (representing NM, H-undetermined [Hund], and rough strains) were all analyzed. We termed the approach MS-H plus WGS-HOT (mass spectrometry-based H-antigen typing plus WGS-based H-antigen, O-antigen, and toxin identification). We proposed that if the approach was successful, a genotype- and phenotype-based method for targeted identification and typing of the important biomarkers of pathogenic E. coli (H antigens, O antigens, and toxins) could be created and used in clinical microbiology laboratories.

**MATERIALS AND METHODS**

**E. coli strains.** All of the E. coli reference strains in this study were obtained from the ISO-certified Enterics Reference Centre at the National Microbiology Laboratory, Public Health Agency of Canada. Sixty clinical strains used in this study for WGS were among 219 clinical isolates received from five Canadian provincial laboratories for public health (Alberta, Manitoba, Quebec, Newfoundland, and Nova Scotia) for routine serotyping and MS-H typing.

**H- and O-antigen serotyping.** For H-antigen serotyping, an ISO-certified method was used (8). Motility induction was performed for a maximum of 2 weeks. Motile strains underwent agglutination reactions with multivalent antiserum pools and formalin-treated bacterial culture. Monovalent antiserum were used to determine the serotype of a particular isolate. For O typing, no motility induction was needed but a similar strategy was used. Agglutination reactions were performed with multivalent antisera pools and formalin-treated bacterial culture. Monovalent antisera were then used to determine the O serotype of a particular isolate.

**Shiga toxin (verotoxin) testing by PCR.** The majority of the clinical isolates were found to contain stx1, and stx2 (the two most common E. coli toxin genes) by provincial laboratories by various PCR methods and reported as either verotoxin positive or stx1/stx2 positive. Some provincial laboratories also tested for hlyA (encoding hemolysin A) and eae (encoding intimin). The stx1, stx2, and hlyA genes of nonreported reference strains and clinical isolates at the National Microbiology Laboratory of Canada were amplified by a previously established protocol (15), with some modifications. In brief, four colonies resulting from overnight growth of freshly cultured cells were mixed with 1 ml of deionized water and boiled for 10 min. The mixture was then centrifuged at 15,000 × g for 10 min, and 1.5 μl of supernatant was mixed with 50 μl of a PCR mixture containing 1 × Mastermix (Bio-Rad) and 5 μM primers targeting the stx1, stx2, or hlyA gene (15). PCRs were performed under the following conditions: initial denaturation at 98°C for 5 min; 35 cycles of denaturation at 98°C for 15 s, annealing at 60°C for 20 s, and extension at 72°C for 20 s; and a final extension at 72°C for 8 min.

**MS-H typing.** Motility induction was not performed. A 10-μl loop of the overnight culture of each clinical isolate was used for flagellum extraction. Each culture was mixed in water, and the flagella were sheared from the cell body by vortexing for 60 s (8). The cell bodies were centrifuged at 16,000 × g for 20 min and the supernatant (containing detached flagella) was applied to a syringe filter (8, 12–14). The flagella were then washed and digested on the filter for 2 h at 37°C. The tryptic digests were flushed out of the filter with water and tested on the Orbitrap-XL (ThermoFisher) liquid chromatography-tandem mass spectrometry (LC-MS/MS) system. MS data were searched against a curated database containing E. coli flagellin sequences by using a Mascot search engine to obtain emPAI (exponentially modified protein abundance index) values (calculated as 10 to the power of [the number of observed peptides divided by the number of observable peptides] minus 1) (12). The following rules were used to assign an H type. (i) A minimum emPAI value of 1 was necessary. (ii) The top hit should have an emPAI value at least twice that of the previous adjacent blank run when the emPAI value of the previous blank run exceeded 1 because of carryover. (iii) Repeated jigsaw cleanups and LC-MS/MS analyses of samples were performed after blank runs showing emPAI values of ≥1. (iv) If the adjacent previous blank run had no significant flagellin identified (fewer than two specific peptides) and the emPAI value of the subsequent sample run was 0.10 to 0.99 because of low flagellum production from “sluggish” isolates, repeat testing should be performed with a larger sample amount to obtain an emPAI value of ≥1 (13, 14).

**WGS-HOT.** The work flow of WGS-HOT, together with MS-H typing, is shown in Fig. 1, and detailed methods are described in the supplemental material. In brief, two to four colonies of overnight E. coli culture were collected and the genomic DNA was extracted by using Epicentre Metagenomic DNA Isolation kits (Mandel Canada). The DNA quality was checked by visualizing electrophoresis products on a 1% agarose gel and quantified with a Qubit DNA quantification system (Invitrogen). At concentrations between 10 and 50 ng/μl, WGS was performed with a Nextera XT DNA Sample Preparation kit (Illumina) and data acquired by 300-bp paired-end sequencing on the Illumina MiSeq with the MiSeq Reagent kit V2 (600 cycles). An *in silico* analysis of the sequences of each isolate was performed, where data were assembled into contigs with SPAdes Assem-
TABLE 1 Results of MS-H plus WGS-HOT analysis of *E. coli* reference strains

| Strain | Serotype | MS-H | H-antigen gene database | O-antigen gene database | Gene cluster | Gene pair 90% | Gene pair 80% | Gene pair 70% | Gene pair 60% | Toxin and virulence factor gene database |
|--------|----------|------|-------------------------|-------------------------|-------------|--------------|--------------|--------------|--------------|----------------------------------------|
| E32511b | O157:HNM | NMd | H7 | O157-fcd | O157-wzx | NA | NA | NA | NA | Stx2, HlyA, intimin, 14 others |
| EDL933b | O157:H7 | H7 | O157-wbdQ | NO | NO | O157-wzx | NA | NA | NA | Stx1, Stx2, HlyA, intimin, 15 others |
| 90-2380b | O157:H7 | H7 | O157-wbdO | O86-gne | O157-wzx | NA | NA | NA | Stx2, HlyA, 12 others |
| H19b | O26:H11 | H11 | H11 | O26-wbuA | O26-wzy | NA | NA | NA | Stx1, HlyA, intimin, 14 others |
| 12-0721a | O26: H11 | H11 | H11 | O26-wzy | O26-wzy | NA | NA | NA | Stx2, HlyA, 7 others |
| B2F1b | O91:H21 | H21 | H21 | O91-wbsB | O91-wzx | NA | NA | NA | Stx2, HlyA, 7 others |
| 86-704a | O15:H27 | H27 | H27 | O15-wzy | O15-wzy | NA | NA | NA | Stx2, HlyA, 7 others |
| H.1.8b | O128:B17 | H2 | H2 | O128-wbsL | NO | O128-wzx | NA | NA | Stx2, HlyA, intimin, 10 others |
| 12-2628b | O103:H25 | H25 | H25 | O103-wbfF | O103-wzx | NA | NA | NA | Stx2, HlyA, intimin, 10 others |
| 12-0747c | O25:H4 | H4 | H4 | O25-rmA | NO | NO | NO | O25-wzy | 6 others |
| 09-0414d | O104:H7 | H7 | H7 | O104-manC | NO | NO | NO | O104-wzx | HlyA, subtilase, 8 others |
| K-12 | Orough:H48 | H48 | H48 | O16-wzy | O16 wzy | NA | NA | NA | 3 others |

* E. *coli* reference strains were identified by serotyping and MS-H and further analyzed by WGS-HOT using curated databases. Parameters were set at a 90% minimum coverage cutoff for the H-antigen, O-antigen gene cluster, and toxin database searches, while various degrees of coverage (60 to 90%) were set for the O-antigen gene pair database.

a Literature has shown that these strains are known to have common toxin genes such as *stx*1, *stx*2, and *hlyA* (15).

b Serotypes were reported earlier (12), and *stx*1, *stx*2, and *hlyA* genes were confirmed by PCR.

c NM, nonmotile.

d Literature has shown that these strains are known to have common toxin genes such as *stx*1, *stx*2, and *hlyA* (15).

e Literature has shown that these strains are known to have common toxin genes such as *stx*1, *stx*2, and *hlyA* (15).

f H types confirmed by matrix-assisted laser desorption ionization–time of flight mass fingerprinting (16).

g H types confirmed by LC-MS/MS (12).

h Non-specific gene based on reference 19.

i NA, not applicable.

j NO, not obtainable.

k Only important toxins (15) were mentioned, and the rest were categorized as “others” (*Stx*1, Shiga-like toxin 1; *Stx*2, Shiga-like toxin 2; *Hly*A, hemolysin A).
Inconsistent with serotype

WGS-HOT tests of five Hund isolates (see Table S4 in the supplemental material) indicated clear H types by MS-H, which were confirmed by WGS. This observation confirmed previous findings indicating that MS-H is more specific than traditional serotyping for identifying H types at the protein sequence level (12). Moreover, for 21 isolates without discordant serotypes and MS-H types, the WGS H types were in better agreement with MS-H (12/21; 57.1%) than with serotyping (4/21; 19.0%) (see Table S5 in the supplemental material). Notably, WGS did not provide results concordant with the phenotypic data in four cases and could not designate an H type for one isolate.

Table 2 summarizes the results of the three H-typing platforms (i.e., serotyping, MS–H, and WGS) for all 60 of the clinical isolates under examination. These data suggest that MS-H typing is more sensitive and accurate than traditional serotyping for NM isolates, but not as sensitive and accurate as WGS. However, WGS-HOT did not identify all the O antigens characterized by WGS, which do not necessarily represent phenotypic H types, as genes responsible for flagellum production are not always expressed (27).

Table 3 presents the results of MS-H plus WGS-HOT analyses of 56 clinical isolates having clear O types by O-antigen serotyping (no O rough strains included; see Tables S2 to S5 in the supplemental material) are summarized in Table 3. When the gene pair database was used for O-antigen identification and typing, most of the top hits (44/56; 78.6%) matched the corresponding serotype designations. The remaining 12/56 isolates (21.4%) were identified as having O types different from those designated by serotyping (2/56, 3.6%), showed no O type at all although a different coverage threshold was applied (5/56, 8.9%), or showed multiple hits with similar confidence (5/56, 8.9%). These multiple hits with high sequence similarity were also reported by others (20, 21, 24, 28); for example, O123/O186, O77/O17, and O117/O107. When the gene cluster database was used, all of the isolates were given O-antigen designations, with 45/56 isolates (80.4%) matching those identified by serotyping when housekeeping genes (e.g., gnd, gne, ugd, wzz, galF, and hisD) were ruled out (19–22, 28, 29). Similar to the results obtained with reference strains, a large number (all but two isolates) of top hits from the gene cluster database were not based on the gene pair comprising wzz and wzy or that comprising wzm and wzt. This phenomenon was exemplified when an O type was revealed for two unidentified isolates (14-8954 and 14-7998) when the gene pair database was applied.

### Table 2 Test summary and comparison of E. coli isolates by serotyping, MS–H typing, and WGS based H typinga

| Serotyping category (no. of isolates) | No. of MS–H types obtained | No. of serotypes concordant with WGS assigned H types | No. of MS–H types concordant with WGS assigned H types | No. of WGS assigned H types in disagreement with serotype and MS–H typesb |
|-------------------------------------|-----------------------------|-----------------------------------------------------|-----------------------------------------------------|---------------------------------------------------------------------|
| Serotype concordant with MS–H typea (17) | 17                          | 17                                                  | 17                                                  | 0                                                                   |
| NM isolates (17)                     | 0                           | NA                                                  | NA                                                  | 6                                                                   |
| Undetermined H type (5)              | 5                           | NA                                                  | 5                                                   | 0                                                                   |
| Serotype different from MS–H type (21) | 21                          | 4                                                   | 12                                                  | 5d                                                                  |

a Tests of all 60 isolates were summarized.
b The isolates tested were randomly chosen from 219 clinical isolates.
c NA, not applicable.
d There was no WGS result for 1 isolate; WGS was performed, but no H type result was obtained.

d### Table 3 Results of MS–H plus WGS-HOT analyses of the O-antigen and toxin genes of 56 clinical E. coli strains

| Antigen ID | No. of isolates/total (%) in O-antigen gene database | No. of isolates/total (%) in toxin/virulence factor gene database |
|------------|------------------------------------------------------|------------------------------------------------------------------|
| Identical to serotype | Gene paira 44/56 (78.6)c | Gene cluster 45/56 (80.4)c | Both databases 53/56 (94.6) | 55/56 (98.2)d |
| Inconsistent with serotypec | 12/56 (21.4)c | 11/56 (19.6)c | 3/56 (5.4) | 1/56 (1.8)d |

a Cumulative results over all coverage levels (60 to 90%).
b Only toxin detection results obtained by the WGS method are shown.
c Multiple hits involving similar O types (two isolates with O123 and O186; two isolates with O77, O17, O44, and O73; and one isolate with O8 and O174) were not included.
d Five of the 12 isolates had no hit at all, another 5 had two or more hits (two isolates with O123 and O186; two isolates with O77, O17, O44, and O73; and one isolate with O8 and O174 for wzm), and two had hits different from the serotype.
e Common genes (gne, gnd, wzz, hisD, galF) were excluded when multiple hits were observed on the basis of current literature reports but were used when a single hit matched the serotype; multiple hits involving similar O types (one isolate with O123 and O186; one isolate with O117 and O107; one isolate with O77, O17, O44, and O73; and one isolate with O8 and O174 for wzm) were not included.
f Seven isolates had multiple “top hits” that were difficult to elucidate on the basis of current literature reports, while four isolates showed hits different from the serotyping results.
g Showed presence of toxin/virulence factor genes.
h Did not show presence of toxin/virulence factor genes.
alone. This observation confirmed that other genes could be useful for O-antigen identification. The combined results of the two O-antigen databases proved complementary, as 94.6% of the O antigens were in agreement with serotyping, compared to 78.6% with the gene pair database alone.

Four (6.7%) of the 60 clinical isolates were identified as O rough (see Table S6 in the supplemental material). When samples were analyzed by using the gene cluster database, O-antigen-related genes were identified in all four isolates, though all were housekeeping genes (galF, gne, or ugd) (19–22). A gene pair database search provided an O type for only one isolate (11-5490), while WGS analysis performed twice on strain 14-6184 with separate sample preparations showed the same result with no specific O-type-related gene identified.

Since two housekeeping genes, galF and ugd, showed some specificity for O-antigen identification, these two genes were compared across three O types (see Table S7 in the supplemental material). The genes did show differences in sequence length (galF, 221 to 784 nucleotides; ugd, 1,128 to 1,167 nucleotides) and nucleic acid composition (galF, 80.96% to 96.08% similarity; ugd, 96.66 to 96.92% similarity).

WGS analysis based on the toxin gene database showed the existence of various toxin genes/virulence factors in 59/60 (98.3%) clinical isolates (Table 3; see Table S6 in the supplemental material). The stx1, stx2, and hlyA genes were identified in the majority of the PCR-confirmed cases (the hlyA gene was identified in all but four isolates). Additional Shiga toxin subtypes 2c, 2d, and 2f and another common toxin, intimin (eae), not routinely tested for by all provincial clinical microbiology laboratories, were found. Subtilase, representing a family of E. coli toxins linked to HUS (30), was detected in one reference strain (09-1414) and two clinical isolates (11-6008 and 14-6184). Reference strain EDL933 and three clinical isolates (14-4602, 14-4603, and 14-4604) were found to have both the stx1 and stx2 genes by WGS. Many virulence factors were also identified by WGS, some of which may not be directly involved in the pathogenicity of the cells. For example, three virulence factors were identified for nonpathogenic strain K-12. Similarly, if only well-verified E. coli toxins (stx1, stx2, hlyA, intimin, subtilase, etc.) were considered, some clinical isolates might be regarded as nonpathogenic.

The identifications of common toxins (stx1, stx2, and hlyA) by WGS-HOT and the CGE platforms SeroTypeFinder and VirulenceFinder were very similar (data not shown). Of the 20 strains analyzed for H antigens, only 1 was incorrectly identified by Galaxy (and accurately identified by SeroTypeFinder). The O antigen of four clinical isolates was correctly determined by Galaxy (two were based on the gene cluster database only) but not SeroTypeFinder, even at the lowest search threshold (see Table S8 in the supplemental material) (24). Neither WGS-HOT nor CGE software was able to identify two O rough isolates.

The performance of the three platforms (serotyping, MS-H typing, and WGS-HOT) is summarized in Table 4.

**DISCUSSION**

This study represents an expansion of MS-H of E. coli (8, 12–14) to MS-H plus WGS-HOT. Initially, MS-H was designed to reduce the duration of outbreak investigations by alleviating the lengthy process of traditional H serotyping. Serotyping is often time-consuming because of the necessary procedure of motility induction. With MS-H, motility induction can be avoided and H types can be obtained after overnight culture. MS-H can also be used to identify isolates shown to be problematic for serotyping (e.g., Hund or O rough strains) because of a variety of factors, such as low-quality antisera and autoagglutination. WGS was originally used to confirm the results of H serotyping and MS-H, as PCR could not obtain all of the H-antigen sequences easily or quickly because of the necessity of using different primer sets (12). We believe that combining MS-H and WGS will dramatically reduce the amount
of time required to identify clinical E. coli H antigens, which will certainly prove useful in outbreak situations.

With a focus on problematic clinical isolates such as serotyping-designated Hund, NM, and rough strains, as well as those with discordant serotyping and MS-H results, this study demonstrated that MS-H is a more sensitive and specific method than serotyping because of its molecular-level identification. Essentially any tryptic peptide that can be ionized in an MS system is identifiable, a stark contrast to the limited number of antigenic epitopes available during serotyping. In addition, some H types (e.g., types H11 and H21 and types H14 and H17) are difficult to differentiate by serotyping, as their flagellar sequences are very similar (8, 12, 14). As MS-H can obtain virtually full sequence coverage of flagellar proteins, differentiation of similar strains does not pose a problem (8, 14). And because MS-H is a phenotypic identification platform, the method can distinguish motile from NM strains. Although WGS analysis also provides H type information, the method is not phenotypic and therefore cannot differentiate between motile and NM strains. The MS-H plus WGS method should be a more confident approach for H-antigen typing, with respect to both phenotypic and genetic identification.

The lack of agreement of the H types of a small number of isolates across the three platforms may be due to factors such as sample preparation during WGS, antibody quality during typing, and the motility status of strains at the time of serotyping or MS-H typing (8, 12–14, 24). On the basis of current data, if both MS-H and WGS are employed and different H types result, the MS-H result should be considered correct, as it reflects the phenotypic existence of the H antigen. To confirm the result, the MS-H procedure can be easily repeated (8, 12–14).

After addressing the issues associated with traditional H serotyping, it became obvious that platforms such WGS could be used to address problems arising with O-antigen serotyping as well. Although O serotyping is faster than H serotyping because there is no need for motility induction, there are many more E. coli O antigens. Serotyping of each of the 182 O antigens is a tedious and complicated procedure, and high-quality antisera remain a necessity. Because O antigens are not proteins but LPS involving various proteins/enzymes to synthesize and transport molecules, molecular approaches such as WGS are ideal for identifying unknown O antigens. Current literature reports indicate that the gene pair comprising wzx and wzy or that comprising wzr and wzl is the best candidate for resolving E. coli O antigens (6, 7, 24). Notably, these gene pairs do not include the O14 and O57 antigens, and as a result, they cannot be typed in this manner (24). This study has demonstrated that other O-antigen gene clusters involved in O-antigen synthesis and transport can be used to complement the gene pair comprising wzx and wzy or that comprising wzr and wzl in the database, ultimately ensuring more confident identification. While using the gene cluster database, it also became apparent that some housekeeping genes (e.g., galF and gndl) showed specificity for some O antigens. This represents the first observation regarding the role of housekeeping genes in O-antigen identification. Although we do recommend keeping housekeeping genes in database searches, a very cautious procedure should be used on the basis of current data collected for E. coli O-antigen typing. First, if a nonhousekeeping gene is identified, it will be given priority over housekeeping genes, and a housekeeping gene that indicates a different O type from the nonhousekeeping gene will be ignored. Second, if a housekeeping gene is the only gene identified in the data output, then different parameters should be used to research the database for a nonhousekeeping gene (e.g., by using a different coverage threshold, as shown by Joensen et al. [24] with SeroTypeFinder). Third, if a housekeeping gene remains the only hit after a second database search with altered parameters, a repeat of the test involving sample preparation and a rerun on the genomic sequencer should be performed to confirm the result.

Incomplete expression of O antigens by O rough isolates is a well-observed phenomenon (9–11). Literature reports have shown a low rate of O-antigen-specific gene identification, likely because of genetic mutations in O-antigen-related genes (24). It will therefore be a challenge to provide O types for all of the rough strains confidently. Further study is required in this area.

Regarding E. coli toxin identification, WGS has proven to be a very powerful platform, particularly for clinical microbiology laboratories that perform routine testing for common toxins such as Shiga toxins, hemolysin, and intimin. Because of the large number of known and unknown toxins (and their related subtypes), the use of traditional identification methods such as PCR is challenging because of the requirement for numerous primer sets and optimizing conditions. Thus far, the strategy of clinical microbiology laboratories has been to identify only common toxins (e.g., Stx1, Stx2, hemolysin A, and intimin), which means that other severe toxins (e.g., subtilase) are going undetected. We recommend that WGS be used to identify the genetic existence of toxins and virulence factors in E. coli outbreak situations, before epidemiological data and other testing results are obtained to finalize their phenotypic presence and potential pathogenicity.

In this study, we found that the Galaxy database search platform gave results very similar to those obtained with web-based SeroTypeFinder and VirulenceFinder for H-antigen and toxin identification, but differences were seen regarding O-antigen typing. Galaxy appears more sensitive, possibly because of differences in the sequence assembly algorithm or other search engine software parameters. Employing the gene cluster database did help increase the number of O antigens identified with Galaxy, but SeroTypeFinder could only employ the gene pair database as a default. We feel that the gene cluster database is very useful for O-antigen typing and should be integrated into O-antigen typing at the molecular level.

In summary, we believe that the MS-H plus WGS-HOT platform, involving phenotypic and genomic H-antigen typing to differentiate motile and NM strains, in combination with genomic typing of O antigens and toxins by WGS, will provide an ideal platform for pathogenic E. coli characterization. With the growing familiarity of MS and WGS techniques; their excellent sensitivity and quick turnaround time, user-friendly hardware, mature databases, and better bioinformatic tools; and the lower cost of reagents, this platform should be very useful in the rapid and accurate identification and typing of pathogenic E. coli.

ACKNOWLEDGMENTS

We thank the Proteomics Core, the Genomics Core, and the Bioinformatics Core at the National Microbiology Laboratory (NML), Public Health Agency of Canada (PHAC), for providing services and Mike Drebot and Matthew Gilmour for their critical support.

We also appreciate Genomics Research and Development Initiative (GRDI) funding from PHAC, which provided financial support for this project.
Wang contributed project ideas and critical writing. K.C. and G.W. contributed clinical isolates and critical writing. A.S., M.D., B.X., C.N., and J.D.K. contributed project ideas and critical writing. K.C. and G.W. contributed project ideas, method design, data summary, and manuscript writing.

REFERENCES

1. Mayer CL, Leibowitz CS, Kurosawa S, Stearns-Kurosawa DJ. 2012. Shiga toxins and the pathophysiology of hemolytic uremic syndrome in humans and animals. Toxins (Basel) 4:1261–1287. http://dx.doi.org/10.3390/toxins4111261.

2. Murray PR, Baron EJ, Pfaffer MA, Tenover FC, Yolken RH (ed). 1999. Manual of clinical microbiology. ASM Press, Washington, DC.

3. Kauffman F. 1947. The serology of the coli group. J Immunol 57:71–100.

4. Kumar HS, Karunasagar I, Tezou T, Shima K, Yamaoka S, Iyoda S, Ma J, Xu B, Reimer A, Knox JD, Wang G. 2015. Rapid, sensitive, and specific Escherichia coli H-antigen typing by matrix-assisted laser desorption ionization–time of flight-based peptide mass fingerprinting. J Clin Microbiol 53:2480–2485. http://dx.doi.org/10.1128/JCM.00593-15.

5. Sha, O, K, Y, C, A, J, D, H. performed MS-H and WGS-HOT sample preparation, database creation, and data analysis. S.M. performed MS-H testing, L.A.M.P., M.D., and A.R. performed serotyping, M.W. and S.T. aided with WGS and data analysis. L.C., J.Y., S.B., S.R., and D.J.M.H. provided clinical isolates and critical writing. A.S., M.D., B.X., C.N., and J.D.K. contributed project ideas and critical writing. K.C. and G.W. contributed project ideas, method design, data summary, and manuscript writing.

O157:H7 serotype, and components of the type 2 Shiga toxin family by multiplex PCR. J Clin Microbiol 40:3613–3619. http://dx.doi.org/10.1128/JCM.40.10.3613-3619.2002.

16. Chui H, Chan M, Hernandez D, Chong P, McCorristor S, Robinson A, Walker M, Peterson LA, Ratnam S, Haldane DJ, Bekal S, Wylie J, Chui L, Westmacott G, Xu B, Drebot M, Naden C, Knox JD, Wang G. 2015. Rapid, sensitive, and specific Escherichia coli H-antigen typing by matrix-assisted laser desorption ionization–time of flight-based peptide mass fingerprinting. J Clin Microbiol 53:2480–2485. http://dx.doi.org/10.1128/JCM.00593-15.

17. Cheng K, Sloan A, McCorristor S, Babiuk S, Bowden TR, Wang G, Knox JD. 2014. Fit-for-purpose curated database application in mass spectrometry-based targeted protein identification and validation. BMC Res Notes 7:444. http://dx.doi.org/10.1186/1756-0500-7-444.

18. Joensen KG, Schueft F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. J Clin Microbiol 52:1501–1510. http://dx.doi.org/10.1128/JCM.03617-13.

19. DebRoy C, Roberts E, Fratamico PM. 2011. Detection of O antigens in Escherichia coli. Annu Health Rev 12:169–185. http://dx.doi.org/10.1161/010.150466252311000193.

20. Wang W, Perepelyov AV, Feng L, Shevelev SD, Wang Q, Senchenkova SN, Han W, Li Y, Shashkov AS, Knirel YA, Reeves PR, Wang L. 2007. A group of Escherichia coli and Salmonella enterica O antigens sharing a common backbone structure. Microbiology 153:2139–2176. http://dx.doi.org/10.1099/mic.0.2006/004192-0.

21. Wang Q, Perepelyov AV, Wen L, Shashkov AS, Wang X, Guo X, Knirel YA, Wang L. 2012. Identification of the two glycosyltransferase genes responsible for the difference between Escherichia coli O107 and O117 O-antigens. Glycobiology 22:281–287. http://dx.doi.org/10.1039/glycob15137.

22. Feng L, Han W, Wang Q, Bastin DA, Wang L. 2005. Characterization of Escherichia coli O86 O-antigen gene cluster and identification of O86-specific genes. Vet Microbiol 106:220–228. http://dx.doi.org/10.1016/j.vetmic.2004.12.021.

23. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A. 2005. Galaxy: a platform for interactive large-scale genome analysis. Genome Res 15:1451–1455. http://dx.doi.org/10.1101/gr.4086505.

24. Joensen KG, Tettschner AM, Iguchi A, Aarestrup FM, Schueft F. 2015. Rapid and easy in silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. J Clin Microbiol 53:2410–2426. http://dx.doi.org/10.1128/JCM.00008-15.

25. Dobbin HS, Hove CJ, Williams CJ, Minnich SA. 2006. The Escherichia coli O157 flagellar regulatory gene flhC and not the flagellin gene flaC impacts colonization of cattle. Infect Immun 74:2894–2905. http://dx.doi.org/10.1128/IAI.74.7.2894-2905.2006.

26. Park JY, Clark P, Londin E, Sponziello M, Kricka LJ, Fortina P. 2015. Clinical exome performance for reporting secondary genetic findings. Clin Chem 61:213–220. http://dx.doi.org/10.1373/clinchem.2014.231456.

27. Ramos Moreano AC, Cabilio Guth BE, Baquerizo Martinez M. 2006. Can the flc PCR-restriction fragment length polymorphism technique replace classic serotyping methods for characterizing the H antigen of enterotoxigenic Escherichia coli strains? J Clin Microbiol 44:1453–1458. http://dx.doi.org/10.1128/JCM.44.4.1453-1458.2006.

28. Chen Y, Yan X, DebRoy C, Fratamico PM, Needleman DS, Li RW, Wang W, Losada L, Brinkac L, Radune D, Toro M, Hegde N, Meng J. 2015. Escherichia coli O-antigen gene clusters of serogroups O62, O68, O131, O140, O142, and O163: DNA sequences and similarity between O62 and O68, and PCR-based serogrouping. Biosensors (Basel) 5:31–58. http://dx.doi.org/10.3390/bios5010051.

29. Guo H, Yi W, Shao J, Lu Y, Zhang W, Song J, Wang PG. 2005. Molecular analysis of the O-antigen gene cluster of Escherichia coli O86:B7 and characterization of the chain length determinant gene (wzz). Appl Environ Microbiol 71:7995–8001. http://dx.doi.org/10.1128/AEM.71.12.7995-8001.2005.

30. Paton AW, Srinamote P, Talbot UM, Wang H, Paton JC. 2004. A new family of potent AB(5) cytotoxins produced by Shiga toxigenic Escherichia coli. J Exp Med 200:35–46. http://dx.doi.org/10.1084/jem.20043092.