Whole-Genome Sequencing Data for Serotyping *Escherichia coli*—It's Time for a Change!

Claire Jenkins

Gastrointestinal Bacteria Reference Unit, Public Health England, London, United Kingdom

The accessibility of whole-genome sequencing (WGS) presents the opportunity for national reference laboratories to provide a state-of-the-art public health surveillance service. The replacement of traditional serology-based typing of *Escherichia coli* by WGS is supported by user-friendly, freely available data analysis Web tools. An [article in this issue](http://dx.doi.org/10.1128/JCM.00008-15) of the *Journal of Clinical Microbiology* (K. G. Joensen, A. M. M. Tetzschner, A. Iguchi, F. M. Aarestrup, and F. Scheut, *J Clin Microbiol*, 53:2410–2426, 2015, <http://dx.doi.org/10.1128/JCM.00008-15>) describes SerotypeFinder, an essential guide to serotyping *E. coli* in the 21st century.

In public health microbiology, whole-genome sequencing (WGS) is a valuable resource used to combat emerging infectious disease and investigate outbreaks caused by a wide range of pathogens, including *Escherichia coli* (1, 2). The unprecedented level of discrimination and the evolutionary context that WGS provides has had a major impact on our understanding of the underlying epidemiology and pathogenicity of bacteria that are a threat to public health (3–5). However, the prospect of implementing WGS for public health surveillance and routine microbial typing can be daunting. Radical changes to conventional thinking, organizational infrastructure, and employee skill sets are required. Additional challenges include maintaining a consistent, high-quality service and producing results that are compatible with historical data and data generated by other stakeholders in the field. With respect to these additional challenges, [Katrine Joensen and colleagues' comprehensive and detailed article](#) in this edition of the *Journal of Clinical Microbiology* reveals why replacing phenotypic serotyping of *E. coli* with *in silico* serotyping using WGS is an easy switch (6).

In the traditional phenotypic serotyping scheme for *E. coli*, antisera to a combination of immunogenic structures, including lipopolysaccharide (LPS) (O) and flagellar (H) antigens, are raised in rabbits. The scheme, first developed in the 1940s (7), has stood the test of time because it is applicable across the *E. coli* species, offers an appropriate level of discrimination for outbreak detection and investigation, and provides a user-friendly designation for taxonomic differentiation and pathogenic groups. However, establishing, maintaining, and developing the scheme (comprising more than 188 O and 53 H antisera) is expensive and laborious and requires specialist resources and expertise. Consequently, phenotypic serotyping is provided by a limited number of reference laboratories worldwide.

Colleagues at the Center for Genomic Epidemiology (CGE) at the Danish Technical University (DTU), Lyngby, Denmark (<http://www.genomicepidemiology.org>), have developed a Web tool, SerotypeFinder, for WGS serotype prediction of *E. coli* based on the O- and H-antigen-processing genes. It is a freely available web tool that requires little or no expertise in bioinformatics to operate. The database was constructed using complete O and H antigen genes from the NCBI nucleotide collection, a comprehensive set of O antigen genes reported by Iguchi et al. (8), and WGS of *E. coli* reference strains from the WHO Collaborating Centre for *E. coli* at the Statens Serum Institute, Copenhagen, Denmark. All

database genes are compared against the genome of each test strain using the BLAST algorithm, and the output is the predicted O and H serotypes, based on the best-matching genes.

SerotypeFinder comprises all valid O types, O1 to O187 (except O14 and O57), and tolerates a wide range of O antigen gene variants, with the *wzx* and *wzy* gene variants clustering according to their O type. Certain pairs of O-antigen-encoding regions have identical or highly similar (>99.8%) *wzx* and/or *wzy* genes and are, therefore, potentially difficult to distinguish using SerotypeFinder, but in practice, the validation data demonstrated that the majority of these difficult identifications could be resolved and were assigned correctly despite the high similarity. Furthermore, the majority of these pairs represent known phenotypic serological cross-reactions and are difficult to differentiate using traditional methods.

Phenotypic H typing can be even more labor-intensive and time-consuming than O typing, as strains of *E. coli* must first be induced to express flagella *in vitro*. This initial step often requires multiple passages through a Craigie tube, which can take several days or even weeks. SerotypeFinder covers all 53 known H antigen genes, including over 100 variants. Sequence similarity of >99% was observed only between H4 and H17 variants, and this relationship was described previously (9).

There are many advantages to adopting the *in silico* WGS serotyping approach for typing *E. coli*. Once implemented, WGS can be faster and more cost-effective than traditional methods. Studies show that increasing numbers of strains of *E. coli* are reported as “O group unidentifiable” due to antisera failing quality control procedures, unresolvable cross-reactions, or novel serogroups (10, 11). *In silico* WGS serotyping avoids the need for the resource-

Accepted manuscript posted online 17 June 2015

Citation Jenkins C. 2015. Whole-genome sequencing data for serotyping *Escherichia coli*—it's time for a change! *J Clin Microbiol* 53:2402–2403. doi:10.1128/JCM.01448-15.

Editor: K. C. Carroll

Address correspondence to claire.jenkins@phe.gov.uk.

Copyright © 2015, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.01448-15

The views expressed in this Commentary do not necessarily reflect the views of the journal or of ASM.

intensive antiserum production process and the inherent quality control issues. Using WGS, problematic phenotypic cross-reactions appear, for the most part, to be resolved, and identifying and establishing novel O groups is much less demanding and operationally complex than producing and verifying new rabbit antisera for the phenotypic scheme. Strains of *E. coli* that are phenotypically untypeable due to lack of expression of O antigens (designated “rough”) or H antigens (designated “nonmotile”) are fully typeable by WGS. WGS data offers valuable insights into the degree of variation in the O- and H-antigen-encoding genes within each O and H type and the significance of cross-reactions between O groups. Furthermore, this approach provides the opportunity to examine the effect of mobile genetic elements, such as prophage, plasmids, and genomic islands, on O-antigen expression and evolution.

Without doubt, the unprecedented level of discrimination provided by WGS will inspire the community to develop and improve typing nomenclature. However, in this early phase of WGS implementation, the option of using the traditional serotype designation will minimize disruption for reference laboratory service users and facilitate data exchange with colleagues in the field. The *in silico* serotyping of *E. coli* using WGS data approach will help keep all our stakeholders “on board” at this time of radical change. The detailed validation and evaluation of SerotypeFinder in the article by Joensen et al. highlight both the positive features and the likely pitfalls of the approach and deliver an essential guide to serotyping *E. coli* in the 21st century.

REFERENCES

1. Dallman TJ, Byrne L, Launders N, Glen K, Grant KA, Jenkins C. 2015. The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing *Escherichia coli* serogroup O26:H11. *Epidemiol Infect* 143:1672–1680. <http://dx.doi.org/10.1017/S0950268814002696>.
2. Jenkins C, Dallman TJ, Launders N, Willis C, Byrne L, Jorgensen F, Eppinger M, Adak GK, Aird H, Elviss N, Grant KA, Morgan D, McLauchlin J. 2015. Public health investigation of two outbreaks of Shiga toxin-producing *Escherichia coli* O157 associated with consumption of watercress. *Appl Environ Microbiol* 81:3946–3952. <http://dx.doi.org/10.1128/AEM.04188-14>.
3. Joensen KG, Scheutz 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>.
4. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, Klammer A, Peluso P, Lee L, Kislyuk AO, Bullard J, Kasarskis A, Wang S, Eid J, Rank D, Redman JC, Steyert SR, Frimodt-Møller J, Struve C, Petersen AM, Krogfelt KA, Nataro JP, Schadt EE, Waldor MK. 2011. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med* 365:709–717. <http://dx.doi.org/10.1056/NEJMoa1106920>.
5. Dallman TJ, Byrne L, Ashton PM, Cowley LA, Perry NT, Adak G, Petrovska L, Ellis RJ, Elson R, Underwood A, Green J, Hanage WP, Jenkins C, Grant K, Wain J. 17 April 2015. Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. *Clin Infect Dis* <http://dx.doi.org/10.1093/cid/civ318>.
6. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz 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>.
7. Kauffmann F. 1947. The serology of the *E. coli* group. *J Immunol* 57:71–100.
8. Iguchi A, Iyoda S, Kikuchi T, Ogura Y, Katsura K, Ohnishi M, Hayashi T, Thomson NR. 2015. A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res* 22:101–107. <http://dx.doi.org/10.1093/dnares/dsu043>.
9. Beutin L, Strauch E, Zimmermann S, Kaulfuss S, Schaudinn C, Männel A, Gelderblom HR. Genetical and functional investigation of fliC genes encoding flagellar serotype H4 in wildtype strains of *Escherichia coli* and in a laboratory *E. coli* K-12 strain expressing flagellar antigen type H48. *BMC Microbiol* 5:4.
10. Byrne L, Vanstone GL, Perry NT, Launders N, Adak GK, Godbole G, Grant KA, Smith R, Jenkins C. 2014. Epidemiology and microbiology of Shiga toxin-producing *Escherichia coli* other than serogroup O157 in England, 2009–2013. *J Med Microbiol* 63:1181–1188. <http://dx.doi.org/10.1099/jmm.0.075895-0>.
11. Jenkins C, Chart H, Willshaw GA, Cheasty T, Tompkins DS. 2007. Association of putative pathogenicity genes with adherence characteristics and fimbrial genotypes in typical enteroaggregative *Escherichia coli* from patients with and without diarrhoea in the United Kingdom. *Eur J Clin Microbiol Infect Dis* 26:901–906. <http://dx.doi.org/10.1007/s10096-007-0388-z>.