



Whole Genome Sequencing for Public Health Surveillance of Shiga Toxin-Producing *Escherichia coli* Other than Serogroup O157

Marie A. Chattaway, Timothy J. Dallman, Amy Gentle, Michael J. Wright, Sophie E. Long, Philip M. Ashton, Neil T. Perry and Claire Jenkins *

Gastrointestinal Bacteria Reference Unit, Public Health England, London, UK

OPEN ACCESS

Edited by:

Chitrita Debroy,
The Pennsylvania State University,
USA

Reviewed by:

Remy Froissart,
Centre National de la Recherche
Scientifique, France
Seamus Fanning,
University College Dublin, Ireland

*Correspondence:

Claire Jenkins
claire.jenkins@phe.gov.uk

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 16 January 2016

Accepted: 16 February 2016

Published: 03 March 2016

Citation:

Chattaway MA, Dallman TJ, Gentle A,
Wright MJ, Long SE, Ashton PM,
Perry NT and Jenkins C (2016) Whole
Genome Sequencing for Public Health
Surveillance of Shiga Toxin-Producing
Escherichia coli Other than Serogroup
O157. *Front. Microbiol.* 7:258.
doi: 10.3389/fmicb.2016.00258

Shiga toxin-producing *Escherichia coli* (STEC) are considered to be a significant threat to public health due to the severity of gastrointestinal symptoms associated with human infection. In England STEC O157 is the most commonly detected STEC serogroup, however, the implementation of PCR at local hospital laboratories has resulted in an increase in the detection of non-O157 STEC. The aim of this study was to evaluate the use of whole genome sequencing (WGS) for routine public health surveillance of non-O157 STEC by comparing this approach to phenotypic serotyping and PCR for subtyping the *stx*-encoding genes. Of the 102 isolates where phenotypic and genotypic serotyping could be compared, 98 gave fully concordant results. The most common non-O157 STEC serogroups detected were O146 (22) and O26 (18). All but one of the 38 isolates that could not be phenotypically serotyped (designated O unidentifiable or O rough) were serotyped using the WGS data. Of the 73 isolates where a flagella type was available by traditional phenotypic typing, all results matched the H-type derived from the WGS data. Of the 140 sequenced non-O157 isolates, 52 (37.1%) harboured *stx1* only, 42 (30.0%) had *stx2* only, 46 (32.9%) carried *stx1* and *stx2*. Of these, *stx* subtyping PCR results were available for 131 isolates and 121 of these had concordant results with the *stx* subtype derived from the WGS data. Of the 10 discordant results, non-specific primer binding during PCR amplification, due to the similarity of the *stx2* subtype gene sequences was the most likely cause. The results of this study showed WGS provided a reliable and robust one-step process for characterization of STEC. Deriving the full serotype from WGS data in real time has enabled us to report a higher level of strain discrimination while *stx* subtyping provides data on the pathogenic potential of each isolate, enabling us to predict clinical outcome of each case and to monitor the emergence of hyper-virulent strains.

Keywords: whole genome sequencing, Shiga Toxin-producing *Escherichia coli*, serotyping, *stx* subtyping

INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) are considered to be a significant threat to public health due to the severity of gastrointestinal symptoms associated with human infection and the risk of cases developing Haemolytic Uraemic Syndrome (HUS; Byrne et al., 2015). STEC are zoonotic; transmission occurs by direct contact with animals or their environment, or by consumption of

contaminated food or water (Byrne et al., 2014). The infectious dose is low (<10 organisms) and person-to-person spread is common, particularly in nursery school settings and in households with young children (Byrne et al., 2015).

In England, the current Standards for Microbiology Investigations protocols are specific for the isolation of non-sorbitol fermenting colonies of *E. coli* serogroup O157 on cefixime tellurite sorbitol MacConkey (CT-SMAC) agar. STEC serogroups other than O157 (non-O157 STEC) are not detected using this method (Byrne et al., 2014). However, since 2012 the implementation of commercial PCR assays for the detection of STEC in faecal specimens from cases with symptoms of gastrointestinal infection, at a twelve local hospital laboratories, has resulted in an increase in the detection of non-O157 STEC (Byrne et al., 2014).

Faecal specimens that are PCR positive for the Shiga Toxin (stx) genes at the local hospital laboratories in England are sent to the Gastrointestinal Bacterial Reference Unit (GBRU) at Public Health England (PHE) for isolation of STEC (Jenkins et al., 2012) and subsequent serotyping (Gross and Rowe, 1985). Recent advances in whole genome sequencing (WGS) have led to the development of a method for high throughput sequencing of bacterial genomes at low cost (Joensen et al., 2014). During 2014, we evaluated the use of WGS for routine public health surveillance of non-O157 STEC by comparing this approach to phenotypic serotyping and PCR for subtyping the stx-encoding genes (Persson et al., 2007).

MATERIALS AND METHODS

All 167 strains of non-O157 STEC isolated during 2014 were phenotypically serotyped by the agglutination of antibodies raised in rabbits to the lipopolysaccharide O antigen and to the flagella H antigen (Gross and Rowe, 1985). Real-time PCR targeting *stx1* and *stx2* and the stx subtyping PCR was performed as previously described (Persson et al., 2007; Jenkins et al., 2012).

Genomic DNA extracted from 140 of the 167 strains of non-O157 STEC was fragmented and tagged for multiplexing with Nextera XT DNA Sample Preparation Kits (Illumina) and sequenced using the Illumina HiSeq 2500. A reference database, *SerotypeFinder*, containing the gene sequences encoding the 180 O antigen groups (*wzx*, *wzy*, *wzm*, and *wzt*) and the 53 H antigens (*fliC*, *flkA*, *flaA*, *flmA*, and *flnA*) was constructed and developed by Joensen et al. (2015). Using the *GeneFinder* tool (Doumith unpublished), FASTQ reads were mapped to the genes in the *SerotypeFinder* database using Bowtie 2 (Langmead and Salzberg, 2012) and the best match to each of the O and H determinants was reported with metrics including coverage, depth, mixture and homology in an XML format for quality assessment. Only *in silico* predictions of serotype that matched to a gene determinant at >80% nucleotide identity over >80% length were accepted. Stx subtyping was performed as described by Ashton et al. (2015). FASTQ sequences were deposited in the National Center for Biotechnology Information Short Read Archive under the bioproject PRJNA248064.

RESULTS

Whole genome sequences were available for 140 of the 167 non-STEC isolates reported in 2014 (Supplementary Table). Of these, 102 had a phenotypically derived serogroup, 25 did not agglutinate with the antisera in the serotyping scheme raised to the known *E. coli* serogroups and were designated “O unidentifiable,” and 13 did not express the O antigen and were designated “O rough.” Of the 102 isolates where phenotypic and genotypic serotyping could be compared, 98 gave fully concordant results (Supplementary Table). The most common non-O157 STEC serogroups detected were O146 (22) and O26 (18). There were 15 strains of STEC O55, all from cases linked to an outbreak in the South of England.

Of the four results that were not fully concordant, two isolates serogrouped as O186 phenotypically but were designated O123/O186 by *in silico* serotyping and one typed as O178 phenotypically and was designated O153/178 using WGS data. There was one mismatch; STEC O74 was identified as STEC O187 when the serotype was derived from the genome. The *in silico* serotyping method failed to type one isolate, STEC O146:H21, due to the short read sequences having low mapping coverage of the O antigen encoding genes. All but one of the 38 isolates that could not be phenotypically serotyped (designated O unidentifiable or O rough) were serotyped using the WGS data (Supplementary Table). The most common WGS derived serotypes that were untypable using the phenotypic approach were O91:H14, O117:H7, and O80:H2.

There were 102 isolates that were processed for H-typing, of which 29 were found to be non-motile and could not be typed. Of the 73 isolates where a flagella type was available by traditional phenotypic typing, all results matched the H-type derived from the WGS data. All the non-motile isolates were typable using the *in silico* serotyping by *in silico* serotyping (Supplementary Table).

Of the 140 sequenced non-O157 isolates, 52 (37.1%) harboured *stx1* only, 42 (30.0%) had *stx2* only, 46 (32.9%) carried *stx1* and *stx2*. Of these, stx subtyping PCR results were available for 131 isolates and 121 of these had concordant results with the stx subtype derived from the WGS data (Supplementary Table).

The most frequently detected stx subtype profile was stx1a only (38), most commonly associated with serotypes O103:H2 (14), O26:H11 (8), and O117:H7 (8) (Supplementary Table). There were 14 isolates with stx1c only, including O146:H21 (5) and O76:H19 (3). All stx1a and stx2b strains belonged to STEC O91:H14 (9). There were 23 isolates that had stx1c and stx2b associated with serogroups O146:H21 (9), O128:H2 (8), O113:H4 (2), and O174:H8 (2). The combination of stx1a and stx2a was found almost exclusively in O26:H11 (7), the one exception being an isolate of STEC O71:H2. Of the 21 isolates harbouring *stx2a* only, 15 were STEC O55:H7 and two were O26:H11. Of the 10 isolates that had *stx2d*, three belonged to O146:H21 (and also had *stx1c*), six were O80:H2 and one was O26:H11. Other rare subtypes (one case of each) included stx2e (O8:H9) and stx2g (O187:H28) (Supplementary Table).

There were 10 cases of HUS in 2014 (including five cases belonging to an outbreak of STEC O55:H7), eight had STEC harbouring *stx2a*, one had STEC O80:H2 carrying *stx2d*, and O103:H2 stx1a was isolated from the tenth case.

DISCUSSION

The results of this study showed WGS provided a reliable and robust one-step process for characterisation of STEC. Previous studies have shown an increasing number of strains of STEC reported as “O group unidentifiable” due to antisera failing quality control procedures, unresolvable cross reactions, lack of expression of O antigens (designated “rough”) or novel serogroups (Jenkins et al., 2003; Byrne et al., 2014). In this study, all but one of the isolates that were previously phenotypically untypable, were serotyped using data derived from the genome.

Of the 10 mismatched results identified in the comparison between the stx subtyping PCR (Scheutz et al., 2012) and the WGS approach (Ashton et al., 2015), all 10 had additional stx subtypes detected by PCR that were not identified in the WGS data (Supplementary Table). Non-specific primer binding during PCR amplification, due to the similarity of the stx2 subtype gene sequences was the most likely cause.

Historically, the most common stx profile of STEC O26 was stx1a but over the last 10 years a more virulent STEC O26 variant harbouring *stx2a* has emerged across Europe (Bielaszewska et al., 2013). The enhanced strain characterisation data that WGS provides facilitated the surveillance of emerging strains of STEC associated with more severe disease (for example STEC O55:H7 stx2a, STEC O26:H11 stx2a, STEC O80:H2 stx2d) and with novel stx profiles (for example STEC O26:H11 stx2d) and enabled us to compare data with colleagues in the field (Mariani-Kurkdjian et al., 2014; Delannoy et al., 2015). Previous studies have also reported an association between stx2a and severe disease (Ethelberg et al., 2004; Byrne et al., 2014). The *eae* gene was detected in 62 (44%) of the 140 non-O157 STEC isolates. All 10 isolates from the HUS cases had *eae*. None of the STEC strains in this data set had *aggR*, previously detected in highly pathogenic STEC variants (Boisen et al., 2015).

Prior to the implementation of WGS, due to limited resources and time constraints, H-typing and stx subtyping of STEC were not routinely reported by GBRU. Deriving the full serotype from

WGS data in real time has enabled us to report a higher level of strain discrimination while stx subtyping provides data on the pathogenic potential of each isolate, enabling us to predict clinical outcome of each case and to monitor the emergence of hyper-virulent strains.

AUTHOR CONTRIBUTIONS

MW and NP isolated the STEC and performed the real-time stx PCR. AG and SL performed the phenotypic serotyping, stx subtyping PCR and extracted the DNA. CJ, MC, MW, and NP implemented the wet lab WGS pipelines and performed analysis. TD and PA implemented the bioinformatics pipelines and performed analysis. CJ, MC, and TD wrote the manuscript.

FUNDING

This work was supported by the National Institute for Health Research and Health Protection Research Unit in Gastrointestinal Infections at the University of Liverpool.

ACKNOWLEDGMENTS

We would like to acknowledge Yoshini Taylor and Dawn Hedges at GBRU, Rediat Twolde in the Infectious Disease Informatics Unit and Cath Arnold in the Genomics Service Unit at PHE. We would very much like to thank Flemming Scheutz at the Statens Serum Institute in Copenhagen and Katrine Joensen at the Danish Technical University for sharing the *SerotypeFinder* database with us.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00258>

REFERENCES

- Ashton, P. M., Perry, N., Ellis, R., Petrovska, L., Wain, J., Grant, K. A., et al (2015). Insight into Shiga toxin genes encoded by *Escherichia coli* O157 from whole genome sequencing. *PeerJ*. 3:e739. doi: 10.7717/peerj.739
- Bielaszewska, M., Mellmann, A., Bletz, S., Zhang, W., Köck, R., Kossow, A., et al (2013). Enterohemorrhagic *Escherichia coli* O26:H11/H-: a new virulent clone emerges in Europe. *Clin. Infect. Dis.* 56, 1373–1381. doi: 10.1093/cid/cit055
- Boisen, N., Melton-Celsa, A. R., Scheutz, F., O'Brien, A. D., and Nataro, J. P. (2015). Shiga toxin 2a and enteroaggregative *Escherichia coli* - a deadly combination. *Gut. Microbes.* 6, 272–278. doi: 10.1080/19490976.2015.1054591
- Byrne, L., Jenkins, C., Launders, N., Elson, R., and Adak, G. K. (2015). The epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009–2012. *Epidemiol. Infect.* 143, 3475–3487. doi: 10.1017/S0950268815000746
- Byrne, L., Vanstone, G. L., Perry, N. T., Launders, N., Adak, G. K., Godbole, G., et al (2014). Epidemiology and microbiology of Shiga toxin-producing *Escherichia coli* other than serogroup O157 in England, 2009–2013. *J. Med. Microbiol.* 63, 1181–1188. doi: 10.1099/jmm.0.075895-0
- Delannoy, S., Mariani-Kurkdjian, P., Bonacorsi, S., Liguori, S., and Fach, P. (2015). Characteristics of emerging human-pathogenic *Escherichia coli* O26:H11 strains isolated in France between 2010 and 2013 and carrying the *stx2d* gene only. *J. Clin. Microbiol.* 53, 486–492. doi: 10.1128/JCM.02290-14
- Ethelberg, S., Olsen, K. E., Scheutz, F., Jensen, C., Schiellerup, P., Enberg, J., et al (2004). Virulence factors for hemolytic uremic syndrome, Denmark. *Emerg. Infect. Dis.* 10, 842–847. doi: 10.3201/eid1005.030576
- Gross, R. J., and Rowe, B. (1985). “Serotyping of *Escherichia coli*,” in *The Virulence of Escherichia coli: Reviews and Methods (Special Publication of the Society for General Microbiology no. 13)*, ed M. Sussman (London: Academic Press), 345–360.
- Jenkins, C., Lawson, A. J., Cheasty, T., and Willshaw, G. A. (2012). Assessment of a real-time PCR for the detection and characterization of verocytotoxigenic *Escherichia coli*. *J. Med. Microbiol.* 61, 1082–1085. doi: 10.1099/jmm.0.041517-0
- Jenkins, C., Willshaw, G. A., Evans, J., Cheasty, T., Chart, H., Shaw, D. J., et al (2003). Subtyping of virulence genes in verocytotoxin-producing *Escherichia coli* other than serogroup O157 associated with disease in the United Kingdom. *J. Med. Microbiol.* 52, 941–947. doi: 10.1099/jmm.0.05160-0

- Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13
- Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and easy *in silico* serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Mariani-Kurkdjian, P., Lemaître, C., Bidet, P., Perez, D., Boggini, L., Kwon, T., et al (2014). Haemolytic-uraemic syndrome with bacteraemia caused by a new hybrid *Escherichia coli* pathotype. *New Microbes New Infect.* 2, 127–131. doi: 10.1002/nmi2.49
- Persson, S., Olsen, K. E., Ethelberg, S., and Scheutz, F. (2007). Subtyping method for *Escherichia coli* shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *J. Clin. Microbiol.* 45, 2020–2024. doi: 10.1128/JCM.02591-06
- Scheutz, F., Teel, L. D., Beutin, L., Piérard, D., Buvens, G., Karch, H., et al (2012). Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J. Clin. Microbiol.* 50, 2951–2963. doi: 10.1128/JCM.00860-12

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Chattaway, Dallman, Gentle, Wright, Long, Ashton, Perry and Jenkins. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.