**Use of whole-genome sequencing for the public health surveillance of *Shigella sonnei* in England and Wales, 2015**

*Shigella* spp., including *Shigella dysenteriae*, *Shigella boydii*, *Shigella flexneri* and *Shigella sonnei*, are the most common cause of bacterial dysentery (bloody diarrhoea) worldwide (Kotloff *et al.*, 1999). Although all species of *Shigella* contribute to the high burden of diarrhoeal disease in low-income regions, *S. sonnei* is the most commonly reported species in middle- and high-income countries (Thompson *et al.*, 2015). In England and Wales, foodborne outbreaks of *S. sonnei* are rare with transmission most commonly associated with person-to-person spread (McDonnell *et al.*, 2013; Morgan *et al.*, 2006; Simms *et al.*, 2015). Historically, schools and nurseries were regarded as the epidemic centres of domestically acquired *S. sonnei* infection (Evans & Maguire, 1996). More recently, outbreaks of *S. sonnei* amongst men who have sex with men (MSM) have been described, and the increasing incidence of *S. sonnei* infection in this community is a challenging public health problem (Morgan *et al.*, 2006; Simms *et al.*, 2015).

Clinically, symptoms of *S. sonnei* are generally regarded as less severe than *S. flexneri*, although more severe clinical manifestation has been documented amongst HIV-positive individuals (Daskalakis & Blaser, 2007). Consequently, Public Health England (PHE) health protection teams do not routinely follow up cases of *S. sonnei* with an exposure history questionnaire in the same way that they do for *S. dysenteriae*, *S. boydii* and *S. flexneri*. However, *S. sonnei* is a significant public health problem in England and Wales. The burden of gastrointestinal (GI) infection caused by *S. sonnei* is higher than the other three species. Between 2007 and 2012, *S. sonnei* was the most commonly isolated species (49.9 %), whereas *S. flexneri*, *S. boydii* and *S. dysenteriae* represented 37.7 %, 7.6 % and 4.34 % of the total number of strains, respectively. *S. sonnei* is more commonly associated with domestically acquired infections and is the most common cause of outbreaks of shigellosis in England and Wales (https://www.gov.uk/government/statistics/shigella-cases-1992-to-2013). Furthermore, there is evidence that emerging public health threats, such as novel transmission routes and increasing antimicrobial resistance, are occurring under the surveillance radar (Hoffmann *et al.*, 2013; Bowen *et al.*, 2015; Simms *et al.*, 2015).

*S. sonnei* has a single O antigen and does not express an H antigen, so it cannot be serotyped. At PHE, phage typing was the typing method of choice since the 1990s (Bentley *et al.*, 1996) but exhibited a low level of strain discrimination, with over 80 % of the isolates submitted to the Gastrointestinal Bacteria Reference Unit (GBRU) between 2007 and 2012 belonging to phage type (PT) 6 or PT2, with PT6 being the most common. Historically, the lack of a robust typing method meant that outbreaks were difficult to detect and investigate, and emerging public health issues were challenging to manage. Recently, whole-genome sequencing (WGS) has been implemented at PHE for the surveillance of bacterial GI pathogens (McDonnell *et al.*, 2013; Dallman *et al.*, 2015; Ashton *et al.*, 2016). We present findings of a preliminary review of the use of WGS for the detection of outbreaks of *S. sonnei* and assess the impact this technology on the public health surveillance of GI disease caused by this important pathogen.

Bacterial isolates of *S. sonnei* from 349 cases were submitted to the GBRU between August 2015 and January 2016, from local and regional hospital laboratories in England and Wales. Epidemiological data on age, sex and region of residence were available from laboratory report forms and recorded in a standardized format in an in-house database, the GastroData Warehouse. Travel history was available for 221 (63.3 %) of cases. All isolates were phage typed using the method described by Bentley *et al.* (1996) and whole genome sequenced.

For WGS, DNA was extracted from cultures of *S. sonnei* for sequencing on the Illumina HiSeq 2500 instrument. High-quality Illumina reads were mapped to the *S. sonnei* reference genome SS46 (Genbank accession: NC_007384.1) using BWA-MEM (Li & Durbin, 2010). Single nucleotide polymorphisms (SNP) were identified using GATK2 (McKenna *et al.*, 2010) in unified genotyper mode. Core genome positions that had a high-quality SNP (>90 % consensus, minimum depth 10×, GQ ≥ 30) in at least one isolate were extracted, and RaxML (Stamatakis, 2014) was used to derive the maximum likelihood phylogeny of the isolates. Hierarchical single linkage clustering was performed on the pairwise SNP difference between all isolates at various distance thresholds (Δ250, Δ100, Δ50, Δ25, Δ10, Δ5 and Δ0). The result of the clustering is an SNP address that can be used to describe the population structure based on clonal groups (Fig. S1, available in the online Supplementary Material). Isolates share the same digit at an SNP address level if their SNP difference is less than the threshold to at least one other isolate in the cluster. If two isolates have matching numbers in each position in the profile, they are said to be matched to the 0 SNP level in comparison to the reference genome; this indicates that they are genetically related to each other. In this study, isolates were organized into clusters on the basis of belonging to the same five-SNP cluster (Fig. S1). FASTQ reads from all sequences in this study can be found at the PHE Pathogens BioProject at the National Center for Biotechnology Information (PRJNA315192).
Of the 349 cases included in this study, 199 (57.0 %) were male, and 150 (43.0 %) were female. A total of 138 cases (39.5 %) reported recent travel abroad prior to onset of symptoms; 83 (23.8 %) specifically reported that they had not travelled, and no travel data were available for 128 cases (36.7 %). The isolates comprised 16 different PTs, the most common being PT6 (71.7 %).

Analysis of the WGS data organized 231 of the 349 isolates into 53 five-SNP clusters, where members of the cluster shared at least one link within the five-SNP cluster. The median number of cases in these clusters was 4.4 and ranged from 2 to 24; 46/53 (86.8 %) clusters comprised less than six cases. It was difficult to identify epidemiological links associated with these small clusters using only the limited epidemiological data available from laboratory report forms, but 29/46 (63.0 %) comprised at least one case reporting recent travel abroad prior to onset of symptoms. Although demographic data associated with each patient were limited, age, sex and travel histories were analysed for the seven clusters comprising six or more cases (Table S1 and highlighted in the Fig. S1). Clusters 4 and 6 appeared to be travel related with five of 12 cases reporting travel to Cape Verde in cluster 4 and four of seven cases reporting travel to Morocco in cluster 6 (Table S1). No epidemiological links were evident from the data available for cluster 5, although none of the cases reported travel.

The remaining four clusters had a high male-to-female ratio with the majority of cases being adult males (cluster 1, 91.7 %, n=22/24; cluster 2, 93.3 %, n=14/15; cluster 3, 100 %, n=10/10; cluster 7, 100 %, n=6/6), and reports of recent travel were restricted to a small number of cases (Table S1). This demographic has been previously shown to be characteristic of clusters linked to sexual transmission amongst the MSM community (Borg et al., 2012; Gilbart et al., 2015; Simms et al., 2015; Baker et al., 2015). Isolates in clusters 1, 2 and 3 exhibited macrolide resistance, those in cluster 3 were also resistant to ceftriaxone and produced the extended-spectrum β-lactamase, CTX-M 27, and isolates in cluster 7 were resistant to ciprofloxacin harbouring mutations in gyrA [83:S-L; 87:D-G] and parC [80:S-I] genes. This preliminary evaluation indicates that WGS provides highly discriminatory molecular typing suitable for detecting and investigating outbreaks of S. sonnei. Evidence from clusters linked to transmission in MSM and travel abroad suggested that isolates of S. sonnei that fell within a five-SNP cluster may share a common transmission route and that epidemiological links – if not immediately apparent – should be sought. The lack of available routine epidemiological data and the fact that, even if exposure history questionnaires were performed, field data were not collated in a standardized format hindered this analysis, and a more systematic approach to epidemiological data collection is required. In the absence of routine detailed epidemiological data collection, WGS combined with basic demographic information will help prioritize cases requiring more detailed and more targeted epidemiological investigation; e.g., the identification of domestically acquired male-dominated clusters would trigger the collection of more detailed sexual history risk factor information.

Of the seven clusters comprising six or more cases, none were associated with outbreaks in schools and nurseries, whereas four were likely to be associated with MSM transmission. This analysis provided further evidence that the common transmission routes driving the dissemination of S. sonnei in England and Wales are changing and that outbreaks in MSM are increasingly associated with antimicrobial resistance. The detailed and comprehensive level of surveillance data that WGS provides will make a positive contribution to clinical guidelines on antibiotic treatment and inform public health interventions.

Prior to the implementation of WGS at PHE, microbiological input into public health surveillance of S. sonnei was limited owing to the lack of a robust typing method. Although rapid and relatively inexpensive, phage typing for S. sonnei provided low-level discrimination (71.7 % of isolates in this study were PT6) and was a poor predictor of strain relatedness. Five of the seven largest five-SNP clusters comprised more than one PT (Table S1). The implementation of WGS delivers a state-of-the-art microbial typing scheme that facilitates the monitoring of novel transmission routes (Ratnakayate et al., 2012; Toro et al., 2015) and emerging antimicrobial resistance (Hoffmann et al., 2013; Bowen et al., 2015), ultimately enabling us to deliver timely and appropriate interventions to reduce the burden of GI disease in England and Wales.

Acknowledgements
We would like to thank Michael Wright, Neil Perry, Vivienne do Nascimento, Amy Gentle and Dawn Hedges in GBRU, and Amy Mikhal in the Gastrointestinal Infections Department at PHE. This work was supported by the National Institutes for Health Research Health Protection Research Unit in Gastrointestinal Infections. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or PHE.

Timothy J. Dallman,1 Marie A. Chattaway,1 Piers Mook,2 Gauri Godbole,1 Paul D. Crook2 and Claire Jenkins1

1Gastrointestinal Bacteria Reference Unit, Public Health England, London, UK

Correspondence: Claire Jenkins Claire.Jenkins1@phe.gov.uk

Abbreviations: GBRU, Gastrointestinal Bacteria Reference Unit; GI, gastrointestinal; MSM, men who have sex with men; PHE, Public Health England; PT, phage type; SNP, single nucleotide polymorphism; WGS, whole-genome sequencing.

References


