The importance of a multifactorial approach for (inter)national surveillance of *Shigella* spp. and entero-invasive *Escherichia coli*

Maaïke J. C. van den Beld (✉ maaïke.van.den.beld@rivm.nl)

National Institute for Public Health and the Environment - RIVM <https://orcid.org/0000-0001-8720-8434>

Frans A.G. Reubsaet

National Institute for Public Health and the Environment - RIVM

Roan Pijnacker

National Institute for Public Health and the Environment - RIVM

Airien Harpal

National Institute for Public Health and the Environment - RIVM

Sjoerd Kuiling

National Institute for Public Health and the Environment - RIVM

Evy M. Heerkens

National Institute for Public Health and the Environment - RIVM

B. J. A. (Dienke) Hoeve-Bakker

National Institute for Public Health and the Environment - RIVM

Ramón C.E.A. Noomen

National Institute for Public Health and the Environment - RIVM

Amber C. A Hendriks

National Institute for Public Health and the Environment - RIVM

Dyogo Borst

National Institute for Public Health and the Environment - RIVM

Han van der Heide

National Institute for Public Health and the Environment - RIVM

A.M.D. (Mirjam) Kooistra-Smid

Certe and University of Groningen, University Medical Center Groningen

John W.A. Rossen

University of Groningen, University Medical Center Groningen

Research article

Keywords: *Shigella*, entero-invasive *E. coli*, EIEC, surveillance, genetic characterization, phenotypic characterization, antimicrobial resistance, virulence, cgMLST, genomic epidemiology

Posted Date: August 21st, 2019

DOI: <https://doi.org/10.21203/rs.2.13263/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: *Shigella* spp. and entero-invasive *Escherichia coli* (EIEC) can cause mild diarrhea to dysentery. In the Netherlands, although shigellosis is a notifiable disease, there is no laboratory surveillance for *Shigella* spp. and EIEC in place. Consequently, the population structure for circulating *Shigella* spp. and EIEC isolates is not known. This study describes the phenotypic and serological characteristics, the phenotypic and genetic antimicrobial resistance profiles, the virulence gene profiles, the classic multi-locus sequence types (MLST) and core genome MLST (cgMLST) types, and the epidemiology of *Shigella* spp. and EIEC isolates collected during a cross-sectional study in the Netherlands in 2016 and 2017. Results: *S. sonnei*, *S. flexneri* and EIEC were predominantly detected in the Netherlands. A substantial part of the characterized isolates was resistant to antimicrobials advised for treatment, i.e., 73% was phenotypically resistant to co-trimoxazol and 19% to ciprofloxacin. Antimicrobial resistance was particularly observed in isolates from male patients who had sex with men or from patients that had travelled to Asia. Furthermore, isolates related to international clusters were also circulating in the Netherlands. Travel-related isolates formed clusters with isolates from patients without travel history, indicating their emergence into the Dutch population. Conclusions: In conclusion, laboratory surveillance using whole genome sequencing for genetic characterization of isolates complements the current epidemiological surveillance, as the latter is not sufficient to detect all (inter)national clusters, emphasizing the importance of multifactorial public health approaches.

Background

MGE) that can be horizontally transferred, including plasmids such as spA or pCERC1, and chromosomal integrons such as the SRL-MDRE island and In2 and the transposon tn7 (5, 15, 16). It was demonstrated that MSM lineages of *S. sonnei* and *S. flexneri* are associated with the presence of the pKSR100 plasmid that contains genes involved in beta-lactam and azithromycin resistance (13). Additionally, vertically transferred chromosomal point mutations mainly conferring resistance to quinolones can be present (8, 17). Multiple studies indicated that the presence of resistance genes or point mutations in whole genome sequences of *E. coli* and *S. sonnei*, accurately predicts phenotypic AMR (18-21).

Enteroinvasive *Escherichia coli* (EIEC) is a pathotype of *E. coli* with similar pathogenicity as *Shigella* spp., and they are genetically similar (2, 22). They can only be distinguished by combining a large number of classical phenotypic tests with classical O-serotyping or in silico analyses of O-antigen genes. However, none of those methods is able to distinguish all isolates accurately (23, 24).

In the Netherlands, as in many other countries, infections with *Shigella* spp. are notifiable by law, while infections with EIEC are not. Epidemiological surveillance of individual shigellosis patients is in place as regulation for control of shigellosis, and source tracing is performed in all cases. However, there is no active laboratory surveillance in place; consequently, the population structure for *Shigella* spp. and EIEC isolates circulating in the Netherlands is not known.

During 2016 and 2017, a cross-sectional study was conducted in the Netherlands with the aim to assess incidence, population structure, disease outcomes and impact on public health of *Shigella* spp. and EIEC. During the study period, 15 participating medical microbiological laboratories sent all their *Shigella* spp. and EIEC isolates to the study group. All isolates were thoroughly characterized, both phenotypically and genotypically, in conjunction with epidemiological data of the patients that were infected. This is a report of the results of the phenotypic and genetic characterization of the isolates.

Methods

Isolates and phenotypic characterization

A total of 414 EIEC and *Shigella* spp. isolates were collected by 15 laboratories in the Netherlands, participating in the cross-sectional Invasive Bacteria *E. coli-Shigella* study (IBESS) performed in 2016-2017 (van den Beld et al., manuscript submitted). All isolates were thoroughly characterized, both phenotypically and genotypically. Identification and *Shigella* and *E. coli* O-serotyping of isolates was performed with phenotypic characterization using classical methods as described before (23). Isolates were called provisional *Shigella* when the species and serotype could not be determined due to auto-agglutination or inconclusive combinations of antisera; furthermore, isolates were called provisional *Shigella* when a serotype could be assigned, but the results of the phenotypical tests deviated from those of the serotype-specific tests. Overall, phenotypic properties of *S. flexneri*, *S. sonnei* and EIEC were compared. In addition, patients were contacted by infectious disease nurses from the public health services Groningen and Amsterdam to collect information on demographics, travel history, sexual behavior, and indicators for high-risk sexual behavior as HIV status, presence of other STI and the use of PrEP using a standardized survey by telephone.

Sequencing and data preparation

Based on the species designations and availability of patient data, 348 of 414 isolates were selected for whole genome sequencing (WGS) using Illumina® technology as described previously (23). Resulting raw reads were processed with an in-house assembly pipeline

(<https://github.com/Papos92>), consisting of quality assessment using FastQC v. 0.11.8 (38) and MultiQC v. 1.7 (39), read trimming using ERNE v. 2.1.1 (40), contamination filtering using CLARK v. 1.2.5.1 (41), assembly using SPAdes v. 3.10.0 (42), and assembly quality assessment using QUASTv. 4.4 (43). Completeness and contamination of assemblies was checked using CheckM v. 1.0.11 (44) (taxonomy_wf: genus '*Shigella*'), draft genomes with good quality, and a completeness higher than 99% and contamination lower than 2% were used in further analysis. All sequences were submitted to the Sequence Read Archive (SRA) under study number PRJEB32617.

Antimicrobial resistance

Phenotypic AMR profiling was performed by participating laboratories of the IBESS study using their own diagnostic protocols. *In silico* resistance profiling was performed to assess the presence of ARGs and chromosomal point mutations. For this purpose, the ResFinder and PointFinder databases and scripts were obtained from the Center for Genomic Epidemiology (CGE) repositories at Bitbucket (<https://bitbucket.org/genomicepidemiology/resfinder/src/master/>). These scripts were integrated in a local pipeline script for batch execution, and were executed using the default analysis settings and the applicable databases. Logistic regression models were used to associate the presence of ARGs with phenotypic resistance. Intermediate phenotypes were not considered. Associations were expressed as odds ratios (OR) with corresponding 95% confidence intervals (CI). Analyses were performed using SPSS version 24.0.0.1.

MLST and cgMLST analysis

A classical MLST and a cgMLST analysis were performed with Ridom SeqSphere⁺, version 3.5.1 (Ridom® GmbH, Münster, Germany). The *E. coli* Warwick MLST scheme, curated by MLST databases of the University of Warwick (45) and the *E. coli* cgMLST genotyping scheme based on the Enterobase *Escherichia/Shigella* cgMLST v1 scheme were used. For context, reference isolates representing *S. sonnei* lineages I, II, III, IV, V, and the subclades of lineage III; IIIa, global III, orthodox Jewish communities associated (OJCA) III, Central Asia associated III, and MSM clades 1 to 4 were added to the cgMLST (13, 15, 17, 25). For *S. flexneri*, isolates were included that represent phylogenetic groups PG1 to PG7, including the PG3 major and minor MSM subclade (13) and *S. flexneri* 3a MSM sublineages A, B, C and Asia and Africa associated sublineages (5). For EIEC, reference isolates representing 3 different STs and 9 serotypes were included (46). Details about used reference genomes were summarized in Additional File 1. Trees were inferred based on cgMLST in Ridom SeqSphere⁺, and visualized using iTOL v4.3.2 (47).

Virulence profiling

For assessment of virulence genes, the VirulenceFinder database for *E. coli* virulence genes was used from the Center for Genomic Epidemiology (CGE) (48). For *Shigella* virulence, genes present in the SHI-1, SHI-2 pathogenicity islands as well as the genes responsible for the T3SS machinery and effectors were used as reference (Additional File 2). Reference genes were indexed based on gene name and accession code obtained from the National Center for Biotechnology Information (NCBI), to make a nucleotide comparison in a local alignment. Both indexing of the reference genes and alignment with the isolates were facilitated by the command line BLAST application, used with default settings and identity cut-offs of 70% (49).

Results

Phenotypic characterization

414 isolates were collected during a two-year period from 411 patients, as three patients had a double-infection with EIEC and *S. flexneri* or *S. sonnei*. The total number of isolated *Shigella* spp. and EIEC in 2016 and 2017 was 204 and 210, respectively. The species distribution in 2016 and 2017 was comparable (χ^2 test, $p = 0.69$). In total, 232 were *S. sonnei*, 104 *S. flexneri*, 64 EIEC, 10 provisional *Shigellae*, 3 *S. boydii* and one isolate was either EIEC or *S. flexneri*, the distinction could not be made (Table 1). No *S. dysenteriae* was identified.

For *S. flexneri*, serotype 2a was mostly identified (51%), followed by serotype 6 (12%), 1c (7%), 3a (7%), 1b (5%), 4av (3%), Xv (3%), Y (3%), 3b (2%), Yv (2%) and 1a (1%). For 6% of *S. flexneri* isolates, the serotype could not be determined due to undescribed combinations of reactions with antisera.

Of the 64 EIEC isolates, 24 (38%) were negative for *E. coli* O1 - O188 antisera. The other 40 isolates were distributed over 16 different O-types, of which 32 (50%) EIEC isolates had O-types that were described as EIEC-associated before (O42, O96, O121, O124, O135, O136, O143, O159 and O164). Additionally, 8 (13%) of EIEC isolates had O-types that were not described as EIEC-associated before (O8, O10, O17, O48, O73, O109 and O141). Results from phenotypic tests for *S. flexneri*, *S. sonnei* and EIEC are summarized in Table 2.

Antimicrobial resistance

A total of 180 out of 248 *Shigella* spp. and EIEC isolates (73%) had phenotypic resistance against co-trimoxazol, 49 out of 264 (19%) had resistance against ciprofloxacin, and 34 (14%) were resistant to both. *In silico* determination of azithromycin resistance genes *erm(B)* and *mphA* was performed, in 30 (9%) out of all 348 genomes *erm(B)* was detected, in 37 (11%) *mphA*, and in 29 (8%) both genes were detected. The detected antimicrobial resistance genes (ARG) and their association with phenotypic resistance are shown in Table 3. Presence of blaTEM-1b as well as the presence ≥ 1 *bla* genes were significantly associated with phenotypic resistance against ampicillin. Furthermore, blaTEM-1b, blaOXA-1 and the presence of ≥ 1 *bla* genes were significantly associated with phenotypic resistance against amoxicillin/clavulanic acid (Table 3). Only one of the isolates phenotypically tested resistant to piperacillin/tazobactam, but no *bla* genes were detected in this isolate. Of the isolates that were phenotypically resistant to 3rd generation cephalosporins, cefotaxime and ceftazidime, respectively 100% and 86% contained one of the *bla*-CTX-M genes or the *bla*DHA-1 gene (Table 3). Phenotypic resistance to aminoglycosides gentamicin and tobramycin was not associated with the presence of *aac(3)-IId* or *aph(3)-Ia* genes. Other ARGs that confer resistance to gentamicin or tobramycin were not detected. Phenotypic resistance to ciprofloxacin was significantly associated with three chromosomal point mutations that are known to confer resistance (17, 21). Two were present in the *gyrA* gene, one mutation on position 83 encoded a leucine instead of a serine and one on position 87 that altered aspartic acid to glycine. The other chromosomal point mutation that was significantly associated with phenotypic ciprofloxacin resistance was found in the *parC* gene, a single mutation at position 80 replaced serine with isoleucine. All but one isolate that displayed resistance to ciprofloxacin possessed two or more chromosomal point mutations, while the presence of plasmid-mediated *qnr* genes or the presence of one chromosomal point mutation was not associated with the resistant phenotype. Phenotypic resistance to trimethoprim perfectly correlated with the presence of one or more *dhfrA* genes. All but one isolate that were phenotypically resistant to co-trimoxazole had one or more *dhfrA* genes, and the presence of one or more *dhfrA* genes and one or more *sul* genes was also significantly associated with co-trimoxazole resistance. None of the ARGs were exclusively found in restricted time periods.

MLST and cgMLST analysis

With classical MLST typing, most *S. sonnei* isolates (96%) were ST152, most *S. flexneri* serotype 1 to 5 isolates (91%) were ST245, and all *S. flexneri* serotype 6 isolates were ST145. In contrast, STs of EIEC isolates were diverse and distributed over 18 known STs, and 5 unknown STs, the latter all consisting of different allele combinations. Of the 18 known STs, 12 were assigned to single EIEC isolates, while ST6 comprises 13 EIEC isolates (21%), ST99 9 isolates (15%), ST4267 8 EIEC isolates (13%), ST245 and ST270 6 (10%) EIEC isolates each, and ST311 3 isolates (5%).

In the cgMLST tree including all isolates, most of the genomes clustered according to their species, although also clusters with mixed species were formed due to deviating phenotypic features or inconclusive serotypes (Figure 1). Three separate cgMLST trees were created for *S. flexneri*, *S. sonnei* and EIEC including context isolates. From 291 of the 348 (84%) sequenced genomes, data about patient demographics, travel history, sexual behavior, and indicators for high-risk sexual behavior as HIV status, presence of other sexually transmitted infections (STIs) and the use of pre-exposure prophylaxis (PrEP) was collected and depicted in the cgMLST trees (Figure 2 to 4).

Genomic epidemiology

S. flexneri isolates clustered predominantly according to their serotype, based on cgMLST. Five clusters were associated with MSM, of which two were MSM clusters described in previous publications, i.e., 3a MSM sublineage A and PG3 major MSM subclade (5, 16)(Figure 2). The other three clusters were labeled flexneri-MSM-1, flexneri-MSM-2 and flexneri-MSM-3. Clusters flexneri-MSM-1 and flexneri-MSM-2 consisted of only MSM, while in cluster 3a MSM sublineage A and PG3 major MSM subclade, 67% and 86% of patients, respectively, reported MSM contact. Other isolates in the last-mentioned clusters were from men that reported not to have MSM contact. Flexneri-MSM-3 was a mixed cluster, consisting half of patients with MSM contact, and the other half were women or men that reported not to have MSM contact (Figure 2). 79% of all isolates in the *S. flexneri* MSM clusters were diagnosed with shigellosis in the Amsterdam region, while the remaining 21% was diagnosed in different regions from the Netherlands. Clusters PG3 major MSM subclade and flexneri-MSM-2 contained both isolates from the Amsterdam region only. Clusters flexneri-MSM-2 and flexneri-MSM-3 were both distantly related to the reference PG3 minor MSM subclade, while flexneri-MSM-1 was not related to any of the MSM reference isolates (Figure 2). PrEP use was only reported by patients infected with isolates in the MSM clusters, and the isolates of only 2 out of 19 patients that reported an HIV infection were situated outside these clusters. The percentage of HIV infections or PrEP use ranged from 43% in the PG3 major MSM subclade cluster to 100% in the 3a MSM sublineage A cluster. All patients with isolates within the MSM clusters had no travel history or they had traveled within Europe (Figure 2). Furthermore, most patients (80%) with *S. flexneri* serotype 6 reported travel to Africa. Three other small clusters were travel-related; a cluster of 2 *S. flexneri* 4av isolates linked to Africa, one cluster of 2 *S. flexneri* 1b isolates linked to Central America and one cluster containing *S. flexneri* Xv and a provisional *Shigella* was related to travel to South America (Figure 2). None of the isolates in our study were closely related to the travel-related references from 3a Africa and 3a Asia sublineages. All MSM or travel-related clusters contained isolates from both 2016 and 2017, indicating that these clusters were not restricted to a specific time period. All isolates in the flexneri-MSM-3 cluster were resistant to ciprofloxacin, and additionally, other isolates in a cluster related to flexneri-MSM-3 were also ciprofloxacin resistant. Two of those isolates were from patients that reported travel to Asia and other isolates were from patients that reported no travel. Both azithromycin resistance genes were present in nine isolates and were only observed in clusters 3a MSM

sublineage A, flexneri-MSM-1 and PG3 major MSM subclades. Seven of these isolates also displayed the *bla-TEM1b* gene, indicating the presence of the MSM-associated pKR S100 plasmid (Figure 2).

In the *S. sonnei* cgMSLT, 3 MSM clusters were found, including isolates related to the earlier described lineage III MSM clade 2 and lineage III MSM clade 4 (13) (Figure 3). An additional MSM-associated cluster was identified that did not relate to any of the reference isolates and was labeled sonnei-MSM-1. The cluster associated with lineage III MSM clade 4 consisted only of MSM, while for isolates related to lineage III MSM clade 2 and sonnei-MSM-1 these percentages were 89% and 80%, respectively. 78% of all isolates in the *S. sonnei* MSM clusters were diagnosed with shigellosis in the Amsterdam region, while the remaining 22% was diagnosed in different regions from the Netherlands. Cluster lineage III MSM clade 4 contained only isolates from the Amsterdam region. Patients that reported to have HIV, other STIs, or using PrEP, were exclusively MSM. In the lineages III MSM clade 2 and MSM clade 4, 50% of patients had HIV or another STI. In cluster sonnei-MSM-1 this percentage was 30%. All patients within the MSM clusters reported no travel history or they had traveled within Europe (Figure 3). Three isolates that were distantly related to lineage IIIa reported travel to South America, the region to which lineage IIIa was associated (25). Four small clusters were related to travel to Central America (n= 4 to 8), four other small clusters (n= 2 to 9) and one large cluster (n=22) were related to travel to Asia, and five clusters were related to travel to Africa (n=4, 8, 13, 14, 33). Furthermore, none of the MSM clusters contained isolates from a restricted time period, while isolates in two out of four non-MSM clusters that were travel-related to Central America were from February to August 2016 and June to October 2016, respectively. Ciprofloxacin resistance was mainly observed in the Asian cluster and sonnei-MSM-1 cluster (Figure 3). Both azithromycin resistance genes *erm(B)* and *mph(A)* were present in eleven isolates, all additionally had the *bla-TEM1b* gene, and all were present in the MSM-associated clusters (Figure 3).

For EIEC, isolates clustered according to their O-types, although there were two clusters that had O135 interspersed by EIEC with O-types O8 and O48 (Figure 4). Additionally, isolates with ST270 and O-type O164 clustered with reference EIEC ST270/O124. MSM associated clusters were not identified and only one patient of whom an EIEC isolate was obtained reported MSM contact. Furthermore, only one of the 32 EIEC-infected patients from whom epidemiological data was collected, reported an HIV infection that was not associated with MSM (Figure 4). Two clusters of EIEC isolates were related to travel to Asia (n=3, 5), one larger cluster (n = 9) was related to travel to Africa and one smaller cluster was related to South America (n=4). The latter only contained isolates cultured from February to May 2016. Although other isolates were also travel-related, no other distinct clusters were found. Phenotypical antimicrobial resistance showed no specific cluster-related pattern. Overall, EIEC isolates were less resistant than *S. flexneri* or *S. sonnei* isolates (Figure 4).

Virulence profiling

In our study, none of the sequenced *Shigella* or EIEC isolates contained genes that encode the Shiga-toxin. *E. coli* virulence genes that were associated with the invasive phenotype that is displayed by *Shigella spp.* and EIEC were present.

For *S. flexneri*, all but one isolate were in possession of the *set* gene located on the SHI-1 island. The *pic* gene was only present in *S. flexneri* 2a or Y, and the *sigA* gene was present in *S. flexneri* serotype 2a, Y, and 6 and with a lower identity percentage in *S. flexneri* serotype 3a and 3b (Figure 2). All isolates that possessed all genes present in the SHI-1 island were from PG3 (Figure 2). Isolates in the 3a MSM sublineage A cluster and *S. flexneri* serotype 6 possessed none of the *shi* genes in SHI-2. Three *S. flexneri* isolates lacked all genes encoding for the T3SS machinery and effectors (Figure 2). One isolate was in possession of the *Osp* genes, but lacked the *mxl-spa* operon, the *ipa-igp* operon and the *virA* gene.

Almost all *S. sonnei* isolates were in possession of the *sigA* and *pic* genes from the SHI-1 island, while the *set* gene was present in approximately half of the isolates (Figure 3). All isolates had all genes present in the SHI-2 pathogenicity island, except for the *shiD* gene, which was present in only two isolates that clustered apart from other isolates in lineage III. More than half of the *S. sonnei* isolates were not in possession of the genes encoding for the T3SS machinery and effectors (Figure 3).

In the analysis of virulence genes of the EIEC isolates, 54 isolates (84%) contained the *set* gene located on the SHI-1 island, all in combination with the *sen (ospD3)* gene encoded on the pINV plasmid (Figure 4). Ten EIEC isolates (16%) harbored no genes encoding for the T3SS machinery or effectors, of which three isolates also contained none of the genes present in the SHI-1 island (Figure 4). The other seven isolates contained the *sigA*, and/or the *pic* genes. The lineage that comprises isolates with ST6 and the lineage that comprises the ST99/O96 and ST4267 isolates did not contain SHI-2 or only a smaller number of genes present in this island. Only 11 EIEC isolates (17%) contained the *shiA* gene on this island, and none contained the *shiE* gene (Figure 4).

Discussion

In the Netherlands, although shigellosis cases are notifiable, there is no active laboratory surveillance of characteristics of *Shigella* and EIEC isolates. Consequently, there is a gap of knowledge about circulating *Shigella* spp. and EIEC isolates and their characteristics and population structure. In our study, circulating *Shigella* spp. and EIEC isolates in the Netherlands during 2016 and 2017 were fully characterized.

Phenotypic characterization

During 2016 and 2017, *S. sonnei* (56%) was the most prevalent species in the Netherlands, followed by *S. flexneri* (24%) and EIEC (15%). Phenotypic properties of the pathotype EIEC were described based on 64 isolates in this study. If EIEC isolates display one of the phenotypic properties that are by definition negative for *Shigella* spp., distinction is uncomplicated. In contrast, when EIEC isolates display the more inactive *Shigella* phenotype, distinction is challenging (26). This challenging identification and distinction of *Shigella* spp. and EIEC was confirmed, because even with the thorough phenotyping and serotyping that was performed, one isolate could not be assigned to the genus *Shigella* or *Escherichia* and ten *Shigella* isolates could not be assigned to a species and were called provisional. Moreover, in the cgMLST tree combining all species, clusters with multiple species were formed, confirming the close genetic relationship among the species of *Shigella* and EIEC that was described before in multiple studies (2, 22, 27, 28).

MLST and cgMLST analysis and genomic epidemiology

The diverse *E. coli* O-types and Warwick MLST types for EIEC isolates showed a large diversity of isolates circulating in the Netherlands. In the cgMLST, EIEC isolates also showed more diversity than *S. flexneri* or *S. sonnei*. This diversity of EIEC isolates was described before, for isolates that were circulating in the United States (22).

In the cgMLST, *S. flexneri* and EIEC isolates clustered mostly according to their serotype. Exceptions within *S. flexneri* are two *S. flexneri* Yv isolates that formed a separate cluster and one *S. flexneri* 2a isolate that deviated from all *S. flexneri* isolates in our study. The separate clustering of the two Yv isolates is probably due to the fact that they relate to the different phylogroups PG1 and PG6 as shown in the cgMLST tree. It was described earlier that serotypes can belong to multiple PGs, although the association of *S. flexneri* Yv with PG1 was not found before (29). However, if one takes into account that *S. flexneri* is able to switch their serotype quite easily due to the exchange of O-antigen genes via horizontal gene transfer (HGT) (30), a plausible hypothesis is that more serotypes per PG will be found if more isolates are sequenced. The clustering of five O164 isolates with the reference EIEC genome ST270/O124 can be explained by strong resemblance between O164 and O124 antigens (31). Additionally, not all EIEC O135 isolates clustered, but were interspersed by *E. coli* O8 and O148, which are not related to O135 (31). Although isolates cluster roughly on serotype-level and serotyping is used for the description of individual isolates, some serotypes form multiple clusters and serotype switching is common. Therefore, more discriminatory techniques as whole genome sequencing provide more information for communication and surveillance purposes or outbreak investigations.

S. flexneri and *S. sonnei* isolates that were MSM-associated, clustered together using cgMLST analysis. These clusters also contained isolates from men that reported no sexual contact with other men or isolates from women. This could possibly be due to spillover to the non-MSM population, or (partially) due to misclassification of MSM as non-MSM. In our study, some MSM clusters were related to earlier described ones, although we also found one MSM-associated *S. sonnei* cluster and three MSM-associated clusters in *S. flexneri* not related to included reference isolates. One of the *S. flexneri* clusters contained only *S. flexneri* serotype 2a, one mostly *S. flexneri* 2a, but also *S. flexneri* serotype Y, and the third cluster contained only *S. flexneri* serotype 1c. Predominantly, *S. flexneri* 2a and *S. flexneri* 3a were associated with MSM before (5, 13), and to our knowledge, this study is the first that associates *S. flexneri* 1c with the MSM population. 78% of all *S. flexneri* and *S. sonnei* isolates within MSM-associated clusters, were isolated in the Amsterdam region, two *S. flexneri* and one *S. sonnei* MSM-associated clusters were entirely formed from isolates from Amsterdam. Two of these clusters were genetically related to other clusters present in the United Kingdom (13, 16). Presumably, the presence of isolates from exclusively Amsterdam in these clusters is merely biased because of the low sample size in combination with overall overrepresentation of MSM isolates from Amsterdam.

Without support of typing of the bacteria, contact tracing and outbreak investigations amongst the MSM population in particular can be complicated due to high numbers of sexual partners and anonymous sex, making it difficult to establish epidemiological links between cases (32). The allocation of isolates from 2016 and 2017 to all *S. flexneri* and *S. sonnei* MSM clusters provides evidence for prolonged circulation of these internationally MSM-associated *Shigella* isolates in the Netherlands. Our study was a snapshot in time, but it is important to monitor these (inter)national patterns for *Shigella* spp. over longer periods to enable outbreak detection, optimal prevention and targeted responses by public health authorities.

Outbreak investigations and other surveillance studies have indicated a large overlap between shigellosis and HIV (8, 11). Our study confirmed this phenomenon for the Dutch situation, as 30% to 100% of patients infected with isolates within MSM-associated clusters reported also an HIV infection, and only three patients that reported HIV were infected with isolates outside the MSM clusters. It was thought that this coexistence of shigellosis amongst MSM and HIV has multiple causes, which can be divided into social causes, as for instance specific sexual practices or the use of social media that might cause serosorting based on HIV status, and biological causes, as for instance susceptibility to infectious diseases or increased shedding of bacteria (8, 11).

While MSM-associated shigellosis is predominantly domestic or acquired from travel to other European countries, shigellosis in the non-MSM population is related to travel outside of Europe. Clusters related to travel were displayed in *S. flexneri* as well as *S. sonnei*. For EIEC, limited data on travel history for patients was available. Within the clusters related to travel, also domestically acquired isolates were present, indicating a further human-to-human transmission of imported isolates in the Netherlands. The emergence of foreign isolates in the Netherlands needs further investigation, for which specific transmission data is essential.

Antimicrobial resistance

In Dutch guidelines, cotrimoxazol, ciprofloxacin and azithromycin are advised for treatment of shigellosis cases (33). Azithromycin was not tested by any of the laboratories, because clinical breakpoints are not known from EUCAST guidelines (34). However, *in silico* determination of azithromycin resistance genes *erm(B)* and *mphA* revealed the detection of *erm(B)* in 9%, *mphA* in 11%, and both genes in 8% of the genomes. When both azithromycin resistance genes were present, *bla-TEM1b* gene was also present in 78% of *S. flexneri* and 100% of *S. sonnei* isolates. This combination of genes was only observed in isolates within the MSM clusters. All genes were described to be present on the pKSR100 plasmid that is associated with horizontal gene transfer (HGT) within MSM lineages before (13). Our study confirms the association of ciprofloxacin resistance with isolates from MSM and travel to Asia (8, 17). Furthermore, the resistance to cotrimoxazol, ciprofloxacin and azithromycin was present throughout the collection period in our dataset, and was predominantly lineage specific. This confirms earlier observations that the acquisition of ARGs through HGT drives the epidemiological outcomes and success of certain lineages (13, 15).

Phenotypic resistance in *Shigella spp.* and EIEC can be predicted with an *in silico* analysis. Our study confirmed earlier observations made in *E. coli* and *S. sonnei*, that correlation of detected ARGs to phenotypic outcome is significant, except for the aminoglycosides (18-21). We found a significant association between ARGs and phenotypes for resistance to ampicillin, cefotaxime, trimethoprim and sulphonamides, as almost all resistant phenotypes (95.6-100%) contained one or more of the associated ARGs, and susceptible phenotypes seldom possessed one of the associated ARGs (0-12.5%). The presence or absence of the plasmid-mediated *qnr* genes or one chromosomal point mutation, predominantly *gyrA* S83L, was not significantly associated with phenotypic resistance to ciprofloxacin. The presence of two or more chromosomal point mutations, however, was significantly associated with phenotypic resistance in all species in our study, a phenomenon that was earlier described for *S. sonnei* alone (17). The presence of point mutation *gyrA* S83L was thought to be a precursor for the full ciprofloxacin resistant phenotype, requiring at least one additional chromosomal point mutation (17, 21). In only 1.7% of phenotypically ceftazidime susceptible isolates, the *bla-CTX-M* gene was detected. In addition, in one of seven isolates displaying a resistant phenotype for ceftazidime none of the *bla* genes or *ampC* mutations was detected. In a previous study, one of 74 *E. coli* isolates displayed an identical phenomenon (19), while in another study no discrepancies were found between phenotype and genotype for ceftazidime resistance (21). The resistance to ceftazidime without a detected *bla* gene may be caused by a, yet unknown resistance mechanism that may be identified if more of such ceftazidime resistant isolates will be characterized. The fact that the presence of *bla* genes were not significantly associated with phenotypic amoxicillin/clavulanic acid resistance, can be explained by the fact that clavulanic acid is known to reduce beta-lactamase activity (35). Similarly, in our study piperacillin/tazobactam susceptible isolates also harbor beta-lactamase genes. However, tazobactam also is a beta-lactamase reducer (36). No association of phenotypic resistance to piperacillin/tazobactam with beta-lactamase genes was found, but this needs to be confirmed using a larger number of samples, as in our study only one resistant isolate was encountered. For the aminoglycosides gentamicin and tobramycin, no association between phenotype and genotypes was observed. Although none of the susceptible isolates contained one of the *aac(3)-IId* or *aph(3)Ia* genes, only low percentages of resistant phenotypes (11.8-13.3%) were in possession of one or more of these genes. Presumably, another resistance mechanism not identified by the methods used in our study causes the resistant phenotypes.

Virulence profiling

Almost all *S. flexneri* and EIEC isolates possessed virulence genes present in the pINV plasmid, while these genes were only detected in approximately half of the *S. sonnei* isolates. It is known that in *S. sonnei*, the pINV plasmid is frequently lost during subculturing (30). Three *S. flexneri* isolates lacked all genes encoding for the T3SS machinery and effectors and one isolate was in possession of the *Osp* genes, but lacked the *mxi-spa* operon, the *ipa-igp* operon and the *virA* gene. This is probably due to the excision of parts of the T3SS region. This phenomenon was described before and is thought to result from the high fitness costs of this region for the bacteria while being outside the human host (37). In our study, 84% of EIEC isolates contained the *set* gene, while an earlier study, analyzing a smaller set of isolates from different geographical origins, described that only 15% of EIEC isolates contained the *set* gene (28). Ten EIEC isolates harbored no genes encoding for the T3SS machinery or effectors, but seven out of these ten isolates contained the *sigA* and/or *pic* genes, indicating that these isolates lost their pINV plasmid. EIEC isolates containing the *shlA* gene in the SHI-2 island were observed, while an earlier study described this gene as absent from all EIEC (28). Some lineages of EIEC were not in possession of the SHI-2 pathogenicity island at all. Explanations for this could be that they might possess another pathogenicity island, like SHI-3 that is only present in *S. boydii*, containing genes involved in the same processes as the genes located on SHI-2 in *S. flexneri* and *S. sonnei*. Another explanation could be that these EIEC isolates are precursors of *Shigella spp.* and are in transition to gain full virulence potential as hypothesized earlier (27). Nonetheless, these EIEC isolates were capable of causing disease, because all isolates were collected from patients with symptoms. From 72% of these patients EIEC was the only detected pathogen (van den Beld et al., manuscript submitted).

Considerations

A strength of this study is that we combined microbiological characteristics of *Shigella spp.* and EIEC isolates with detailed epidemiological data of the patients. In addition, our study is representative for the Netherlands, as isolates from laboratories geographically distributed over the whole country were included.

Limitations of this study are that epidemiological data was collected from patients, and was therefore not an objective measurement. Although this probably does not have a major effect on the reported sexes of patients or travel history, MSM contact and HIV or STI status might be underreported. Furthermore, for EIEC isolates, the cluster formation was not as distinct as for *S. flexneri* and *S. sonnei*, probably due to the diversity of the isolates and to limited availability of epidemiological data. Moreover, as not all *Shigella spp.* and EIEC isolates detected in the Netherlands in 2016 and 2017 were available for this study, the observed clusters probably comprise more isolates. Therefore, the clusters observed during this study are the alleged “tip of the iceberg”.

Conclusions

During 2016 and 2017 predominantly *S. sonnei*, *S. flexneri* and EIEC were detected in our study. Isolates related to MSM-associated clusters from other countries were circulating, and had an overlap with patients that reported HIV infection and with antimicrobial resistance to azithromycin and ciprofloxacin. Travel-related isolates clustered together, sometimes with domestically acquired isolates, indicating further transmission of imported isolates. A substantial part of the characterized isolates was resistant to one or more of the first- and second-line antimicrobials for treatment. Identification with phenotypic methods and serotyping is challenging, as EIEC had no specific key characteristics and serotype switching is common in *S. flexneri*. In the Netherlands, thorough shigellosis case investigations are standardly conducted, which results in a comprehensive knowledge of epidemiological data. However, the current guidelines in which no laboratory surveillance of *Shigella spp.* is conducted, is not sufficient to detect all national and international clusters due to the low resolution of serotyping and due to the challenging contact investigations of MSM groups in particular. This study emphasized that epidemiological and laboratory surveillance are complementary to each other. Furthermore, multifactorial public health approaches for (inter)national surveillance purposes and outbreak investigations are important, particularly when combined with thorough characterization of isolates using techniques with high discriminatory power such as whole genome sequencing.

Abbreviations

AMR Antimicrobial resistance

ARG Antimicrobial resistance gene

CGE Center for Genomic Epidemiology

cgMLST Core genome multi-locus sequence typing

CI Confidence Interval

DNA Deoxyribonucleic acid

EIEC entero-invasive Escherichia coli

HGT Horizontal gene transfer

HIV Human Immunodeficiency Virus

IBESS Invasive Bacteria *E. coli-Shigella* Study

MGE mobile genetic elements

MLST multi-locus sequence typing

MMLs medical microbiological laboratories

MSM men who have sex with men

OR odds ratio

PCR polymerase chain reaction

PG phylogroup

pINV large invasion virulence plasmid

PrEP pre-exposure prophylaxis

SHI *Shigella* island

Spp. species

SRA Sequence Read Archive

ST sequence type

STI sexually transmitted infection

T3SS Type III secretion system

WGS whole genome sequencing

Declarations

Ethics approval and consent to participate

The IBESS-study was registered as observational study under number 23481 in the Dutch Trial Register. Patients were informed about the study and subjected to a single survey after their consent, to collect additional clinical and epidemiological data. In case of minors, one of the parents or caretakers was asked to participate in the survey. The medical ethics review board (METC) in Utrecht, the Netherlands, stated that this study was not subject to “medical research with human subjects” laws (protocol number 15-414/C). Data handling complied with the Dutch Personal Data Protection Act and with the EU General Data Protection Regulation.

Consent for publication

Not applicable.

Availability of data and material

The sequence data generated during the current study is available in the Sequence Read Archive (SRA) with study number PRJEB32617 (<https://www.ncbi.nlm.nih.gov/sra/>). Patient data other than provided in the Additional File 1 are not publicly available due to research participants' privacy.

Competing interests

The authors declare that they have no competing interests.

Funding

Data collection from patients was supported by the research fund of the Dutch National Institute for Public Health and environment (RIVM) for local Public Health Services.

Authors' contributions

MB, FR, MKS, and JR conceptualized the project and designed experiments. AH, SK, EH, DHB, RN and ACAH performed experiments. MB, DB, and HH analyzed the data. MB, RP, FR, MKS and JR interpreted results. FR, MKS and JR supervised the project. MB wrote the manuscript. All authors read, reviewed and approved the final manuscript.

Acknowledgements

The IBESS group provided isolates and patient data, and consists of the following contributors from the Netherlands:

- M. J. C. van den Beld, National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven and Department of Medical Microbiology and Infection Prevention, University of Groningen, University Medical Center Groningen, Groningen
- E. Warmelink, Public Health Service GGD Groningen, Groningen.
- M. D. Kooistra-Smid, Certe, Department of Medical Microbiology, Groningen and Department of Medical Microbiology and Infection Prevention, University of Groningen, University Medical Center Groningen, Groningen.
- A. W. Friedrich, Department of Medical Microbiology and Infection Prevention, University of Groningen, University Medical Center Groningen, Groningen.
- F. A. G. Reusbaet, National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven.
- D. W. Notermans, National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven.
- M. W. F. Petrignani, Public health service GGD Amsterdam, Amsterdam.
- T. Waegemaekers, Public health service GGD Gelderland-Midden, Arnhem.
- J. W. A. Rossen, Department of Medical Microbiology and Infection Prevention, University of Groningen, University Medical Center Groningen, Groningen.
- A. P. van Dam, Amsterdam Health Service, Amsterdam.
- S. Svraka-Latifovic, CBSL, Tergooi, Hilversum.
- J. J. Verweij, Elisabeth-TweeSteden Hospital, Laboratory for Medical Microbiology and Immunology, Tilburg.
- L. E. S. Bruijnesteijn van Coppenraet, Isala, Laboratory for Medical Microbiology and Infectious diseases, Zwolle.
- K. Waar, Izore, Centre for Infectious Diseases Friesland, Leeuwarden.

- M. Hermans, Jeroen Bosch Ziekenhuis, Laboratorium Medische Microbiologie, 's-Hertogenbosch.
- D. L. J. Hess, LabMicTA, Laboratory for Medical Microbiology and Public Health, Hengelo.
- L.J.M. van Mook, Microvida location Amphibia, Breda.
- M. C. Bergmans, Microvida location Bravis, Roosendaal.
- R. R. Jansen, OLVG, Medical Microbiological Laboratory, Amsterdam.
- J. H. B. van de Bovenkamp, PAMM Laboratory for Medical Microbiology, Veldhoven.
- A. Demeulemeester, SHL-group, Etten-Leur.
- E. Reinders, St. Antonius Ziekenhuis, Medical Microbiology and Immunology, Nieuwegein.
- F. M. Linssen, Zuyderland Medical Centre, Medical Microbiology, Heerlen.
- All adjacent Public Health Services .

References

1. Mattock E, Blocker AJ. How Do the Virulence Factors of *Shigella* Work Together to Cause Disease? Front Cell Infect Microbiol. 2017;7:64.
2. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. Nat Rev Microbiol. 2004;2(2):123-40.
3. Gray MD, Lacher DW, Leonard SR, Abbott J, Zhao S, Lampel KA, et al. Prevalence of Shiga toxin-producing *Shigella species* isolated from French travellers returning from the Caribbean: an emerging pathogen with international implications. Clin Microbiol Infect. 2015;21(8):765 e9- e14.
4. Lima IF, Havt A, Lima AA. Update on molecular epidemiology of *Shigella* infection. Curr Opin Gastroenterol. 2015;31(1):30-7.
5. Baker KS, Dallman TJ, Ashton PM, Day M, Hughes G, Crook PD, et al. Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: a cross-sectional study. Lancet Infect Dis. 2015;15(8):913-21.
6. Pijnacker R, Friesema IHM, Franz E, Van Pelt W. Trends van shigellosemeldingen in Nederland, 1988-2015. Infectieziekten Bulletin. 2017;28 (4):121-8.
7. Wu HH, Shen YT, Chiou CS, Fang CT, Lo YC. Shigellosis outbreak among MSM living with HIV: a case-control study in Taiwan, 2015-2016. Sexually transmitted infections. 2019;95(1):67-70.
8. Ingle DJ, Easton M, Valcanis M, Seemann T, Kwong JC, Stephens N, et al. Co-circulation of multidrug-resistant *Shigella* among men who have sex with men, Australia. Clin Infect Dis. 2019(Epub ahead of print).
9. Bowen A, Grass J, Bicknese A, Campbell D, Hurd J, Kirkcaldy RD. Elevated Risk for Antimicrobial Drug-Resistant *Shigella* Infection among Men Who Have Sex with Men, United States, 2011-2015. Emerg Infect Dis. 2016;22(9):1613-6.
10. Toro C, Arroyo A, Sarria A, Iglesias N, Enriquez A, Baquero M, et al. Shigellosis in subjects with traveler's diarrhea versus domestically acquired diarrhea: implications for antimicrobial therapy and human immunodeficiency virus surveillance. Am J Trop Med Hyg. 2015;93(3):491-6.
11. Mohan K, Hibbert M, Rooney G, Canvin M, Childs T, Jenkins C, et al. What is the overlap between HIV and shigellosis epidemics in England: further evidence of MSM transmission? Sexually transmitted infections. 2018;94(1):67-71.
12. Hoffmann C, Sahly H, Jessen A, Ingiliz P, Stellbrink HJ, Neifer S, et al. High rates of quinolone-resistant strains of *Shigella sonnei* in HIV-infected MSM. Infection. 2013;41(5):999-1003.
13. Baker KS, Dallman TJ, Field N, Childs T, Mitchell H, Day M, et al. Horizontal antimicrobial resistance transfer drives epidemics of multiple *Shigella species*. Nat Commun. 2018;9(1):1462.
14. Mook P, McCormick J, Bains M, Cowley LA, Chattaway MA, Jenkins C, et al. ESBL-Producing and Macrolide-Resistant *Shigella sonnei* Infections among Men Who Have Sex with Men, England, 2015. Emerg Infect Dis. 2016;22(11):1948-52.
15. Holt KE, Baker S, Weill FX, Holmes EC, Kitchen A, Yu J, et al. *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nat Genet. 2012;44(9):1056-9.
16. Baker KS, Dallman TJ, Field N, Childs T, Mitchell H, Day M, et al. Genomic epidemiology of *Shigella* in the United Kingdom shows transmission of pathogen sublineages and determinants of antimicrobial resistance. Sci Rep. 2018;8(1):7389.

17. Chung The H, Rabaa MA, Pham Thanh D, De Lappe N, Cormican M, Valcanis M, et al. South asia as a reservoir for the global spread of ciprofloxacin-resistant *Shigella sonnei*: a cross-sectional study. *PLoS Med*. 2016;13(8):e1002055.
18. Tyson GH, McDermott PF, Li C, Chen Y, Tadesse DA, Mukherjee S, et al. WGS accurately predicts antimicrobial resistance in *Escherichia coli*. *J Antimicrob Chemother*. 2015;70(10):2763-9.
19. Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, et al. Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *J Antimicrob Chemother*. 2013;68(10):2234-44.
20. Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agerso Y, Lund O, et al. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *J Antimicrob Chemother*. 2013;68(4):771-7.
21. Sadouki Z, Day MR, Doumith M, Chattaway MA, Dallman TJ, Hopkins KL, et al. Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of *Shigella sonnei* isolated from cases of diarrhoeal disease in England and Wales, 2015. *J Antimicrob Chemother*. 2017;72(9):2496-502.
22. Pettengill EA, Pettengill JB, Binet R. Phylogenetic analyses of *Shigella* and enteroinvasive *Escherichia coli* for the identification of molecular epidemiological markers: whole-genome comparative analysis does not support distinct genera designation. *Front Microbiol*. 2015;6:1573.
23. van den Beld MJC, de Boer RF, Reubsæet FAG, Rossen JWA, Zhou K, Kuiling S, et al. Evaluation of a culture dependent algorithm and a molecular algorithm for identification of *Shigella spp.*, *Escherichia coli*, and enteroinvasive *E. coli* (EIEC). *J Clin Microbiol*. 2018;56:e00510-18.
24. Chattaway MA, Schaefer U, Tewolde R, Dallman TJ, Jenkins C. Identification of *Escherichia coli* and *Shigella Species* from Whole-Genome Sequences. *J Clin Microbiol*. 2017;55(2):616-23.
25. Baker KS, Campos J, Pichel M, Della Gaspera A, Duarte-Martinez F, Campos-Chacon E, et al. Whole genome sequencing of *Shigella sonnei* through PulseNet Latin America and Caribbean: advancing global surveillance of foodborne illnesses. *Clin Microbiol Infect*. 2017;23(11):845-53.
26. van den Beld MJ, Reubsæet FA. Differentiation between *Shigella*, enteroinvasive *Escherichia coli* (EIEC) and noninvasive *Escherichia coli*. *Eur J Clin Microbiol Infect Dis*. 2012;31(6):899-904.
27. Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella spp.* *Infect Immun*. 2004;72(9):5080-8.
28. Hazen TH, Leonard SR, Lampel KA, Lacher DW, Maurelli AT, Rasko DA. Investigating the relatedness of enteroinvasive *Escherichia coli* to other *E. coli* and *Shigella* isolates by using comparative genomics. *Infect Immun*. 2016;84(8):2362-71.
29. Connor TR, Barker CR, Baker KS, Weill FX, Talukder KA, Smith AM, et al. Species-wide whole genome sequencing reveals historical global spread and recent local persistence in *Shigella flexneri*. *Elife*. 2015;4:e07335.
30. The HC, Thanh DP, Holt KE, Thomson NR, Baker S. The genomic signatures of *Shigella* evolution, adaptation and geographical spread. *Nat Rev Microbiol*. 2016;14(4):235-50.
31. DebRoy C, Fratamico PM, Yan X, Baranzoni G, Liu Y, Needleman DS, et al. Comparison of O-antigen gene clusters of all O-serogroups of *Escherichia coli* and proposal for adopting a new nomenclature for O-typing. *PLoS One*. 2016;11(1):e0147434.
32. Gilbert VL, Simms I, Jenkins C, Furegato M, Gobin M, Oliver I, et al. Sex, drugs and smart phone applications: findings from semistructured interviews with men who have sex with men diagnosed with *Shigella flexneri* 3a in England and Wales. *Sexually transmitted infections*. 2015;91(8):598-602.
33. SWAB. Optimaliseren van het antibioticabeleid in Nederland XVIII: SWAB richtlijn antimicrobiële therapie voor acute infectieuze diarree 2014 [Available from: [http://www.swab.nl/swab/cms3.nsf/uploads/B5B9ED1BD30F42DFC1257CB80019C398/\\$FILE/Herziene%20SWAB%20richtlijn%20Acute%20Diarree.pdf](http://www.swab.nl/swab/cms3.nsf/uploads/B5B9ED1BD30F42DFC1257CB80019C398/$FILE/Herziene%20SWAB%20richtlijn%20Acute%20Diarree.pdf)].
34. EUCAST. Clinical breakpoints and dosing 2019 [Available from: http://www.eucast.org/clinical_breakpoints/].
35. Wise R, Andrews JM, Bedford KA. In vitro study of clavulanic acid in combination with penicillin, amoxycillin, and carbenicillin. *Antimicrob Agents Chemother*. 1978;13(3):389-93.
36. Cullmann W, Stieglitz M. Antibacterial activity of piperacillin and tazobactam against beta-lactamase-producing clinical isolates. *Chemotherapy*. 1990;36(5):356-64.
37. Pilla G, McVicker G, Tang CM. Genetic plasticity of the *Shigella* virulence plasmid is mediated by intra- and inter-molecular events between insertion sequences. *PLoS Genet*. 2017;13(9):e1007014.
38. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-8.

39. Brown J, Pirrung M, McCue LA. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*. 2017.
40. Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM. An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS One*. 2013;8(12):e85024.
41. Ounit R, Lonardi S. Higher classification sensitivity of short metagenomic reads with CLARK-S. *Bioinformatics*. 2016;32(24):3823-5.
42. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455-77.
43. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-5.
44. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25(7):1043-55.
45. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, et al. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol*. 2006;60(5):1136-51.
46. Cowley LA, Osegun DR, Chattaway MA, Dallman TJ, Jenkins C. Phylogenetic comparison of enteroinvasive *Escherichia coli* isolated from cases of diarrhoeal disease in England, 2005-2016. *J Med Microbiol*. 2018;67:884-8.
47. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019;47:W256-W9.
48. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol*. 2014;52(5):1501-10.
49. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.

Tables

Table 1. Isolates and their identification, sequence status and patient data availability.

Species	n total	n (%) sequenced ^a	n (%) patient data available	n (%) sequenced and data
<i>S. dysenteriae</i>	0	0	0	0
<i>S. flexneri</i>	104	87 (84)	79 (76)	79 (76)
<i>S. boydii</i>	3	2 (67)	2 (67)	2 (67)
<i>S. sonnei</i>	232	190 (82)	168 (72)	168 (72)
Provisional <i>Shigella</i>	10	6 (60)	8 (80)	5 (50)
EIEC	64	62 (97)	33 (52)	32 (50)
EIEC/ <i>S. flexneri</i>	1	1 (100)	1 (100)	1 (100)
Total	414	348 (84)	291 (70)	287 (69)

^aAll EIEC isolates were sequenced, except for two that were not available anymore. Other selection for sequencing was based on definitive species identification and data availability.

Table 2. Phenotypic traits of *S. sonnei*, *S. flexneri* and EIEC, in percentage of positives.

Phenotypic trait	<i>S. sonnei</i> (n=232)	<i>S. flexner</i> (n=104)	EIEC (n=64)
Motility ^a	0	0	30
LDC ^a	0	0	45
ODC	98	0	41
ADH	2	5	6
Esculin ^a	0	0	8
Indole	0	16	77
Gas from D-glucose	0	0	72
Indole + gas from D-glucose ^a	0	0	59
ONPG	90	1	89
Fermentation of:			
D-glucose	99	99	100
lactose	2	1	69
D-sucrose	2	0	44
D-xylose	39	8	84
D-mannitol	81	96	97
dulcitol	0.4	0	34
D-sorbitol	0.4	5	88
Salicin ^a	0	0	5
D-trehalose	100	82	97
D-raffinose	0.4	8	45
glycerol	9	3	50

^aTests used for distinction

Table 3. Phenotypic resistance of isolates, and the presence of associated antimicrobial resistance genes.

	Resistant phenotype		Sensitive phenotype		OR (95% CI) ^a		Resistant phenotype		Sensitive phenotype		OR (95% CI) ^a
	n	%	n	%			n	%	n	%	
Ampicillin (n = 241)	109	45	132	55		Gentamicin (n = 243)	17	7.0	226	93.0	
<i>blaTEM-1b</i>	46	42.2	2	1.5	47.5 (11.2-201-8)	<i>aac(3)-IId</i>	1	5.9	0	0	
<i>blaTEM-1c</i>	2	1.8	0	0		<i>aph(3)-Ia</i>	1	5.9	0	0	
<i>blaTEM-30</i>	1	0.9	0	0		<i>≥ 1 of aac or aph gene</i>	2	11.8	0	0	
<i>blaDHA-1</i>	1	0.9	0	0		Tobramycin (n = 238)	15	6.3	223	93.7	
<i>blaOXA-1</i>	55	50.5	0	0		<i>aac(3)-IId</i>	1	6.7	0	0	
<i>blaCTX-M-15</i>	10	9.2	0	0		<i>aph(3)-Ia</i>	1	6.7	0	0	
<i>blaCTX-M-32</i>	1	0.9	0	0		<i>≥ 1 of aac or aph gene</i>	2	13.3	0	0	
<i>blaCTX-M-55</i>	2	1.8	0	0		Ciprofloxacin (n = 264)	49	18.6	215	81.4	
<i>≥ 1 of bla genes</i>	106	97.2	2	1.5	2296.7 (376.8-13998.4)	<i>qnrB19</i>	0	0	12	5.6	
Amoxicillin/clavulanic acid (n = 227)	57	25	170	75		<i>qnrB4</i>	0	0	1	0.5	
<i>blaTEM-1b</i>	19	33.3	23	13.5	3.2 (1.6-6.5)	<i>qnrS1</i>	4	8.2	12	5.6	
<i>blaTEM-1c</i>	0	0	2	1.2		<i>gyrA S83A^b</i>	0	0	1	0.5	
<i>blaTEM-30</i>	1	1.8	0	0		<i>gyrA S83L^b</i>	48	98.0	36	16.7	238.7 (31.9-1785.5)
<i>blaDHA-1</i>	0	0	1	0.6		<i>gyrA D87G^b</i>	28	57.1	2	0.9	142.0 (31.6-638.3)
<i>blaOXA-1</i>	39	68.4	14	8.2		<i>gyrA D87Y^b</i>	2	4.1	3	1.4	
<i>blaCTX-M-15</i>	2	3.5	2	1.2	24.1 (11.0-52.8)	<i>gyrA D87N^b</i>	18	36.7	0	0	
<i>blaCTX-M-32</i>	1	1.8	0	0		<i>parC S80I^b</i>	48	98.0	1	0.5	10272.0 (631.3-167142.7)
<i>blaCTX-M-55</i>	1	1.88	0	0		<i>parE S458A^b</i>	12	24.5	0	0	
<i>≥ 1 of bla genes</i>	56	98.2	39	22.9	188.1 (25.2-1403.1)	<i>≥ 1 of qnr genes</i>	4	8.2	25	11.6	
Piperacillin/Tazobactam (n = 227)	1	0.4	226	99.6		<i>1 point mutation</i>	1	2.0	40	18.6	
<i>blaTEM-1b</i>	0	0	43	19.0		<i>≥ 2 of point mutations</i>	48	98.0	1	0.5	10272.0 (631.3-167142.7)
<i>blaTEM-1c</i>	0	0	2	0.9		<i>≥ 1 of genes/mutations</i>	49	100	64	29.8	
<i>blaTEM-30</i>	0	0	1	0.4		Trimethoprim (n =181)	157	86.7	24	13.3	
<i>blaOXA-1</i>	0	0	49	21.7		<i>dfrA1</i>	131	83.4	2	8.3	55.4 (12.3-250.2)
<i>blaCTX-M-15</i>	0	0	9	4.0		<i>dfrA14</i>	21	13.4	1	4.2	
<i>blaCTX-M-32</i>	0	0	1	0.4		<i>dfrA17</i>	13	8.3	0	0	
<i>blaCTX-M-55</i>	0	0	1	0.4		<i>dfrA7</i>	3	1.9	0	0	
<i>≥ 1 of bla genes</i>	0	0	97	42.9		<i>dfrA8</i>	1	0.6	0	0	
Cefotaxime (n = 241)	13	5.4	228	94.6		<i>≥ 1 of dfrA genes</i>	157	100	3	12.5	
<i>blaCTX-M-15</i>	10	76.9	0	0		Trimethoprim/sulphonamide (cotrimoxazole) (n = 248)	180	72.6	68	27.4	
<i>blaCTX-M-32</i>	1	7.7	0	0		<i>Sul1</i>	24	13.3	1	1.5	10.3 (1.4-77.8)
<i>blaCTX-M-55</i>	2	15.4	0	0		<i>Sul2</i>	166	92.2	8	11.8	88.9 (35.5-222.6)
<i>≥ 1 of blaCTX-M genes</i>	13	100	0	0	n.c. ^c	<i>Sul3</i>	1	0.6	0	0	
Ceftazidime (n = 242)	7	2.9	235	97.1		<i>dfrA1</i>	143	79.4	36	52.9	3.4 (1.9-6.2)
<i>blaDHA-1</i>	1	14.3	0	0		<i>dfrA14</i>	30	16.7	0	0	
<i>blaCTX-M-15</i>	3	42.9	4	1.7	43.3 (7.2-260.4)	<i>dfrA17</i>	18	10.0	1	1.5	
<i>blaCTX-M-32</i>	1	14.3	0	0		<i>dfrA5</i>	1	0.6	0	0	
<i>blaCTX-M-55</i>	1	14.3	0	0		<i>dfrA7</i>	5	2.8	0	0	
<i>≥ 1 of blaDHA/CTX-M genes</i>	6	85.7	4	1.7	346.5 (33.5-3584.1)	<i>dfrA8</i>	1	0.6	0	0	
						<i>≥ 1 of sul genes</i>	172	95.6	9	13.2	140.9 (52.0-382.1)
						<i>≥ 1 of dfrA genes</i>	179	99.4	37	54.4	150.0 (19.8-1133.4)
						<i>≥ 1 of dfrA and ≥ 1 sul genes</i>	172	95.6	3	4.4	465.8 (119.9-1810.0)

^aIf significant, odds ratio (OR) with 95% confidence interval (95% CI) are displayed. ^bChromosomal point mutation at position n. ^cNon-calculable because it perfectly predicts phenotypic resistance.

Figures

Figure 1

Core genome MLST tree of all isolates with species designations 348 isolates, distance based on comparing 2315 alleles using the Enterobase Escherichia/Shigella cgMLST v1 scheme. Missing values are an own category. Grey squares = results of decisive phenotypic tests or serology, box with border only = negative, filled square = positive. Phenotypic/serologic tests from inner to outer ring: motility, lysine decarboxylase, combination of gas and indole, esculin, salicin fermentation, and inconclusive Shigella serology.

Figure 2

Core genome MLST tree of *S. flexneri*, including context isolates 101 isolates, distance based on comparing 2315 alleles using the Enterobase Escherichia/Shigella cgMLST v1 scheme. Missing values are an own category. Red text = MSM-associated clusters. Black text = serotype; prov = provisional Shigella. Qnr genes left to right = qnrB19, qnrB4, qnrS1; SHI-1 left to right = sigA, pic, set ; SHI-2 left to right = iucA, iucB, iucC, iucD, iutA, shiA, shiB, shiD, shiE; T3SS effectors left to right = ipa-igp operon, virA, ospB, ospC1, ospC3, ospD3, ospE1, ospE2, ospF, ospG. Further features are explained in the legend within the figure.

Figure 3

Core genome MLST tree of *S. sonnei*, including context isolates 203 isolates, distance based on comparing 2315 alleles using the Enterobase Escherichia/Shigella cgMLST v1 scheme. Missing values are an own category. Red text = MSM-associated clusters. Black text = serotype; prov = provisional Shigella. Qnr genes left to right = qnrB19, qnrB4, qnrS1; SHI-1 left to right = sigA, pic, set ; SHI-2 left to right = iucA, iucB, iucC, iucD, iutA, shiA, shiB, shiD, shiE; T3SS effectors left to right = ipa-igp operon, virA, ospB, ospC1, ospC3, ospD3, ospE1, ospE2, ospF, ospG. Further features are explained in the legend within Figure 2.

Figure 4

Core genome MLST tree of EIEC, including context isolates 71 isolates, distance based on comparing 2315 alleles using the Enterobase Escherichia/Shigella cgMLST v1 scheme. Missing values are an own category. Red text = MSM-associated clusters. Black text = serotype; prov = provisional Shigella. Qnr genes left to right = qnrB19, qnrB4, qnrS1; SHI-1 left to right = sigA, pic, set ; SHI-2 left to right = iucA, iucB, iucC, iucD, iutA, shiA, shiB, shiD, shiE; T3SS effectors left to right = ipa-igp operon, virA, ospB, ospC1, ospC3, ospD3, ospE1, ospE2, ospF, ospG. Further features are explained in the legend within Figure 2.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFile1V2.xlsx](#)
- [AdditionalFile2.xlsx](#)