Shigella spp. and entero-invasive Escherichia coli
van den Beld, Maaike

DOI:
10.33612/diss.101452646

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 9

Summarizing discussion, future perspectives and conclusions
Summarizing discussion

Shigella spp. and the E. coli pathotype entero-invasive Escherichia coli (EIEC) cause diarrheal disease characterized by the invasion of human epithelial cells and resulting in watery or bloody diarrhea, called shigellosis or dysentery [1, 2]. Both bacteria evolved from commensal E. coli via convergent evolution, and are related to such an extent that they should be classified as one species together with other E. coli [3-5]. The detection of Shigella spp. and EIEC from fecal samples by culture is insensitive and demanding, especially in the case of EIEC [6, 7]. When an isolate is obtained, distinction between Shigella spp. and EIEC can be complicated as EIEC can be both E. coli-like and Shigella-like regarding their biochemical profile [8-10]. While accurate detection and identification with molecular approaches is common practice, this does not allow distinction between Shigella spp. and EIEC [11, 12]. Moreover, even after molecular detection, culturing the bacteria remains pivotal. Dutch case definitions for shigellosis require isolation of Shigella spp. by culture. Furthermore, culturing is also necessary for antimicrobial susceptibility testing and surveillance purposes [13]. In the Netherlands, as in many countries, control and epidemiological surveillance of infections with Shigella spp. are regulated in public health guidelines employed by local and national health authorities, while these guidelines do not regulate EIEC infections [13, 14]. Because distinction in the laboratory is complex, execution of the shigellosis control regulations can be challenging for physicians and laboratories [15, 16]. In this thesis, we aimed to provide tools for a better, more distinct diagnosis of Shigella spp. and entero-invasive Escherichia coli (EIEC) infections. In addition, we discussed the challenges of current public health guidelines for these infections and suggested potential improvements.

Taxonomy and current diagnostics

Modern molecular techniques confirm the relatedness of Shigella spp. and E. coli [3, 17]. However, genomic taxonomy studies, including all typestrains of the species, were never performed. In part I of this thesis, we assessed the taxonomic status and current diagnostics of Shigella spp. and EIEC. In chapter 2, a thorough genomic taxonomy study with all typestrains of the species within the genera Shigella and Escherichia showed that all Shigella species and E. coli share a genetic similarity above the established species boundaries of 98.7% 16S rRNA gene similarity, 70% digital DNA-DNA hybridization and 95% average nucleotide identity (ANI) and amino-acid identity (AAI). In contrast, all other species within the genus Escherichia displayed similarities with the typestrains of Shigella spp. and E. coli below species boundaries. This confirms conclusions from other studies stating that the separate genus and species status of Shigella is taxonomically incorrect [3-5]. Furthermore, in silico evaluation of eight published diagnostic assays for Shigella spp. and EIEC demonstrated that these identification methods could not distinguish between the two genera if other Shigella spp. and EIEC isolates were included than in the original studies. The incorrect classification of Shigella spp. is one of the reasons for the subsequent adversities in diagnostics that impedes the conduction of public health measurements. Therefore, we explored practical proposals for reclassification. These proposals comprehend either the incorporation of Shigella spp. into the species E. coli, or the incorporation of the pathotype EIEC into the genus Shigella. Both options comply with taxonomic rules, and will facilitate both diagnostics as well as the application of public health guidelines for medical microbiological laboratories (MMLs) and public health authorities.

In clinical diagnostics, a large variety of methods are available for the detection and identification of Shigella spp. and EIEC. As notifications from MMLs towards health authorities were not uniform, the comparability of the current culture dependent and molecular methods used by MMLs in the Netherlands for detection and identification of Shigella spp. and EIEC was questioned and needed elaboration [15]. Therefore, in chapter 3, we evaluated the status of currently used diagnostics by the organization of a collaborative trial. This trial comprised a survey about used culture and molecular techniques and a standardized sample set containing DNA of an S. flexneri serotype 2a isolate and stool samples spiked with this isolate, both in ascending concentrations. The results of the survey showed that the 16 participating MMLs used a wide variety of selective agar plates, enrichment broths, identification techniques and serotyping methods for the detection and identification of Shigella spp. or EIEC by culture. Significantly, 11 out of 16 MMLs employed no protocol for the detection and culturing of EIEC at all. In addition, all MMLs used the ipaH gene as a target in their molecular diagnostic approaches. However, they all used a unique combination of input volumes, DNA extraction methods, PCR-platforms, master mixes, ampiclon sizes, primers and probes, and PCR reaction volumes. Nonetheless, this variety had no consequences for the qualitative molecular detection, in contrast to the semi-quantitative detection. The different results in the molecular semi-quantitative techniques demonstrated that the use of a universal cut-off in Ct-value of molecular detected Shigella spp. or EIEC cannot function as a prioritization for the employments of public health measurements, as suggested in the beginning of the molecular diagnostic era [15]. Ct-values are only indicative for quantification. Conceivably, cut-off values based on actual concentration or colony forming units (CFUs) could function as prioritization for public health actions, but requires local calibrations for each molecular method used by MMLs.

Optimization of diagnostics

As there is a need to distinguish Shigella spp. and EIEC, the opportunities for optimization of detection and subsequent identification of Shigella spp. and EIEC were explored in part II of this thesis. In chapter 4, a culture dependent diagnostic algorithm was proposed based on earlier described virulence targets, and phenotypical and serological features of Shigella spp., E. coli and EIEC [18]. This culture dependent algorithm was compared to a molecular algorithm based on detection of the ipaH gene in conjunction with O-serotype specific genes of Shigella spp. For discrepancy analysis, whole genome sequencing (WGS) was used. The
comparison demonstrated that the molecular algorithm was fast and 100% accurate but had a limited resolution, because it did not allow for definitive identification of S. boydii, EIEC and S. dysenteriae isolates other than serotype 1. On the other hand, the culture dependent algorithm had a high resolution as it could identify all species and serotypes, but was more time-consuming and labor-intensive. It identified all S. sonnei and noninvasive E. coli correctly, and 92% of S. dysenteriae, 85% of S. flexneri, 93% of S. boydii and 90% of EIEC isolates. Isolates that showed discrepant results were analyzed using WGS, which partly solved species and serotype assignment. However, it also complicated the species and serotype assignment partly by adding yet another assignment instead of confirming the identification of one of the two other methods, probably because these isolates were originally misidentified. This demonstrates the complexity of distinguishing Shigella spp. from each other and from EIEC. Regardless of this complexity, the molecular and the culture dependent algorithms both performed relatively well and appeared to be complementary to each other. The molecular algorithm is more suitable to be applied as rapid screening used by routine diagnostic laboratories, since exact identification is not necessary for treatment of patients. Subsequently, isolates of which a more thorough identification is required, for instance for (local) surveillance purposes, can be send to a reference laboratory that employs the culture dependent algorithm.

For the identification of most bacteria, such extended techniques as described in chapter 4 are not applied to routine clinical diagnostics. Currently, most MMLs use Matrix-Assisted Laser-Desorption Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) for the identification of cultured bacteria in a routine diagnostic setting. Therefore, we investigated the applicability of MALDI-TOF MS for the identification of Shigella spp. and EIEC in chapter 5. As commercially available databases are not able to distinguish Shigella spp. from E. coli in general, alternative databases and approaches were explored [19]. Previous studies already examined potential alternatives as assigning biomarkers, developing mathematical classification algorithms or designing custom-made databases [20-22]. However, none of them analyzed EIEC isolates. Moreover, they used mass ranges other than those used in routine settings or used general biomarkers not assigned to specific species. In chapter 5, a custom-made database was developed, biomarkers were assigned and classifiers were developed with machine learning using Shigella spp., E. coli, and EIEC isolates. We showed that none of these alternative approaches for MALDI-TOF MS was suitable for their distinction. In conclusion, an identification algorithm with MALDI-TOF MS using commercially databases for the identification of the Shigella spp./E. coli complex is feasible. However, for subsequent distinction, other test methods are necessary.

Incidence, epidemiology, clinical implications and impact on public health
Current public health measurements only regulate control of culture-confirmed shigellosis, while shigellosis that is only detected with molecular methods and EIEC infections are not regulated at all. In part III, we organized a Dutch multicenter cross-sectional study, the Invasive Bacteria E. coli-Shigella Study (iBESS), and collected samples and metadata during 2016 and 2017. To investigate if there is biological reason for their different public health approaches, we assessed and compared the incidence, risk factors for infection, symptoms and severity of disease, degree of secondary infections and socio-economic consequences of EIEC infections and Shigella spp infections in chapter 6. No differences were found regarding incidence and the degree of secondary infections, although differences regarding risks for infection, symptoms, severity of disease and socio-economic consequences were observed. The predominant differences between infections with Shigella spp. and EIEC were related to the source of infection. Patients with EIEC were more likely to report the ingestion of contaminated food or water, and less likely to be men who have sex with men (MSM). Furthermore, although the overall percentage of patients that were hospitalized was comparable, patients with Shigella spp. infection were more likely to have a lengthy stay. The presence of symptoms and disease severity showed ambivalent results and no clear evidence for a more severe course for one over the other was determined. We concluded that the above-described differences provide insufficient divergence in clinical outcomes and impact on public health to justify different approaches to control Shigella spp. and EIEC infections. Additionally, risk factors for infection, symptoms and severity of disease, the degree of secondary infections and socio-economic consequences of shigellosis cases with culture confirmed infections were compared to shigellosis cases that were detected with molecular methods only. Again, predominantly sources of infection differentiated, as culture-negative cases were more likely to report travelling and ingestion of contaminated food or water and less likely to report MSM contact than culture-positive cases. Moreover, it appeared that, in our setting, culture-negative cases suffered from a slightly more severe course of disease and were longer absent from work. Therefore, our findings do not provide biological evidence for the current case-definition of shigellosis in the Netherlands, in which the isolation of Shigella spp. is required. Based on the evidence gathered, we presented recommendations for potential optimization of shigellosis guidelines that include EIEC infections and infections only detected with molecular methods.

Another potential optimization for managing shigellosis could be the inclusion of predictive diagnostic genetic determinants for disease outcomes in individual patients that could guide public health control efforts, independent from their species designation. For example, genetic markers were used to update laboratory criteria for shiga-toxin producing E. coli (STEC), another pathotype of E. coli, to guide employment of control measures [23]. In chapter 7, we used genome wide association studies (GWAS) to assess the presence of predictive genetic determinants in Shigella spp. and EIEC isolates, using data from the iBESS study. However, genetic determinants associated with specific symptoms or disease severity were not observed for shigellosis patients, suggesting that the divergence in presence of symptoms and severity of disease was depending on other factors than genetic variation in
isolates recovered from patients, as discussed below in future perspectives. Because a specific virulence gene profile for infecting isolates is not present, this cannot be used to predict outcomes in individual patients and for guiding control measures. To verify the accuracy of the used methods, the feature genus was also associated with genetic variants as benchmark. For this feature, we observed significant related gene combinations and small sequences (k-mers), indicating that our algorithms were performing accurately.

In the Netherlands, each case of culture-confirmed shigellosis is followed up by source tracing, conducted by public health services. The National Institute for Public Health and the Environment (RIVM) collects and combines all epidemiological surveillance data. In contrast to this surveillance based on patient data, laboratory surveillance is not in place. Consequently, the phenotypic and genetic characteristics of circulating isolates are unexplored. To close this gap of knowledge, a thorough characterization of isolates collected during the IBESS-study was performed in chapter 8. Phenotypic features, phenotypic resistance, presence of resistance and virulence genes, multi locus sequence types (MLST), core genome (cg)MLST and genomic epidemiology were assessed for Shigella spp. and EIEC isolates collected in 2016 and 2017. In these years, S. sonnei, S. flexneri and EIEC were the most isolated species in descending order. The application of the culture dependent algorithm described in chapter 4 of this thesis on the isolates collected during IBESS confirmed the troublesome identification of Shigella spp. and EIEC, as it was not feasible to identify or serotype all isolates. Furthermore, phenotypic resistance to advised antimicrobials by Dutch guidelines, i.e., co-trimoxazole and ciprofloxacin, was frequently observed. A significant association between resistance to the second line drug ciprofloxacin and patients that reported MSM contact or travel to Asia was found, as has been previously reported by studies in other countries [24-29]. When inferring a cgMLST tree comprising all isolates, some clusters contained multiple species, confirming again the close genetic relationship of Shigella spp. and EIEC as described before in chapter 2, 4 and 5 of this thesis and by many other researchers [3, 30-33]. Isolates from patients that travelled to the same regions clustered together, interspersed with some domestic isolates, indicating secondary transmission of these imported isolates. Additionally, we were able to link Dutch isolates to international clusters, predominantly MSM associated [24, 34-36]. Overall, the benefits of a multifactorial public health approach that comprises epidemiological surveillance as well as laboratory surveillance with the use of discriminatory techniques as WGS were illustrated. Currently, only epidemiological data is collected, together with data about the infecting isolates based on serotyping performed by the MML that notified the shigellosis case towards health authorities. However, as resolution of serotyping is low, and serotype switching amongst S. flexneri is common [37, 38], WGS is superior and should be used for surveillance purposes. Moreover, WGS is essential for the detection of international clusters or MSM related clusters, in particular because contact investigations amongst MSM are challenging.

**Future perspectives**

**Recommendations for clinical diagnostics**

We suggest using molecular-based methods for the detection of Shigella spp. and EIEC in fecal samples in routine clinical diagnostics. These methods are faster, cheaper and more accurate and sensitive than detection with culture methods [6, 39]. The ipaH gene is the best target for this, because of its multiplicity nature and its unique presence in Shigella spp. and EIEC [40, 41]. Despite the usefulness of the ipaH gene for detection, it is not suitable for distinction between these species. Therefore, we suggest additional targeting of the O-antigen genes of Shigella spp. for this purpose (chapter 4). Although in chapter 6 we demonstrated that the sensitivity of direct detection of the Shigella O-antigen genes is limited, the specificity for S. sonnei and S. flexneri is high, and useful to guide culture-based approaches. The limited sensitivity does not influence the limit of detection, as for this the ipaH gene is used. After the direct molecular screening, guided culture should be performed for resistance profiling and because this is currently required according to public health guidelines. Guided culture increases the number of obtained isolates compared to unguided culture [6, 39]. After obtaining an isolate, we advise to use MALDI-TOF with commercially available databases for identification to confirm the identity up to the level of Shigella spp. / E. coli. Since further distinction is not possible with MALDI-TOF (chapter 5), identification should be completed by PCR targeting the ipaH gene and a few phenotypic tests as motility, LDC, gas formation and indole in conjunction with Shigella serotyping using polyvalent antisera [18]. If a higher resolution of typing is desired, for instance for surveillance purposes, the isolate should be sent to a local reference laboratory that performs WGS analysis.

Using WGS analysis for routine identification and in silico typing is possible, but not feasible to perform in most MMLs in the Netherlands due to high costs and longer turnaround times. However, larger MMLs in for instance university medical centers often have their own sequence facility, and perform fast and accurate WGS in a routine setting. If WGS data is available and isolates from these MMLs were sent to health authorities for surveillance purposes, we suggest that the WGS data is also shared. It is a waste of public funds if the same WGS data is generated twice only for different research questions. For this purpose, an online platform should be designed in which sequences of pathogens and clinical data of patients can be shared in a safe environment without compromising the patients’ privacy.

**Consequences for public health approaches**

Our multicenter cross-sectional study into the differences in incidence, epidemiology, clinical implications and impact on public health of infections with EIEC compared to infections with Shigella spp., does not provide evidence that justify the absence of control guidelines for EIEC infections, while these are in place for shigellosis. Moreover, this thesis (chapter 6) and other studies confirmed the pathogenic potential of EIEC [42, 43] or their ability to cause
food related outbreaks [44-47]. Based on this biological evidence, we suggest reconsidering the distinct public health approach for cultured *Shigella spp.* and EIEC. Conveniently, if public health actions for infections with *Shigella spp.* and EIEC would be synchronized, the challenging distinction of the two organisms (chapter 2 to 5) will become redundant, providing more straightforward diagnostics and subsequent application of public health guidelines for MMLs and public health authorities.

In addition, our multicenter cross-sectional study IBEES demonstrated that culture-negative cases do not suffer from a less severe clinical picture than culture-positive cases and the degree of secondary infections and health care usage is similar (chapter 6). Furthermore, most of the differences in culture-negative and culture-positive cases were predominantly observed with regard to risk factors, supporting earlier observations from two smaller similar Australian studies [48, 49]. Moreover, other research groups demonstrated, using case control studies, that the presence of the *ipaH* gene in fecal samples was associated with diarrhea [7, 43, 50]. In some endemic areas, only presence of high concentrations of the *ipaH* gene was associated with diarrhea, suggesting that a cut-off based on locally calibrated Ct-values should be used in those areas [51]. In addition, it was shown by using shotgun metagenomics that the sequence composition and quantity of *Shigella spp.* in culture positive cases is similar to that of culture negative but *ipaH* gene positive cases [52]. Finally, the case definitions for shigellosis of the European Union (EU), the United States of America (USA) and Australia recently incorporated positive molecular detection as a probable case. In Australia, these probable cases are obligatory notifiable to health authorities, while the guidelines of the EU and the USA allow the interpretation of notification conditions to their member countries or states [53-55]. The biological evidence provided by this thesis and other studies mentioned above, does not support the current case definition that requires isolation of *Shigella spp.* for the employment of public health measurements. We suggest reconsidering the case definition and notification obligation for shigellosis in the Netherlands, and incorporating molecular detection of *Shigella spp.* (and EIEC) into the guidelines.

From this thesis (chapter 8) and other surveillance studies, it became clear that in particular monitoring of isolates associated with MSM or travel is essential for public health purposes [28, 56, 57]. We suggest to install laboratory surveillance consisting of WGS typing of isolates, to be able to trace the (inter)national circulating *Shigella spp.* or EIEC isolates and perform outbreak detection.

**Guidance for further research**

More research should be done in the field of taxonomy of *Shigella spp.*, *E. coli* and EIEC in particular. In this thesis, options for reclassification were explored (chapter 2). However, more isolates per species should be sequenced and analyzed using genomic taxonomy in future studies, to correct for intra-species variation. Second, to validate the separate species status of all *Escherichia* species, the genomes of the typestrains of *E. albertii* and *E. hermannii* should be sequenced as these are not available in public databases yet. With data from these proposed taxonomic studies complemented to chapter 2, a formal request of opinion for reclassification of *Shigella spp.* should be submitted to the Judicial Commission of the International Committee on Systematics of Prokaryotes to support one of the options explored in chapter 2 of this thesis.

For diagnostics, future studies should focus on the use of sequencing techniques. First, using GWAS as described in chapter 7, associations of genetic content with genus allocations were demonstrated. This presence was not further explored because it was not our scope at the time. However, Random Forest methods associating presence or absence of multiple genes or k-mers based GWAS can be explored for improving the distinction of the species of *Shigella* or EIEC. Second, genome sequences of *Shigella spp.* and EIEC are routinely used for surveillance purposes, but not for identification. Therefore, methods for deriving the *Shigella* species and O-type from WGS data should be validated or developed. Third, more research into metagenomics approaches is required to determine their usability in the diagnostics of shigellosis. Some studies that use metagenomics approaches to detect *Shigella spp.* or to assess the microbiome composition that influences risks for shigellosis were performed, although the similarity of commensal *E. coli* to *Shigella spp.* is a complicating factor [52, 58]. However, even if all our suggestions about the incorporation of molecular techniques and infections with EIEC into shigellosis guidelines would be executed, laboratories still need to perform reflexive or guided culturing. This is because obtaining isolates is necessary for WGS typing for surveillance purposes and outbreak detection, but primarily essential for antimicrobial resistance profiling as multidrug resistance in *Shigella spp.* is a major global public health concern [59]. If typing and resistance profiling of specific pathogens directly from fecal samples using metagenomics, e.g., by using long-read sequencing, will become feasible in the future for routine diagnostic settings, we anticipate a shift in the diagnostics of enteric pathogens, in which the culturing of bacteria will no longer play an important part. Holding on to culture confirmation for public health approaches, has not only a paralyzing influence on the development of molecular diagnostic methods, but also results in an under-estimation of the number of cases.

In addition to diagnostics of the infecting pathogens, also host factors should be more explored in relation to disease severity. It is known that *Shigella spp.* and EIEC manipulate the inflammatory responses of the immune system or the features present in human epithelial cells to be able to access and spread into submucosal sites for infection [2]. Nevertheless, most research on this topic was done using only *S. flexneri*, indicating that more research using other species and serotypes should be performed [2, 60]. In addition, host factors are probably also the answer to questions about the differences in patient outcomes as for instance the development of bloody diarrhea or a more severe clinical
picture. We proved in chapter 7 that specific pathogen associated features of isolates circulating in the Netherlands did not influence these differences nor the underlying disease and medication of the patients. This indicates that other host factors are responsible for diversity in disease outcomes. We speculate that these differences are probably caused by variations in immune responses, as this response plays a significant role in the infection cascade of Shigella spp. and EIEC. Variation in immune responses are influenced by patient dependent factors as well as by variation in the gut microbiome, and both areas need to be further explored [61, 62].

In chapter 6 we showed that the most prominent differences between Shigella and EIEC infections as well as between culture-positive and culture-negative cases were observed in risk factors as traveling, ingestion of contaminated food or water and having MSM contact. This indicates that the focus on future research with regard to optimization of public health guidelines regarding shigellosis should be on the development and application of prevention strategies that efficiently serve the different risk groups.

Furthermore, the use of laboratory surveillance in combination with epidemiological surveillance should be investigated further to determine the added value of WGS in monitoring, cluster detection and outbreak investigation in the surveillance of shigellosis. Currently, a pilot study into laboratory surveillance is conducted during 2019 and 2020 at the Dutch National Institute for Public Health and the Environment in collaboration with the Amsterdam public health service and three MMLs in Amsterdam, the Netherlands. During this pilot study, the added value of laboratory surveillance in addition to epidemiological surveillance is assessed. Additionally, WGS based surveillance can be used to reveal the extent of spillover of Shigella spp. from MSM populations to non-MSM populations. The increase of shigellosis amongst MSM is of growing concern in the Netherlands [63]. Therefore, in addition to spillover of bacteria, also the spillover or exchange of mobile genetic elements that often encode MDR should be monitored. Because the position of these MGE in the bacteria cannot be determined using short-read sequencing, long read sequencing is essential to establish their presence and transmission potential [64].

The guidance for further research into Shigella spp. and EIEC as suggested above, is just a selection of potential beneficiary studies that only focusses on taxonomy, diagnostics, the resulting disease outcomes and public health. The versatility of unexplored areas of interest indicate that still not all aspects about the disease and its pathogen are unraveled almost 2,500 years after the first description of dysentery by Hippocrates, illustrating the fascinating complexity of these perfectly human adapted bacteria.

**Conclusions resulting from this thesis**

- The classification of Shigella spp. as a separate genus is incorrect, proven by genomic taxonomy with all relevant type strains. This incorrect classification is one of the origins for the challenging distinction in diagnostics (Chapter 2).
- Current used diagnostics for Shigella spp. and EIEC in the Netherlands are represented by a large variety of culture dependent and molecular techniques. Nevertheless, this variety is not affecting the qualitative detection (Chapter 3).
- Even after optimization of diagnostic methods, distinction between Shigella spp. and EIEC remains complex and laborious (Chapters 4 and 5).
- No supportive biological evidence was obtained to justify the current dissimilarities in public health approaches or case definitions when comparing infections with EIEC to shigellosis and culture-positive to culture-negative shigellosis cases (Chapter 6).
- Using GWAS, no genetic determinants were identified in Shigella spp. and EIEC isolates from patients that had a predictive value for symptoms or disease severity, and can therefore not accommodate the prioritization of public health actions (Chapter 7).
- Laboratory surveillance of Shigella spp. and EIEC has added value for epidemiological surveillance, as inter(national) clusters are more conveniently detected and traced using WGS (Chapter 8).

