CLINICAL MICROBIOLOGY - REVIEW

Historical, current, and emerging tools for identification and serotyping of *Shigella*

Fatima Bachir Halimeh[1,2] · Rayane Rafei[1] · Marwan Osman[1,3] · Issmat I. Kassem[4] · Seydina M. Diene[2] · Fouad Dabboussi[1] · Jean-Marc Rolain[2] · Monzer Hamze[1]

Abstract

The *Shigella* genus includes serious foodborne disease etiologic agents, with 4 species and 54 serotypes. Identification at species and serotype levels is a crucial task in microbiological laboratories. Nevertheless, the genetic similarity between *Shigella* spp. and *Escherichia coli* challenges the correct identification and serotyping of *Shigella* spp., with subsequent negative repercussions on surveillance, epidemiological investigations, and selection of appropriate treatments. For this purpose, multiple techniques have been developed historically ranging from phenotype-based methods and single or multilocus molecular techniques to whole-genome sequencing (WGS). To facilitate the selection of the most relevant method, we herein provide a global overview of historical and emerging identification and serotyping techniques with a particular focus on the WGS-based approaches. This review highlights the excellent discriminatory power of WGS to more accurately elucidate the epidemiology of *Shigella* spp., disclose novel promising genomic targets for surveillance methods, and validate previous well-established methods.

Keywords *Shigella* spp. · *Escherichia coli* · Identification · Serotyping · Whole-genome sequencing

---

Fatima Bachir Halimeh and Rayane Rafei contributed equally to this work.

Responsible Editor: Beatriz Ernestina Cabilio Guth

Highlights:
- Genetic and phenotypic closeness between *Shigella* and *Escherichia coli* is a diagnostic issue.
- Many molecular algorithms are supposed to differentiate between enteroinvasive *E. coli* and *Shigella.*
- These algorithms need more validation for routine integration in diagnosis.
- Genoserotyping is a promising alternative for standard *Shigella* serotyping.
- Whole-genome sequencing is the future gold-standard method for both *Shigella* diagnosis and surveillance.

✉ Monzer Hamze
   mhamze@monzerhamze.com

[1] Laboratoire Microbiologie Santé et Environnement (LMSE), Doctoral School of Sciences and Technology, Faculty of Public Health, Lebanese University, Tripoli, Lebanon

[2] Aix-Marseille University, IRD, APHM, MEPHI, IHU-Méditerranée Infection, Faculté de Médecine Et de Pharmacie, 19–21 boulevard Jean Moulin, 13385 Marseille CEDEX 05, France

[3] Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14850, USA

[4] Center for Food Safety and Department of Food Science and Technology, University of Georgia, 1109 Experiment Street, Griffin, GA 30223-1797, USA

Overview

Shigella has been known for a long time by its clinical manifestation, "bacillary dysentery," even before its identification by Kiyoshi Shiga as the causative agent during a severe Japanese outbreak [1]. *Shigella* ranks currently as the second leading etiology of diarrhea-associated mortality and is responsible annually for approximately 212,438 deaths, of which 63,713 are children younger than 5 years and 74,402 individuals older than 70 [2]. Although primarily a disease of developing countries, *Shigella* infection, or shigellosis, remains a public health issue across the globe, with nearly 500,000 cases in the USA annually [3]. *Shigella* is

transmitted through the oral–fecal pathway with a low infectious dose of 10 to 100 cells enough to result in disease [4]. Shigellosis is a non-systematic, enteric, and acute infection characterized by colonic epithelium destruction responsible for bloody diarrhea, sometimes accompanied by mucus, abdominal pain, and fever [5]. Occasionally, *Shigella* can lead to invasive infections such as meningitis, osteomyelitis, spleen abscess, and sepsis, occurring mainly in malnourished and HIV patients [6].

Shigella species are intracellular Gram-negative, facultative aerobic, and non-sporulating bacilli belonging to the *Enterobacteriaceae* family [7]. Despite the close relatedness to *Escherichia coli*, the genus *Shigella* and its four species were formally validated in 1954 [8]. Since their discovery, several studies have attempted to classify *Shigella* spp. accurately within the *Enterobacteriaceae* family. In 1982, based on 192 morphological, biochemical, and phenotypic properties, Dodd and Jones determined that *Shigella* spp. fell into a major distinct cluster more closely related to *Yersinia* and *Proteus*/*Providencia* species than to *E. coli* [9]. This phenotypic classification strengthened the traditional separation of *Shigella* as a separate genus. However, with the rolling of the molecular era, many ambiguities have been raised about the precise taxonomical position of *Shigella*, while the dilemma of the relatedness to *E. coli* persisted. *Shigella* spp. and *E. coli* appeared as "one species genetically" with DNA–DNA hybridization (DDH) experiments that revealed 80–90% similarity in identity [10]. After that, the sequencing analysis of eight housekeeping genes grouped *Shigella* into three major clusters, and a limited number of outliers were found to evolve independently from multiple non-pathogenic *E. coli* ancestors [11]. Even with the availability of completely sequenced genomes and the inclusion of either more housekeeping genes in analysis or the whole set of conserved genes (also known as the core genomes) have confirmed the previous notion that *Shigella* spp. were intermixed with *E. coli* [12], [13]. Currently, there is a consensus that *Shigella* spp. belong to the *E. coli* species, but the nomenclature has been kept for historical and medical reasons [14]. However, this consensus is still a contentious issue where some researchers claim that the four species of *Shigella* genus are not clones of *E. coli*, but members of the *Escherichia* genus, at the same footing with *E. coli* [14–16].

Although challengeable in clinical microbiology laboratories, differentiating *Shigella* from *E. coli* can be guided by many distinctive morphological features. Indeed, more than 80% of *E. coli* strains are motile, able to decarboxylate lysine, ferment many sugars, are indole positive, and produce gas from D-glucose. Nevertheless, *Shigella* are non-motile, unable to decarboxylate lysine, do not produce acid from salicin or hydrolyze esculin, ferment few sugars, and do not produce gas from D-glucose, with the exception of *Shigella flexneri* serotype 6 and *Shigella dysenteriae* 3. In addition,

Shigella sonnei strains can ferment lactose slowly and can be mucate positive [17, 18]. However, the real identification "problem" is more evident in differentiating *Shigella* from a group of *E. coli* variants called "inactive *E. coli*," which includes the enteroinvasive *E. coli* (EIEC) pathovar. Indeed, EIEC shares with *Shigella* some biochemical properties, among them: negativity to lactose, immobility, inability to decarboxylate lysine (lysine decarboxylase negative), and absence of gas production [19, 20]. Nevertheless, some exceptions are present in the case of EIEC wherein some strains for instance those belonging to serotype O124:H30 are mostly mobile, decarboxylate lysine, and can also ferment lactose [21–23]. Besides, *Shigella* spp. and EIEC might also be considered as a single pathotype based on several reasons, including (1) the acquisition of a similar virulence plasmid (pINV) that mediates invasion of host cells; (2) intracellular survival by inactivation of sets of genes as an adaptive mechanism; and (3) the frequent assumption that EIEC is evolving to a *Shigella*-like phenotype, which is reflected in the similar patterns of gene expression [17, 18, 24, 25]. Despite having similar disease symptoms, EIEC is generally less virulent than *Shigella* with a higher infectious dose and reduced potential for propagated person-to-person transmission [26–28]. While *Shigella dysenteriae* type 1 can only cause HUS (hemolytic uremic syndrome) among *Shigella*, EIEC is not known to trigger this syndrome [18, 29]. Besides, EIEC can also be differentiated from *Shigella* by a minimal number of tests, including motility, mucate, salicin fermentation, esculin hydrolysis, and the combined positivity of indole production and gas formation from D-glucose, and acetate and Christensen citrate utilization. EIEC isolates may be positive for one or more of these tests, but *Shigella* are generally negative [19, 30, 31]. Subsequently, a guide that combined biochemical, physiological, and serological features was designed for the daily identification of EIEC, *E. coli*, and *Shigella* in diagnostic laboratories [19].

Amid the technological evolution, an accurate technique able to differentiate between *Shigella* spp. and *E. coli*, especially EIEC, continues to be a significant diagnostic challenge due to their genetic and phenotypic relatedness as indicated above. The distinction is paramount because Shigellosis is a mandatory notifiable disease in most countries, whereas EIEC is not [32]. The correct identification will also elucidate the epidemiology of *Shigella* spp. and their trends of developing antimicrobial resistance, which will facilitate treatment regimens based on different antimicrobial susceptibility profiles [33]. In addition, the differentiation between *Shigella* spp. allows for a better understanding of each species' unique epidemiology such as the prevalence of *S. flexneri* in low- and middle-income countries versus *S. sonnei*, which is more dominant in high-income countries [6]. The correct identification at the serotype level is a cornerstone for determining the spatial–temporal distribution

of circulating serotypes, understanding the differences in disease burden across countries, and tracking the potential emergence of novel serotypes, investigating outbreaks, and critically evaluating implemented policies for vaccine development and disease containment. Here, we discuss both phenotypic and molecular identification techniques used to differentiate *Shigella* from EIEC and to identify *Shigella* at the species level (Table 1, Fig. 1). Subsequently, different countries and laboratories can choose an identification method that suits diagnostic capabilities. This review also highlights inherent loopholes in *Shigella*'s phenotypical serotyping strategy and summarizes the proposed molecular serotyping alternatives.

Identification techniques

Phenotypic identification techniques

Biochemical test systems

Commercial biochemical identification systems are based on one of five different technologies or a combination thereof: pH-based reactions, utilization of carbon sources, enzyme-based reactions, visual detection of bacterial growth, and/or detection of volatile or non-volatile fatty acids[47]. There are multiple tests dedicated to *Enterobacteriaceae* identification. These tests are categorized into manual such as API 20E, RapiD 20E (BioMérieux, Marcy-l'Étoile, France), RapID ONE, and Micro-ID (Remel, San Diego, California, USA); and automated such as BD Phoenix 100 ID/AST system NID panel (Becton Dickinson, New Jersey, USA), Vitek 2 (BioMérieux), and MicroScan Neg ID Type 2 (Beckman Coulter, California, USA). In terms of their effectiveness in identifying *Shigella*, Api 20E that has been largely accepted in the last decades in the clinical microbiology laboratories failed to identify 3 to 10% of *Shigella* strains [47–49], while BD Phoenix misidentified nearly 17% of *Shigella* isolates and defined them as *E. coli* [36, 50, 51]. As for Vitek 2, it repeatedly misidentified a commensal inactive *E. coli* as *S. sonnei* [52]. However, the evaluation studies' reliability required more analysis because the tested identification system (that is conventional or commercial biochemical methods) was generally questionable regarding its ability to separate *E. coli* and *Shigella* [51, 53].

Serotyping

The four *Shigella* species (subgroups) are divided into serotypes and subserotypes based on their O antigen. *S. dysenteriae* (subgroup A) has 15 serotypes, *S. flexneri* (subgroup B) has 18 serotypes, *Shigella boydii* (subgroup C) has 20 serotypes, and *S. sonnei* (subgroup D) has a single serotype

[4]. Notably, a confirmed identification of *Shigella* spp. must be based on both serological and biochemical profiles [54]. Traditionally, serotyping was performed using *in house* or commercial antisera to LPS O-antigen that are divided into polyvalent and monovalent antisera. The polyvalent antisera contain antibodies for multiple *Shigella* serotypes and can subsequently determine *Shigella* subgroups, while the monovalent antisera contain serotype-specific antibodies [54]. Among the most common commercialized *Shigella* serotyping antisera kits are Wellcolex Color *Shigella* Kit (Thermo Fisher Scientific Inc., Massachusetts, USA) and Vision Polyvalent *Shigella* Antisera (ProLab Diagnostics Inc., Ontario, Canada) providing polyvalent antisera. Meanwhile, some companies provide both polyvalent and monovalent antisera such as BioRad Laboratories Inc. (California, USA), Deben Diagnostics Ltd. (Ransomes industrial estate, UK), and Denka Seiken Co., Ltd. (Tokyo, Japan). While serotyping was admittedly regarded as the gold standard for *Shigella* species identification [36], it is considered laborious, time consuming, and impractical for a large number of samples. Also, additional issues lessen the usefulness of such an approach. First, many intra- and inter-species cross-reactions are observed, and commercial antisera are ideally 91% accurate [55]. During a cohort test, 28% of *S. sonnei* isolates were misidentified by conventional serotyping techniques, and additional tests such as PCR analysis of *ipaH* and *lacY* genes or repeated serotyping were used to resolve this discrepancy [36]. Indeed, inherent similarities between *E. coli* and *Shigella* O-antigens hinder the reliability of serotyping. Of 34 distinct O antigens identified in *Shigella*, 21 are identical or very similar to those described in *E. coli* [18]. EIEC O112ac is similar or identical to *S. dysenteriae* 2/*S. boydii* 15/*S. boydii* 1, EIEC O124 to *S. dysenteriae* 3/provisional *Shigella* serovar 3615.53, EIEC O136 to *S. dysenteriae* 3/*S. boydii* 1, EIEC O143 to *S. boydii* 8, EIEC O152 to provisional *Shigella* serovar 3341:55, EIEC O135 to *S. flexneri*, and EIEC O164 to *S. dysenteriae* 3. Second, occasional provisional *Shigella* serovars, which are biochemically indistinguishable from *Shigella* spp. but fail to agglutinate with standard commercial antisera, are problematic [56, 57]. This may be due to morphologic transition from smooth to untypable rough strains without O antigens, accounting for 6 to 10% of annual *Shigella* cases in the USA [55]. The presence of capsular antigens may also prevent *Shigella* strains from reacting with the antisera [54]. In addition, the emergence of novel and atypical serotypes able to escape host immunity responses can also be explained by serotype conversion phenomena mediated by either temperate bacteriophages or plasmids carrying serotype encoding genetic elements [58–60]. The diagnosis techniques for these non-serotypeable *Shigella* are discussed further below. Third, distinct connections between biochemical features, serotypes, and phylogenetic relationships are

Table 1 Summary of the most frequently used methods for *Shigella* identification

| Technique		Target	Ability to distinguish between *Shigella* and *E. coli*	References
Phenotypic	Biochemical tests	The enzymatic activity and the ability to ferment some sugars	Inability to distinguish between *Shigella* spp. and other genera, *Hafnia*, *Providencia*, atypical *E. coli*, and non-lactose fermenting or anaerogenic *E. coli*	[34]
	Serology	The O antigen present in the surface of bacteria	Occurrence of intra- and inter-species cross-reactions, occasional provisional *Shigella* serovars, serotype conversion, and untypable strains, not illustrative of phylogenetic relationships	[35]
	MALDI-TOF MS	The spectra of ribosomal proteins signature	Inability of conventional MALDI-TOF MS assays to discriminate between *Shigella* and *E. coli*. Advanced software or customized reference library could distinguish inactive and other non-lactose-fermenting *E. coli* from *Shigella* spp.	[36–38]
Molecular	PCR techniques	*ipaH*, and *lacY* *uidA*, and *lacY*	Differentiation between *Shigella* (*ipaH*+, *lacY*−), EIEC (*ipaH*+, *lacY*+), non-invasive *E. coli* (*ipaH*−, *lacY*+); however, recent data demonstrate low performance Differentiation between *Shigella* (*uidA*+, *lacY*−), and *E. coli*, including EIEC (*uidA*+, *lacY*+); however, recent data demonstrate low performance	[27, 39, 40]
	Single locus sequence-based	*16S rRNA* gene	Inability to distinguish *E. coli* from *Shigella* spp. due to the high sequence similarity (>99%)	[41]
		rpoB gene	Inability to distinguish *E. coli* from *Shigella* spp. due to the high sequence similarity (>99%)	[42]
		gyrB gene	Capability to differentiate between *E. coli* and *Shigella* spp., but further large validations are needed	[43]
	WGS	K-mers	Accurately differentiate *E. coli* from *Shigella* to the species level, but some EIEC isolates can be misidentified as *S. flexneri* or *S. boydii*	[44]
		SNP markers	Accurately differentiate EIEC from *Shigella* spp. with typing ability among strains of the same species	[27]
		Average of nucleotide identity	Most of the ANI-based delineation tools are unable to delineate *Shigella* spp. from *E. coli*. Notably, the GGDC web tool with the ANI-f1 formula showed some power	[45]
		Cg-MLST: conserved *loci* shared between *Shigella* and *E. coli* species	The cgMLST-based clustering differentiates between *E. coli* and *Shigella* spp. with high-resolution capabilities	[46]
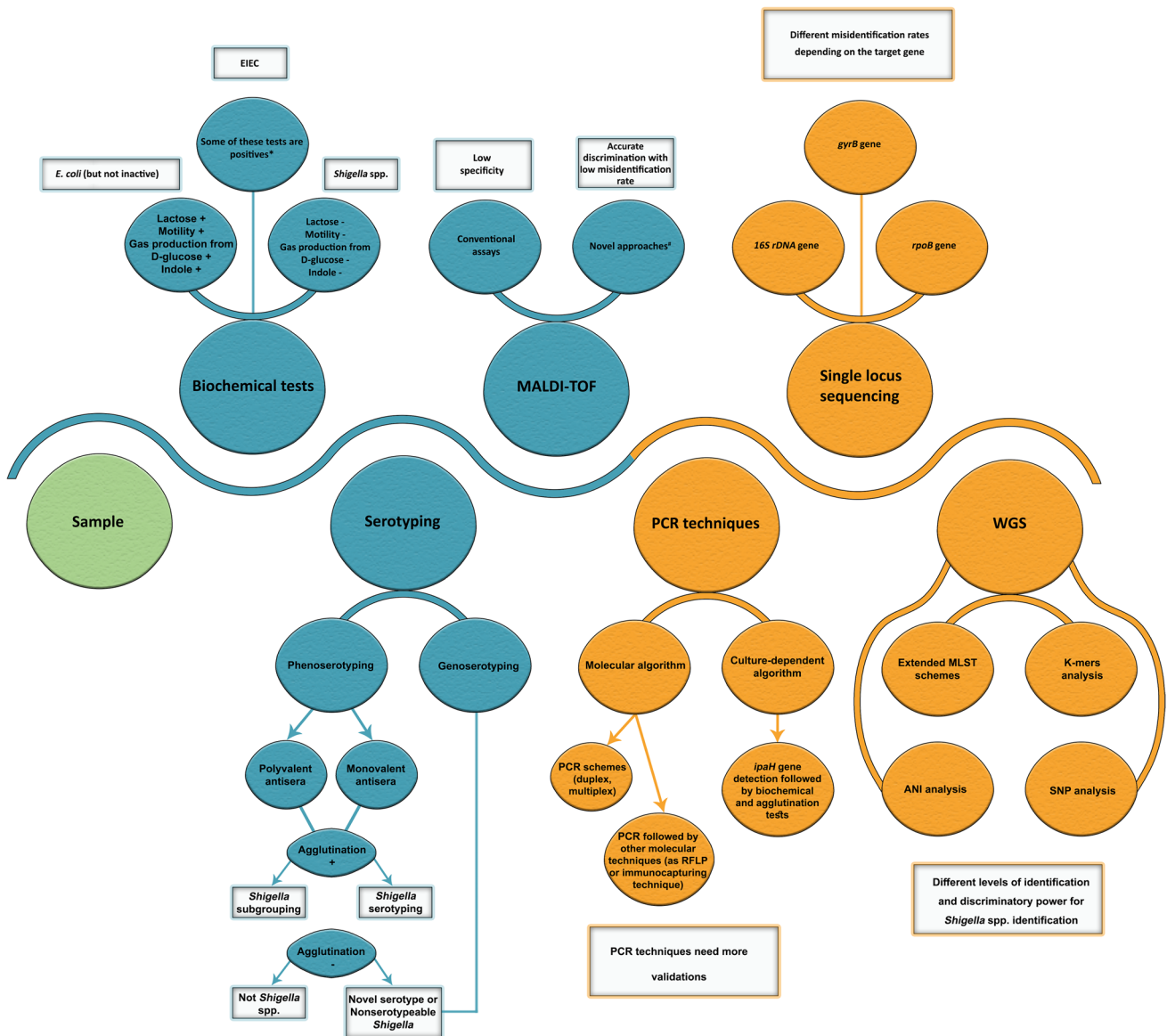
Fig. 1 *Shigella* identification methods and strategies. Blue: phenotypic techniques; yellow: molecular techniques. *EIEC can be differentiated from *Shigella* by a number of tests, which include motility, mucate, and salicin fermentation, esculin hydrolysis, the combined positivity of indole production and gas formation from D-glucose, and acetate utilization. #Novel approaches of MALDI-TOF include ClinProTools, referenced library, and short-term lactose incubation

not readily notable. For example, phenotypic variability observed within a particular serotype rises when increasing the testing of isolates, while the presence of serotypes that are genetically and serologically related but unrelated phylogenetically, falling into distinct *Shigella* clades, convolute further the usefulness of serotyping [57, 58].

MALDI-TOF MS

Matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) has recently been recognized as a rapid, cost-effective, high-throughput, and reliable microbial identification tool with broader applicability to a large spectrum of microorganisms [37]. Despite this versatility, conventional MALDI-TOF assays using MALDI-TOF Biotyper (Bruker Daltonics, Bremen, Germany), and Vitek 2 MS systems (bioMérieux, Marcy l'Etoile, France), failed to distinguish *Shigella* spp. from *E. coli* due to the high degree of similarity between their spectra [36, 61]. However, studies suggesting the use of a specialized automated algorithm (ClinProTools) or customized reference library reflecting the genetic diversity of *Shigella* and *E. coli* outperformed routine MALDI-TOF assays by enabling accurate discrimination between *E. coli* and *Shigella* with

misidentification rates reaching approximately 3% [36, 38]. Recently, an approach merging biochemical methods and the MALDI-TOF assays seems to be interesting. By adding a short-term incubation in a high-lactose fluid medium before MALDI-TOF analysis, Ling et al. identified seven novel differential MS peaks serving as biomarkers to reliably identify these related bacteria with nearly 98% accuracy [62].

Molecular identification techniques

PCR-based identification techniques

In the last decades, several PCR-based identification techniques were developed to differentiate between *Shigella* spp., *E. coli*, and EIEC [39, 40]. However, PCR development faced challenges in selecting appropriate targets that allow an accurate differentiation between these targets, where some PCR assays were unable to separate *Shigella* from EIEC [63, 64]. Generally, PCR identification schemes for *Shigella* often target plasmid virulence genes such as *virA*, *ial*, *she*, and *tuf*, which are vulnerable to horizontal gene transfers; potentially leading to false-positive and false-negative results [65]. One of the primary gene targets commonly integrated into PCR schemes is *ipaH*, a multicopy gene encoding a virulence factor and located on both the chromosome and the large invasion plasmid pINV, which is exclusively found in *Shigella* and EIEC isolates [65, 66]. Recently, *ipaH* amplification was used as the first step in two different algorithms proposed for the *Shigella* identification to differentiate between the *Shigella*/EIEC (*ipaH+*) and non-invasive *E. coli* (*ipaH−*). The culture-dependent algorithm was followed by profiling phenotypical, biochemical, and serological features. In contrast, the molecular algorithm targeted additionally the *wzx* genes of *S. sonnei* phase I, *S. flexneri* serotypes 1–5, *S. flexneri* 6, and *S. dysenteriae* serotype 1 [57]. After analysis with whole-genome sequencing (WGS), the culture-dependent algorithm succeeded in identifying 100% of *S. dysenteriae*, *S. sonnei*, and non-invasive *E. coli* isolates, but only 85% of *S. flexneri*, and 93% of *S. boydii* and EIEC. While the molecular algorithm fully identified all targeted species or serotypes, it could not precisely detect the *ipaH* positive serotypes, with none assessed *wzx*, and binned them into a single group as either EIEC, *S. boydii*, *S. sonnei* phase II, or *S. dysenteriae* serotypes 2–15 [57].

To differentiate *Shigella* from EIEC, one PCR scheme amplifies, in addition to *ipaH*, a lactose permease encoding gene (*lacY*) present in *E. coli* including EIEC [40]. This scheme enabled the differentiation of EIEC O121 and O124 groups and *Shigella* but could not classify EIEC O164 group consistently. Another scheme delineates *Shigella* from *E. coli*, including EIEC, by targeting the β-glucuronidase-encoding gene (*uidA*) commonly found in *E. coli* and *Shigella* spp. and *lacY* which is specifically observed in *E. coli*

strains [39]. However, the accuracy of this PCR approach appeared not to be as excepted. While it correctly identified in silico 100% of *S. sonnei*, it failed to define 8% of *S. flexneri*, 14% of *S. boydii*, 20% of *S. dysenteriae*, 23% of non-invasive *E. coli*, and 38% of EIEC isolates [27]. Moreover, the utility of *lacY* can be questioned because while *S. flexneri* and *S. boydii* lack the *lac* genes (*Y*, *A*, and *Z*), other *Shigella* spp. possess some *lac* genes. *S. dysenteriae* has *lacA* and *lacY* and *S. sonnei* has all the *lac* genes. However, they cannot ferment lactose due to the lack of permease activity [67]. Furthermore, after 4-h enrichment of the sample in a growth medium, a conventional pentaplex PCR could identify *Shigella* at genus level and differentiate between *S. flexneri*, *S. sonnei*, and *S. dysenteriae* by amplifying the specific targets *invC*, *rfc*, *wbgZ*, and *rfpB*, respectively, with an internal control (*ompA*) [68]. Notably, most of the targets in this pentaplex PCR are located on mobile elements, making them vulnerable to horizontal gene transfer, and thus limiting the usefulness of such a PCR scheme in identification. Interestingly, a new proposed phylogenomic-based multiplex PCR assay by Sahl et al. was able to identify unknown *Shigella* isolates and classify them into appropriate phylogenetic clades [69]. However, when the primers were tested on a considerable genetically diverse isolate collection, they could not phylogenetically differentiate *Shigella* [27]. To override the issue of targeting plasmid virulence genes, Kim et al. designed primers targeting novel genetic markers identified through comparative genomics that can differentiate *Shigella* from diarrheagenic *E. coli*, including EIEC, and identify the four *Shigella* spp [65]. Additional steps can be added to PCR such as restriction fragment length polymorphism (PCR–RFLP) [70] or an immunocapturing technology [71] to increase either the sensitivity and/or specificity of detection. However, some of these methods are relatively expensive, technically demanding, and require special equipment which complicate application as diagnostic or epidemiological tools.

Single locus sequence-based identification techniques

16S rRNA gene sequencing Although representing 0.1% of the coding part of a microbial genome, 16S rRNA gene sequencing has been recognized as a highly useful tool in bacterial classification and has been widely used to provide genus and species identification for isolates. However, its usefulness is impaired by its low discriminatory power and poor resolution to distinguish between closely related bacteria [43, 72]. The reported 16S rRNA gene sequence similarities between *E. coli* and *Shigella* spp. exceed 99%; reaching up to 99.8% with *S. flexneri*, 99.9% with *S. sonnei*, and 99.7% with *S. boydii* [43]. Subsequently, 16S rRNA gene sequencing is not considered a reliable tool for differentiating between *E. coli* and *Shigella* spp. because they were

intermingled together in the 16S rRNA gene–based phylogenetic tree. Using Sanger sequencing, only 26.7% of the *E. coli* strains were correctly identified, compared to 33.3% as *S. sonnei* and 40% as *S. dysenteriae* [41]. This was achieved using a species finder, a web-based tool for prokaryotic species identification based on the similarity of 16S rRNA gene sequences (https://cge.cbs.dtu.dk/services/SpeciesFinder/) with the known reference sequences available at the center of genomic epidemiology. However, the latter revealed a notably inadequate performance in comparison to Kmer-Finder (another in silico tool, discussed below) and *gyrB* sequence analysis because only 74% of non-serotypeable *Shigella* were reliably identified to the species level [43].

***rpoB* sequencing** Being a single copy protein-encoding housekeeping gene, *rpoB* can be more advantageous than 16S rRNA gene in microbial identification. While *rpoB* is deemed a high-resolution marker able to reveal molecular variation down to the population level, it has an overlapping similarity between closely related isolates such as *Shigella* and *E. coli* [73]. The *rpoB* sequence similarities between *E. coli* and *Shigella* spp. exceed 93% [73], reaching up to 99.8% with *S. flexneri*, 99.4% with *S. sonnei*, and 99.78% with *S. boydii*. However, Devanga Ragupathi et al. revealed that *rpoB* and another housekeeping gene malate–lactate dehydrogenase (*mdh*) accurately identified *Shigella* and different *E. coli* virotypes [67].

***gyrB* sequencing** Compared with the 16S rRNA gene, *gyrB* that encodes the β subunit protein of the DNA gyrase (Topoisomerase Type II) seems to have a more significant evolutionary divergence, with an ability to distinguish between closely related species. With regards to *E. coli* and *Shigella* spp., the *gyrB* similarity percentages between *E. coli* and either *S. sonnei*, *S. flexneri*, or *S. boydii* were 98.1, 97.8, and 98%, respectively, being lower than those obtained with 16S rRNA gene analysis, which hints seemingly for the comparative accurateness of *gyrB* gene sequence analysis [74]. Many studies revealed an outperformance of *gyrB* in comparison to 16S rRNA gene sequencing [43, 75, 76], where the identification results of *gyrB* sequencing were highly congruent to the KmerFinder tool with 100% identification of non-serotypeable isolates at the species level [43].

Whole-genome sequencing Cumulative data generated from many genetic and intergenic regions could provide a more in-depth resolution of the identity of an isolate. Therefore, decoding bacterial genomes via WGS is an up-and-coming desirable technology, particularly with its precipitously decreasing cost, and is predicted to replace conventional microbial diagnostic workflow and become a public health resource for global surveillance [77]. WGS can be followed by multiple analyses to identify, serotype, classify, and even

understand their pathogenesis *Shigella* spp. Numerous enticing WGS-based approaches were assessed concerning their ability to differentiate *Shigella* and *E. coli* including K-mers, whole-genome single nucleotide polymorphism (SNP), and average nucleotide identity (ANI) [27, 44, 45]. Notably, the common limitation hindering the complete integration of WGS, especially in low-income countries, is the investment requirement (equipment, reagents) of the WGS platforms and experts in bioinformatic analysis.

K-mers-based approaches

K-mers-based species identification tools (such as KmerFinder the online tool available at https://cge.cbs.dtu.dk/services/KmerFinder/) split the WGS data of unknown isolate into relatively short oligomers of a defined length k, then compare the resulting content of k-mers to a set of k-mers derived from a collection of reference genomes [43, 44, 78]. The similarity between the query and reference sets is expressed as a percentage value indicating the portion of common kmers. The kmer-based identification predicted 98.4% of 1982 *Shigella* and *E. coli* isolates in agreement with traditional biochemistry and serology schemes. The 25 discrepant results revealed either the superiority of kmer approach over the traditional schemes when an non-functional O antigen biosynthesis genes in *S. flexneri* could conventionally misidentify them as *S. boydii*, or the inferiority of kmer notably for 10 EIEC isolates misidentified as *S. flexneri* or *S. boydii* by the kmer-derived identification [44].

SNP-based approaches

SNP-based approaches catch only informative genetic signatures in both gene-encoding and intergenic regions, thus omitting the inclusion of genetically conserved meaningless data [79]. Therefore, SNP, generally considered stable and reproducible molecular markers, can provide additional strain differentiation at a thorough level, which is ultimately essential for outbreak investigation and surveillance strategies of important pathogens such as *Shigella* spp [27, 80]. Beyond the typing scope, SNP analysis can also draw the true phylogeny of *Shigella* spp. and decipher their enigmatic relations with EIEC [27]. In addition, SNPs are valuable markers for developing rapid, accurate, and discriminative diagnostic methods [80, 81]. Based on in silico analysis of eight *Shigella* genomes, 24 informative SNPs were selected from nine genes (*gapA*, *lpxC*, *sanA*, *thrB*, *yaaH*, *ybaP*, *ygaZ*, *yhbO*, and *ynhA*) and were found useful in identifying *Shigella* spp. as well as providing some resolving power among individual strains within the same species [81]. When analyzing a comprehensive genomic collection of *Shigella* and EIEC, Pettengill et al. revealed that their phylogenetic profile does not resemble their distinct genera designation [27].

Besides, they identified a panel of 254 SNP markers able to classify EIEC and *Shigella* isolates into phylogenetic clades from WGS data, rather than classifying them into genus and species [27].

ANI-based approaches

The average nucleotide identity (ANI) between two genomes has been suggested as a valid alternative to the wet-laboratory DDH methods for species delineation. Genomes can be defined as members of the same species if sharing ≥ 95–96% ANI [82]. Various ANI-based approaches and software are currently available to compare the genomes in silico [45]. To reduce the high computational requirements of ANI-based approaches, a novel method known as the whole genome parameter (WGP) was proposed for the delineation of bacterial genomes using four statistical parameters calculated from numerical representations of whole bacterial genomes (phase signal and cumulated phase signal) [45]. However, when these methods were tested for their ability to delineate *Shigella* spp. from *E. coli*, the majority including the WGP failed because the tested *Shigella* isolates yielded high similarity values with *E. coli* (~ 97.84%) above the delineation species threshold of WGP that is set at 96% [45]. Although these results mirrored the inability of traditional DDH to separate *Shigella* and *E. coli*, the usefulness of the ANI approach must be validated on a large diverse sample of *E. coli* and *Shigella* spp. rather than a small sample [45, 83, 84]. Notably, the Genome-To-Genome Distance Calculator (GGDC) web tool with the ANI-f1 formula, one of the tested ANI-based delineation tools, showed some power in differentiating *E. coli* from *Shigella* spp. and in generating ANI-f1 values under the species delineation threshold (70%) for most of the comparisons between *E. coli* and *Shigella* strains [45].

Extended MLST schemes

Multi locus sequence typing (MLST) is a sequence-based genotyping technique based on sequencing several housekeeping genes. Three MLST schemes that were developed originally for *E. coli* have also been applied to *Shigella*; the Achtman scheme includes seven housekeeping genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh recA*, and *purA*), while the Pasteur scheme contains eight genes (*dinB*, *icdA*, *pabB*, *polB*, *putP*, *trpA*, *trpB*, and *uidA*), and the Whitmann scheme targets 15 genes (*arcA*, *aroE*, *aspC*, *clpX*, *cyaA*, *dnaG*, *fadD*, *grpE*, *icdA*, *lysP*, *mdh*, *mtlD*, *mutS*, *rpoS*, *uidA*) [85–87]. Although sequence types (ST) were assigned by MLST schemes regardless of the species identity as either *E. coli* or *Shigella*, categorizing isolates into ST seems to mirror *Shigella* classification [44] where the majority of the isolates within the same species had closely related STs belonging to the same clonal complexes (CCs). Chattaway et al. suggested the combined use of kmer and MLST to differentiate *E. coli* from *Shigella* [44]. However, some CCs can encompass many species as in the case of CC288 that is membered by *S. boydii* and *S. dysenteriae* isolates [44].

Thanks to the advent of WGS, the MLST schemes with usually seven genes can be extended to encompass more *loci* distributed over the chromosome (i.e., WgMLST or Whole genome MLST), or conserved *loci* shared among most of the isolates of the same species (cgMLST or core genome MLST). Undoubtedly, these new MLST facets give a more in-depth genomic insight and a higher resolution than the conventional MLST, especially for closely related bacteria [46, 79, 88]. For *Shigella* and *E. coli*, the same cgMLST (targeting 2513 core genes) and wgMLST (targeting 25,002 coding genes) schemes are available at the publicly accessible Enterobase database (http://enterobase.warwick.ac.uk/species/index/ecoli). BioNumerics proposed another wgMLST scheme based on Enterobase but with modifications (17,350 target genes—2513 core genes and 14,837 accessory genes) for both *Shigella* and *E. coli* [89]. These recent schemes are mostly used for epidemiologic investigations of *E. coli* and rarely for typing *Shigella* [90–93.] In addition to their promises in typing, the cgMLST technique demonstrates a capability in resolving the discrepancies raised between the culture-dependent and molecular-dependent algorithms proposed by Van den Beld et al. This is done by configuring the cgMLST-based clustering of inconclusive isolates with reference strains [57]. However, the ability of cgMLST in species allocation requires additional investigation. While cgMLST clustered most of the *Shigella* and EIEC genomes according to their species, it formed some clusters with mixed-species due to their deviating phenotypic features [94].

Genoserotyping

The issues mentioned above of phenotypic serotyping spurred the development of several molecular techniques allowing the detection and characterization of isolates at the genetic level, regardless of whether the genetic material was expressed. Monitoring disease burden requires fast and high-throughput methods that facilitate identification and surveillance at the serotype level. Generally, molecular serotyping techniques are considered as fast methods generating a deluge of objective information in a relatively short period because of their high-throughput capabilities. Although WGS will complement or replace *Shigella*'s conventional serotyping soon, it is not ready for routine use in most clinical microbiological laboratories.

In brief, molecular serotyping was firstly applied to *Shigella* by Coimbra et al., proposing the use a restriction method (*rfb*-RFLP) with the enzyme MboII of an amplified

region that harbors O-antigen encoding genes (known as the *rfb* cluster) to decipher the serotype-specific *rfb* polymorphism [95]. This technique had shown a closer resolution to traditional serotyping scheme generating discernible O-antigen patterns for each serotype except for *S. boydii* 12, which showed two distinct patterns, and *S. flexneri* serotypes 1–5, X and Y, which all gave the same indistinguishable pattern [95]. A dynamic software (Molecular serotyping tool) was then developed to ensure a quick identification at the serotype level and compare the *rfb*-RFLP patterns of clinical isolates to those in a database encompassing profiles of 171 previously known *Shigella* and *E. coli* [96].

Furthermore, many multiplex PCR schemes have been established for *Shigella* serotyping as quick and affordable methods, especially for *S. flexneri*. Sun et al. developed a single tube multiplex PCR assay with eight sets of primers targeting O-antigen synthesis and modification genes, which allowed the identification of 14 out of 15 serotypes of *S. flexneri* (except serotype Xv) with a high agreement (97.8%) with traditional slide agglutination methods [97]. This conventional PCR was also upgraded to a real-time version [98]. Evaluation studies proved its full correlation with WGS and its outperformance over traditional methods as discrepancies between phenotypic and genotypic techniques were attributed to the presence of novel genotypes, non-specific cross-reactions, or genetic modifications in O-antigen synthesis or modification genes [99]. In addition, two other multiplex PCR assays could efficiently determine the 19 serotypes of *S. flexneri* recognized so far, where "PCR A" defined serotype genes and "PCR B" identified serotype 7 specific genes and group antigenic factors genes [100]. To resolve the PCR multiplex-associated problems, particularly the differentiation between similar-sized bands, Li et al. developed a DNA microarray able to simultaneously detect 34 distinct O-antigen *Shigella* forms with high sensitivity and specificity [101]. However, these methods that rely on O-antigen specific biosynthetic genes must be complemented with biochemical tests for a reliable differentiation because many *Shigella* serotypes share identical O-antigen with commensal *E. coli*, in addition to the high level of observed recombination among serotype-specific genes mostly encoded on mobile genetic elements [35, 58]. WGS provides new insights into the *Shigella* phylogeny that has never been tackled before. By performing in silico molecular serotyping based on Sun et al.'s (2011) scheme, Connor et al. revealed that the serotype weakly predicted the phylogenetic relationships between strains of *S. flexneri*, where each of the seven identified phylogenetical groups encompassed two or more serotypes [58, 97]. In addition, WGS places the whole genetic repertoire under scrutiny and offers the ability to interrogate many genes simultaneously. Indeed, analyzing a sole genetic marker could mislead the identification at both species and serotype levels due to the considerable genomic variability of the individual genomic targets [55].

WGS could maintain backward compatibility with historical data by providing a framework for in silico genome-derived serotyping, along with its ability to identify novel serotypes. For example, after an in-depth examination of 259 *Shigella* genomes belonging to 53 serotypes, Wu et al. have recently developed an automated pipeline, ShigaTyper, able to quickly identify and predict 59 serotypes from Illumina paired-end reads with high accuracy (98.2%) [55]. Likewise, Ventola et al. proposed two novel tools to be implemented in the National Reference of *Salmonella* and *Shigella* of Belgium for *Shigella* surveillance. The first tool consisted of a cost-effective Luminex assay based on a modular multiplex oligonucleotide ligation-PCR procedure targeting five genetic markers for species identification and 11 serotype markers for *S. sonnei* and *S. flexneri*, in a single test [82, 102]. The second tool is a WGS-based workflow for automated prediction of *Shigella* serotypes, focusing on gene functionality [82].

Conclusion and general considerations

Identifying bacterial pathogens at genus, species, and strain levels is indispensable in supporting appropriate diagnosis and treatment, assessing the disease burden, tracking sources, performing traceback investigations, and disclosing changes in the frequency of phylogenetic groups in humans/animal disease and environmental niches. Concerning *Shigella* identification and serotyping, this remains a daunting task, especially in developing countries 75. Essentially, the hardship might arise from the taxonomic ambiguities lurking behind the separation of *E. coli* and *Shigella* because of their genetic relatedness. With the beginning of the WGS era, it is necessary to reconsider *Shigella*/*E. coli* based on phylogenetic criteria with or without renaming of genera and species to better serve medicine and science interests. Pending a more refined taxonomic concept for *E. coli* and *Shigella*, clinical microbiological laboratories should select the most appropriate identification tests to set *Shigella* apart from *E. coli*, more particularly EIEC, in terms of trade-offs between their advantages and disadvantages as discussed earlier in the review. In laboratories with low-resources settings, clinical symptoms and phenotypic tests can be used to differentiate *Shigella* from EIEC and even to serotype *Shigella*. When possible, an algorithm merging both phenotypic and molecular tests might help elucidate an isolate's real identity. In high-resource settings, WGS could serve as an "all-in-one test" for both identification and serotyping of *Shigella* spp., and also for disclosing novel genomic markers and validating previously well-established methods on extensive diverse genomic collections. When considering

WGS as a technique, one should pay close attention to the questions at hand and select the most appropriate analytical approaches. WGS-based approaches with low resolution and speciation objectives such as those based on K-mers and ANI can draw the proper (or real) borders between these different taxonomic entities; that we call species. However, these approaches (e.g., ANI) must be verified thoroughly on a representative collection. Otherwise, WGS-based approaches with finer resolution (usually denoted as typing approaches) such as those based on SNP, cgMLST, or wgMLST can go far within the same species and dissect the borders between serotypes, clones, and even isolates. Therefore, the question that must be answered in future studies is at what level of resolution should a typing approach with finer classifications (number of SNP or different alleles, for example) show clades that share characteristics attributable to *Shigella* but not to *E. coli* in regards for example to clinical symptoms, infectious doses, and biochemical features. Undeniably, WGS (will) become the future gold standard for the *Shigella* surveillance and epidemiologic investigations, particularly with the steadily decreasing cost of sequencing platforms and the growing number of user-friendly bioinformatics tools and pipelines. Meanwhile, appropriate backward compatibility should be maintained to harmonize the data between the different stakeholders and establish firm bridges with historical data.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Lampel KA, Formal SB, Maurelli AT (2018) A brief history of Shigella. EcoSal Plus 8(1). https://doi.org/10.1128/ecosalplus.ESP-0006-2017
2. Baker S, The HC (2018) Recent insights into Shigella. Curr Opin Infect Dis 31(5):449–454. https://doi.org/10.1097/QCO.0000000000000475
3. McCrickard LS, Crim SM, Kim S, Bowen A (2018) Disparities in severe shigellosis among adults – Foodborne diseases active surveillance network, 2002–2014. BMC Public Health 18(1):221. https://doi.org/10.1186/s12889-018-5115-4
4. Levine MM, Kotloff KL, Barry EM, Pasetti MF, Sztein MB (2007) Clinical trials of Shigella vaccines: two steps forward and one step back on a long, hard road. Nat Rev Microbiol 5(7):540–553. https://doi.org/10.1038/nrmicro1662
5. Mattock E, Blocker AJ (2017) How do the virulence factors of Shigella work together to cause disease? Front Cell Infect Microbiol 7. https://doi.org/10.3389/fcimb.2017.00064
6. Kotloff KL, Riddle MS, Platts-Mills JA, Pavlinac P, Zaidi AKM (2018) Shigellosis. Lancet Lond Engl 391(10122):801–812. https://doi.org/10.1016/S0140-6736(17)33296-8
7. Sansonetti PJ III (2001) Shigellosis: from symptoms to molecular pathogenesis. Am J Physiol-Gastrointest Liver Physiol 280(3):G319–G323. https://doi.org/10.1152/ajpgi.2001.280.3.G319
8. Parker CT, Tindall BJ, Garrity GM (2019) International code of nomenclature of prokaryotes. Int J Syst Evol Microbiol 69(1A):S1–S111. https://doi.org/10.1099/ijsem.0.000778
9. Dodd CE, Jones D (1982) A numerical taxonomic study of the genus Shigella. J Gen Microbiol 128(9):1933–1957. https://doi.org/10.1099/00221287-128-9-1933
10. Kang C-H, Nam Y-D, Chung W-H et al (2007) Relationship between genome similarity and DNA–DNA hybridization among closely related bacteria. J Microbiol Biotechnol 17(6):945–951
11. Pupo GM, Lan R, Reeves PR (2000) Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics. Proc Natl Acad Sci U S A 97(19):10567–10572. https://doi.org/10.1073/pnas.180094797
12. Lukjancenko O, Wassenaar TM, Ussery DW (2010) Comparison of 61 sequenced Escherichia coli genomes. Microb Ecol 60(4):708. https://doi.org/10.1007/s00248-010-9717-3
13. Reeves PR, Liu B, Zhou Z et al (2011) Rates of mutation and host transmission for an Escherichia coli clone over 3 years. PLoS ONE 6(10):e26907. https://doi.org/10.1371/journal.pone.0026907
14. Zuo G, Xu Z, Hao B (2013) Shigella strains are not clones of Escherichia coli but sister species in the genus Escherichia. Genomics Proteomics Bioinforma 11(1):61–65. https://doi.org/10.1016/j.gpb.2012.11.002
15. Hao B, Qi J (2004) Prokaryote phylogeny without sequence alignment: from avoidance signature to composition distance. J Bioinform Comput Biol 2(1):1–19. https://doi.org/10.1142/s0219720004000442
16. Zuo G, Hao B (2015) CVTree3 web server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy. Genomics Proteomics Bioinforma 13(5):321. https://doi.org/10.1016/j.gpb.2015.08.004
17. Lan R, Reeves PR (2002) Escherichia coli in disguise: molecular origins of Shigella. Microbes Infect 4(11):1125–1132
18. Ud-Din A, Wahid S (2015) Relationship among Shigella spp. and enteroinvasive Escherichia coli (EIEC) and their differentiation. Braz J Microbiol 45(4):1131–1138. https://doi.org/10.1590/s1517-83822014000400002
19. van den Beld MJC, Reubsaet FAG (2012) Differentiation between Shigella, enteroinvasive Escherichia coli (EIEC) and noninvasive Escherichia coli. Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol. 31(6):899–904. https://doi.org/10.1007/s10096-011-1395-7
20. Versalovic J, Carroll KC, Funke G, Jorgensen JH, Landry ML, Warnock DW (2011) Manual of clinical microbiology, 10th edition. 10th ed. Am Soc Microbiol. https://doi.org/10.1128/9781555816728
21. Gibotti A, Tanaka TL, Oliveira VR, Taddei CR, Martinez MB (2004) Molecular characterization of enteroinvasive Escherichia coli ipa genes by PCR-RFLP analysis. Braz J Microbiol 35(1–2):74–80. https://doi.org/10.1590/S1517-83822004000100012
22. Pasqua M, Michelacci V, Di Martino ML et al (2017) The intriguing evolutionary journey of enteroinvasive E. coli (EIEC) toward pathogenicity. Front Microbiol 8. https://doi.org/10.3389/fmicb.2017.02390
23. Beutin L, Gleier K, Kontny I, Echeverria P, Scheutz F (1997) Origin and characteristics of enteroinvasive strains of Escherichia coli (EIEC) isolated in Germany. Epidemiol Infect 118(3):199–205. https://doi.org/10.1017/s0950268897007413

24. Le Gall T, Darlu P, Escobar-Páramo P, Picard B, Denamur E (2005) Selection-driven transcriptome polymorphism in Escherichia coli/Shigella species. Genome Res 15(2):260–268. https://doi.org/10.1101/gr.2405905

25. Octavia S, Lan R (2015) Chapter 65 – Shigella and Shigellosis: genetics, epidemiology and pathogenesis. In: Tang Y-W, Sussman M, Liu D, Poxton I, Schwartzman J (eds) Molecular medical microbiology (second edition). Academic Press, pp 1147–1168. https://doi.org/10.1016/B978-0-12-397169-2.00065-2

26. Liu D (2015) Chapter 64 - Diarrhoeagenic Escherichia coli. In: Tang Y-W, Sussman M, Liu D, Poxton I, Schwartzman J (eds) Molecular medical microbiology (second edition). Academic Press, pp 1133–1146. https://doi.org/10.1016/B978-0-12-397169-2.00064-0

27. Pettengill EA, Pettengill JB, Binet R (2015) 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 6:1573. https://doi.org/10.3389/fmicb.2015.01573

28. The HC, Thanh DP, Holt KE, Thomson NR, Baker S (2016) The genomic signatures of Shigella evolution, adaptation and geographical spread. Nat Rev Microbiol 14(4):235–250. https://doi.org/10.1038/nrmicro.2016.10

29. Mark TC (2008) Enterohaemorrhagic Escherichia coli and Shigella dysenteriae type 1-induced haemolytic uraemic syndrome. Pediatr Nephrol Berl Ger 23(9):1425–1431. https://doi.org/10.1007/s00467-008-0820-3

30. Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR (2004) Molecular evolutionary relationships of enteroinvasive Escherichia coli and Shigella spp. Infect Immun 72(9):5080–5088. https://doi.org/10.1128/IAI.72.9.5080-5088.2004

31. Lampel KA (2014) ESCHERICHIA COLI | Enteroinvasive Escherichia coli. In: Encyclopedia of food microbiology. Second Edition. pp 718–721. https://www.elsevier.com/books/encyclopedia-of-food-microbiology/batt/978-0-12-384730-0

32. van den Beld MJC, Warmelink E, Friedrich AW et al (2019) Incidence, clinical implications and impact on public health of infections with Shigella spp. and entero-invasive Escherichia coli (EIEC): results of a multicenter cross-sectional study in the Netherlands during 2016–2017. BMC Infect Dis 19(1):1037. https://doi.org/10.1186/s12879-019-4659-y

33. Dekker JP, Frank KM (2015) Salmonella, Shigella, and Yersinia. Clin Lab Med 35(2):225–246. https://doi.org/10.1016/j.cll.2015.02.002

34. Hawkey PM (2006) Identification of Enterobacteriaceae. In: Principles and practice of clinical bacteriology. John Wiley & Sons, Ltd, pp 341–345. https://doi.org/10.1002/9780470017968.ch27

35. Liu B, Knirel YA, Feng L et al (2008) Structure and genetics of Shigella O antigens. FEMS Microbiol Rev 32(4):627–653. https://doi.org/10.1111/j.1574-6976.2008.00114.x

36. Khot PD, Fisher MA (2013) Novel approach for differentiating Shigella species and Escherichia coli by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 51(11):3711–3716. https://doi.org/10.1128/JCM.01526-13

37. Rahi P, Prakash O, Shouche YS (2016) Matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry (MALDI-TOF MS) based microbial identifications: challenges and scopes for microbial ecologists. Front Microbiol 7:1359. https://doi.org/10.3389/fmicb.2016.01359

38. Paauw A, Jonker D, Roeselers G et al (2015) Rapid and reliable discrimination between Shigella species and Escherichia coli using MALDI-TOF mass spectrometry. Int J Med Microbiol IJMM 305(4–5):446–452. https://doi.org/10.1016/j.ijmm.2015.04.001

39. Pavlovic M, Luze A, Konrad R et al (2011) Development of a duplex real-time PCR for differentiation between E. coli and Shigella spp. J Appl Microbiol 110(5):1245–1251. https://doi.org/10.1111/j.1365-2672.2011.04973.x

40. Løbersli I, Wester AL, Kristiansen Å, Brandal LT (2016) Molecular differentiation of Shigella spp. from enteroinvasive E. Coli. Eur J Microbiol Immunol 6(3):197–205. https://doi.org/10.1556/1886.2016.00004

41. Chen L, Cai Y, Zhou G et al (2014) Rapid Sanger sequencing of the 16S rRNA gene for identification of some common pathogens. PLoS ONE 9(2):e88886. https://doi.org/10.1371/journal.pone.0088886

42. Vos M, Quince C, Pijl AS, de Hollander M, Kowalchuk GA (2012) A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. PLoS ONE 7(2):e30600. https://doi.org/10.1371/journal.pone.0030600

43. Dhiviya Prabaa MS, Naveen Kumar DR, Yesurajan IF, Anandan S, Kamini W, Balaji V (2017) Identification of nonserotypeable Shigella spp. using genome sequencing: a step forward. Future Sci OA 3(4):FSO229. https://doi.org/10.4155/fsoa-2017-0063

44. Chattaway MA, Schaefer U, Tewolde R, Dallman TJ, Jenkins C (2017) Identification of Escherichia coli and Shigella species from whole-genome sequences. J Clin Microbiol 55(2):616–623. https://doi.org/10.1128/JCM.01790-16

45. Maderankova D, Jugas R, Sedlar K, Vitek M, Skutkova H (2019) Rapid bacterial species delineation based on parameters derived from genome numerical representations. Comput Struct Biotechnol J 17:118–126. https://doi.org/10.1016/j.csbj.2018.12.006

46. Kimura B (2018) Will the emergence of core genome MLST end the role of in silico MLST? Food Microbiol 75:28–36. https://doi.org/10.1016/j.fm.2017.09.003

47. O'Hara CM (2005) Manual and automated instrumentation for identification of Enterobacteriaceae and other aerobic gram-negative bacilli. Clin Microbiol Rev 18(1):147–162. https://doi.org/10.1128/CMR.18.1.147-162.2005

48. Holmes B, Willcox WR, Lapage SP (1978) Identification of Enterobacteriaceae by the API 20E system. J Clin Pathol 31(1):22–30. https://doi.org/10.1136/jcp.31.1.22

49. Ekwall E, Dimander M (1982) Comparison of micro-ID with API 20E for identification of Salmonella, Shigella and Yersinia species. Eur J Clin Microbiol 1(3):134–137. https://doi.org/10.1007/bf02019612

50. Carroll KC, Glanz BD, Borek AP et al (2006) Evaluation of the BD Phoenix automated microbiology system for identification and antimicrobial susceptibility testing of Enterobacteriaceae. J Clin Microbiol 44(10):3506–3509. https://doi.org/10.1128/JCM.00994-06

51. Snyder JW, Munier GK, Johnson CL (2008) Direct comparison of the BD phoenix system with the MicroScan WalkAway system for identification and antimicrobial susceptibility testing of Enterobacteriaceae and nonfermentative gram-negative organisms. J Clin Microbiol 46(7):2327–2333. https://doi.org/10.1128/JCM.00075-08

52. Gupta S, Aruna C, Muralidharan S (2011) Misidentification of a commensal inactive Escherichia coli as Shigella sonnei by an automated system in a critically ill patient. Clin Lab 57(9–10):767–769

53. O'Hara CM, Miller JM (2000) Evaluation of the MicroScan rapid neg ID3 panel for identification of Enterobacteriaceae and some common gram-negative nonfermenters. J Clin Microbiol 38(10):3577–3580

54. van der Ploeg CA, Viñas MR, Terragno R, Bruno SB, Binsztein N. Laboratory protocol: serotyping of Shigella spp. Geneva. Published online June 26, 2010. http://www.antimicrobial-resistance.dk/data/images/protocols/gfn_shigellaserotypification-final-29-06-10.pdf

55. Wu Y, Lau HK, Lee T, Lau DK, Payne J (2019) In silico serotyping based on whole-genome sequencing improves the accuracy of Shigella identification. Appl Environ Microbiol 85(7). https://doi.org/10.1128/AEM.00165-19

56. Muthuirulandi Sethuvel DP, Devanga Ragupathi NK, Anandan S, Veeraraghavan B (2017) Update on: Shigella new serogroups/serotypes and their antimicrobial resistance. Lett Appl Microbiol 64(1):8–18. https://doi.org/10.1111/lam.12690

57. van den Beld MJC, de Boer RF, Reubsaet FAG et al (2018) Evaluation of a culture-dependent algorithm and a molecular algorithm for identification of Shigella spp., Escherichia coli, and enteroinvasive E. coli. J Clin Microbiol 56(10). https://doi.org/10.1128/JCM.00510-18

58. Connor TR, Barker CR, Baker KS et al (2015) Species-wide whole genome sequencing reveals historical global spread and recent local persistence in Shigella flexneri. eLife 4:e07335. https://doi.org/10.7554/eLife.07335

59. Sun Q, Lan R, Wang J et al (2013) Identification and characterization of a novel Shigella flexneri serotype Yv in China. PLoS ONE 8(7). https://doi.org/10.1371/journal.pone.0070238

60. Allison GE, Verma NK (2000) Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri. Trends Microbiol 8(1):17–23. https://doi.org/10.1016/s0966-842x(99)01646-7

61. Almuzara M, Barberis C, Traglia G, Famiglietti A, Ramirez MS, Vay C (2015) Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry for species identification of nonfermenting Gram-negative bacilli. J Microbiol Methods 112:24–27. https://doi.org/10.1016/j.mimet.2015.03.004

62. Ling J, Wang H, Li G et al (2019) A novel short-term high-lactose culture approach combined with a matrix-assisted laser desorption ionization-time of flight mass spectrometry assay for differentiating Escherichia coli and Shigella species using artificial neural networks. PLoS ONE 14(10):e0222636. https://doi.org/10.1371/journal.pone.0222636

63. Theron J, Morar D, Du Preez M, Brözel VS, Venter SN (2001) A sensitive seminested PCR method for the detection of Shigella in spiked environmental water samples. Water Res 35(4):869–874. https://doi.org/10.1016/S0043-1354(00)00348-1

64. Jackson MP (1991) Detection of Shiga toxin-producing Shigella dysenteriae type 1 and Escherichia coli by using polymerase chain reaction with incorporation of digoxigenin-11-dUTP. J Clin Microbiol 29(9):1910–1914

65. Kim H-J, Ryu J-O, Song J-Y, Kim H-Y (2017) Multiplex polymerase chain reaction for identification of Shigellae and four Shigella species using novel genetic markers screened by comparative genomics. Foodborne Pathog Dis 14(7):400–406. https://doi.org/10.1089/fpd.2016.2221

66. Ashida H, Sasakawa C (2015) Shigella IpaH family effectors as a versatile model for studying pathogenic bacteria. Front Cell Infect Microbiol 5:100. https://doi.org/10.3389/fcimb.2015.00100

67. Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Inbanathan FY, Veeraraghavan B (2018) Accurate differentiation of Escherichia coli and Shigella serogroups: challenges and strategies. New Microbes New Infect 21:58–62. https://doi.org/10.1016/j.nmni.2017.09.003

68. Ojha SC, Yean Yean C, Ismail A, Banga Singh K-K. A pentaplex PCR assay for the detection and differentiation of Shigella species. BioMed Res Int. https://doi.org/10.1155/2013/412370

69. Sahl JW, Morris CR, Emberger J et al (2015) Defining the phylogenomics of Shigella species: a pathway to diagnostics. J Clin Microbiol 53(3):951–960. https://doi.org/10.1128/JCM.03527-14

70. Kingombe CIB, Cerqueira-Campos M-L, Farber JM (2005) Molecular strategies for the detection, identification, and differentiation between enteroinvasive Escherichia coli and Shigella spp. J Food Prot 68(2):239–245. https://doi.org/10.4315/0362-028x-68.2.239

71. Peng X, Luo W, Zhang J, Wang S, Lin S (2002) Rapid detection of Shigella species in environmental sewage by an immuno-capture PCR with universal primers. Appl Environ Microbiol 68(5):2580–2583. https://doi.org/10.1128/AEM.68.5.2580-2583.2002

72. Janda JM, Abbott SL (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol 45(9):2761. https://doi.org/10.1128/JCM.01228-07

73. Adékambi T, Shinnick TM, Raoult D, Drancourt M (2008) Complete rpoB gene sequencing as a suitable supplement to DNA–DNA hybridization for bacterial species and genus delineation. Int J Syst Evol Microbiol 58(Pt 8):1807–1814. https://doi.org/10.1099/ijs.0.65440-0

74. Fukushima M, Kakinuma K, Kawaguchi R (2002) Phylogenetic analysis of Salmonella, Shigella, and Escherichia coli strains on the basis of the gyrB gene sequence. J Clin Microbiol 40(8):2779–2785. https://doi.org/10.1128/jcm.40.8.2779-2785.2002

75. Halimeh FB, Rafei R, Diene S, et al. Challenges in identification of enteroinvasive Escherichia coli and Shigella spp. in Lebanon. Acta Microbiol Immunol Hung. Published online March 30, 2020:1–7. https://doi.org/10.1556/030.2020.01102

76. Nochi Z, Sahebekhtiari N, Kharaziha P et al (2009) Comparison of 16S rRNA, 23S rRNA andgyrB genes sequences in phylogenetic relationships of Shigella isolates from Iran. Ann Microbiol 59(3):615–622. https://doi.org/10.1007/BF03175154

77. Wetterstrand KA. DNA sequencing costs: data. Genome.gov. Published January 21, 2020. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data. Accessed 21 Jan 2020

78. Uelze L, Grützke J, Borowiak M et al (2020) Typing methods based on whole genome sequencing data. One Health Outlook 2(1):3. https://doi.org/10.1186/s42522-020-0010-1

79. Rafei R, Osman M, Dabboussi F, Hamze M (2019) Update on the epidemiological typing methods for Acinetobacter baumannii. Future Microbiol 14:1065–1080. https://doi.org/10.2217/fmb-2019-0134

80. Chattaway MA, Greig DR, Gentle A, Hartman HB, Dallman TJ, Jenkins C (2017) Whole-genome sequencing for national surveillance of Shigella flexneri. Front Microbiol 8. https://doi.org/10.3389/fmicb.2017.01700

81. Hayford AE, Mammel MK, Lacher DW, Brown EW (2011) Single nucleotide polymorphism (SNP)-based differentiation of Shigella isolates by pyrosequencing. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 11(7):1761–1768. https://doi.org/10.1016/j.meegid.2011.07.015

82. Ventola E, Bogaerts B, De Keersmaecker SCJ et al (2019) Shifting national surveillance of Shigella infections toward geno-serotyping by the development of a tailored Luminex assay and NGS workflow. MicrobiologyOpen 8(8):e00807. https://doi.org/10.1002/mbo3.807

83. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9(1):1–8. https://doi.org/10.1038/s41467-018-07641-9

84. Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106(45):19126–19131. https://doi.org/10.1073/pnas.0906412106

85. Wirth T, Falush D, Lan R et al (2006) Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol

60(5):1136–1151. https://doi.org/10.1111/j.1365-2958.2006.05172.x

86. Weihong Qi, Lacher DW, Bumbaugh AC et al (2004) EcMLST: an online database for multi locus sequence typing of pathogenic Escherichia coli. In: Proceedings. 2004 IEEE Computational Systems Bioinformatics Conference, 2004. CSB 2004. IEEE, pp:499–500. https://doi.org/10.1109/CSB.2004.1332482

87. Jaureguy F, Landraud L, Passet V et al (2008) Phylogenetic and genomic diversity of human bacteremic Escherichia coli strains. BMC Genomics 9:560. https://doi.org/10.1186/1471-2164-9-560

88. Maiden MCJ, Jansen van Rensburg MJ, Bray JE et al (2013) MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol 11(10):728–736. https://doi.org/10.1038/nrmicro3093

89. BioNumerics. Escherichia coli – Shigella schema for whole genome sequencing typing. http://www.applied-maths.com/sites/default/files/extra/Release-Note-Eschericha-coli-Shigella-schema.pdf

90. Campos-Madueno EI, Bernasconi OJ, Moser AI et al (2020) Rapid increase of CTX-M-producing Shigella sonnei isolates in Switzerland due to spread of common plasmids and international clones. Antimicrob Agents Chemother 64(10). https://doi.org/10.1128/AAC.01057-20

91. Díaz-Jiménez D, García-Meniño I, Herrera A et al (2020) Genomic characterization of Escherichia coli isolates belonging to a new hybrid aEPEC/ExPEC pathotype O153:H10-A-ST10 eae-beta1 occurred in meat, poultry, wildlife and human diarrheagenic samples. Antibiot Basel Switz 9(4). https://doi.org/10.3390/antibiotics9040192

92. Hammerum AM, Porsbo LJ, Hansen F et al (2020) Surveillance of OXA-244-producing Escherichia coli and epidemiologic investigation of cases, Denmark, January 2016 to August 2019. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 25(18). https://doi.org/10.2807/1560-7917.ES.2020.25.18.1900742

93. Rumore J, Tschetter L, Kearney A et al (2018) Evaluation of whole-genome sequencing for outbreak detection of verotoxigenic Escherichia coli O157:H7 from the Canadian perspective. BMC Genomics 19(1):870. https://doi.org/10.1186/s12864-018-5243-3

94. van den Beld MJC, Reubsaet FAG, Pijnacker R et al (2019) The importance of a multifactorial approach for (inter)national surveillance of Shigella spp. and entero-invasive Escherichia coli. [Preprint]. https://doi.org/10.21203/rs.2.13263/v1

95. Coimbra RS, Grimont F, Grimont PA (1999) Identification of Shigella serotypes by restriction of amplified O-antigen gene cluster. Res Microbiol 150(8):543–553. https://doi.org/10.1016/s0923-2508(99)00103-5

96. Coimbra RS, Artiguenave F, Jacques LSRZ, Oliveira GC (2010) MST (molecular serotyping tool): a program for computer-assisted molecular identification of Escherichia coli and Shigella O antigens. J Clin Microbiol 48(5):1921–1923. https://doi.org/10.1128/JCM.00357-10

97. Sun Q, Lan R, Wang Y et al (2011) Development of a multiplex PCR assay targeting O-antigen modification genes for molecular serotyping of Shigella flexneri ∇. J Clin Microbiol 49(11):3766–3770. https://doi.org/10.1128/JCM.01259-11

98. Gentle A, Ashton PM, Dallman TJ, Jenkins C (2016) Evaluation of molecular methods for serotyping Shigella flexneri. J Clin Microbiol 54(6):1456. https://doi.org/10.1128/JCM.03386-15

99. Brengi SP, Sun Q, Bolaños H et al (2019) PCR-based method for Shigella flexneri serotyping: international multicenter validation. J Clin Microbiol 57(4). https://doi.org/10.1128/JCM.01592-18

100. van der Ploeg CA, Rogé AD, Bordagorría XL, de Urquiza MT, Castillo ABC, Bruno SB (2018) Design of two multiplex PCR assays for serotyping Shigella flexneri. Foodborne Pathog Dis 15(1):33–38. https://doi.org/10.1089/fpd.2017.2328

101. Li Y, Cao B, Liu B et al (2009) Molecular detection of all 34 distinct O-antigen forms of Shigella. J Med Microbiol 58(1):69–81. https://doi.org/10.1099/jmm.0.000794-0

102. Wuyts V, Roosens NHC, Bertrand S, Marchal K, De Keersmaecker SCJ. Guidelines for optimisation of a multiplex oligonucleotide ligation-PCR for characterisation of microbial pathogens in a microsphere suspension array. BioMed Res Int. https://doi.org/10.1155/2015/790170