Phylogenetic and Evolutionary Analysis Reveals the Recent Dominance of Ciprofloxacin-Resistant \textit{Shigella sonnei} and Local Persistence of \textit{S. flexneri} Clones in India

Dhiviya Praba Muthuirulandi Sethuvel,a Ankur Mutreja,b,c Agila Kumari Pragasam,a Karthick Vasudevan,a Dhivya Murugan,a Shalini Anandan,a Joy Sarojini Michael,a Kamini Wallia,d Balaji Veeraraghavan,a

\textsuperscript{a}Department of Clinical Microbiology, Christian Medical College, Vellore, India  
\textsuperscript{b}Department of Medicine, University of Cambridge, Cambridge, United Kingdom  
\textsuperscript{c}Wellcome Sanger Institute, Hinxton, United Kingdom  
\textsuperscript{d}Division of Epidemiology and Communicable Diseases, Indian Council of Medical Research, New Delhi, India

\textbf{ABSTRACT} \textit{Shigella} is the second leading cause of bacterial diarrhea worldwide. Recently, \textit{Shigella sonnei} seems to be replacing \textit{Shigella flexneri} in low- and middle-income countries undergoing economic development. Despite this, studies focusing on these species at the genomic level remain largely unexplored. Here, we compared the genome sequences of \textit{S. flexneri} and \textit{S. sonnei} isolates from India with the publicly available genomes of global strains. Our analysis provides evidence for the long-term persistence of all phylogenetic groups (PGs) of \textit{S. flexneri} and the recent dominance of the ciprofloxacin-resistant \textit{S. sonnei} lineage in India. Within \textit{S. flexneri} PGs, the majority of the study isolates belonged to PG3 within the predominance of serotype 2. For \textit{S. sonnei}, the current pandemic involves globally distributed multidrug-resistant (MDR) clones that belong to Central Asia lineage III. The presence of such epidemiologically dominant lineages in association with stable antimicrobial resistance (AMR) determinants results in successful survival in the community.

\textbf{IMPORTANCE} \textit{Shigella} is the second leading cause of bacterial diarrhea worldwide. This has been categorized as a priority pathogen among enteric bacteria by the Global Antimicrobial Resistance Surveillance System (GLASS) of the World Health Organization (WHO). Recently, \textit{S. sonnei} seems to be replacing \textit{S. flexneri} in low- and middle-income countries undergoing economic development. Antimicrobial resistance in \textit{S. flexneri} and \textit{S. sonnei} is a growing international concern, specifically with the international dominance of the multidrug-resistant (MDR) lineage. Genomic studies focusing on \textit{S. flexneri} and \textit{S. sonnei} in India remain largely unexplored. This study provides information on the introduction and expansion of drug-resistant \textit{Shigella} strains in India for the first time by comparing the genome sequences of \textit{S. flexneri} and \textit{S. sonnei} isolates from India with the publicly available genomes of global strains. The study discusses the key differences between the two dominant species of \textit{Shigella} at the genomic level to understand the evolutionary trends and genome dynamics of emerging and existing resistance clones. The present work demonstrates evidence for the long-term persistence of all PGs of \textit{S. flexneri} and the recent dominance of a ciprofloxacin-resistant \textit{S. sonnei} lineage in India.

\textbf{KEYWORDS} antimicrobial resistance, multidrug resistance, Central Asia III, \textit{Shigella}, pangenome

\textit{Shigella} is ranked as the second leading cause of bacterial diarrhea worldwide and the third leading cause of death in children less than 5 years old (1). This has been categorized as a priority pathogen among enteric bacteria by the Global Antimicrobial


Editor Krishna Rao, University of Michigan—Ann Arbor

Copyright © 2020 Muthuirulandi Sethuvel et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Balaji Veeraraghavan, vbalaji@cmcvellore.ac.in.

Received 13 June 2020

Accepted 9 September 2020

Published 7 October 2020
Resistance Surveillance System (GLASS) of the World Health Organization (WHO) (2). Among the four species of *Shigella*, the most common cause of endemic shigellosis and the most frequently isolated species in India is *Shigella flexneri*, followed by *S. sonnei*. *S. dysenteriae* and *S. boydii* are now relatively uncommon. Studies from various parts of India have shown that the overall rate of isolation of *Shigella* species ranges from 3 to 6% among diarrheal patients (3). We have documented a prevalence of 4.8% to 4.6% between 2014 and 2015 and have continued to observe the same rate up to today at our tertiary care center in Vellore, India (4). In recent years, *S. sonnei* appears to be replacing *S. flexneri* in low- and middle-income countries undergoing economic development (5).

Several epidemics of *Shigella* have been reported from Asian countries such as Bangladesh, Sri Lanka, Maldives, Nepal, Bhutan, and Myanmar. In India, epidemics have been reported from Vellore, southern India; Andaman and Nicobar Islands, eastern India; and Chandigarh, northern India. Outbreaks caused by *Shigella* are not uncommon in India and have been reported from various parts of the country like West Bengal, Kerala, and Maharashtra. *Shigella* species are the third most common bacterial agents causing traveler’s diarrhea among enteric infections reported to FoodNet, and India is one of the most common travel destinations for picking up this infection (3).

Antimicrobial resistance (AMR) in *S. flexneri* and *S. sonnei* is a growing international concern, specifically with the international dominance of the multidrug-resistant (MDR) lineage of *S. sonnei* (6). A previous phylogenetic analysis predicted that South Asia would be the hub for the international spread of ciprofloxacin-resistant *S. sonnei* (6). However, only a limited number of *Shigella* isolates from India have been included in previous studies. Therefore, understanding the expansion of this lineage and the phylogenetic relationship with isolates from within and outside Asia, including India, is critical.

Genomic studies focusing on *S. flexneri* and *S. sonnei* in India remain largely unexplored. Understanding the genome sequences of antimicrobial-resistant pathogens can enhance our knowledge of the molecular identity of resistance traits and their mechanism of dissemination within the microbial population. Here, we compared the genome sequences of *S. flexneri* and *S. sonnei* isolates from India with publicly available genomes of global strains to understand the introduction and expansion of drug-resistant strains in India. Bayesian phylogenetic analysis was performed, in particular for *S. sonnei* isolates from India, to demonstrate the evolution of ciprofloxacin-resistant *S. sonnei* clones in India. In addition, virulence, resistance, and plasmid profiles of the isolates were analyzed and correlated with previously defined *S. flexneri* phylogenetic groups (PGs) and *S. sonnei* lineages.

**RESULTS**

The whole-genome phylogenetic tree was constructed using 106 *S. flexneri* and 82 *S. sonnei* genomes sequenced in this study and 60 *S. flexneri* and 362 *S. sonnei* genomes from previous studies. All *S. flexneri* and *S. sonnei* sequences were mapped to the reference sequences (GenBank accession numbers CP000038 and NC_004337.2).

**Phylogenetic analysis of *S. flexneri*.** The phylogenetic comparison of Indian *S. flexneri* isolates against isolates available from other South Asian countries (Pakistan and Bangladesh) showed that Indian isolates were distributed across all the previously defined phylogenetic groups (PG1, -2, -3, -5, -6, and -7) with the exception of PG4 (Fig. 1). The majority of Indian isolates clustered within PG3 (70%), followed by PG1 (22%) and PG2 (7%). Each PG contained multiple serotypes. However, the common serotypes found within PG2, PG3, and PG5 were *S. flexneri* serotypes 3a, 2a, and 5a, respectively. The core single nucleotide polymorphism (SNP)-based comparisons between global and Indian strains suggested high genetic diversity within the *S. flexneri* PG3 isolates, with a median pairwise SNP difference of 97 SNPs (interquartile range [IQR], 88 to 105). The intragroup comparison of PG3 isolates from India showed a median difference of 78 SNPs (IQR, 47 to 100) (see Fig. S1 in the supplemental material).
Variation in the distribution of virulence determinants among the PGs was also evident. Notably, the SHI-1 pathogenicity island (PAI), which is known to carry pic, sigA, and the Shigella enterotoxin 1 (ShET-1) set1A and/or set1B genes, was exclusively seen in PG3 isolates with the exception of a single isolate of PG1. SHI-1 was present in various serotypes within PG3 but was predominantly seen in serotype 2a. Additionally, the SHI-2 PAI known to carry the iucABCD and iutA genes was found in isolates of all PGs. Similarly, sat, a serine protease autotransporter, was observed in all PGs. The sen gene encoding enterotoxin ShET-2 was seen only in the PG2 and PG7 isolates. Entero-bactin genes (entBCEFS and fepABCDG) that encode iron siderophores were present across PG3, PG4, PG5, and PG7 and were also found in a few isolates of PG1. Notably, the fimE gene was absent in all PG3 isolates except one.

Screening of known AMR genes revealed highly variable AMR gene distributions both within and across PGs (Fig. 1). Genes encoding resistance to earlier first-line antimicrobials, including streptomycin (strA and/or strB), trimethoprim/sulfamethoxazole (dfrA1 and sul), chloramphenicol (catA1), and tetracycline (tetA and/or tetB), were distributed across all the PGs. Among beta-lactamases, the blaOXA-1 gene was present in all PGs with the exception of PG4 and PG7. Similarly, blaTEM-18 was present across all PGs with the exception of PG4, whereas blaCTX-M-15 was seen only in PG2, PG3, and PG6.
The AmpC beta-lactamase gene *bla*<sub>DHA</sub> was identified in two isolates, one each from PG1 and PG7. In addition, the *bla*<sub>CMY-60</sub> gene was identified in PG3 alone. These AmpC genes were observed only in the Indian isolates. Furthermore, the analysis of fluoroquinolone resistance mechanisms showed the presence of both plasmid-mediated quinolone resistance (PMQR) genes and mutations within the quinolone resistance-determining region (QRDR). The PMQR gene *qnr*S1 was identified in all PGs with the exception of PG4, while *qnr*B4 was seen in two isolates, one each from PG1 and PG7. The most frequently observed QRDR mutations were S83L in *gyr*A and S80I in *par*C as a result of replacements of serine by leucine and isoleucine, respectively. Macrolide resistance genes were recognized, such as *mph*A in PG2 and PG7, *erm*B and *mph*E in PG2, and *msr*E in PG3.

Additional plasmid analysis showed that IncFII was the most predominant plasmid across the PGs, followed by ColRNAI (Fig. 1). Among the Inc plasmid types, IncX3 was found in isolates belonging to PG1, PG2, and PG3, whereas IncY/IncL/M/IncI2 plasmids were identified only in PG3 isolates. Furthermore, IncFIC was seen only in PG1, while IncFIB was seen in all PGs with the exception of PG4.

**Phylogenetic analysis of S. sonnei.** The phylogenetic analysis of *S. sonnei* revealed that the ciprofloxacin-resistant isolates were grouped within the Central Asia III lineage. All Indian isolates, including those from Vellore, belonged to the Central Asia III lineage. The pairwise SNP comparison between global and Indian strains suggested high genomic identity within the *S. sonnei* Central Asia III lineage, with a median SNP difference of 8 SNPs (IQR, 5 to 11). The intragroup comparison of Central Asia lineage III isolates from India showed a median difference of 6 SNPs (IQR, 5 to 9) (Fig. S1). The majority of the isolates in the Central Asia III lineage carried genes encoding resistance to streptomycin, trimethoprim/sulfamethoxazole, and tetracycline, as described above. Most of the isolates (69%) within this lineage carried triple mutations (*gyr*A-S83L,D87G and/or D87N and *par*C-S80I) in the QRDRs (Fig. 2). Double and single mutations were observed in 1% and 19% of the isolates, respectively. Only one isolate carried a PMQR gene, *qnr*B. Central Asia lineage III had a different plasmid profile than the isolates of other lineages. Other lineages included ciprofloxacin-susceptible isolates with varying AMR and plasmid profiles. In addition, analysis of virulence determinants among the *S. sonnei* isolates showed a varying profile of virulence genes in comparison to *S. flexneri*. Notably, *S. sonnei* isolates were found to carry only the *sig*A gene within the SHI-1 PAI while harboring multiple virulence genes within the SHI-2 PAI. Furthermore, enterotoxin (*sen*) and enterobactin (*ent*BCEFS and *fep*ABCDG) genes were identified among the isolates.

As there was a strong temporal signature for the observed mutations in the global *S. sonnei* population, we explored the temporal structure of the *S. sonnei* isolates that belonged to the Central Asia III lineage from India using Bayesian phylogenetic methods. The root-to-tip analysis revealed a strong correlation ($R^2$, 0.7543) between the time of isolation and distance from the root, suggesting temporal clocklike evolution in the *S. sonnei* Central Asia III lineage from India (Fig. 3). The median substitution rate of the *S. sonnei* population was estimated to be $2.083 \times 10^{-3}$ substitutions per base per year in this study. This time-scaled phylogenetic reconstruction demonstrated the sequential accumulation of QRDR mutations among the *S. sonnei* population. The most recent common ancestor (MRCA) of the ciprofloxacin-resistant triple mutant clade in India was estimated to be from the year 2000 (95% highest posterior density [HPD], 1999.6 to 2004.4). Bayesian hierarchical clustering using core SNPs segregated the ciprofloxacin-resistant isolates that had a triple mutation in a QRDR into a separate clade (Fig. 4). Mutation *gyr*A-S83L was estimated to occur in $\sim$1996 (95% HPD, 1995 to 2001.1), mutation *gyr*A-D87Y was estimated to occur in $\sim$1996 (95% HPD, 1996 to 1999.8), and triple mutations S83L, D87Y/G, and S80I were estimated to occur in $\sim$2003 (95% HPD, 2003 to 2006.8). This implies that the ciprofloxacin-resistant population expanded rapidly after 2004 and has since been sustained.
Apart from fluoroquinolone resistance, other antimicrobial resistance genes (ARGs) were also identified in the Central Asia III lineage. These included genes encoding resistance to previous first-line antimicrobials, such as *strA/B, tetA/B, dfrA1, sulII*, and *aph(3)-lb/aph(6)-Id*, conferring resistance to streptomycin, tetracyclines, trimethoprim, sulfonamide, and aminoglycosides, respectively. Other important ARGs were those belonging to extended-spectrum beta-lactamases (ESBLs) such as the *blaCTX-M* and *blaTEM* families. Three *blaCTX-M* and two *blaTEM* variants were identified; these included *blaCTX-M-15*, *blaCTX-M-14*, *blaCTX-M-55*, and *blaTEM-1A* and *blaTEM-1B*, respectively. Also, *mdfA*, an MDR transporter gene, was identified. In addition, we found the *mphA* gene, a macrolide resistance gene, among the Central Asia III *S. sonnei* isolates. Analysis of major plasmids among the isolates showed that Col-type, IncB/O/K/Z, IncFIA, and IncFIB plasmids were prevalent. Among the Indian Central Asia III *S. sonnei* isolates, genes conferring resistance to first-line antimicrobials and a single MDR transporter gene were widely seen, and the Col-type plasmid was the most common. The contrasting genomic features and epidemiological distributions of *S. flexneri* and *S. sonnei* are given in Table 1.

**Pangenome analysis.** The pangenome analysis of 106 *S. flexneri* strains identified 17,506 orthologs with 2,698 core genes (15%) and 14,808 accessory genes (85%). These included 452 soft core genes, 1,628 shell genes, and 12,728 cloud genes. A total of 9,845 unique genes were identified among the *S. flexneri* pangenome, with the FC533 strain belonging to PG3 having the largest number of unique genes (7,090). Of these genes, most of them were found in the accessory genome, and 26% of them were for hypothetical proteins.

In *S. sonnei*, the pangenome analysis of 82 isolates identified 11,478 orthologous groups with 662 core genes (6%) and 10,816 accessory genes (94%). These included 689...
soft core genes, 3,262 shell genes, and 6,865 cloud genes. Notably, 14 isolates lost a few genes in their core genome and formed a separate cluster in the tree. The majority of these missing genes are found to encode hypothetical proteins (54%) with unknown functions and also include various transcriptional regulatory proteins. A total of 4,351 unique genes were identified, with the highest numbers (2,432) identified in the FC1793 strain belonging to the Central Asia III lineage. Like *S. flexneri*, most of the genes were found in the accessory genome, and 70% of them were for hypothetical proteins. The core and accessory gene compositions of *S. flexneri* and *S. sonnei* genomes are shown in Fig. 5 and 6.

DISCUSSION

Here, we provide evidence for the long-term persistence of all PGs of *S. flexneri* and the recent dominance of the ciprofloxacin-resistant *S. sonnei* lineage in India. Unlike *S. sonnei*, the replacement of one particular *S. flexneri* PG over the other was not evident; instead, we saw that the older PGs have persisted along with the newer ones and continued causing disease, as reported previously. This was evidenced by the fact that every PG comprises at least one isolate collected since the early 1990s. Likewise, previous studies evidenced the longer-term colonization of *S. flexneri* in locations of endemicity with diverse populations of both antimicrobial-resistant and -susceptible strains cocirculating where selection for antimicrobial resistance is less favorable, thus resulting in substantially less international transmission (7, 8). The variation in the prevailing species in different settings might be due to specific transmission patterns (9).

The epidemiological species shift has been documented worldwide, mainly in Asian, Latin American, and Middle Eastern countries (5, 10). We reported an increasing *S.
sonnei prevalence in Vellore, from 20% in 2014 to 34% in 2015, in our previous study (4). Similar studies from other parts of India have also highlighted the rising dominance of S. sonnei in regions that were traditionally dominated by S. flexneri (11). Although the reasons behind this changing species-level epidemiology are poorly understood,

TABLE 1 Genomic and epidemiological features of S. flexneri versus S. sonnei

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geographical distribution(s)</strong></td>
<td>Low- and middle-income countries</td>
</tr>
<tr>
<td><strong>Multidrug resistance phenotype</strong></td>
<td>AMP + SXT + NAL/CIP</td>
</tr>
<tr>
<td><strong>Genotypic resistance determinants</strong></td>
<td>dhfrA1, sulI/sulll</td>
</tr>
<tr>
<td><strong>Trimethoprim/sulfamethoxazole</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Ampicillin</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Ciprofloxacin</strong></td>
<td>QRDR mutations (gyrA and parC)</td>
</tr>
<tr>
<td><strong>PMQR—qnrB/S and aac(6′)Ib-cr</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Efflux—mdfA, acrA/B, and tolC</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Ceftaxime/cefixime</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Azithromycin</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Mobile genetic elements</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Plasmid type</strong></td>
<td>IncF</td>
</tr>
<tr>
<td><strong>Integron class(es)</strong></td>
<td>1 and 2</td>
</tr>
<tr>
<td><strong>PAI(s)</strong></td>
<td>SHI-1, SHI-2, and SRL</td>
</tr>
<tr>
<td><strong>Epidemiological feature(s)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Existing terms</strong></td>
<td>PGs</td>
</tr>
<tr>
<td><strong>Nature of spread</strong></td>
<td>Limited global spread</td>
</tr>
<tr>
<td><strong>Dominant clone in India</strong></td>
<td>PG3</td>
</tr>
</tbody>
</table>

*S. flexneri* | *S. sonnei* |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>dhfrA1, sulI/sulll</td>
<td>dhfrA1, sulI/sulll</td>
</tr>
<tr>
<td>blaOXA-1, blaTEM-1</td>
<td>blaTEM-1</td>
</tr>
<tr>
<td>QRDR mutations (gyrA and parC)</td>
<td>QRDR mutations (gyrA and parC)</td>
</tr>
<tr>
<td>PMQR—qnrB/S and aac(6′)Ib-cr</td>
<td>PMQR—qnrB/S and aac(6′)Ib-cr</td>
</tr>
<tr>
<td>Efflux—mdfA, acrA/B, and tolC</td>
<td>Efflux—mdfA, acrA/B, and tolC</td>
</tr>
<tr>
<td>Cefotaxime/cefixime</td>
<td>Cefotaxime/cefixime</td>
</tr>
<tr>
<td>Azithromycin</td>
<td>Azithromycin</td>
</tr>
<tr>
<td>Mobile genetic elements</td>
<td>Mobile genetic elements</td>
</tr>
<tr>
<td>Plasmid type</td>
<td>IncF</td>
</tr>
<tr>
<td>Integron class(es)</td>
<td>1 and 2</td>
</tr>
<tr>
<td>PAI(s)</td>
<td>SHI-1, SHI-2, and SRL</td>
</tr>
<tr>
<td>Epidemiological feature(s)</td>
<td></td>
</tr>
<tr>
<td>Existing terms</td>
<td>PGs</td>
</tr>
<tr>
<td>Nature of spread</td>
<td>Limited global spread</td>
</tr>
<tr>
<td>Dominant clone in India</td>
<td>PG3</td>
</tr>
</tbody>
</table>

AMP, ampicillin; SXT, trimethoprim/sulfamethoxazole; NAL, nalidixic acid; CIP, ciprofloxacin; PAI, pathogenicity island; PGs, phylogenetic groups.
previous study by Thompson et al. presented that this could be due to several causes: (i) improved water quality prevents the passive immunization derived from *Plesiomonas shigelloides*, normally found in contaminated water; (ii) the amoeba *Acanthamoeba castellanii* has been shown to phagocytize *S. sonnei* efficiently and symbiotically and

![Diagram](http://msphere.asm.org/)

**FIG 5** Core and accessory gene composition of 106 *S. flexneri* isolates extracted using Roary. A total of 2,698 core genes were shared by all strains, while 9,845 unique genes were identified.

**FIG 6** Core and accessory gene composition of 82 *S. sonnei* isolates extracted using Roary. A total of 662 core genes were shared by all strains, while 4,351 unique genes were identified.
can thus withstand highly chlorinated environments where *S. flexneri* is not able to grow; and (iii) *S. sonnei* has a better ability to acquire antimicrobial resistance than *S. flexneri* (5).

In the present study, we have compared *S. flexneri* isolates mainly from three Southeast Asian countries for exploring the relationship of strains in this confined region. Regionally, *S. flexneri* appears to be highly endemic, and there is limited evidence of international dissemination reported within the last 30 years. Our phylogenetic analysis reveals that *S. flexneri* lineages are distinct and highly diverse compared to *S. sonnei*, as previously observed (9).

Previously, seven phylogenetic groups (PGs) of *S. flexneri* were defined by Connor et al. (9). Of the PGs, PG1, -2, -4, and -6 are reported to be the oldest lineages, while PG3 and PG5 are more recent. In this study, *S. flexneri* isolates were distributed across all the PGs except PG4 and were not specific to locations, unlike *S. sonnei* isolates, which are location specific and have wide global spread. The majority of the isolates belonged to PG3, which includes multiple serotypes of *S. flexneri*, with serotype 2 being predominant. This presence of multiple serotypes in a single PG shows that the core genome remains stable but that serotype switching has been common, as reported previously (9). The presence of different serotypes in a single cluster confirms phylotyping as a more relevant method for public health surveillance and outbreak investigation rather than serotyping. This ensures that unless it is a protein-based vaccine carefully calibrated for stability across the phylogeny, vaccines against *Shigella* should continue to target the most common serotypes and not be lineage specific, as happens to be the case.

Recently, India developed the first indigenous vaccine against human shigellosis, licensed by the Indian Council of Medical Research (ICMR), India, and developed by the National Institute of Cholera and Enteric Diseases (NICED). OmpA of *S. flexneri* 2a coated with chitosan-alginate microparticles, which act as a delivery system, has been used as the vaccine target. OmpA of *S. flexneri* 2a was reported to be cross-reactive and induces strong protective immunity against prevalent serogroups of shigellae in animal models. Since the most prevalent species in India is found to be *S. flexneri* serotype 2, as observed in the present study and other studies in the country, this new vaccine could be a promising vaccine candidate against shigellosis in India if approved (12, 13).

Our study reveals variability in the composition of virulence factors among *S. flexneri* isolates. Notably, the copresence of the SHI-1 PAI, which was known to carry three essential virulence genes, along with enterobactin genes and antimicrobial resistance genes was exclusively observed in PG3, which is predominantly composed of *S. flexneri* serotype 2a isolates. This may account for the enhanced virulence and international dominance of this serotype. Furthermore, AMR has been shown to be a strong influence on the recent evolutionary history of many bacterial pathogens (7). The sustained presence of resistance genes for first-line antimicrobials in *S. flexneri* over the decades might be an essential factor for the successful maintenance of lineages in locations of endemicity, including India.

With regard to *S. sonnei*, there are five distinct lineages reported globally. The current pandemic involves globally distributed MDR clones that belong to lineage III. In particular, strains of the ciprofloxacin-resistant Central Asia III lineage are widespread. Although the Central Asia III lineage has spread internationally, its expansion in India has not been previously studied. Here, we compared Indian isolates against a global collection to provide insights into its expansion in our country. Our study provides evidence for the dominance of this lineage mainly in South India, as the other isolates from India included in the analysis have no location data available in the database. In addition, a few isolates from Africa, Europe, and America were found within this lineage and clustered closely with the Indian isolates, which indicates travel-related spread. Previous studies have also confirmed that India was the most commonly reported travel destination associated with ciprofloxacin-resistant *S. sonnei* and suggested that South Asia remains the primary source or hub for ciprofloxacin-resistant strains (6).

Furthermore, double/triple mutations in the QRDR in Central Asia lineage III isolates were consistent with previous findings and could be the reason behind the increased
fitness and global expansion of this lineage alongside the overall synergy between PMQR qnr genes and chromosomal mutations that were noted in other studies from India (11). Interestingly, the copresence of qnr genes and mutations in gyrA and parC genes were identified in only one isolate in our study. Our data show the temporal introduction of the gyrA-S83L mutation in 1996 and the sequential accumulation of secondary mutations leading to the rapid expansion of ciprofloxacin-resistant clones within the Central Asia III lineage in the early 2000s. Our study also indicates that the earliest isolate (1990) within the Central Asia clade was from South Asia, which supports a previous hypothesis suggesting that South Asia was the likely origin of this lineage (6).

The emergence of resistance to previous first-line antimicrobials such as ampicillin, trimethoprim/sulfamethoxazole, and nalidixic acid during this intervening period made ciprofloxacin the drug of choice to treat drug-resistant shigellosis. Following this, due to the intensive use of ciprofloxacin, resistance to this drug increased from 0.6% in 1998 to 2000 to 29% in 2007 to 2009 in Asia and Africa (14). This further limited the treatment options and made third-generation cephalosporins and macrolides the drugs of choice. With the ESBL-producing strains that have left azithromycin as a last resort (1, 15, 16), the recent emergence of azithromycin resistance in S. sonnei and S. flexneri serotype 3a, particularly in men who have sex with men (MSM) communities, is no surprise (8).

Furthermore, the analysis of core and accessory genes of S. flexneri and S. sonnei revealed variation in composition. S. sonnei was found to have a larger accessory genome than S. flexneri. Strains FC533 and FC1793 were found to have the largest numbers of unique genes of the S. flexneri and S. sonnei strains compared, respectively. The core genome of S. flexneri remains stable, suggesting the strength of these genomes in adapting to evolutionary pressures for persistence, while few isolates of S. sonnei were missing certain core genes. Previous studies have shown that the loss of genes is a characteristic of an intracellular pathogenic lifestyle. In Shigella, genes that are more prone to deletion are generally associated with pathogenesis, and most of these deleted genes are found to be involved in cellular metabolism (17). Insights into these genes are out of the scope of this study and can be studied in the future; however, it would be interesting to assess whether this gene loss in S. sonnei provides any survival benefit to this pathogen.

While our study was limited by the number of Shigella sequences representing different regions of India, data from previous studies and this study show that Shigella species from different geographical locations share common AMR and virulence patterns. This indicates that the genetic contents of the isolates are based merely on the lineages circulating in the region. The presence of epidemiologically dominant lineages associated with stable AMR determinants results in successful survival in the community. The global dissemination of this lineage is more likely facilitated by frequent travel between other parts of the world.

MATERIALS AND METHODS

**Bacterial isolates.** Totals of 106 S. flexneri and 82 S. sonnei strains isolated from stool specimens from patients with diarrhea or dysentery between 1990 and 2017 at the Department of Clinical Microbiology, Christian Medical College, Vellore, India, were included in the study. All isolates were retrieved from the archives and confirmed by a standard protocol (18). Serotype identification was done by the slide agglutination test using polyvalent and monovalent antisera (Denka, Seiken, Japan).

**Antimicrobial susceptibility testing.** Antimicrobial susceptibility testing of the isolates against ampicillin (10 µg), trimethoprim/sulfamethoxazole (1.25/23.75 µg), nalidixic acid (30 µg), norfloxacin (10 µg), ceftoxime (30 µg), cefotaxime (5 µg), and azithromycin (15 µg) was performed using the Kirby-Bauer disc diffusion method. The results were interpreted using breakpoints recommended by Clinical and Laboratory Standards Institute guidelines (19). The quality control strains used were Escherichia coli ATCC 35218 and E. coli ATCC 25922.

**Whole-genome sequencing.** Genomic DNA was extracted from a culture grown overnight using the QIAamp DNA minikit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. The quantity and quality of the DNA were analyzed using the Qubit 3.0 fluorometer (Thermo Fisher, USA) and the Nanodrop spectrophotometer (Thermo Fisher, USA). Paired-end genomic libraries were prepared with unique indexing of each DNA sample and sequenced using the short-read Illumina HiSeq V4 platform according to the manufacturer’s guidelines.

**Publicly available data.** S. flexneri and S. sonnei isolates obtained from human stool samples used in other whole-genome sequence analyses were included for comparative analysis in this study. The
Phylogenetic Analysis of S. flexneri and S. sonnei

associated studies (along with the European Nucleotide Archive [ENA] study accession numbers) are as follows: discovery of sequence diversity in Shigella spp. (PRJEB2128), phylogenomics of Shigella spp. (PRJEB2818), the Shigella C1 D8 project (PRJEB2976), S. flexneri from around the world (PRJEB2542), global fluoroquinolone-resistant S. sonnei genomic study (PRJNA320210), and intercontinental dissemination of azithromycin-resistant Shigella (PRJEB2846). Our study included representatives of S. flexneri phylogenetic groups (PGs) and S. sonnei lineages for analysis (see Table S1 in the supplemental material).

SNP-based phylogenetic analysis. Phylogenetic analyses of S. flexneri and S. sonnei genomes were performed against publicly available global genomes. All S. flexneri and S. sonnei sequences were mapped to the S. flexneri 2a strain 301 and S. sonnei 55046 reference genomes, respectively (GenBank accession numbers CP00038 and NC_004337), using SMALT (version 0.7.4), and SNPs were called against the reference and filtered using SAMtools (20). This resulted in alignments of 5,066 SNPs for S. sonnei and 67,454 SNPs for S. flexneri, which were used for phylogenetic inference. A maximum likelihood (ML) tree was constructed using RAxML v0.7.4 under the generalized time reversible (GTR) substitution model, and bootstrap replicates were determined (21). Pairwise SNP analysis was performed using pairwise_snp_differences (30).

The temporal signal in the ML phylogeny for S. sonnei was investigated using TempEst (http://tree.bio.ed.ac.uk/). The relationship between the root-to-tip distance and the time of isolation was analyzed.

The temporal phylogenetic structure of S. sonnei was determined using Bayesian Evolutionary Analysis by Sampling Trees (BEAST) v.1.10 (22). The recombination-free core-genome SNP alignment file generated by Gubbins was used as the input to compute the mean evolutionary rate of the genomes and time of the most recent common ancestor (MRCA) (23). Trees were visualized in association with metadata using the Web-based Interactive Tree of Life (iTOL) (24). The study isolates were assigned to previously described lineages or PGs based on the clustering described previously by Connor et al. (9). The sequences were also analyzed for virulence, antimicrobial resistance (AMR), and mobile elements.

Pangenome analysis. Pangenome analysis of S. flexneri and S. sonnei isolates from India was performed using Roary v.3.11.2 with default settings (25). The genomes were annotated using Prokka (26). The genes that were common in all compared strains (core genes) and accessory genes were extracted and used to construct a phylogenetic tree. Core genes (99% ≤ strains ≤ 100%), soft core genes (95% ≤ strains < 99%), shell genes (15% ≤ strains < 95%), cloud genes (0% ≤ strains ≤ 15%), and total genes (0% ≤ strains ≤ 100%) were calculated.

Identification of virulence, resistance, and plasmids. Genome data were analyzed for the presence of virulence determinants using VirulenceFinder 1.5 (27). Acquired antimicrobial resistance genes and chromosomal mutations in the QRDR were identified using ResFinder 2.1 (28) with a 90% threshold for identity and 60% minimum length coverage. The presence of plasmids was analyzed using PlasmidFinder 1.3 (29) with a 95% threshold for identity.

Data availability. We declare that all data that support the findings of this study are available within the paper (and its supplemental material) and from publicly available repositories. Raw sequences of Illumina reads generated from this study have been deposited in the ENA under project accession number PRJEB23045. Accession numbers for the sequences submitted to GenBank are provided in Table S1. All published tools used in this work are referenced in Materials and Methods. Further details are available from the corresponding author upon reasonable request.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, TIF file, 0.2 MB.
TABLE S1, XLSX file, 0.04 MB.

ACKNOWLEDGMENTS

We acknowledge the Institutional Review Board of the Christian Medical College, Vellore (83-I/11/13), for approving the study. We thank the sequencing team at the Wellcome Trust Sanger Institute for sequencing the samples.

The study was supported by the Indian Council of Medical Research, New Delhi (reference number AMR/TF/55/13ECDDII dated 23 October 2013).

We declare that no competing interests exist.

REFERENCES


