Differential gene expression profile of *Shigella dysenteriae* causing bacteremia in an immunocompromised individual

Dhiviya Prabaa Muthuirulandi Sethuvel¹, Naveen Kumar Devanga Ragupathi¹, Marilyn M Ninan¹, Joy Sarojini Michael¹, Shalini Anandan¹ & Balaji Veeraraghavan*¹

¹Department of Clinical Microbiology, Christian Medical College, Vellore 632004, India

*Author for correspondence: Tel.: +91 944 221 0555; vbalaji@cmcvellore.ac.in

**Aim:** *Shigella* species has varying levels of virulence gene expression with respect to different sites of infection. In this study, the differential gene expression of *S. dysenteriae* in response to its site of infection was analyzed by transcriptomics. **Methods:** This study includes four clinical *Shigella* isolates. Transcriptomics was done for the stool and blood samples of a single patient. Isolates were screened for the presence of antimicrobial resistance genes. **Results:** The majority of genes involved in invasion were highly expressed in the strain isolated from the primary site of infection. Additionally, antimicrobial resistance (*dhfr*¹A, *sulII*, *bla*OXA, *blaCTX-M-1* and *qnrS*) genes were identified. **Conclusion:** This study provides a concise view of the transcriptional expression of clinical strains and provides a basis for future functional studies on *Shigella* spp.

**Lay abstract:** *Shigella* infection is restricted to the gastrointestinal tract and rarely causes fatal extra-intestinal complications like bacteremia. There are limited studies available from India on molecular characterization of *Shigella* spp. In this study, we characterized four *Shigella* isolates obtained from bloodstream infections. *Shigella* spp. isolated from the stool and blood of one representative patient was further sequenced to study the differential gene expression profile. The differential protein expression by *S. dysenteriae* observed in this study demonstrates that it has a specific response to particular intracellular environments. Further, the *in vivo* mechanism of *Shigellae* invasion are difficult to fully study until the intracellular environment is mimicked *in vitro*. To the best of our knowledge, this is the first Indian study that compared the gene expression profile of clinical *Shigella* strains.

First draft submitted: 23 October 2019; Accepted for publication: 6 January 2020; Published online: 29 January 2020

**Keywords:** gene expression • IcsA • invasive • RNA-Seq analysis • *Shigella*

Diarrheal disease is the second leading cause of mortality in children according to WHO [1]. *Shigella* spp. is one of the important causes of dysentery globally and causes severe and occasionally life-threatening diarrheal infection. In Asia, it is estimated that there are 125 million infections and 14,000 deaths due to shigellosis annually [2]. Clinically, the infection may lead to rare but potentially fatal extra-intestinal complications like bacteremia. Though, bacteremia due to *Shigella* spp. is rare, it is reported in 0.4–7% of the cases. Notably, young age, malnutrition and immunosuppression are known to be the risk factors for *Shigella* spp. bacteremia [3].

Bacteria have developed various mechanisms to adhere to the organ surfaces. Some bacteria can adopt an intracellular lifestyle and get internalized inside various host cell types to replicate. Finally, pathogenic bacteria can get access to deeper tissues using various mechanisms to cross mucosal barriers and access the bloodstream, which is a gateway for all host organs [4].

Pathogens showing a variable expression of virulence factors have been observed. In fact, the expression of virulence factors depends largely on the environmental conditions. This expression of virulence genes is induced under conditions similar to those found at the site of invasion. Studies have demonstrated that a temperature of 37°C is a favorable growth condition for bacteria in intestinal epithelial cells, but bacteria grown at 30°C can be phenotypically avirulent and noninvasive [5]. The bacterium can be found either in the intestinal lumen,
inside epithelial cells, phagocytes or in the bloodstream. The expression level of virulence factors in these different locations varies accordingly in order to counteract different host defense mechanisms, as reported earlier by Ribet and Cossart [4].

In this study, Shigella strains causing bacteremia were characterized using RNA-sequencing to identify genes that are differentially expressed based on the site of infection. The genes responsible for invasion, virulence, stress, antimicrobial resistance (AMR) and other genes involved in cellular metabolism are also discussed.

### Materials & methods

#### Strains

This study reports four cases of *Shigella* bacteremia diagnosed between the years 2015 and 2018. The identified isolates include two isolates each of *S. flexneri* serotype 2 and *S. dysenteriae* serotype 9. The isolates were confirmed by standard biochemical tests [6]. The isolate was serotyped using commercial antisera as per the manufacturer’s instructions (Denka Seiken, Tokyo, Japan). For transcriptome analysis, stool (FC3355) and blood (BA42767) samples of the sole patient were studied further. Patient’s symptoms, clinical diagnosis and outcome were detailed in Table 1. The term invasive (sterile site) and noninvasive (nonsterile site) refer to the pathogen isolation site in this study.

#### Antimicrobial susceptibility testing

Antimicrobial susceptibility testing of isolates against ampicillin (10 μg), trimethoprim/sulphamethoxazole (1.25/23.75 μg), nalidixic acid (30 μg), ciprofloxacin (5 μg), norfloxacin (10 μg), ofloxacin (5 μg), cefpodoxime (10 μg), cefepime (30 μg), cefotaxime (30 μg), cefixime (5 μg), azithromycin (15 μg), imipenem (10 μg), meropenem (10 μg), amikacin (30 μg), gentamicin (10 μg), netilmicin (30 μg) and piperacillin/tazobactam (100/10 μg) was performed using Kirby–Bauer disc diffusion method. The results were interpreted using breakpoints recommended by the Clinical and Laboratory Standards Institute Guidelines 2018 [7]. Quality control strains used were *Escherichia coli* ATCC 35218, *Escherichia coli* ATCC 25922 and *Pseudomonas aeruginosa* ATCC 27853 for the antibiotics tested.

#### AMR genes PCR

Genomic DNA was extracted using the QiaSymphony DNA extraction platform (Qiagen, Hilden, Germany). The isolates were screened for the presence of AMR (dhfr1A, sulII, blaOXA, blaCTX-M-1 and qnrS) genes by PCR as described earlier [8,9].

#### RNA isolation

Total RNA was extracted using RNeasy Mini kit (Cat#74106, Qiagen, GmbH, Germany) according to the manufacturer’s recommendations. The RNA was checked using the Qubit® 3.0 Fluorometric Quantitation kit (Invitrogen, Merelbeke, Belgium).

#### RNA-sequencing & analysis

The invasive traits of selected isolates were studied by comparing the differential gene expression profile of the strain isolated exclusively from stool and blood specimen concurrently by transcriptomics. RNA-sequencing procedure was performed according to the manufacturer’s instructions using Ion Torrent (PGM) sequencer with 400-bp read chemistry (Life Technologies, CA, USA) [10]. The quality and quantity of each library was determined at each
Transcriptomic analysis of *Shigella* causing bacteremia  

Table 2. Antimicrobial resistance profile of the study isolates.

| Sample ID | Species       | Resistant profile | AMR genes                        |
|-----------|---------------|-------------------|----------------------------------|
| BA12827   | *S. flexneri* 2 | R– CD, CP, MS - GEN, AK, P/T | blaqXX, sul1, dhfr1a, qnrS, blaCTX-M-1 |
| BA42767   | *S. dysenteriae* 9 | R – NAL, GEN, AK | –                                |
| BA21871   | *S. dysenteriae* 9 | R – SXT, MS - CP, OFL, TAX, FEP, NAL | sul1, dhfr1a, qnrS                 |
| BA10746   | *S. flexneri* 2 | R – AMP            | –                                |

AMP: Ampicillin; AK: Amikacin; BA: Blood; CIP: Ciprofloxacin; CD: Cefpodoxime; FC: Feces; FEP: Cefepime; GEN: Gentamicin; MS: Moderate susceptible; NAL: Nalidixic acid; OFL: Ofloxacin; P/T: Piperacillin/tazobactam; R: Resistant; SXT: Trimethoprim/sulfamethoxazole; TAX: Cefotaxime.

step with a Qubit® 3.0 Fluorometer. *De novo* assembly using AssemblerSPAdes and annotation through RNA-Seq analysis was performed in PATRIC, the bacterial bioinformatics database and analysis resource.

**Statistical analysis**

In this study, greater than twofold changes in the gene expression level between two variables was considered significant. Results were analyzed for correlation and tested for significance by Student’s t-test (p < 0.05). SPSS 16.0 and Microsoft Excel 2007 (IL, USA) were used for the statistical evaluation.

**Results & discussion**

*Shigella* infection is in the majority confined to the GI tract which invades the colonic mucosa but rarely penetrates further into deeper tissues [11]. This study discusses four cases of *Shigella* bacteremia. The risk factors observed in these patients were diabetes, malignancy and immunosuppressant therapy. Previous literatures on mechanisms of pathogenesis have been described for *S. flexneri* However, the present study has shown the invasion process of *S. dysenteriae* serotype 9.

**AMR gene PCR**

Among the four isolates studied, only two isolates harbored AMR genes that codes for β-lactams, trimethoprim/sulfamethoxazole, fluroquinolones and cephalosporins, whereas no AMR genes were identified in the other two isolates. The resistance genes obtained in this study were found to be a common profile seen in the genus. The results were given in Table 2. AMR was generally more common in *Shigella* than in other enteric bacteria [12].

**Differential gene expression analysis**

*S. dysenteriae* serotype 9 obtained from stool and blood specimen of the single patient was studied. In RNA-Seq analysis, significant fold change was observed between non-invasive Sd_FC3355 and invasive Sd_BA42767 strains for the genes involved in invasion, virulence, motility and other cellular processes. Totally 56 genes were differentially expressed between the strains. Of these, few genes were expressed only in invasive strain Sd_BA42767 like csp, dcm, hisE and enterotoxin genes with reduced expression, this showed the significance of these genes in the invasive phenotype of the strain. The majority of the genes (44/56 genes) were highly expressed in non-invasive isolate from the gut, which is the primary site of invasion for *Shigella* infection. Genes with no expression data were excluded from the analysis. The genes analysed were given as a supplementary material.

**Motility-associated genes**

*Shigella* pathogenesis involves bacterial invasion and spread through colonic mucosa [13]. *Shigella* spp. are able to move through the cytoplasm of host cells and into adjacent cells by polymerizing actin [14] which is mediated by IcsA (virG), encoded on the 220-kb virulence plasmid [15,16]. We observed that IcsA protein was expressed only in noninvasive *Shigella* isolate (Table 3). This correlates with the fact that IcsA is required for inter- and intracellular spreading of *Shigella* within the host intestinal epithelium. VirK gene, which is required for post-transcriptional regulation of icsA expression, has also been reported.

**Virulence/invasion associated genes**

*Shigella* virulence plasmid is an essential virulence determinant of the species and encodes the molecular machinery necessary for tissue invasion and intracellular survival. The virulence plasmid encodes the 30 kb Mxi-Spa type III secretion system (T3SS) and invasion plasmid antigens (Ipa proteins) required for invasion of the colonic and rectal
| Genes          | Product                                                                 | Fold change |
|---------------|-------------------------------------------------------------------------|-------------|
| SDY_0834/ipaH | Invasion plasmid antigen/Internalin, putative                          | 6           |
| SDY_1062/ipaH | Invasion plasmid antigen/Internalin, putative                          | 2           |
| SDY_2001/ipaH | Invasion plasmid antigen/Internalin, putative                          | 3           |
| SDY_2003/ipaH | Invasion plasmid antigen/Internalin, putative                          | 11          |
| SDY_2753/ipaH | Invasion plasmid antigen/Internalin, putative                          | 26          |
| SDY_P003/ospB | Hypothetical protein                                                   | 5           |
| SDY_P004/phon2| Hypothetical protein                                                   | 18          |
| SDY_P010/ospD2| Enterotoxin                                                             | 40          |
| SDY_P023/ospD1| OspD1                                                                   | 361         |
| SDY_P025/ipgB2| Putative chaperone (ipgB2)                                             | 1672        |
| SDY_P037/ipaH4.5| Invasion plasmid antigen/Internalin, putative                        | 20          |
| SDY_P038/ipaH7.8| Invasion plasmid antigen/Internalin, putative                        | 23          |
| SDY_P045/ipaH1.4| Invasion plasmid antigen/Internalin, putative                        | 21          |
| SDY_P055/ospC1| Hypothetical protein                                                   | 15          |
| SDY_P056/ospD3| Enterotoxin                                                             | 16          |
| SDY_P070/ospC2| Hypothetical protein                                                   | 1111        |
| SDY_P099/ipaH9.8| Invasion plasmid antigen/Internalin, putative                        | 17          |
| SDY_P109/virK | Virulence factor VirK                                                 | 365         |
| SDY_P110/ospC3| Hypothetical protein                                                   | 224         |
| SDY_P110/ospC3| Hypothetical protein                                                   | 1170        |
| SDY_P130/ospC3| Hypothetical protein                                                   | 76          |
| SDY_P136/ospC3| Hypothetical protein                                                   | 54          |
| SDY_P140/ospC3| Hypothetical protein                                                   | 179         |
| SDY_P145/ospC3| Hypothetical protein                                                   | 1244        |
| SDY_P150/ospC3| Hypothetical protein                                                   | 1582        |
| SDY_P160/ospC3| Hypothetical protein                                                   | 937         |
| SDY_P169/ospC3| Hypothetical protein                                                   | 613         |
| SDY_P170/ospC3| Hypothetical protein                                                   | 495         |
| SDY_P171/ospC3| Hypothetical protein                                                   | 3399        |
| SDY_P172/ospC3| Hypothetical protein                                                   | 3235        |
| SDY_P173/ospC3| Hypothetical protein                                                   | 1611        |
| SDY_P186/ospC3| Hypothetical protein                                                   | 1450        |
| SDY_P191/ospC3| Hypothetical protein                                                   | 1457        |
| SDY_P192/ospC3| Hypothetical protein                                                   | 1476        |
| SDY_P193/ospC3| Hypothetical protein                                                   | 334         |
| SDY_P194/ospC3| Hypothetical protein                                                   | 355         |
| SDY_P195/ospC3| Hypothetical protein                                                   | 160         |
| SDY_P196/ospC3| Hypothetical protein                                                   | 483         |
| SDY_P197/ospC3| Hypothetical protein                                                   | 308         |
| SDY_P198/ospC3| Hypothetical protein                                                   | 233         |
| SDY_P199/ospC3| Hypothetical protein                                                   | 116         |
| SDY_P200/ospC3| Hypothetical protein                                                   | 71          |
| SDY_P201/ospC3| Hypothetical protein                                                   | 28          |

0: Not expressed.
Table 3. Gene expression profile of the two selected isolates represented in fold change (cont.).

| Genes          | Product                                                                 | Fold change |
|----------------|-------------------------------------------------------------------------|-------------|
| SDY_P193/spa40 | Type III secretion innermembrane protein (YscU, SpaS, EsCU, HrcU, SsaU, homologous to flagellar export components) | 33          |
| SDY_P211/virA | Hypothetical protein                                                    | 43          |
| SDY_P214/icsA | Hypothetical protein                                                    | 469         |
| SDY_P224/icsP | Protease VII (Omptin) precursor (EC 3.4.23.49)                          | 461         |
| SD1617,4624/  | Virulence factor MviM                                                   | 0           |
| SD1617,3340/  | Enterotoxin                                                             | 0           |
| /              | Enterotoxin                                                             | 0           |
| SD1617,0737/ivB| Acetolactate synthase large subunit (EC 2.2.1.6)                        | 81          |
| SD1617,0939/ivD| Dihydroxy-acid dehydratase                                             | 91          |
| SD1617,0940/ivA| Threonine dehydratase biosynthetic (EC 4.3.1.19)                       | 59          |
| SD1617,0942/ivC| Ketol-acid reductoisomerase (EC 1.1.1.86)                               | 133         |
| SD1617,0938/ivE| Branched-chain amino acid aminotransferase (EC 2.6.1.42)               | 58          |
| SD1617,3738/ivN| Acetolactate synthase small subunit (EC 2.2.1.6)                       | 41          |
| /              | IvBN operon leader peptide                                             | 0           |
| SDY_2022/phoP | Transcriptional regulatory protein PhoP                                  | 222         |
| SDY_2023/phoQ | Sensor histidine kinase PhoQ (EC 2.7.13.3)                              | 136         |
| SDY_3003/barA | Signal transduction histidine-protein kinase BarA (EC 2.7.13.3)         | 24          |
| SDY_1104/uvrY | BarA-associated response regulator UvrY (= GacA = SirA)                 | 261         |
| SDY_2892/csrA | Carbon storage regulator                                               | 1021        |
| SD1617,4387/hsF| Histidinol-phosphate synthase cyclase subunit (EC 4.1.3)                | 0           |
| SD1617,4390/hsB| Histidinol-phosphatase (EC 3.1.3.15)/imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19) | 28          |
| SD1617,4391/hsC| Histidinol-phosphate aminotransferase (EC 2.6.1.9)                      | 0           |
| SD1617,4388/hsA| Phosphoribosylformimino-5-aminomidazole carboxamide ribotide isomerase (EC 5.3.1.16) | 0           |
| SD1617,4392/hsD| Histidinol dehydrogenase (EC 1.1.1.23)                                 | 0           |
| SD1617,4386/hsE| Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19)/phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) | 0           |
| SD1617,4393/hsG| ATP phosphoribosyltransferase (EC 2.4.2.17) - HisGl                    | 0           |
| SDY_221S/hsH | Imidazole glycerol phosphate synthase amidotransferase subunit (EC 2.4.2) | 0           |
| SD1617,4262/  | Cold shock protein of CSP family - CspA (naming convention as in E. coli) | 0           |
| SDY_2381/cspD | Cold shock protein CspD                                                | 0           |
| SDY_0546/cspE | Cold shock protein CspE                                                | 0           |
| SD1617,4774/  | Cold shock protein of CSP family - CspC (naming convention as in E. coli) | 0           |
| SDY_4448/groES | Heat shock protein 60 family co-chaperone GroES                         | 3531        |
| SDY_4449/groEL | Heat shock protein 60 family chaperone GroEL                            | 3994        |
| SDY_4172/ibpB | 16 kDa heat shock protein B                                            | 146         |
| SDY_4173/ibpA | 16 kDa heat shock protein A                                            | 190         |
| SDY_2787/grpE | Heat shock protein GrpE                                               | 491         |
| /              | Heat shock protein C                                                   | 6           |
| SDY_3677/hslO | 33 kDa chaperon (Heat shock protein 33) (HSP33)                         | 274         |
| SD1617,5932/dcm| DNA-cytosine methyltransferase (EC 2.1.1.37)                            | 0           |
| SDY_4150/uhpA | Transcriptional regulatory protein UhpA                                 | 0           |
| SDY_4659/creB | Response regulator CreB of two-component signal transduction system CreBC | 0           |
| SDY_4658/creA | Conserved uncharacterized protein CreA                                  | 0           |
| SDY_4477/evgA | Positive transcription regulator EvgA                                   | 176         |
| SDY_4478/evgS | Hybrid sensory histidine kinase in two-component regulatory system with EvgA | 12          |
| SDY_3723/hydH | Sensor protein of zinc sigma-54-dependent two-component system         | 68          |
| SDY_3727/hydG | Response regulator of zinc sigma-54-dependent two-component system     | 76          |

0: Not expressed.
**Table 3. Gene expression profile of the two selected isolates represented in fold change (cont.).**

| Genes       | Product                                                                 | Fold change |
|-------------|-------------------------------------------------------------------------|-------------|
| SDY_1275/narL | Nitrate/nitrite response regulator protein NarL                         | Sd_FC33SS: 66  | Sd_BA42767: 19 |
| SDY_1276/narX | Nitrate/nitrite sensor protein NarX                                      | Sd_FC33SS: 31  | Sd_BA42767: 34 |
| SDY_3874/ glmL | Nitrogen regulation protein NtrB (EC 2.7.13.3)                          | Sd_FC33SS: 69  | Sd_BA42767: 67 |
| SDY_3875/ glmG | Nitrogen regulation protein NtrC                                        | Sd_FC33SS: 82  | Sd_BA42767: 106 |
| SDY_3214/ygiX | Two-component system response regulator QseB                           | Sd_FC33SS: 25  | Sd_BA42767: 33 |
| SDY_3213/qseC | Sensory histidine kinase QseC                                           | Sd_FC33SS: 21  | Sd_BA42767: 24 |
| SDY_0856/rcsC | Sensor histidine kinase RcsC (EC 2.7.13.3)                              | Sd_FC33SS: 32  | Sd_BA42767: 32 |
| SDY_0857/rcsB | DNA-binding capsular synthesis response regulator RcsB                 | Sd_FC33SS: 870 | Sd_BA42767: 899 |
| SDY_1824/rstA | Transcriptional regulatory protein RstA                                 | Sd_FC33SS: 86  | Sd_BA42767: 54 |
| SDY_1825/rstB | Sensory histidine kinase in two-component regulatory system with RstA   | Sd_FC33SS: 60  | Sd_BA42767: 56 |
| SDY_2184/yfhA | Transcriptional response regulatory protein GlrK                         | Sd_FC33SS: 44  | Sd_BA42767: 41 |
| SDY_2186/yfhK | Sensor histidine kinase GlrK                                            | Sd_FC33SS: 64  | Sd_BA42767: 48 |
| SDY_2443/dcuA | C4-dicarboxylate transporter DcuA                                       | Sd_FC33SS: 225 | Sd_BA42767: 250 |
| SDY_2186/baeR | Response regulator BaeR                                                | Sd_FC33SS: 53  | Sd_BA42767: 34 |
| SDY_2187/baeS | Sensory histidine kinase BaeS                                           | Sd_FC33SS: 6   | Sd_BA42767: 1  |
| SDY_1046/vsr | Very-short-patch mismatch repair endonuclease (Guanine–Thymine [G–T] specific) | Sd_FC33SS: 0 | Sd_BA42767: 28 |
| SDY_1047/yedA | Uncharacterized innermembrane transporter YedA                          | Sd_FC33SS: 0  | Sd_BA42767: 0  |
| SDY_1048/yedl | Innermembrane protein Yedl                                             | Sd_FC33SS: 0  | Sd_BA42767: 0  |
| SDY_1970/yobF | Uncharacterized protein YobF                                            | Sd_FC33SS: 30  | Sd_BA42767: 2205 |

0: Not expressed.

epithelial cells and cell-to-cell spread of the bacteria, resulting in the symptoms of bacillary dysentery [17,18]. *Shigella* pathogenesis mainly relies on the Mxi-Spa T3SS and its effector proteins [19]. The invasion plasmid antigen (*ipaH*) gene, which was reported to be carried by all four *Shigella* species, was found to be highly expressed in invasive isolate in this study, whereas *ipaD*, a host injection protein was expressed only in noninvasive isolate. Further, *ipgA, B, C, D, F* known to facilitate local invasion in to epithelial cells, were also expressed only in noninvasive isolates (Table 3). Therefore, the virulence plasmid is the key molecular signature of *Shigella* spp. pathogenesis and is fundamental for initiating infection and manipulating the immune response of the host [18].

PhoQ/PhoP is a two-component system that governs virulence and regulates several cellular activities in *Shigella* spp. [20]. In the present study, PhoP was highly expressed in invasive isolate, whereas PhoQ showed no significant difference in the expression level. In addition, BarA-UvrY two-component system was shown to have increased expression in invasive isolate. This system also controls the activity of CsrA (carbon storage regulator) protein which regulates carbon metabolism, flagellar biosynthesis and biofilm formation. This process has been previously reported in uropathogenic *E. coli* [21]. We observed that CsrA protein was upregulated in noninvasive isolate.

**Stress-associated genes**

Bacteria have developed a number of mechanisms to adapt the changing environmental conditions within the cells. One such mechanism is the production of small cold shock proteins (Csp) to counteract the sudden temperature downshift. CspS have been shown to contribute to osmotic, oxidative, starvation, pH and ethanol stress tolerance as well as to host cell invasion [22]. CspA is a major cold shock protein, first described in *E. coli* [22] was found to have significant differences in the expression level between the invasive and noninvasive isolate. Similarly, CspD and CspE proteins showed significant differences in their expressions, whereas CspC was highly and solely expressed in invasive isolate (Table 3). Another defense mechanism against various environmental stresses is the production of heat shock proteins. Heat shock proteins that are important for cell survival and are usually related to the virulence of the pathogens have been expressed in both the isolates [23].

**Genes involved in metabolism**

In this study, *ile* proteins such as *ileA, B, C, D, E* and *N* involved in amino acid biosynthesis showed significantly increased expression in invasive isolate. Histidine (*his*) proteins like *hisA, B, C, D, E, F, G* and *H* were found to have...
Transcriptomic analysis of Shigella causing bacteremia  
Short Communication

Significant upregulation in invasive isolate. Further, member of the two-component regulatory system NtrB/NtrC and other regulator proteins like NarL and NarX involved in the regulation of nitrogen was expressed in both the isolates with no significant difference in the expression level. Similarly, several other genes such as (ygiX, qseC, rcsC, rcsB, rrsA, rrsB, yfbA, yfbK, dcuA, baeR, baeS, vsr) were present but showed no significant difference between the isolates.

**Cellular process & signaling**

During Shigella infection, certain effector proteins promote cell survival. IpgD which associated with increased intracellular bacterial replication [24] was highly and solely expressed in noninvasive isolate as expected. Further ospC and virA were also found to be expressed in noninvasive isolate [24]. DNA methylation is an important component in numerous cellular processes and plays an important role in regulating gene expression [25,26]. DNA cytosine methyltransferase protein was only slightly expressed in invasive isolate in this study.

**Uncharacterized genes**

Two genes encoding uncharacterized proteins were identified. Uncharacterized innermembrane transporter YedA gene was not expressed in the study isolates, which has been previously identified as hypothetical protein in S. dysenteriae strain Sd197. Another gene named YobF, which is a small protein with no known function showed significantly increased expression in invasive isolate. Yet the functions of these genes remain obscure.

**Conclusion & future perspective**

Shigella spp. is a highly contagious pathogen and humans are the only reservoir that spreads through fecal–oral contamination. The invasive ability of this pathogen is a key determinant in the establishment of the disease. The invasive phenotype of Shigella spp. is linked to the expression of various effector/regulatory genes. The differential protein expression by S. dysenteriae serotype 9 observed in this study suggests that it has a specific response to particular intracellular environment. Notably, many uncharacterized genes with unknown functions demonstrate the complexity of the regulatory network in S. dysenteriae. These genes needs to be further characterized to understand unidentified strategies for infection and successful survival of this pathogen. Further, the *in vivo* mechanism of S. dysenteriae invasion are difficult to fully study until the intracellular environment is mimicked *in vitro*. To the best of our knowledge, this is the first Indian study that compares the gene expression profile of clinical S. dysenteriae serotype 9 with respect to their invasion.

**Executive summary**

- Most of the earlier studies on mechanisms underlying pathogenesis was derived from Shigella flexneri. However, the present study shows the invasion process of Shigella dysenteriae serotype 9.
- RNA sequencing was done to study the differential expression of genes involved in the invasion process of the pathogen with the respect to the infection site.
- On virulence analysis, enterotoxin gene (set) and invasion associated genes such as ipaH and ial was identified in two, one and three isolates, respectively.
- For antimicrobial resistance, only two isolates harbored genes that codes for β-lactams, trimethoprim/sulfamethoxazole, fluoroquinolones and cephalosporins resistance.
- RNA-Seq analysis showed significant fold change between noninvasive Sd_FC3355 and invasive Sd_BA42767 strains for the genes involved in invasion, virulence, motility and other cellular processes.
- Majority of the genes (44/56 genes) were highly expressed in noninvasive isolate, which is the primary site of invasion for Shigella spp. Few genes were expressed only in invasive isolate Sd_BA42767, which shows the significance of these genes in the invasive phenotype of the strain.
- This study explores that Shigella spp. has a specific response to particular intracellular environment.
- The identification of genes with uncharacterized functions demonstrates the complexity of the regulatory network in S. dysenteriae.

**Author contributions**

B Veeraraghavan and S Anandan conceptualized the study. DP Muthuirulandi Sethuvel and NK Devanga Ragupathi analyzed, interpreted data and wrote the manuscript. DP Muthuirulandi Sethuvel carried out bench work and generated data. B Veeraraghavan, MM Ninan and JS Michael critically revised and approved the manuscript. All authors read and approved the manuscript.
Acknowledgments
The authors gratefully acknowledge the Institutional Review Board for approving the study and the department of Clinical Microbiology for providing lab space and facilities.

Financial & competing interests disclosure
The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.
No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research
The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Open access
This work is licensed under the Creative Commons Attribution 4.0 License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

Nucleotide sequence accession number
The raw sequence data were submitted to the National Center for Biotechnology Information Sequence Read Archive under Accession No. SRR6031691 (BA42767) and SRR6031692 (FC3355).

References
Papers of special note have been highlighted as: • of interest; •• of considerable interest
1. Sambe-Ba B, Espie E, Faye ME et al. Community-acquired diarrhea among children and adults in urban settings in Senegal: clinical, epidemiological and microbiological aspects. BMC Infect. Dis. 13, 580 (2013).
2. Thompson CN, Duy PT, Baker S. The rising dominance of Shigella sonnei: an intercontinental shift in the etiology of bacillary dysentery. PLoS Negl. Trop. Dis. 9, e0003708 (2015).
3. Jain S, Sharma M, Gupta R, Shree N, Kumar M. Multidrug resistant Shigella flexneri: a rare case of septicemia in an infant. J. Clin. Diagn. Res. 8, DD03 (2014).
4. Ribet D, Cossart P. How bacterial pathogens colonize their hosts and invade deeper tissues. Microb. Infect. 17, 173–183 (2015).
•• Provides the detailed description of pathogenesis of Shigella in humans.
5. Zhao G, Zhu L, Feng E et al. A novel anti-virulence gene revealed by proteomic analysis in Shigella flexneri 2a. Proteome Sci. 8, 30 (2010).
6. Bopp CA, Brenner FW, Fields PL et al. Escherichia, Shigella, and Salmonella. In: Manual of Clinical Microbiology (8th Edition). Murray PR, Baron EJ, Jorgensen J, Pfaller MA, Yolken RH (Eds). American Society for Microbiology, DC, USA, 654–671 (2003).
7. CLSI, Clinical and Laboratory Standards Institute. Performance standards for antimicrobial susceptibility testing; twenty-fourth informational supplement M100-S28. Clinical and Laboratory Standards Institute, MI, USA (2018).
8. Thong KL, Hoe SLL, Puthucheary SD, Yasin RM. Detection of virulence genes in Malaysian Shigella species by multiplex PCR assay. BMC Infect. Dis. 5, 8 (2005).
• Describes the application of multiplex PCR for the detection of important Shigella virulence genes.
9. Anandan S, Sethuvel DPM, Gajendiren R, Verghese VP, Walia K, Veeraraghavan B. Molecular characterization of antimicrobial resistance in clinical Shigella isolates during 2014 and 2015: trends in South India. Germs 7, 115 (2017).
• Reports the prevalence of various antimicrobial resistance genes among Shigella spp. in India.
10. Castro TL, Seyffert N, Ramos RT et al. Ion Torrent-based transcriptional assessment of a Corynebacterium pseudotuberculosis equi strain reveals denaturing high-performance liquid chromatography a promising rRNA depletion method. Microb. Biotechnol. 6, 168–177 (2013).
11. Carayol N, Van Nhieu GT. The inside story of Shigella invasion of intestinal epithelial cells. Cold Spring Harb. Perspect. Med. 3, a016717 (2013).
12. Mamatha B, Rituparna C. Decreased susceptibility to antimicrobials among Shigella flexneri isolates in Manipal, South India – a 5 year hospital based study. Southeast Asian J. Trop. Med. Public Health 43, 1447 (2012).
13. Goldberg MB, Therriot JA. Shigella flexneri surface protein IcsA is sufficient to direct actin-based motility. Proc. Natl Acad. Sci. USA 92, 6572–6576 (1995).
14. Magdalena J, Goldberg MB. Quantification of Shigella IcsA required for bacterial actin polymerization. *Cell Motil. Cytoskeleton* 51, 187–196 (2002).

15. Purdy GE, Fisher CR, Payne SM. IcsA surface presentation in *Shigella flexneri* requires the periplasmic chaperones DegP, Skp, and SurA. *J. Bacteriol.* 189, 5566–5573 (2007).

16. Teh MY, Morona R. Identification of Shigella flexneri IcsA residues affecting interaction with N-WASP, and evidence for IcsA–IcsA co-operative interaction. *PLoS ONE* 8, e55152 (2013).

17. Yang F, Yang J, Zhang X et al. Genomic signatures of Shigella evolution, adaptation and geographical spread. *Nat. Rev. Microbiol.* 14, 235 (2016).

18. Yang SC, Hung CF, Aljuffali IA, Fang JY. The roles of the virulence factor IpaB in Shigella spp. in the escape from immune cells and invasion of epithelial cells. *Microbiol. Res.* 181, 43–51 (2015).

20. Cai X, Zhang J, Chen M et al. The effect of the potential PhoQ histidine kinase inhibitors on Shigella flexneri virulence. *PLoS ONE* 6, e23100 (2011).

21. Palaniyandi S, Mitra A, Herren CD et al. BarA-UvrY two-component system regulates virulence of uropathogenic *E. coli* CFT073. *PLoS ONE* 7, e31348 (2012).

22. Keto-Timonen R, Hietala N, Palonen E, Hakakorpi A, Lindstrom M, Korkeala H. Cold shock proteins: a minireview with special emphasis on Csp-family of enteropathogenic Yersinia. *Front. Microbiol.* 7, 1151 (2016).

23. Harikrishnan H, Ismail A, Singh KKB. Temperature-regulated expression of outer membrane proteins in *Shigella flexneri*. *Gut Pathog.* 5, 38 (2013).

24. Lu R, Herrera BB, Eshleman HD et al. Shigella effector OspB activates mTORC1 in a manner that depends on IQGAP1 and promotes cell proliferation. *PLoS Pathog.* 11, e1005200 (2015).

25. Lim DH, Maher ER. DNA methylation: a form of epigenetic control of gene expression. *Obstet. Gynecol.* 12, 37–42 (2010).

26. Gomez-Eichelmann MC, Levy-Mustri A, Ramirez-Santos J. Presence of 5-methylcytosine in CC (A/T) GG sequences (Dcm methylation) in DNAs from different bacteria. *J. Bacteriol.* 173, 7692–7694 (1991).