Proteomic Analysis of Virulence Profiles in Clinical Strains of *Shigella flexneri* in Manaus – Amazon State

Fernanda Batalha

Programa de Pós Graduação em Biologia da Interação Patógeno – Hospedeiro, Instituto Leônidas e Maria Deane – ILMD- Fiocruz

Isabelle B. Cordeiro

Universidade Federal do Amazonas

Paula Taquita

Programa de Pós Graduação em Biologia Celular e Molecular (PGBMC/IOC/FIOCRUZ)

João Victor Verçosa

Universidade Federal do Amazonas

Alessandra Cavalcante

Programa de Pós Graduação em Biologia da Interação Patógeno – Hospedeiro, Instituto Leônidas e Maria Deane – ILMD- Fiocruz

Ivanildes Santos

Universidade Federal do Amazonas

Daniel Figueiredo

Programa de Pós Graduação em Biologia da Interação Patógeno – Hospedeiro, Instituto Leônidas e Maria Deane – ILMD- Fiocruz

Ruth Moura Souza

Programa de Pós Graduação em Biologia da Interação Patógeno – Hospedeiro, Instituto Leônidas e Maria Deane – ILMD- Fiocruz

Paloma Dantas

Universidade Federal do Amazonas

Patrícia Orlandi (✉ patricia.orlandi@fiocruz.br)

Instituto Aggeu Magalhães – IAM – Fiocruz

Research Article

Keywords: Amazônia, effector proteins, proteome, *Shigella*, virulence

Posted Date: March 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-358446/v1>

Abstract

Background: *Shigella* is a Gram-negative bacterium and belongs to Enterobacteriaceae family. These bacteria have been described as responsible for many diarrheic infections around the world and affects children as 5 years old. As a striking feature of these bacteria, we can say about the invasive capacity and the severe damage in the intestine of the host. Epidemiological studies conducted during 2007 to 2009 by our research group, Diagnosis and Control of Infectious Diseases of the Amazon - DCDIA, identified *Shigella* as the 4th most frequent bacterial pathogen in children with diarrhea treated in public hospitals in Manaus - AM. To understand the mechanisms of pathogenesis of these clinical strains, and to describe the mechanisms of cellular invasion, this study proposes the identification of the proteome of two isolates, through the mass spectrometry coupled to liquid chromatography. The clinical strains were submitted to experimental conditions that mimic the epithelial cellular contact in the host, using the inductor Congo Red, in order to investigate which proteins are being produced by this pathogen. Results: The proteomic profile of *Shigella* strain 201 reveals 386 intracellular proteins cultivated in LB medium and 189 intracellular proteins cultivated using the Congo Red inductor. For the M90T strain, a total of 470 intracellular proteins were detected in LB medium and 383 intracellular proteins cultivated with Congo Red. The findings reveal that proteins exclusively induced by Congo Red in the clinical strain are related to virulence processes, such as IpaC and IpaD proteins, which have already been extensively investigated in the literature. Conclusions: However, new target proteins are pointed out, such as Hmp, YkfE, AepA, MobC, MetK, OsmY, LptA and LuxS which are classified as proteins predicted as pathogenic, based on our analyses. Although such proteins are involved in the virulence of enteric pathogens, their functions are still little explored or inexistent for the *Shigella* genus, mainly in the northern region of Brazil. We expected this work to reveal the mechanisms underlying the isolated clinical strains and elucidate new effectors and how they modulate the pathogenesis of these bacteria.

Background

WHO reports that diarrheal diseases mainly affect children under 5 years of age that do not have access to basic sanitation and clean water. It is estimated that around 1.5 million deaths occur annually due to diarrheal diseases and are considered a disease of child morbidity and mortality (1–3). The successive episodes of diarrhea compromise the socioeconomic status of endemic areas and the physical and intellectual development of affected children. The consequences of long-term continuous infections are drastic and are manifested, for example, by deficiencies in growth and increased susceptibility to other pathogens and chronic diseases such as diabetes (4,5)

Among the main pathogens causing diarrhea, *Shigella* is observed in patients with this infectious condition (6). This genus has shown constant resistance to antimicrobials in different countries and a higher incidence of cases in low-income countries, in addition, the spread of strains with a high capacity to transfer virulence factors and capable of causing severe damage to the host have been worrying factors, since there is still no effective vaccine against *Shigella* spp. Although this context is global, in Brazil, more specifically in the Northern region of the country, few studies aim to understanding the

different aspects of this pathogen, either in epidemiological surveillance, in understanding mechanisms related to the pathogenesis or in strategies for its diagnosis, control and treatment.

Epidemiological studies conducted in 2007-2009 by our research group, Diagnosis and Control of Infectious Diseases of the Amazon - DCDIA, identified *Shigella* as the 4th most frequent bacterial pathogen in children with diarrhea treated in public hospitals in Manaus - AM. The group also verified the presence of virulence genes among the isolated samples, which were related to the symptoms presented by the patients as Shet1B and Shet2 genes. Besides these, other genes were also detected, such as IpaBCD, IpaH7.8, set-1A, set-1B, sen/ospD3, virF, e invE (7).

Based on previous researches of the group, among the evaluated isolates, we highlight the immunogenic potential of the strain *S. flexneri* strain 201, which showed greater invasive capacity, greater lethality and morphological changes in the infected tissue such as hemorrhage, intense cell infiltration and with destruction of bronchial and alveolar epithelia, when compared to the reference strain in the study. In addition, this isolate presented a gene expression similar to the reference strain for the main effectors associated with the cell invasion process (8,9). However, since it is not yet known the totality of the effectors that are secreted and their respective functions, the whole mechanism of *Shigella* pathogenesis continues to be the target of new investigations, mainly through the proteomics approach (10).

Since not all the effectors that are secreted by this pathogen and that a part of a panel of proteins with the most different functions are known (11). Studies using in vivo and in vitro models, attenuated or mutant strains have contributed to the development of strategies aimed at the production of vaccines against shigellosis has provided a basis for the understanding of the pathogenesis of this microorganism (12,13).

In this study, we propose the identification of proteins expressed by these clinical strains with different virulence profiles during contact with Congo Red (CR) using the proteomic approach. We believe that the identification of the proteomic profile has the potential to provide knowledge on how the genetic differences presented by the strains of the study reflect on the proteome, giving the basis for future research in understanding the mechanism of invasion by these clinical strains.

Methods

Bacterial strain and Growth Condition

Shigella flexneri 5a M90T used as reference strain and *Shigella flexneri* strain 201 isolated from the patient as previously described (7), obtained from Hospital in Manaus. The isolates were kept frozen in 60% glycerol at -80°C before to the experiments. They were subcultivated in Luria-Bertani (LB) medium grown, pH 7.0, at 37°C with constant agitation for 24 hours. Before each experiment, the colonies were inoculated in fresh nutrient media (50 mL) with a 1.5×10^8 CFU of the overnight culture, with or without the Congo Red (CR) stain and grown at 37 °C with vigorous agitation.

Bacterial Growth curve

The growth (measured in $\log_{10}$ CFU/mL) of *Shigella spp.* strains were performed in triplicate in LB medium broth. The growth curve of the *Shigella flexneri* 5a M90T control was compared with those *Shigella flexneri* 201 strains, both with (0,01%) and without CR staining. OD_{600nm} was measured at 1-hour intervals until the decline phase.

Protein extraction, precipitation and quantification

For intracellular proteins: the cells were pelleted and washed with 3 mL (Tris-HCl 50mM, pH 7.5), centrifuged for 15 min, 42727 xg at 4 °C. Then, the pellet was resuspended in 1 mL of lysis buffer (7M Urea, 2M thiourea, 4%CHAPS, 50 mM DTT – Dithiothreitol, 1 mM PMSF - Phenylmethylsulfonyl fluoride) and sonicated in three cycles of 30 s, at 4°C. The samples were incubated for 1 hour in ice, then centrifuged for 30 min, 114462 xg at 4°C. Each 100 µL of supernatant was redistributed into microtubes and precipitated with 5V of methanol. The pellet was resuspended in 1 mL methanol, centrifuged for 15 min, 12.000 rpm at 4°C twice. After washes, the pellet was dried in room temperature and stored at -20°C. Three samples were collected to perform quantification using 2D Quant Kit (GE Healthcare), as manufacturer's instructions.

1-DE gel electrophoresis and analysis

For 1-DE gel analysis, 15 µg of intracellular protein were loaded on to an 12% SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) gel of size 7.2 x 8.6 cm and 1 mm thick (BIO-RAD, USA). Gels were then stained with Coomassie blue R-350 and the revealed lanes were cut in a linear fashion from top to the bottom into 9-12 slices of approximate size 0.4 cm x 1.2 cm x 1 mm and subjected to in-gel digestion.

In-gel digestion of intracellular proteins

The gel slices were submitted to in-gel digestion protocol (14) with modifications. Briefly, the gel slices were destained in 500 µL of solution 1 (50% methanol, 2,5% acetic acid) for 3 hours at room temperature and dehydrated with 200 µL of acetonitrile for 5 minutes. The gel slices were incubated in 50 µL DTT 10 mM at room temperature for 30 min, then 50 µL Iodoacetamide (IAA) 50 mM for same time and temperature, in the dark. The gel slices were washed with ammonium bicarbonate 100 mM for 10 min. The solution was removed, and the gel slices were dehydrated in 200 µL of acetonitrile for 10 minutes and rehydrated with ammonium bicarbonate 100 mM for 10 minutes. This step was repeated once and after dehydrating the samples with acetonitrile, the residual volume was evaporated in freeze drier. A trypsin solution (20ng/µL) was added in each microtube and the slices were rehydrated for 30 minutes in ice. The solution excess was removed and for in-gel digestion, the samples were incubated overnight with 20 µL ammonium bicarbonate 50 mM at 37°C. In the other day, the slices were submitted to peptides extraction. The slices were incubated in 10 µL formic acid 5% solution for 10 minutes and the supernatant transferred to another microtube. The second step was incubated the gel slices in 12 µL 5%

formic acid/50% acetonitrile solution for 10 minutes and the supernatant transferred to another microtube, with previously supernatant. The samples were dried and stored at -20°C until mass spectrometry analysis.

Mass spectrometry and data analysis

For protein analysis, peptides (4.5 µL) were separated by C18 (100 µm x 100 mm) RP-nanoUPLC (nanoAcquity, Waters) coupled with a Q-ToF Premier mass spectrometer (Waters) with nanoelectrospray source at a flow rate of 0.6 µL/min. The gradient was 2–90% acetonitrile in 0.1% formic acid over 45 min. The nanoelectrospray voltage was set to 3.5 kV, a cone voltage of 30 V and the source temperature was 100°C. The instrument was operated in the 'top three' mode, in which one MS spectrum is acquired followed by MS/MS of the top three most-intense peaks detected. After MS/MS fragmentation, the ion was placed on exclusion list for 60 s and for the analysis of endogenous cleavage peptides, a real time exclusion was used. The spectra were acquired using software MassLynx v.4.1 and the raw data files were converted to a peak list format (mgf) without summing the scans by the software Mascot Distiller v.2.3.2.0, 2009 (Matrix Science Ltd.) and searched against *Shigella flexneri* Uniprot database (297073 sequences, 85153106 residues) using Mascot engine v.2.3.01 (Matrix Science Ltd.), with carbamidomethylation as fixed modifications, oxidation of methionine as variable modification, one trypsin missed cleavage and a tolerance of 0.1 Da for both precursor and fragment ions.

Bioinformatic analysis

To predict the cellular localization of proteins, the UNIPROT database was used together with the aid of Protein Subcellular Localization Prediction Tool (PSORTdb) in version 3.0.2 (Yu et al. 2010), based on the protein FASTA format identified. To compare the total proteins identified in both strains, Venn diagrams were constructed using the InteractiVenn program available at <<http://www.interactivenn.net/>> (15).

Protein interaction prediction was made using the computational tool Search Tool for Retrieval of Interacting Genes (STRING), available at <<https://string-db.org/>>. Identification of proteins was performed using the Panther Data Base Classification System (Panther DB), available at <http://www.pantherdb.org/> and Cello2go (16). Potentially pathogenic proteins were predicted using Predict Pathogenic Proteins in Metagenomic Datasets (MP3) software, following the following parameters: threshold score: -0.2, which sets the sensitivity of 82.53%, specificity of 86.97% and precision of 86.02%.

Results

Growth curves of *Shigella* spp. wild type (M90T) and clinical strain 201.

Figure 1a shows the growth curve of *Shigella* spp. wild-type strains (M90T), and Figure 1b, the growth curve of the strain *S. flexneri* 201 in liquid culture medium LB. The sequential sampling data were used from the assays done every 1 hour for 35 hours. It can be observed in both Figures (1a and 1b) that *S. flexneri* 201 strain and *S. flexneri* M90T strain grow similarly until 10 hours. In both conditions, it can be

observed that the decline phase is about 30 hours, but in Figure 1a, *S. flexneri* 201 strain in CR condition showed a significant rate of decline in relation to LB condition, with a p-value of 0.0052. For protein extraction, the 10-hour growth culture was collected, extracted and analyzed in one-dimensional gel electrophoresis, as previously described in Methods.

Proteomic profiles of *Shigella* spp. strain 201 and M90T

Figure 2 shows the one-unidimensional gel profiles of M90T and strain 201, and the selected lanes to excised and submitted to tryptic digestion. In total, twelve lanes were excised for LB condition in M90T and eleven lanes, for the CR condition. For the 201 strain, thirteen lanes were excised for both LB and CR condition. The total proteins identified in mass spectrometry experiments are shown in Table 01.

Table 01. Total protein identified in wild and clinical strain of *S. flexneri* by mass spectrometry.

Organism	LB total proteins	CR total proteins	Common proteins between conditions
<i>S. flexneri</i> M90T	470	383	239
<i>S. flexneri</i> strain 201	386	189	111

After the identification of the proteins, the results were analyzed and compared, and the proteins that appears in two independent experiments were considered for analyses. Figure 3 shows the comparative proteins between the conditions and the bacterial strain. This study aims to understand the biological mechanisms by which the *S. flexneri* 201 strain has more virulence, and which are the possible metabolic pathways presented in CR conditions. All proteins compared were analyzed by the Veen diagram shown in Figure 3 (C) and the gene ontologies were researched with proteins from both strains and conditions.

Biological processes and molecular functions in both bacteria strain cultivated in LB and CR medium

Using the STRING program, biological processes, molecular functions and cellular components were mapped for each experimental condition. For the M90T condition grown in the LB medium, 470 proteins with Uniprot entries were identified, 299 proteins with an annotated gene name and of these, 110 proteins mapped in the STRING platform (Figure 4). Of these 149 proteins mapped, 125 biological processes related to the identified proteins were observed. For the M90T condition grown in the environment with the presence of the inductor VC, it was identified 383 proteins with entries in Uniprot, 249 mapped with the gene name and only 134 in the STRING platform. As a result, 149 biological processes were described for the M90T- CR condition.

In the case of the clinical strain 201 grown in medium LB, 386 proteins were identified at Uniprot, with only 260 proteins mapped with the gene name and of these, only 153 mapped at STRING. The analysis resulted in 141 biological processes mapped. Finally, for the clinical strain 201 grown in medium with the inductor CR, 189 proteins were identified with entry in Uniprot, being only 136 mapped with gene name

and 75 evaluated in STRING (Additional file 1: Table S1). The result for this condition was the identification of 63 associated biological processes.

From these results, the exclusive proteins of *S. flexneri* 201 grown in CR were evaluated together with the biological processes mapped by STRING and Cello2GO in search of a more holistic understanding of the metabolism of this bacterium and how it differs from the wild strain in relation to its virulence capacity (Figure 5). As observed, in both strains, in molecular function categories we can observe the response to stress and pathogenesis categories. The following results discuss some proteins with implications on the virulence of the bacterium, host infection and adaptation in aerobic and anaerobic conditions. Using the Predict Pathogenic Proteins in Metagenomic Datasets (MP3), Figure 6 indicates the main proteins predicted to be involved with virulence of the *S. flexneri* 201 strain grown in CR medium.

Discussion

The S-adenosimeticin cycle is present in the clinical strain 201 VC

The mapped pathways demonstrate that biological translation processes (GO:0006412), biosynthetic organonitrogenic compound pathways (GO:1901566), peptide metabolic processes (GO:0006518), cellular biosynthetic processes (GO:0044249) are the most prevalent in all conditions. For comparative purposes, it was observed if there were exclusive processes for condition 201 VC and it was detected that the biological processes S-adenosylmethionine cycle (GO:0033353) and S-adenosylhomocysteine metabolic process (GO:0046498) appeared only in this condition. The proteins related to this process identified in the 201 CR strain were the S-adenosylmethionine synthase proteins (A0A090NDY8; MetK) and S-ribosylhomocysteine lyase (A0A090NLU7; LuxS). Although the MetK protein was also identified in the M90T LB condition, the biological process did not appear because there was no identification of the LuxS protein. For the other conditions (M90T CR and 201 LB) there was no identification of the MetK protein, but they identified the LuxS protein. Works of PARVEEN & CORNELL (2011) report the cycle of activated methyl (Activated Methyl Cycle – AMC). In the metabolic pathway, methionine is converted to S-adenosylmethionine (SAM) by the enzyme SAM synthetase (MetK). The methyl group of S-adenosylmethionine to other acceptors results in S-adenosylhomocysteine (SAH). This event has been reported as an important flag in the reactions involving the quorum sensing in bacteria, using an inducer (autoinducer2: AI-2)(17).

Works indicate that Gram-Negative bacteria secrete three types of signaling molecules, called self-inducers. When the accumulation of these autoinducers occurs as a function of cell density (Quorum sensing), a cascade of signaling is activated leading to modifications in the patterns of gene expression and adaptive responses (18–24); .Another important fact demonstrates that the inhibition of an enzyme related to this process (Methylthioadenosine/S-adenosylhomocysteine nucleosidase) limits the synthesis of self-inducers, decreasing biofilm formation and attenuating virulence. The presence of two proteins associated with this metabolic pathway may suggest their participation in the virulence of the clinical strain 201. Reviews of (25) report that strains of *Neisseria meningitidis* have similar cellular machinery in

colonizing their hosts, but the difference in expression of some of their genes and their regulation determines their virulence capacity. The work suggests that genes involved in responses to oxidative stress and glutathione metabolism during infection may be key in determining strain virulence. Although not exclusive to the CR condition, proteins such as Thiol peroxidase (TPX; A0A090NWB6) and Thioredoxin/glutathione peroxidase (BtuE; A0A127GKP1) were identified in the present proteome and are proteins associated with oxidative stress and glutathione metabolism.

Additional experiments to validate the expression of these proteins in the CR condition should guide future discussions if the clinical strain 201 has pathogenicity mechanisms like those described for *N. meningitidis*.

Associated with these factors, the authors discuss that metabolic adaptations of the strains are fundamental for the bacterium to infect its host efficiently, reporting previous findings that associate with the term "nutritional virulence", i.e., they are specific mechanisms that the bacteria perform.

The protein Flavohemoprotein expressed in the strain *S. flexneri* 201 CR is responsible for detoxification of nitric oxide and is associated with virulence in bacteria

Among the exclusive proteins of the condition of the clinical strain with the CR inductor, the expression of the flavohemoprotein protein (Hmp; A0A1S9KCV2) was observed. In aerobic conditions, this protein is associated to the detoxification of nitric oxide (ON), protecting the bacteria from several nitrogenous compounds, and playing a central role in the bacterial response to nitrosative stress. This protein can reduce NO to N₂O (nitrous oxide) under anaerobic conditions (26). The literature reports that the body's defense cells are able to produce nitric oxide to prevent microbial infection. Some bacteria have already been described as capable of overlapping this host defense mechanism, producing proteins related to detoxification to ON and this is related to the virulence capacity of these bacteria, as already observed in *Mycobacterium tuberculosis*, *Neisseria meningitidis*, *Vibrio cholerae*, *Salmonella enterica* serovar *Typhimurium*, *Pseudomonas aeruginosa*, and enterohemorrhagic *Escherichia coli* (EHEC) (27).

The lysozyme type C inhibitor (YkfE) overlaps the host's defense mechanisms and is expressed exclusively by *S. flexneri* 201 CR.

One of the mechanisms studied to combat bacterial invasion is the expression of lysozyme by the host, which results in the hydrolysis of the peptidoglycan bacterial wall. In contrast, bacteria have developed systems to superimpose the host's defense mechanisms, and one of these mechanisms lies in the expression of lysozyme inhibitors. There are 3 bacterial defense systems: modification of the peptidoglycan wall to resist hydrolysis by lysozyme, modifications in the load and integrity of the envelope and expression of inhibitors of this enzyme(28). In the present work, we identified the expression of the protein C-lysozyme inhibitor (YkfE or Ivy; A0A0F6MA48) exclusively in the clinical strain grown in the medium with the VC inductor.

This protein has already been described in other bacteria as *Pseudomonas aeruginosa* and has been reported as a virulence factor for Gram-negative bacteria. Inhibition occurs through a loop protrusion in Ivy protein that occludes the active site of lysozyme via a lock-and-key mechanism (29). Its presence exclusively in the condition of the clinical strain with the CR inductor suggests that this may be a mechanism that potentiates the virulence of this strain.

The expression of the exoenzyme AepA is only present in condition 201 CR and is associated with the virulence of bacteria.

Our proteomics studies demonstrated that the strain 201 grown in medium with the CR inductor is capable of expressing a protein called Exoenzymes regulatory protein (AepA; D6BD12). In the literature there are few reports of AepA protein, but some publications discuss that AepA protein has been described as a transcriptional activator of the enzymes pectate lyase, cellulase, polygalacturonase and protease in *Erwinia carotovora* subsp. *carotovora* (30,31). Reviews indicate that the synthesis of exoenzymes and enzyme secretion is related to virulence aspects in bacteria. In *E. carotovora*, it was reported that the expression of exoenzymes is closely related to certain compounds of the plant that it infects and that this response is dependent on the presence of the genes *aepA* and *aepB* (32).

The MobC protein facilitates the transmission of virulence factors and is expressed exclusively by *S. flexneri* 201 CR

This protein is part of the transfer system called MOB, which is composed of six families, including MobC. This system acts in the conjugation process, transferring moving elements, such as plasmids to a donor cell. In order to have the transference, the DNA is cleaved by the relaxases, which are proteins of multiple domains, allowing the mobilization of the genetic elements, called conjugative plasmids.

Several plasmids of MobC are found in different classes of bacteria, as in gamaproteobacteria, firmicutes and tenicutes. Clinical Isolates of *E. coli*, *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* share such genetic elements as well as other enterobacteria, thus suggesting that such mechanism facilitates the transmission of virulence factors.

Furthermore, MobC counterparts are identified in *E. coli* and *Klebsiella pneumoniae*, for example, as well as other bacteria that share plasmids, which contain relaxases, involved in antibiotic resistance (33).

Genetic studies indicate that bacteria such as *Yersinia*, *Salmonella* and *Klebsiella* share similar sequences of transfer elements, such as MobC, indicating genomic flow between enterobacteria (34).

A recent study identified the MobC gene present in a plasmid of virulence in *Shigella flexneri*, such plasmid confers resistance to antibiotics of clinical importance, however, the study did not determine the ability of this plasmid to be transferred by conjugating (35). Studies relating MobC to virulence mechanisms are still scarce, especially in bacteria that cause damage to the intestinal microbiota, since little research is directed to the identification and characterization of conjugative systems, as well as the low frequency of the MobC family in relation to other components of the MOB family (36).

Although this family contributes to a greater mobilization of plasmids, many proteins of this family are not recorded (37).

***S. flexneri* 201 CR also expresses lipoprotein OsmY related to carrier biogenesis related to bacterial virulence.**

The protein OsmY (Osmotically inducible protein Y) is a lipoprotein found in Gram-negative and contributes to adaptation to the gastrointestinal osmotic stress environment, acting as a membrane stabilizer (38,39). Another recently discovered role is inhibition of protein aggregation, although studies have been proven only in vitro to date (40).

Although this protein is conserved and identified and predicted as virulent by our study, the literature lacks information related to bacterial virulence, especially for the genus of *Shigella* sp.

However, there is evidence that OsmY is involved in the biogenesis of auto carriers, such as type AIDA - I (adhesins involved in diffuse adherence), at *E.coli*. The AIDA - I protein is involved in bacterial virulence, as in biofilm formation, and in other virulence phenotypes as adhesion (41,42). The literature states that AIDA-I, already cited above, is a counterpart to the Flu protein (A0A0H2V1S1), found exclusively in the CR condition of the strain *S. flexneri* 201. Flu protein, also called ag43, is an external membrane protein belonging to the autotransport family and is mainly identified in uropathogenic strains of *E.coli* and *S.flexneri*. Being AIDA-I homologous to Flu protein, its role in biofilm formation and secretion and translocation processes is reinforced, through in vitro assays (43). Just as OsmY is involved in the biogenesis of AIDA-I, it is also in ag43, since deficient mutants of OsmY tend to be more sensitive to proteolytic action, for example (42).

Variants of ag43 present in uropathogenic strains of *E.coli* indicate that it contributes to the persistence of the species in the urinary tract in patients with urinary infection. However, the mechanisms that contribute to such persistence, as well as its direct involvement in the virulence of other pathogens, such as *Shigella* spp. are not yet fully clarified (44).

1. *flexneri* 201 CR expresses the LptA protein and this is associated with resistance to antibiotics.

Another potential target for therapeutic action against Gram-negatives is the Lpt translocation system, more specifically one of the subunits that make up this system, which is the Lipopolysaccharide export system protein (LptA; A0A090NL57), also known as YhbN.

This system works on the transport of Lipopolysaccharide (LPS) to the external membrane, which besides composing the bacterial cell structure, contributes to resistance to attacks by external agents, such as antibiotics (45).

Thus, biosynthesis and transport of lipopolysaccharides have been targeted for new drugs, as in *E.coli*.

Compounds that interact specifically as LptA can damage the LPS transport system, consequently affecting growth and altering bacterial morphology. It is suggested that this system may also be harmed by other Gram-negative bacteria, such as *Shigella spp.* (45).

LptA was identified to interact with another set of proteins also classified as pathogenic, such as YdgH (A0A090NGX6), the latter identified in the strains *S. flexneri* M90T LB and *S. flexneri* 201 CR. However, there is a scarcity of data on this protein, although its domain is preserved among the enterobacteria (46).

Conclusions

The identification of proteins in bacteria of clinical importance has been essential, especially in recent years, for the understanding of various molecular mechanisms, especially in the pathogenesis of diseases. The protein profile identified in this study provides unpublished data about the strains studied, thus adding new findings that contribute to previous research done by our group.

This lack of data is related to the fact that studies of this size occur with known standards. Studies that carry out the genetic and proteomic characterization of clinical isolates are rare, but they are of enormous importance since they help us to elucidate the pathogenicity of a microorganism in practice. The *Shigella flexneri* 201 CR of this study showed the secretion of several proteins related to pathogenicity, quorum sensing, adhesion, resistance to antibiotics and evasion of the immune system. Although many of these proteins are characterized for other enterobacteria, little is known about their role in the pathogenicity of *Shigella* and some of the proteins highlighted in this work are still little studied in the literature, mainly in processes related to virulence, ie Bacteria are microorganisms fast and dynamic evolution. We believe that these findings with clinical strains can help in regional vaccine study, since there is a strong relationship between phenotype and genotype and expression of the same.

Some of the proteins highlighted in this work are still little studied in the literature, mainly in processes related to virulence, ie, there is a lack of information, especially on the pathogenesis of *Shigella spp.*, Which limits the deepening of discussions on how such proteins are promoting increased virulence mechanisms, for example. In this sense, we believe that in silico prediction of pathogenic proteins presented here, opens ways for new investigations, not only for *Shigella* species, but for other enterobacteria, thus contributing to the elucidation of virulence of this genus, which is still responsible for the death of thousands of people.

Declarations

Acknowledgments

We thank those who contributed in some way to this work especially the DCDIA group.

Mass Spectrometry Group at Brazilian Biosciences National Laboratory (LNBio), CNPEM, Campinas, Brazil for their support with the mass spectrometry data and analysis.

Authors' contributions

FB, IC , PO designed the research and performed experiments, FB, IC, JV, RM, IS, DJ, PT, PD and AC contributed to data analysis and interpretation. FB e IC wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The Comitê de Ética em Pesquisa da Universidade Federal do Amazonas (CEP/UFAM) approved the permission to collect samples in hospitals, under protocol number 266/2006.

Funding

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES and Fundação de Amparo à Pesquisa do Estado do Amazonas – FAPEAM funding support. Instituto Leônidas e Maria Deane – Fiocruz Amazônia and Universidade Federal do Amazonas – UFAM had support reagents, material, and laboratory facilities.

Supporting institutions did not participate in the writing of this paper, nor did it read or approve it.

Availability of data and materials

The datasets used and analyzed in the study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Competing interests

The author(s) declare that they have no competing interests.

Abbreviations

AM: Amazonas

CR: Congo Red

DCDIA: Diagnosis and Control of Infectious Diseases of the Amazon

DTT: Dithiothreitol

EHEC: Enterohemorrhagic *Escherichia coli*

LB: Luria-Bertani

LPS: Lipopolysaccharide

IAA: Iodoacetamide

MP3: Pathogenic Proteins in Metagenomic Datasets

Panther DB – Panther Data Base Classification System

PMSF: Phenylmethylsulfonyl fluoride

PSORT–Protein Subcellular Localization Prediction Tool

WHO: World Health Organization

References

1. UNICEF. Committing to Child Survival: A Promise Renewed. 2013 Progress Report. New York; 2013.
2. WHO. World Health statistics 2014. World Heal Organ [Internet]. 2014;180. Available from: <http://www.who.int/mediacentre/news/releases/2014/world-health-statistics-2014/en/>
3. The HC, Thanh DP, Holt KE, Thomson NR, Baker S. The genomic signatures of Shigella evolution, adaptation and geographical spread. Nat Rev Microbiol [Internet]. 2016;14(4):235–50. Available from: <http://www.nature.com.abc.cardiff.ac.uk/nrmicro/journal/v14/n4/full/nrmicro.2016.10.html>
4. Das JK, Tripathi A, Ali A, Hassan A, Dojosoendy C, Bhutta ZA. Vaccines for the prevention of diarrhea due to cholera , shigella , ETEC and rotavirus. BMC Public Health [Internet]. 2013;13(Suppl 3):S11. Available from: <http://www.biomedcentral.com/1471-2458/13/S3/S11>
5. DEFEATDD [Internet]. Available from: <https://www.defeatdd.org/cascading-burden/#1>
6. Schroeder GN, Hilbi H. Molecular pathogenesis of Shigella spp.: Controlling host cell signaling, invasion, and death by type III secretion. Clin Microbiol Rev. 2008;21(1):134–56.
7. Cruz CBN Da, Souza MCS De, Serra PT, Santos I, Balieiro A, Pieri FA, et al. Virulence factors associated with pediatric shigellosis in Brazilian Amazon. Biomed Res Int. 2014;2014.
8. Serra PT. Análise investigativa dos fatores de virulência de cepas selvagens de Shigella spp. in vivo e seu potencial inflamatório. Dissertação (Mestrado em Imunologia Básica e Aplicada) - Universidade Federal do Amazonas; 2013.
9. Dantas PI de S. Variabilidade genética dos principais fatores de virulência em amostras Clínicas de Shigella flexneri. Dissertação (Mestrado em Biotecnologia) - Universidade Federal do Amazonas; 2016.
10. Yu K, Wang Z, Zhou F, Jiang J, Liu Y, Hu M, et al. Quantitative analysis of Shigella flexneri protein expression under acid stress. Proteomics. 2017;201600381:1–10.
11. Maffei B, Francetic O, Subtil A. Tracking Proteins Secreted by Bacteria: What's in the Toolbox? Front Cell Infect Microbiol [Internet]. 2017;7(May):1–17. Available from:

<http://journal.frontiersin.org/article/10.3389/fcimb.2017.00221/full>

12. Dautin N. Serine protease autotransporters of Enterobacteriaceae (SPATEs): Biogenesis and function. *Toxins (Basel)*. 2010;2(6):1179–206.
13. Barry EM, Pasetti MF, Sztein MB, Kotloff KL, Levine Mucosal MM, Levine MM, et al. Progress and pitfalls in Shigella vaccine research. *Nat Publ Gr [Internet]*. 2013;10(10):245–255. Available from: <http://dx.doi.org/10.1038/nrgastro.2013.12>
14. Shevchenko A, Wilm M, Vorm O, Mann M. Techniques in Protein Chemistry. *J Am Chem Soc*. 1964;86(8):1653–1653.
15. Heberle H, Meirelles GV, Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*. 2015;16(1):169.
16. Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, et al. CELLO2GO: A web server for protein subCELLular lOcalization prediction with functional gene ontology annotation. *PLoS One*. 2014;9(6).
17. Parveen N, Cornell KA. Methylthioadenosine/S-adenosylhomocysteine nucleosidase, a critical enzyme for bacterial metabolism. *Mol Microbiol*. 2011;79(1):7–20.
18. Winzer K, Hardie KR, Burgess N, Doherty N, Kirke D, Holden MTG, et al. LuxS: Its role in central metabolism and the in vitro synthesis of 4-hydroxy-5-methyl-3(2H)-furanone. *Microbiology*. 2002;148(4):909–22.
19. Chen X, Schauder S, Potier N, Dorsselaer A Van, Pelczar A, Bassler BL, et al. Structural identification of a bacterial quorum-sensing signal containing boron. *Nature*. 2002;5(3):545–9.
20. Ng WL, Bassler BL. Bacterial quorum-sensing network architectures. *Annu Rev Genet*. 2009;43:197–222.
21. Waters CM, Bassler BL. Quorum sensing: Cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol*. 2005;21:319–46.
22. Zhu J, Miller MB, Vance RE, Dziejman M, Bassler BL, Mekalanos JJ. Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *Proc Natl Acad Sci U S A*. 2002;99(5):3129–34.
23. Sun J, Daniel R, Wagner-Döbler I, Zeng AP. Is autoinducer-2 a universal signal for interspecies communication: A comparative genomic and phylogenetic analysis of the synthesis and signal transduction pathways. *BMC Evol Biol*. 2004;4:1–11.
24. Miller MB, Skorupski K, Lenz DH, Taylor RK, Bassler BL. Parallel quorum sensing systems converge to regulate virulence in *Vibrio cholerae*. *Cell*. 2002;110(3):303–14.
25. Schoen C, Kischkies L, Elias J, Joseph Ampattu B. Metabolism and virulence in *Neisseria meningitidis*. *Front Cell Infect Microbiol*. 2014;4(AUG):1–16.
26. Karlinsey JE, Bang IS, Becker LA, Frawley ER, Porwollik S, Robbins HF, et al. The NsrR regulon in nitrosative stress resistance of *Salmonella enterica* serovar Typhimurium. *Mol Microbiol*. 2012;85(6):1179–93.

27. Robinson JL, Adolfsen KJ, Brynildsen MP. Deciphering nitric oxide stress in bacteria with quantitative modeling. *Curr Opin Microbiol* [Internet]. 2014;19(1):16–24. Available from: <http://dx.doi.org/10.1016/j.mib.2014.05.018>
28. Ragland SA, Criss AK. From bacterial killing to immune modulation: Recent insights into the functions of lysozyme. *PLOS Pathog*. 2017;13.
29. Abergel C, Monchois V, Byrne D, Chenivresse S, Lembo F, Lazzaroni JC, et al. Structure and evolution of the Ivy protein family, unexpected lysozyme inhibitors in Gram-negative bacteria. *Proc Natl Acad Sci*. 2007;104(15):6394–9.
30. MURATA H, CHATTERJEE A, LIU Y, CHATTERJEE AK. Regulation of the Production of Extracellular Pectinase, Cellulase, and Protease in the Soft Rot Bacterium *Erwinia carotovora* subsp. *carotovora*: Evidence that aepH of *E. carotovora* subsp. *carotovora* 71 Activates Gene Expression in *E. carotovora* subsp. *car*. *Appl Environ Microbiol*. 1994;60(9):3150–9.
31. Barnard AML, Salmond GPC. Quorum sensing in *Erwinia* species. *Anal Bioanal Chem*. 2007;387:415–423.
32. Lengeler JW, Drews G, Schlegel HG. Biology of the Prokaryotes. *Biol Prokaryotes*. 1999;
33. Garcillán-Barcia MP, Francia MV, De La Cruz F. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol Rev*. 2009;33(3):657–87.
34. Wilharm G, Lepka D, Kerrinnes T, Skiebe E, Hahn B, Fruth A. Adding to *Yersinia enterocolitica* gene pool diversity: Two cryptic plasmids from a biotype 1A isolate. *J Biomed Biotechnol*. 2009;2009.
35. Parajuli P, Rajput MI, Verma NK. Plasmids of *Shigella flexneri* serotype 1c strain Y394 provide advantages to bacteria in the host. *BMC Microbiol*. 2019;19(1):1–10.
36. Smillie C, Garcillan-Barcia MP, Francia M V., Rocha EPC, de la Cruz F. Mobility of Plasmids. *Microbiol Mol Biol Rev*. 2010;74(3):434–52.
37. Garcillán-Barcia MP, Cuartas-Lanza R, Cuevas A, de la Cruz F. Cis-Acting Relaxases Guarantee Independent Mobilization of MOBQ4 Plasmids. *Front Microbiol*. 2019;10(November):1–11.
38. Radford D, Strange P, Lepp D, Hernandez M, Rehman MA, Diarra MS, et al. Genomic and proteomic analyses of *Salmonella enterica* serovar Enteritidis identifying mechanisms of induced de novo tolerance to ceftiofur. *Front Microbiol*. 2018;9(SEP):1–20.
39. Wolfe AJ. The Acetate Switch. *Microbiol Mol Biol Rev*. 2005;69(1):12–50.
40. Lennon CW, Thamsen M, Friman ET, Cacciaglia A, Sorgenfrei FA, Wasik MA, et al. Folding optimization in vivo uncovers new chaperones. *J Mol Biol*. 2016;427(18):2983–94.
41. Vo JL, Mart GC. Autotransporter Adhesins in *Escherichia coli* Pathogenesis. *Proteomics*. 2017;1600431:1–10.
42. Yan Z, Hussain S, Wang X, Bernstein HD, Bardwell JCA. Chaperone OsmY facilitates the biogenesis of a major family of autotransporters. *Mol Microbiol*. 2019;112(5):1373–87.
43. Woude MW Van Der, Henderson IR. Regulation and Function of Ag43 (Flu). *Annu Rev Microbiol*. 2008;62:153–9.

44. Lüthje P, Annelie B. Ag43 Promotes Persistence of Uropathogenic Escherichia coli Isolates in the Urinary Tract. J Clin Microbiol. 2010;48(6):2316–7.

45. Zhang X, Li Y, Wang W, Zhang J, Lin Y, Hong B, et al. Identification of an anti-Gram-negative bacteria agent disrupting the interaction between LPS transporters LptA and LptC. Int J Antimicrob Agents [Internet]. 2018; Available from: <https://doi.org/10.1016/j.ijantimicag.2018.11.016>

46. Eletsky A, Michalska K, Houliston S, Zhang Q, Daily MD, Xu X, et al. Structural and Functional Characterization of DUF1471 Domains of Salmonella Proteins SrfN, YdgH / SssB, and YahO. PLoS One. 2014;9(7).

Figures

Figure 1

Growth curve of *Shigella flexneri*. (A) Growth curve of *Shigella flexneri* strain 201 in LB and CR conditions. (B) Growth curve of *Shigella flexneri* 5a M90T (wild type) in both LB and CR conditions. The curves were obtained by measuring the culture absorbance every one hour for 35 hours

Figure 2

Proteomic profiles of *Shigella flexneri* 5a M90T and strains 201. (a) *Shigella flexneri* 5a M90T and strain *Shigella flexneri* 201 (b) in both LB and CR conditions. MW: Molecular mass standard. Segments of the gel were excised as indicated by the lines in in color. Each lane was cut in small pieces and submitted to the in-gel digestion protocol. The triplicate are showed for each condition (LB and CR) and each strain (M90T and strain 201) identified as L1, L2, L3, L4, L5 and L6.

Figure 3

Venn diagrams of *S. flexneri* M90T and *S. flexneri* strain 201. (A) two hundred and thirty- nine proteins were identified in M90T in both conditions (LB and CR). Two hundred and thirty-one proteins were uniquely identified in LB condition and one hundred and forty four proteins were uniquely identified in CR condition. (B) One hundred eleven proteins were identified in strain 201 in both conditions (LB and CR). Two hundred seventy-five proteins were uniquely identified in LB condition and seventy-eight proteins were uniquely identified in CR condition. (C) To better understand which mechanisms are activated under *S. flexneri* strain 201 CR conditions, Venn diagram between all conditions was obtained. Forty-three

proteins are exclusive of *S. flexneri* strain 201. The comparative list of proteins in all conditions is presented in Supplemental Table (Additional file 1: Table S1).

Figure 4

String pathway analysis in both strains in CR medium. (A) String pathway analysis of *S. flexneri* 201 CR without ribosomal proteins and (B) String pathway analysis of *S. flexneri* M90T CR.

Figure 5

Molecular function of exclusive proteins identified in both strains in CR medium. (A) Molecular function of exclusives proteins identified in S.flexneri strain 201 cultured in CR medium. (B) Biological process of exclusives proteins identified in S.flexneri strain 201 cultured in CR medium. (C) Molecular function of exclusives proteins identified in S.flexneri M90T in CR medium. (D) Biological process of exclusives proteins identified in S.flexneri M90T cultured in CR medium

Figure 6

Predicted pathogenic proteins os Shigella flexneri. (A) Schematic view of the pathway identified in S. flexneri 201 CR and (B) the heat map predicted by The MP3 tool in the exclusive proteins identified in S. flexneri 201 CR.

Supplementary Files

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

- [Additionalfile1.docx](#)