Salmonella Typhi, Paratyphi A, Enteritidis and Typhimurium core proteomes reveal differentially expressed proteins linked to the cell surface and pathogenicity

Sara Saleh, Sandra Van Puyvelde, An Staes, Evy Timmerman, Barbara Barbé, Jan Jacobs, Kris Gevaert, Stijn Deborggraeve

1 Department of Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium, 2 VIB Center for Medical Biotechnology, Ghent, Belgium, 3 Department of Biomolecular Medicine, Ghent University, Ghent, Belgium, 4 Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium, 5 Department of Microbiology and Immunology, KU Leuven, Leuven, Belgium

* sdeborggraeve@itg.be

Abstract

Background

Salmonella enterica subsp. enterica contains more than 2,600 serovars of which four are of major medical relevance for humans. While the typhoidal serovars (Typhi and Paratyphi A) are human-restricted and cause enteric fever, non-typhoidal Salmonella serovars (Typhimurium and Enteritidis) have a broad host range and predominantly cause gastroenteritis.

Methodology/Principle findings

We compared the core proteomes of Salmonella Typhi, Paratyphi A, Typhimurium and Enteritidis using contemporary proteomics. For each serovar, five clinical isolates (covering different geographical origins) and one reference strain were grown in vitro to the exponential phase. Levels of orthologous proteins quantified in all four serovars and within the typhoidal and non-typhoidal groups were compared and subjected to gene ontology term enrichment and inferred regulatory interactions. Differential expression of the core proteomes of the typhoidal serovars appears mainly related to cell surface components and, for the non-typhoidal serovars, to pathogenicity.

Conclusions/Significance

Our comparative proteome analysis indicated differences in the expression of surface proteins between Salmonella Typhi and Paratyphi A, and in pathogenesis-related proteins between Salmonella Typhimurium and Enteritidis. Our findings may guide future development of novel diagnostics and vaccines, as well as understanding of disease progression.
Author summary

With an estimated 20 million typhoid cases and an even higher number of non-typhoid cases the health burden caused by salmonellosis is huge. Salmonellosis is caused by the bacterial species *Salmonella enterica* and over 2500 different serovars exist, of which four are of major medical relevance for humans: Typhi and Paratyphi A cause typhoid fever while Typhimurium and Enteritidis are the dominant cause of non-typhoidal *Salmonella* infections. The proteome is the entire set of proteins that is expressed by a genome and the core proteome are all orthologous proteins detected in a given sample set. In this study we have investigated differential expression of the core proteomes of the *Salmonella* serovars Typhi, Paratyphi A, Typhimurium and Enteritidis, as well as the regulating molecules. Our comparative proteome analysis indicated differences in the expression of surface proteins between the serovars Typhi and Paratyphi A, and in pathogenesis-related proteins between Typhimurium and Enteritidis. Our findings in proteome-wide expression may guide the development of novel diagnostics and vaccines for *Salmonella*, as well as understanding of disease.

Introduction

The gram-negative bacterial genus *Salmonella* is divided in two species, *Salmonella enterica* and *Salmonella bongori*. Only the *Salmonella enterica* subspecies *enterica* is of clinical relevance for humans and is further classified into more than 2,600 serovars. The human restricted serovar Typhi (STY) and the closely related serovar Paratyphi A (SPTA) cause enteric fever [1], while the generalist serovars Typhimurium (STM) and Enteritidis (SENT) are the most important causes of non-typhoidal salmonellosis [2]. Enteric fever is a systemic disease that affects more than 27 million people worldwide and leads to more than 200,000 deaths annually [3,4]. While STY and SPTA both cause a systemic disease, SPTA causes a milder disease with a shorter incubation time [5]. In the last 20 years, the number of infections with SPTA has significantly increased in Asia [6]. The global burden of non-typhoidal *Salmonella*, a common cause of food poisoning that is usually characterized by localized gastroenteritis, is even higher with an estimated 93.8 million cases and 155,000 deaths each year [2]. Moreover, invasive non-typhoidal *Salmonella* has emerged as an important cause of bloodstream infection in Sub-Saharan Africa in both adults and children, and the incidence of invasive non-typhoidal *Salmonella* is estimated at 3.4 million cases with more than 600,000 deaths each year [7].

Comparative genomics of *Salmonella enterica* has revealed specific genetic fingerprints associated with invasive disease and host adaptation [8,9]. A comparative analysis of 8 typhoidal and 27 non-typhoidal *Salmonella* genomes demonstrated presence of typhoid-specific protein families which include virulence factors such as Vi polysaccharide pilus related proteins [10]. In addition, an *in silico* comparative analysis of *Salmonella* genomes identified 469 genes involved in the central anaerobic metabolism which was intact in gastrointestinal pathogens (SENT and STM among others) but decaying in extra-intestinal pathogens, such as STY and SPTA. This metabolic advantage might have a role in competing with other bacteria in the inflamed gut, thereby enhancing transmission of the gastrointestinal pathogens [11]. However, not all phenotypic differences in typhoidal and non-typhoidal *Salmonella* can be explained by presence or absence of functional genes. Investigating differential expression of the core proteomes (defined as all orthologous proteins quantified in a given sample set) between *Salmonella* serovars [12], and the regulating molecules involved, can reveal additional insights in the
adaptations to different host environments and pathogenesis, as well as reveal the expression of potential vaccine and diagnostic targets.

In the last decade, mass spectrometry (MS) based proteomics has advanced rapidly and provides a comprehensive view on the proteins that are expressed by an organism. In clinical microbiology laboratories, MALDI-TOF MS is routinely used for bacterial genus and species identification [13]. In research, proteomics was used to characterize the proteomes of *Salmonella* Typhimurium and Enteritidis under specific in vitro culture conditions mimicking the phagosome [14,15], to identify proteins that were expressed by *Salmonella* Typhimurium isolated from infected macrophages [16], and to study antimicrobial resistance and virulence in *Salmonella* Typhimurium [17–19]. Next to proteome analysis within single serovars, comparative proteome studies have been conducted to assess the proteome variability between different *Salmonella* serovars. However, these studies used laboratory reference strains which may not represent the currently circulating clinical strains [20–22].

Here, we conducted a comparative analysis of the core proteomes of the clinically most relevant *Salmonella enterica* serovars: Typhi, Paratyphi A, Typhimurium and Enteritidis, using 20 *Salmonella* strains isolated from patients covering various geographical origins, as well as one reference strain per serovar. Our findings show that differential expression of the core proteome of the typhoidal serovars is mainly related to cell surface components and, for the non-typhoidal serovars, to pathogenicity.

**Methods**

**Bacterial strains and growth conditions**

Five clinical isolates per *Salmonella* serovar Typhi, Paratyphi A, Typhimurium and Enteritidis were selected from the strain collection at the clinical laboratory of the travel clinic of the Institute of Tropical Medicine, Antwerp, Belgium for shotgun proteome analysis. One ATCC reference strain for each *Salmonella* serovar was added to the sample set and for the *Salmonella* Typhi reference strain, a clinical strain was certified (Table 1). Given that the burden of typhoid fever and invasive non-typhoidal salmonellosis is highest in Asia and Africa respectively, we have selected representative strains from different countries covering both continents. All in vitro incubation was done at 37˚C. Minimum and maximum temperatures were recorded and ranged between 35˚C and 37˚C. As all clinical strains have been isolated from patients, the strains were revived from Microbank cryogenic vials (Pro-Lab Diagnostics) on blood agar (BD Columbia Agar, 5% sheep blood) and grown overnight at 37˚C. Single colonies were sub-cultured on MacConkey agar (BD MacConkey II Agar) and grown overnight at 37˚C. Colonies were further solubilized into 3 ml of synthetic growth medium and supplemented with 1% glucose (Teknova HI-DEF Azure Media) until the OD was 0.06, and 250 µl of this suspension was inoculated into 5 ml of synthetic medium supplemented with 1% glucose and grown at 37˚C with shaking at 220 rpm until mid-log phase (OD 0.5-OD 0.6). The Teknova HI-DEF Azure synthetic medium (S1 File) is based on the medium described by Neidhardt et al. [23].

**Protein extraction and in-solution digestion**

Upon harvesting the bacteria, duplicate samples of 1 ml were taken from each culture and centrifuged at 5000 x g for 10 min at 4˚C and the cell pellets were washed twice with phosphate buffered saline (PBS). Duplicate samples are thus further considered as technical replicates. Proteins were extracted from the bacterial pellets with the Qproteome Bacterial Protein Prep Kit (Qiagen) following the manufacturer’s instructions. Briefly, after snap-freezing on dry ice, bacterial cell pellets were thawed on ice for 15 minutes. Cell pellets were re-suspended 750 µl of lysis buffer supplemented with lysozyme and Benzonase Nuclease, all included in the extraction
kit. EDTA-free protease inhibitor (Roche) was added to a final concentration of 2%. After incubation on ice for 30 minutes, lysates were centrifuged at 14,000 for 30 minutes to pellet the cellular debris, and the supernatant was collected. The protein concentration was determined with the BCA Protein Assay Kit (Pierce) (S1 Table). Proteins were reduced with 15 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) and alkylated with 30 mM iodoacetamide (IAM) for 15 min in the dark while shaking at 37˚C. The buffer was exchanged to digestion buffer (50 mM ammonium bicarbonate, pH 7.9) using G-25 illustra NAP-5 gel filtration columns (GE Healthcare). The eluates were then heated at 99˚C for 5 min, put immediately on ice and, after cooling, sequencing grade modified trypsin (Promega) was added to a 1:100 trypsin to protein ratio upon which digestion proceeded at 37˚C for 16 h. The trypsin activity was stopped by adding 60 μl of 10% trifluoroacetic acid (TFA) (0.6% final concentration).

**LC-MS/MS analysis**

The peptide mixtures were subjected to LC–MS/MS analysis using an Ultimate 3000 RSLC nano LC (Thermo Scientific, Bremen, Germany) in-line connected to a Q Exactive mass spectrometer.
spectrometer (Thermo Fisher Scientific). The sample mixture was first loaded on a trapping column (made in-house, 100 μm internal diameter (I.D.), 20 mm long, filled with 5 μm C18 Reprosil-HD beads, Dr. Maisch, Ammerbuch-Entringen, Germany). After flushing from the trapping column, the peptides were loaded on an analytical column (75 μm I.D., 400 mm long and filled with 3 μm C18 Reprosil-HD beads (Dr. Maisch)) packed in the needle PicoFrit SELF/P PicoTip emitter (PF360-75-15-N-5 (NewObjective, Woburn, USA)). Peptides were loaded with loading solvent (0.1% TFA in water) and separated with a linear gradient from 98% solvent A (0.1% formic acid in water) to 40% solvent B (0.1% formic acid in water/acetonitrile, 20/80 (v/v)) in 130 min at a flow rate of 300 nL/min. This was followed by a 15 min wash reaching 99% solvent B. The mass spectrometer was operated in data-dependent, positive ionization mode, automatically switching between MS and MS/MS acquisition for the 10 most abundant peaks in a given MS spectrum. The source voltage was 3.4 kV and the capillary temperature was at 275˚C. One MS1 scan (m/z 400−2000, AGC target 3 × 10^6 ions, maximum ion injection time 80 ms) acquired at a resolution of 70,000 (at 200 m/z) was followed by up to 10 tandem MS scans (resolution 15,700 at 200 m/z) of the most intense ions fulfilling the defined selection criteria (AGC target 5 × 10^4 ions, maximum ion injection time 60 ms, isolation window 2 Da, fixed first mass 140 m/z, spectrum data type: centroid, underfill ratio 2%, intensity threshold 1.7x10^4, exclusion of unassigned 1, 5–8, >8 charged precursors, peptide match preferred, exclude isotopes: on, dynamic exclusion time 20 s). The HCD collision energy was set to 25% normalized collision energy and the polydimethylcyclosiloxane background ion at 445.120025 Da was used for internal calibration (lock mass). The mass spectrometry proteomics data have been deposited to the PRIDE Archive (http://www.ebi.ac.uk/pride/archive/) via the PRIDE partner repository with the data set identifier PXD011154 (username: reviewer00797@ebi.ac.uk; password: hN5SqXtY).

MS data processing

Raw MS files were analyzed by MaxQuant [24] version 1.5.0.25 and MS/MS spectra were searched against the translated protein sequences of the annotated genomes of Salmonella Typhi CT18 (NCBI accession number AL513382.1) [25], Paratyphi A ATCC 9150 (CP000026.1) [26], Typhimurium 14028S (CP001363.1) [27], and Enteritidis PT4/P125109 (AM933172.1) [28]. The following parameters were applied for the database search: enzyme specificity was set to trypsin/P allowing for a maximum of two missed cleavages; carbamidomethylation of cysteine was set as a fixed modification; methionine oxidation, N-terminal formylation on the protein level and conversion of N-terminal glutamine to pyroglutamate were set as variable modifications. The first search for precursor ions was performed with a mass tolerance of 20 ppm for calibration, while 6 ppm was applied for the main search. For protein identification, at least two unique peptides were required per protein group and the minimum peptide length was set to 7. The false discovery rate for peptide and protein identification was set to 1%. The minimum score threshold for both modified and unmodified peptides was set to 30. MS runs were analyzed with the “match between runs” option between samples of a given serovar. For matching, a retention time window of 42 s was selected. Protein quantification was based on the MaxQuant label-free (MaxLFQ) algorithm. For all other parameters, default settings were applied as advised by the developers.

Comparative analysis of core proteomes

The MaxQuant output file “proteinGroups.txt” was loaded into Perseus 1.5.0.8. The protein entries were filtered to remove potential contaminants, reverse hits and proteins only identified by site. Then, the LFQ intensities were log2 transformed and data were filtered for
proteins containing a minimum number of valid values in 9 out of 12 samples. The log2 transformed data were then normalized by subtracting the median per sample within the dataset. To compare the different Salmonella serovars we used orthology mapping. Orthologous genes within the four serovars were retrieved from the Orthologous Matrix (OMA) database [29] with NCBI Taxonomy IDs 220341 (STY), 295319 (SPTA), 550537 (SENT) and 588858 (STM). Statistical significant differences in LFQ intensities were assessed using a two-sided t-test with Bonferroni adjusted P values using R. Proteins were considered differentially expressed if they showed a minimal 2-fold change in their overall levels with an adjusted P-value lower than 0.05. Principal component analysis (PCA) was done in Perseus 1.5.0.8 using default settings as advised by the developers.

Functional enrichment analysis
Differentially expressed proteins were subjected to gene ontology (GO) term enrichment to investigate biological processes, molecular function and cellular compartment using the Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics resources 6.7 [30]. Briefly, we have uploaded the differentially expressed core proteins as an input list and performed GO term enrichment analysis against a background list with default settings (count threshold is 2 and EASE threshold is 0.1).

Regulatory network analysis
To infer regulatory interactions that can explain differential expression profiles we used the PheNetic web server (http://bioinformatics.intec.ugent.be/phenetic/#/index) with default settings (Cost is 0.1, Pathlength is 4 and k-best paths is 20) and upstream run mode [31]. Input data consisted of the available interaction network for Salmonella Typhimurium LT2 (http://bioinformatics.intec.ugent.be/phenetic/index.html#/network), the list of detected proteins that are shared by two groups, and the list of differentially expressed proteins with \( P < 0.05 \).

Ethics statement
The clinical Salmonella isolates were obtained through the project “Surveillance of antimicrobial resistance among consecutive blood culture isolates in tropical settings”, within the Third Framework Agreement between the Belgian Directorate of Development Cooperation (DGD) and the Institute of Tropical Medicine (ITM), Antwerp, Belgium. The partner institutes involved in this surveillance project that provided strains were: Sihanouk Hospital Centre of Hope, Phnom Penh, Cambodia and Institut National de Recherche Biomédicale, Kinshasa, Democratic Republic of the Congo. Ethical approval for the Microbiological Surveillance was granted by the Institutional Review Board at the ITM in Antwerp, by the Ethics Committees of the Antwerp University (Belgium). Ethical approval for the Microbiological Surveillance Study was granted by the Institutional Review Board of ITM, the Ethics Committee of Antwerp University and the competent ethical committees from the DR Congo and Cambodia respectively. The Salmonella Typhimurium isolate from The Gambia was received from the Medical Research Council (MRC) Keneba, MRC The Gambia. Ethical approval was granted by the Gambia Government/MRC Joint Ethics Committee. The remaining strains were obtained from patients presenting at the travel clinic at ITM, ethical approval was granted by the Institutional Review Board of ITM. All isolates and subsequent biological samples were anonymized.
Results

Salmonella proteins identified by LC-MS/MS

The reference genomes of STY, SPTA, SENT and STM used in our analysis contain 4,600, 4,095, 4,318 and 5,372 protein-encoding genes, respectively. In total, 3596 orthologous genes in the four serovars were retrieved from the OMA database and 1,414, 1,558, 1,222 and 1,099 proteins were detected by LC-MS/MS analysis in the STY, SPTA, SENT and STM strains, respectively. Protein detection in technical replicates showed Pearson correlation coefficients higher than 0.92 for all samples, except for the STM strain from Ethiopia with a Pearson correlation of 0.86 (S2 Table). Intra-serovar PCA of the LFQ intensities of expressed proteins show little variation in expression levels between strains within the same serovar (S2 File). However, in order to conduct reliable intra-serovar comparisons, more strains should have been included per serovar.

In total, 418 orthologous proteins were detected in all serovars (Fig 1) and expression levels in the typhoidal (STY and SPTA) and non-typhoidal (STM and SENT) Salmonella serovars were compared by PCA of the LFQ intensities (Fig 2A). The first two components capture ~72% of the variability in the dataset and show that the typhoidal serovars do not separate from the non-typhoidal serovars based on the observed variability in LFQ intensities. When we compared the typhoidal with the non-typhoidal Salmonella strains, a total of 128 proteins showed a minimal 2-fold change in their overall levels with an adjusted $P$-value lower than 0.05 (S3 Table). GO term enrichment of these 128 proteins showed that all GO terms with a $P$ value lower than 0.05 are related to translation and structural components of the ribosomes (Table 2).

Differentially expressed proteins in Salmonella Typhi (STY) and Paratyphi A (SPTA) are associated with the cell surface

A set of 810 core proteins were detected in Typhi and Paratyphi A and their LFQ intensities were used as input for PCA (Fig 2B). The first two components allow a clear separation of the STY from the SPTA strains, covering 80% of the total variation in expression levels. In addition, the PCA shows that clinical isolates do not separate from the reference strains in both serovars. A total of 230 proteins with a minimal 2-fold change in their overall levels and an adjusted $P$-value lower than 0.05 were considered significantly differentially expressed between STY and SPTA strains (S4 Table). GO functional enrichment analysis of these proteins indicated an enrichment of biological pathways that are related to carbohydrate and polysaccharide biosynthesis and metabolism, as well as the external encapsulating structure (Table 2). We have plotted our differential expression data set on the wide interaction network for Salmonella Typhimurium LT2. Using the upstream run mode, PheNetic searches for regulatory mechanisms that can explain our observed data set. The inferred sub-network (Fig 3) shows that many differentially expressed proteins are connected to each other by outer membrane, stress and carbohydrate metabolism regulatory proteins such as CpxR, YjeB and CRP, which are not necessarily differentially expressed themselves, but might have a post-translational serovar-specific effect. Moreover, the small regulatory RNAs OmrA and OmrB connect differentially expressed proteins involved in carbohydrate metabolism.

Differentially expressed proteins in Salmonella Typhimurium (STM) and Enteritidis (SENT) are associated with pathogenicity

A set of 465 core proteins were detected in all strains of STM and SENT. PCA of the LFQ intensities of these proteins showed a clear separation of the STM isolates from the SENT
isolates based on the observed protein expression levels where the first two components cover ~80% of the total variation in expression levels (Fig 2C). The PCA also shows that the reference strains and the clinical isolates do not separate in STM and SENT. A total of 192 proteins with a minimal 2-fold change in their overall levels and an adjusted *P*-value lower than 0.05 were considered significantly differentially expressed between STM and SENT strains (S5 Table).

Fig 1. Venn diagram of the orthologous proteins detected by LC-MS/MS in 6 *Salmonella* Typhi, 6 *Salmonella* Paratyphi A, 6 *Salmonella* Enteritidis and 6 *Salmonella* Typhimurium strains.

https://doi.org/10.1371/journal.pntd.0007416.g001

Fig 2. Principal component analysis (PCA) separate serovars based on LFQ intensities. The PCA plots show that the first and second principle components capture ~72% of the variability among the *Salmonella* serovars Typhi (STY), Paratyphi A (SPTA), Typhimurium (STM) and Enteritidis (SENT) (A), 80% of the variability between the serovars STY and SPTA (B), and ~80% of the variability between the serovars STM and SENT (C). Reference strains for each serovar are presented in red.

https://doi.org/10.1371/journal.pntd.0007416.g002
GO enrichment analysis of these proteins showed that all GO terms with \( P < 0.05 \) are related to pathogenesis (Table 2). The inferred subnetwork (Fig 4) revealed that the flagellar biosynthesis sigma factor FliA and the flagellar transcriptional regulators FlhD and FlhC (STM1924.S) connect the upregulated flagellar synthesis and motility proteins in STM. HilA, the main regulator of *Salmonella* Pathogenicity Island 1 (SPI-1), is possibly involved in the upregulation of the type 3 secretion system (T3SS) structural protein Prgl and effector protein SipA in STM.

### Discussion

The genomes of typhoidal and non-typhoidal *Salmonella* have a high level of similarity with more than 98% of sequence identity [32]. However, these two groups cause different diseases, host-pathogen interactions and immune responses. Here, we conducted the first comprehensive analysis of the proteomes of the *Salmonella* serovars Typhi, Paratyphi A, Typhimurium and Enteritidis using five clinical isolates that cover different geographical regions and one reference strain per *Salmonella* serovar. We have compared the expression levels of proteins from the core proteome under in vitro conditions and identified regulators that may help to explain the differences between different *Salmonella* serovars.

Table 2. Gene ontology functional enrichment analysis of differentially expressed core proteins between typhoidal and non-typhoidal, Typhi and Paratyphi A, Enteritidis and Typhimurium *Salmonella*.

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Count</th>
<th>%</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typhoidal versus non-typhoidal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOTERM_M_F_FAT</td>
<td>GO:0005198--structural molecule activity</td>
<td>38</td>
<td>29.46</td>
<td>2.56E-10</td>
</tr>
<tr>
<td>GOTERM_M_F_FAT</td>
<td>GO:0003735--structural constituent of ribosome</td>
<td>37</td>
<td>28.68</td>
<td>3.13E-10</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0006412--translation</td>
<td>42</td>
<td>32.56</td>
<td>5.99E-07</td>
</tr>
<tr>
<td>GOTERM_CC_FAT</td>
<td>GO:0005840--ribosome</td>
<td>37</td>
<td>28.68</td>
<td>9.52E-06</td>
</tr>
<tr>
<td>GOTERM_CC_FAT</td>
<td>GO:0030529--ribonucleoprotein complex</td>
<td>37</td>
<td>28.68</td>
<td>2.17E-05</td>
</tr>
<tr>
<td>GOTERM_M_F_FAT</td>
<td>GO:0003723--RNA binding</td>
<td>29</td>
<td>22.48</td>
<td>2.39E-05</td>
</tr>
<tr>
<td>GOTERM_CC_FAT</td>
<td>GO:0043232--intracellular non-membrane-bounded organelle</td>
<td>38</td>
<td>29.46</td>
<td>9.00E-05</td>
</tr>
<tr>
<td>GOTERM_CC_FAT</td>
<td>GO:0043228--non-membrane-bounded organelle</td>
<td>38</td>
<td>29.46</td>
<td>9.00E-05</td>
</tr>
<tr>
<td>GOTERM_M_F_FAT</td>
<td>GO:0019843--rRNA binding</td>
<td>21</td>
<td>16.28</td>
<td>2.82E-04</td>
</tr>
<tr>
<td>GOTERM_M_F_FAT</td>
<td>GO:0000049--tRNA binding</td>
<td>7</td>
<td>5.42</td>
<td>0.048571</td>
</tr>
<tr>
<td>STY versus SPTA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0016051--carbohydrate biosynthetic process</td>
<td>10</td>
<td>4.4</td>
<td>5.55E-04</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0008610--lipid biosynthetic process</td>
<td>10</td>
<td>4.4</td>
<td>0.001</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0034637--cellular carbohydrate biosynthetic process</td>
<td>8</td>
<td>3.52</td>
<td>0.006</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0000271--polysaccharide biosynthetic process</td>
<td>7</td>
<td>3.08</td>
<td>0.009</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0009103--lipopolysaccharide biosynthetic process</td>
<td>6</td>
<td>2.64</td>
<td>0.014</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0008653--lipopolysaccharide metabolic process</td>
<td>6</td>
<td>2.64</td>
<td>0.014</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0005976--polysaccharide metabolic process</td>
<td>7</td>
<td>3.08</td>
<td>0.016</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0042426--cellular polysaccharide metabolic process</td>
<td>6</td>
<td>2.64</td>
<td>0.026</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0036929--cellular polysaccharide biosynthetic process</td>
<td>6</td>
<td>2.64</td>
<td>0.026</td>
</tr>
<tr>
<td>GOTERM_CC_FAT</td>
<td>GO:0030312--external encapsulating structure</td>
<td>7</td>
<td>3.08</td>
<td>0.033</td>
</tr>
<tr>
<td>STM versus SENT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOTERM_CC_FAT</td>
<td>GO:0019861--flagellum</td>
<td>8</td>
<td>4.14</td>
<td>0.037</td>
</tr>
<tr>
<td>GOTERM_CC_FAT</td>
<td>GO:0042995--cell projection</td>
<td>8</td>
<td>4.14</td>
<td>0.037</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0009405--pathogenesis</td>
<td>9</td>
<td>4.66</td>
<td>0.044</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pntd.0007416.t002
The classification of the four serovars into typhoidal and non-typhoidal groups is largely based on clinical presentation, with systemic and gastrointestinal disease, respectively. However, PCA of the LFQ intensities of the 418 detected proteins shared by all four serovars did not separate the typhoidal from the non-typhoidal serovars. Out of these 418 detected core proteins, 128 were significantly differentially expressed between typhoidal and the non-typhoidal serovars. However, GO analysis showed enrichment for proteins involved in translation and ribosomal activity, and thus largely represent the housekeeping machinery of the bacterial cells. PCA showed that the LFQ intensities of the reference and clinical isolates within the STY, SPTA, STM and SENT serovars do not cluster separately, and the reference strains can thus be considered as representative for the serovar.

Further analysis showed that 230 proteins were differentially expressed between STY and SPTA. GO analysis revealed that proteins involved in carbohydrate and lipopolysaccharide metabolism, and proteins involved in external encapsulating structures were most enriched. The regulators in the sub-network analysis connecting the differentially expressed proteins are implicated in the cell envelope stress response and in polysaccharide metabolism. For example, OmrA/B connect Dld and SdaB, two proteins that are involved in transport of sugars and carbohydrate biosynthesis in *E.coli*, respectively. It is plausible that a serovar-specific effect acts at the sRNA-level, which is not detected in our proteomic analysis. CpxR that is known to have a role in the response to alterations in the cell envelope in *Salmonella* [33], explains the expression of Psd and LpxA required for phospholipid and glycolipid metabolism, respectively.

Fig 3. Phenetic sub-network inference analysis of differential protein expression in STY versus SPTA. 122 out of 230 differentially expressed proteins are shown in this sub-network. Red nodes represent proteins with higher expression in SPTA versus STY. Green nodes represent proteins with higher expression in STY versus SPTA. The more intense the color, the higher the level of differential expression. Gray nodes have no differential expression. The color of the edge indicates the interaction type with blue referring to metabolic, green to protein-protein and red to protein-DNA interactions.
RpoS, RpoE and RpoH are involved in the stress response to different environmental conditions and contribute to Salmonella virulence [36–38]. CRP regulates the transcription of different operons involved in the transport of sugars and in catabolic functions [39], and FruR is required for carbohydrate metabolism [40]. The observation that cell surface proteins are significantly differently expressed between STY and SPTA is relevant for the diagnosis of Salmonella as well as for vaccination purposes. While the reference diagnostic method for typhoid fever is microbiological culture (blood, bone marrow or stool) and subsequent serotyping, rapid diagnostic tests (RDTs) have been developed and are commercially available for STY antigen and antibody detection [41]. However, diagnostic accuracy of the current RDTs is low, ranging from 31–97% [42] and more performant RDTs are urgently needed, including RDTs for SPTA. It has recently been shown that Salmonella antigen-based RDTs can be successfully applied to blood culture broths for Salmonella identification [43]. Three currently available typhoid vaccines are recommended by the WHO: an oral vaccine based on a live attenuated mutant strain of STY Ty21a (Ty21a), the injectable Vi capsular polysaccharide (ViCPS) vaccine and the typhoid conjugate vaccine (TCV) (http://www.who.int/immunization/policy/position_papers/typhoid/en/). However, these Typhi vaccines do not provide protection against paratyphoid fever caused by SPTA [44], and hence, a vaccine that protects against typhoid and paratyphoid fever would be of high value. When selecting antigens for developing new diagnostics or vaccines for both STY and SPTA, one should take into account that

Fig 4. Phenetic sub-network inference analysis of differential protein expression in STM versus SENT. 78 out of 192 differentially expressed proteins are shown in the sub-network. Red nodes represent proteins with higher expression in SENT versus STM. Green nodes represent proteins with higher expression in STM versus SENT. The more intense the color, the higher the level of differential expression. Gray nodes have no differential expression. The color of the edge indicates the interaction type with blue referring to metabolic and orange to protein-DNA interactions.

https://doi.org/10.1371/journal.pntd.0007416.g004

[34,35].
although encoded in both serovars, membrane proteins can be differentially expressed between both serovars and this should be tested in vitro and in vivo.

Upon comparing the proteomes of STM and SENT, 465 core proteins were detected, of which 192 were differentially expressed between the two serovars. GO enrichment analysis revealed that flagellar proteins and proteins involved in pathogenesis were most differentially expressed between both serovars. Among the higher expressed proteins in STM over SENT, six proteins are directly related to *Salmonella* pathogenicity island 1-encoded Type III secretion system (InvJ, SipA, SipD, SipC, PrgI, SipB). The T3SS-1 is an important virulence machinery that controls penetration of the gut epithelium during the infection by injecting effector proteins directly into the cytoplasm of epithelial cells through a needle-like appendages [45]. The regulator proteins InvJ and PrgI are known to be involved in needle and inner rod assembly [46], while SipA induces actin cytoskeletal rearrangements [47] and the translocases SipB and SipC form a translocation pore into the host cell membrane which is connected to the needle complex [48]. The sub-network also shows that HilA is possibly involved in the observed activation of the invasion proteins (SipA and PrgI) in STM. In addition, in the inferred sub-network the regulators FlhC (STM1924.S), FlhD and FliA were identified as regulators that connect 8 differentially expressed flagellar proteins (FlgL, FliD, FlgE, FlgM, FlgK, FlgD, FlgN, FlgG), showing higher expression profiles in Typhimurium strains. Besides their role in motility, flagellins were shown to stimulate both the innate and adaptive immune system and to cause inflammation upon STM infection [49]. Moreover, loss of flagellin expression in *Salmonella* has been linked to increased virulence in mice [50].

Some limitations in our study should be considered. The *Salmonella* strains were grown in standard in vitro conditions which may not be representative for protein expression in the infected host [51]. The addition of glucose to the medium may have induced catabolite repression. However, the addition of glucose as carbon source in needed to permit the growth of bacteria. Moreover, growth temperatures ranged between 35˚C and 37˚C and may have impacted expression levels. For instance, pathogenicity related gene expression is known to be temperature-sensitive [52]. In addition, the protein extraction procedure might have minorly affected the observed protein profiles although all steps have been performed on ice or 4˚C. However, all strains have been grown using the same in vitro culture conditions and underwent the same extraction procedure and any possible effects are thus very likely averaged out in the comparative analysis. In addition, our mass spectrometry set-up is not as sensitive as the newest instruments currently available, and we captured around 20 to 40% of the proteomes. Poorly expressed proteins in the standard in vitro culture conditions used may thus have been missed, such as virulence related proteins [53]. Finally, the aim of our study was to conduct a comparative analysis of orthologous proteins shared between the four *Salmonella* serovars, and as such, we do not present information on serovar-specific (non-orthologous) proteins.

In conclusion, to the best of our knowledge this is the first study that compared the core proteomes of a large panel of clinical *Salmonella* isolates, covering the four clinically most relevant *Salmonella enterica* serovars: Typhi, Paratyphi A, Typhimurium and Enteritidis. Our comparative proteome analysis indicated differences in the expression of surface proteins between STY and SPTA, and in pathogenesis-related proteins between STM and SENT. Our insights may guide future developed of novel diagnostics and vaccines, and understanding of disease progression.

**Supporting information**

S1 File. Composition of Hi-Def Azure medium.
(DOCX)
S2 File. Principal component analysis (PCA) of LFQ intensities of expressed proteins within the same serovar.

(DOCX)

S1 Table. Protein concentration in μg/ml.

(XLSX)

S2 Table. Pearson correlation coefficients between two biological replicates.

(XLSX)

S3 Table. Differentially expressed proteins between non-typhoidal (NTS) and typhoidal Salmonella strains based on log2 fold change of LFQ intensity levels (log2 FC).

(XLSX)

S4 Table. Differentially expressed proteins between S. Paratyphi (SPTA) and S. Typhi (STY) based on log2 fold change of LFQ intensity levels (log2-FC).

(XLSX)

S5 Table. Differentially expressed proteins between S. Enteritidis (SENT) and S. Typhimurium (STM) based on log2 fold change of LFQ intensity levels (log2-FC).

(XLSX)

Acknowledgments

We thank Tessa De Block for the technical assistance in the study. The partner institutes involved in this surveillance project that provided strains were: Sihanouk Hospital Centre of Hope, Phnom Penh, Cambodia and Institut National de Recherche Biomédicale, Kinshasa, Democratic Republic of the Congo. The stool isolate from The Gambia was received from the Medical Research Council (MRC) Keneba, MRC The Gambia, Keneba, The Gambia. The remaining strains were obtained from the travel clinic at ITM.

Author Contributions

Conceptualization: Sara Saleh, Sandra Van Puyvelde, Jan Jacobs, Stijn Deborggraeve.

Data curation: Sara Saleh, Sandra Van Puyvelde, Kris Gevaert, Stijn Deborggraeve.

Formal analysis: Sara Saleh, Sandra Van Puyvelde, Stijn Deborggraeve.

Funding acquisition: Jan Jacobs, Stijn Deborggraeve.

Investigation: Sara Saleh, Sandra Van Puyvelde, An Staes, Kris Gevaert, Stijn Deborggraeve.

Methodology: Sara Saleh, Sandra Van Puyvelde, An Staes, Evy Timmerman, Barbara Barbé, Jan Jacobs, Kris Gevaert, Stijn Deborggraeve.

Project administration: Stijn Deborggraeve.

Resources: Stijn Deborggraeve.

Software: Sara Saleh, An Staes, Evy Timmerman.

Supervision: Kris Gevaert, Stijn Deborggraeve.

Validation: Sara Saleh, Stijn Deborggraeve.

Visualization: Sara Saleh, Sandra Van Puyvelde.

Writing – original draft: Sara Saleh, Sandra Van Puyvelde, Stijn Deborggraeve.
Writing – review & editing: Sara Saleh, Sandra Van Puyvelde, Barbara Barbé, Jan Jacobs, Kris Gevaert, Stijn Deborggraeve.

References


