Original Article

Microarray analysis of virulence gene profiles in *Salmonella* serovars from food/food animal environment

Wen Zou[1], Sufian F. Al-Khaldi[2], William S. Branham[3], Tao Han[3], James C. Fuscoe[3], Jing Han[4], Steven L. Foley[4], Joshua Xu[5], Hong Fang[5], Carl E. Cerniglia[4], Rajesh Nayak[4]

[1]*Division of Personalized Nutrition and Medicine, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079, United States*
[2]*Division of Microbiological Studies, Food and Drug Administration, Center for Food Safety and Applied Nutrition, 5100 Paint Branch Parkway, College Park, MD 20740, United States*
[3]*Division of Systems Toxicology, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079, United States*
[4]*Division of Microbiology, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079, United States*
[5]*Z-Tech Corporation, an ICF International company at the National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079, United States*

Abstract

Introduction: Rapid, accurate and inexpensive analysis of the disease-causing potential of foodborne pathogens is an important consideration in food safety and biodefense, particularly in developing countries. The objective of this study is to demonstrate the use of a robust and inexpensive microarray platform to assay the virulence gene profiles in *Salmonella* from food and/or the food animal environment, and then use ArrayTrack™ for data analysis.

Methodology: The spotted array consisted of 69 selected *Salmonella*-specific virulence gene probes (65bp each). These probes were printed on poly-L-lysine-coated slides. Genomic DNA was digested with *Sau*3AI, labeled with Cy3 dye, hybridized to the gene probes, and the images were captured and analyzed by GenePix 4000B and ArrayTrack™, a free software developed by Food and Drug Administration (FDA) researchers.

Results: Nearly 58% of the virulence-associated genes tested were present in all *Salmonella* strains tested. In general, genes belonging to *inv, pip, prg, sic, sip, spa* or *ttr* families were detected in more than 90% of the isolates, while the *iacP*, *avrA, invH, rhuM, sirA, sopB, sopE* or *sugR* genes were detected in 40 to 80% of the isolates. The gene variability was independent of the *Salmonella* serotype.

Conclusions: This hybridization array presents an accurate and cost-effective method for evaluating the disease-causing potential of *Salmonella* in outbreak investigations by targeting a selective set of *Salmonella*-associated virulence genes.

Key words: *Salmonella*; virulence associated genes; microarray; pathogenicity

J Infect Dev Ctries 2011; 5(2):094-105.

(Received 15 July 2010 - Accepted 15 September 2010)

Introduction

It has been estimated that approximately 1.3 billion cases of salmonellosis occur worldwide each year [1]. *Salmonella* infections can largely depend on the immune status of the host and virulence factors such as *Salmonella* pathogenicity islands (SPIs), plasmids, toxins, fimbriae and flagella [2,3]. The nature of infections caused by *Salmonella* in humans and animals could depend on a variety of bacterial and host factors, and their complex interactive environment [4]. Multiple SPIs encode type III secretion systems that transport bacterial proteins (SptP and SopE) into the cytosol of the target host cells facilitating the uptake of the bacterium by the host cells [5-7]. The SPI genes are also essential for *Salmonella* to proliferate within host cells and cause systemic infections [8,9]. They also encode the *mgtBC* operon and are required for intra-macrophage survival. *Salmonella* serovars also harbor plasmids that contain virulence-associated genes, such as *spvRABCD* [10]. In addition, the toxin-encoding genes, fimbria-encoding genes, and the flagella-encoding gene system play diverse roles in *Salmonella* pathogenesis [11]. The genomic

reservoir of *Salmonella* species contains horizontally transferred genetic elements, including some virulence genes that may play roles in pathogenicity and disease development [12]. The characterization of virulence-associated genes is important in identifying *Salmonella* pathogenicity, understanding the potential transfer mechanisms, and developing an efficient detection method in epidemic disease control.

DNA microarrays have demonstrated great potential for analysis of gene expression, genotyping, pathway analysis, monitoring changes in genomic DNA, and host-pathogen interaction [7,12,13]. Microarray techniques have been useful in high-throughput genetic profiling of pathogenic microorganisms [14,15]. This technology can also detect the presence or absence of thousands of genes simultaneously by a single genomic hybridization step [16,17]. Spotted DNA microarray platforms can be cost-effective and easy to reproduce in a laboratory setting with basic infrastructure [18]. Furthermore, interpretation of microarray data is easier to automate and standardize than that of gel-based technologies [19]. The objective of this study was to demonstrate the usefulness of a robust microarray platform to detect the virulence-associated gene profiles in different *Salmonella* serovars from food and/or the food animal environment, and the use of ArrayTrack™, a free software developed by researchers at the Food and Drug Administration (FDA) for data analysis.

Table 1. *Salmonella* isolates used in this study

Strain	Serotype	Institute	Source	Comments
ATCC14028	Typhimurium	FDA-National Center for Toxicological Research	ATCC14028	(+) control strain
VV302 (Δ*hilA*-523) SVM725 (Δ*invF*)	Typhimurium	Washington University	Dr. Virginia L. Miller's Lab	(-) control strain
1 12	Enteritidis Heidelberg	FDA- Office of Regulatory Affairs	Eggs	Chicken egg house
180 181 182 187 188 193 197 219	Miami Anatum Tallahassee Newport Inverness Manhattan Florida Give	FDA-Center for Food Safety and Applied Nutrition	Water Water Water Water Filter sand Water Water Ditch water	Roma tomato outbreak strains
524 528 580 585 589 601 606 614 801 802 804 805 806 807	Muenster Anatum Heidelberg Heidelberg Heidelberg Senftenberg Senftenberg Worthington Untypeable *Salmonella* Untypeable *Salmonella* Untypeable *Salmonella* Untypeable *Salmonella* Untypeable *Salmonella* Muenster	West Virginia University	Litter Waterers Waterers Litter Litter Feed Feed Leftover feed Turkey Litter Waterer Waterer Waterer Swab	Turkey farm

Methodology

Bacterial strains and DNA preparation

A total of 24 *Salmonella enterica* isolates were analyzed in this study (Table 1). *S. enterica* subspecies *enterica* serovar Enteritidis (*S*. Enteritidis) strain1 and *S.* Heidelberg strain 12 were obtained from the FDA's Office of Regulatory Affairs culture collection and the outbreak isolates were obtained from the FDA's Center for Food Safety and Applied Nutrition's investigation of the Roma tomato outbreak. These isolates were obtained as pure cultures for this study. *Salmonella* isolates from turkey farms were part of our previous study [20]. Microarray validation was conducted using *S.* Typhimurium ATCC 14028 as the positive control strain for the probes used in this study and *S.* Typhimurium VV302 (Δ*hil*A-523) and *S.* Typhimurium SVM725 (Δ*inv*F) as the negative controls. The bacterial genomic DNA was isolated from a freshly grown (18 to 24 hours) bacterial culture using a Wizard Genomic DNA Purification kit (Promega Corporation, Madison, WI, USA) and quantified by measuring the absorbance at 260 nm (NanoDrop ND-1000 V3.3 spectrophotometer, NanoDrop Technologies, Wilmington, DE, USA) for microarray and PCR experiments.

Selection of target genes and probe design

Sixty-nine gene-specific oligonucleotide probes (Operon Technologies Alameda, CA, USA) were designed based on the open reading frame sequences of *Salmonella* pathogenicity islands (SPI-1 to SPI-5) virulence-associated genes [9] from the National Center for Biotechnology Information (NCBI) GenBank (Bethesda, MD, USA) and the Array Designer software (Premier Biosoft International, Palo Alto, CA, USA), using *Salmonella enterica* Typhimurium LT2 genome NC_003197 as the reference genome. To increase the efficacy of our microarray chip, two additional probes, *celB* (5'-GGGGATCCAGCTGAAT GGACAGGTGGTGATCAAGAAGGTTGGAATTT CGTCAATGAAATG-3') and *celF* (5'-TGGTGCCGGTAGTCTTAACACTTACAAGGGT TATGTTGACAACATTTCTAGAACTATTCG-3') (MWG Biotech AG, Ebersberg, Germany) were added as the negative gene controls. The *celB* and *celF* were designed from the cellulase encoding gene (Accession number: U57818) and the 1,4-beta-D-glucan-cellobiohydrolase encoding gene (Accession number: U97154) from the fungi *Orpinomyces sp*. PC-2, respectively [21]. These probes exhibited no homology to the *Salmonella* genome sequences in the GenBank database.

Microarray printing and processing

Salmonella oligonucleotide probes (50 µM each) were printed on poly-L-lysine-coated slides (Erie Scientific, Portsmouth, NH, USA) using SMP3 printing pins (TeleChem International, Sunnyvale, CA, USA) on an OmniGrid^{TM} 100 Microarrayer (GeneMachines, San Carlos, CA, USA). Printed slides were baked at 80°C for 1 hour and UV cross-linked (UV Stratalinker 2400, Stratagene, La Jolla, CA). The slides were then treated with a blocking solution of 3X SSC (1X solution is 150 mM sodium chloride and 15 mM sodium citrate, pH 7.0), 0.1% SDS (sodium dodecyl sulfate) and 1% BSA (bovine serum albumin) using gentle agitation for 5 minutes at 50°C and washed with Milli-Q water four times for 1 minute each at room temperature. The slides were placed in boiling Milli-Q water for 2 minutes, followed by a 1 minute wash in ethanol at room temperature. The slides were dried in a microarray high speed centrifuge (TeleChem International, Sunnyvale, CA, USA) and stored at room temperature with low humidity until used.

Eight identical arrays were printed on each microarray chip for simultaneous duplicate analysis of four different bacterial isolates. Each virulence-associated gene was represented by two identical probes in different locations within the array to minimize systematic errors.

Fluorescence labeling of genomic DNA, microarray hybridization and slide scanning

Genomic DNA from each test isolate was digested with *Sau*3AI (Promega, Madison, WI, USA) at 37^{o}C for 2 hours. The digested DNA was precipitated and chemically labeled using the Micromax Cy3 labeling dye (Perkin-Elmer Life Science, Inc., Boston, MA, USA) according to the manufacture's protocol. The labeled DNA was purified using PCR purification kit (Qiagen Sciences, MA, USA) and resuspended in 1X hybridization buffer (5X Denhardt's solution, 6X SSC, and 0.1%Tween 20) at 90^{o}C for 2 minutes, followed by incubation with the same buffer for 60 to 90 minutes at 60^{o}C for hybridization to occur. After hybridization, the arrays were washed and scanned by GenePix 4000B (Molecular Devices, Sunnyvale, CA). Fluorescent images were captured and analyzed using the GenePix Pro 6.0 software. Each template hybridization experiment was conducted at least three

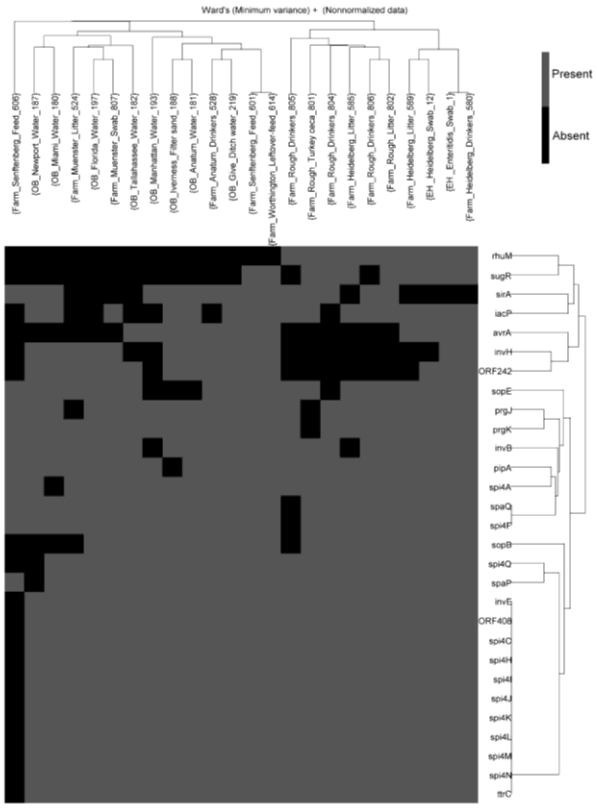
times with the genomic DNA isolated at different times and each of the hybridizations was evaluated on at least three microarray slides. The genomic DNA of *S.* Typhimurium ATCC 14028 was used as a positive control on every chip, hybridized to a randomly selected array along with the test strains. Genomic DNA from *S.* Typhimurium VV302 (Δ*hil*A-523) and *S.* Typhimurium SVM725 (Δ*inv*F) were added to assess the performance of the microarray chip.

Data normalization and analysis

The fluorescence intensity for each probe was calculated by subtracting the median value of the local background intensities. For normalization of the Cy3 probe signals, the local background-subtracted intensity of each probe was divided by the median value of the total signals from all probes in the same array. Any spot which showed a ratio greater than 1.0 was counted as positive. Ratios less than 1.0 were then standardized by comparison with those of the corresponding probes of the positive control strain, *S.* Typhimurium ATCC 14028, in the same chip. The standardized ratios of the two duplicate probes on each of the two chips (4 spots total) for each gene were averaged as the measurement of the signal strength for the gene. A ratio of < 0.8 indicated absence and a ratio > 0.9 indicated presence of a probe sequence. Values between 0.8 and 0.9 were classified as uncertain. The two threshold cutoffs, 0.8 and 0.9, were determined by evaluating the data from all 69 virulence-associated gene probes for the *Salmonella* isolates.

Hierarchical clustering analysis (HCA) was performed with ArrayTrack™, an open sharing software developed by researchers at the US Food and Drug Administration. ArrayTrack™ is a bioinformatics tool that provides an integrated environment for genomic data management, analysis, and interpretation with a focus on microarray data. This software is freely available to the scientific community through the FDA website (http://www.fda.gov/ArrayTrack). The user manual and tutorials are available from the website.

Figure 1. A flag-based two-way hierarchical clustering analysis of the 24 *Salmonella* isolates based on all 29 virulence genes that varied between isolates.



Ward's minimum variance algorithm was used for clustering analysis. The clustering dendrogram on the top represents grouping by the isolates and their origin [egg house (EH), farm or outbreak (OB)], while the dendrogram displayed on the right of the figure represents grouping based on the 29 variable virulence genes. Gray boxes indicate presence of a gene and black boxes indicate absence of a gene.

PCR verification of microarray results

PCR was used to screen the all the bacterial isolates for the presence of following genes: *purR, rmbA, rhuM, sugR, spi4H, ttrB, iacP, avrA, prgK, invH* and *sopE* (Table 2). Most genes selected for PCR screening were either absent or elicited weak hybridization signals. The *16S rRNA* and *purR* genes were used as positive controls and *celB* as a negative control for PCR analysis. A typical PCR (25 µl) contained 0.6 pmol/l of each primer, 12.5 µl 2X PCR MasterMix (Core System Kit, Promega) and 40 ng of purified genomic DNA. PCR amplification was conducted by incubating the samples at 94^{o}C for 5 minutes, followed by 30 cycles of 94^{o}C for 30 seconds, variable melting temperatures (Table 2) for 30 seconds and 72^{o}C for 30 seconds. A 10 µl aliquot

of PCR products was loaded on a 2% agarose E-gel (Invitrogen, Carlsbad, CA, USA) and separated according to the manufacture's instructions. Whenever there was a discrepancy between the PCR and microarray results for a particular gene target, representative PCR-positive products were sequenced to verify the PCR results. The PCR amplicons were labeled using the BigDye Terminator Cycle Sequencing Kit (version 3.1, Applied Biosystems, Foster City, CA, USA) and separated using an ABI Prism 310 Automatic Sequencer (Applied Biosystems) to confirm the identity of the PCR product.

Table 2. Primers used for PCR amplification of virulence-associated genes in *Salmonella*

Target gene	Accession No.	Length (bp)	Direction	Tm (^{o}C)	Primer sequence (5'→3')	Reference
iacP	U40013 (SPI-1)	176	F R	57 57	CACCTCTTGTATTGCCGTTG GGCATATATCCGCAAAGGTC	This study
avrA	AF013573 (SPI-1)	170	F R	55 59	AATGGAAGGCGTTGAATCTG GAGCTGCTTTGGTCCTCAAC	This study
invH	U84286 (SPI-I)	153	F R	57 57	AGCAACTGGCCAACGCAAAT TGCAGTCTTTCATGGGCAGCAA	This study
prgK	U21676 (SPI-I)	217	F R	57 53	TTGAACAGCGACTGGAACAG TCATAATCCACATCGGCAAA	This study
sopE	AF043239 (SPI-I)	186	F R	59 55	ATTGTTGTGGCGTTGGCATCGT AATGCGAGTAAAGATCCGGCCT	This study
rhuM	AF106566 (SPI-3)	222	F R	59 57	CATCGGCTGTACCCGACTAT CAGCACGCTGATGAATGAGT	This study
sugR	AF106566 (SPI-3)	152	F R	59 57	ACTGCTCGCTTGTGGTATCAGGAA TCTTACCGTGCTGTGGATGGTTCA	This study
spi4H	AF06080869 (SPI-4)	154	F R	54 53	CGCTGACGGTCGTTATCG TCAATGCTCAGACGGACTTC	[9]
rmbA	AF106566 (SPI-3)	454	F R	54 54	AGCCTTCACAAATTGTCCATTG TCCGTATAGTTAAGCGTTCGTC	[9]
ttrB	AJ224978 (SPI-2)	608	F R	56 57	ATGTGGACGGGAGTCAATATGG GTGGCGATGCGGCTATGG	[9]
purR	AF040636	629	F R	55 55	CGCTTCCCCTTTTCCTCAAG CCATCACCACCATCGGAATATG	[9]
16S rRNA	Z49264	300	F R	54 54	CCTGGCTCAGATTGAACGC AGTGTGGCTGGTCATCCTC	[9]
celB	U57818	206	F R	55 53	TGGAGTGGTCATTTTGGTGA TCAGCAGCAATTTGTTTCCA	This study

Results

Distribution of virulence-associated genes in Salmonella *serovars*

The distribution of 69 virulence-associated genes in *Salmonella* is summarized in Table 3. The positive control genes, *16S rRNA* and purine nucleotide synthesis repressor (*purR*), were detected in all *Salmonella* strains, while only background hybridization signals were observed on spots of the negative controls, *celB*, *celF*, 2X SSC, printing buffer and blanks (data not shown). Overall, 58% of the virulence-associated genes (40/69) were present in all 24 *Salmonella* isolates tested while the remaining genes were variable in their distribution. In this study, regardless of the serotype, *Salmonella* isolates exhibited most variability among the *iacP*, *avrA*, *sopE*, *sirA* and *invH* genes belonging to the SPI-1 class (Table 3). The *Salmonella* acyl carrier protein-encoding gene, *iacP*, was present in 71% (17/24) of the strains. The *avrA* gene, which encodes the secreted effector protein in *Salmonella*, was detected in 50% of the strains. The *sirA* gene was detected in 63% of *Salmonella* isolates. This gene encodes a two-component response regulator of the FixJ family that has a positive regulatory influence on the expression of Type III secretion genes involved with epithelial cell invasion and the elicitation of bovine gastroenteritis. Thirteen *Salmonella* isolates (54%) were positive for *invH*, which encodes the protein required for entry of the bacteria into cultured epithelial cells. The SPI-2 associated virulence genes were well conserved in most *Salmonella* serotypes; *ORF242* was found to be variable in *S.* Heidelberg and absent from all untypeable *Salmonella* rough isolates (Table 3).

The SPI-3 genes were conserved among all serotypes tested, with the exception of *rhuM* and *sugR*. These two genes were found to be absent from most serotypes, with the exception of *S.* Heidelberg, *S.* Enteritidis and untypeable *Salmonella* rough isolates (Table 3). The *sugR* and *rhuM* genes located in *S. enterica* SPI-3 and encoding the putative ATP binding protein and a cytoplasmic protein, respectively, were detected in 42% of the isolates. None of the *Salmonella* outbreak strains showed positive signals with the *sugR* and *rhuM* probes on the microarray chip (Table 3). The SPI-3-associated *mgtB* and *mgtC* genes, which are important for intracellular *Salmonella* replication, were present in all serotypes. The SPI-4 associated genes were present in all *Salmonella* serotypes; these genes have sequence similarity to genes required for survival in macrophages. Interestingly, the *spi4C, spi4H* to *spi4N* and *spi4Q* genes were absent from *S.* Senftenberg strain 606 isolated from the poultry farm. Lastly, with the exception of *sopB* and *pipA* genes, the SPI-5 associated genes were conserved in all *Salmonella* serotypes. The virulence-associated effector protein-encoding gene, *sopB*, was detected in 79% of the isolates.

Overall, there appeared to be no major differences in the virulence-associated gene profiles of the poultry (egg houses and farm) and outbreak strains, nor any differences among the multiple isolates of serovars Heidelberg, Muenster, Anatum or untypeable isolates (Table 3). However, the two isolates of *S.* Senftenberg (601 and 606) displayed differences for 13 virulence genes (*iacP*, *ORF242*, *ORF408*, *sopB*, *spi4H/I/J/K/L/M/N/Q* and *ttrC*); these genes were absent from isolate 606.

Figure 1 illustrates a flag-based two-way HCA based on 29 virulence genes that showed variability among the 24 *Salmonella* isolates. In general, the outbreak and farm isolates appear to group in two distinct clusters. The HCA showed that the five untypeable *Salmonella* isolates were sub grouped with one *S.* Heidelberg (strain 585); all six strains were farm isolates, while *S.* Enteritidis strain 1 and *S.* Heidelberg strain 12 isolates from egg houses were sub grouped with the two *S.* Heidelberg farm isolates (580 and 589). The outbreak strains were grouped in a distinct cluster and shared profiles with six isolates from the farm (strains 606, 524, 807, 528, 601 and 614).

Validation of the microarray signals by PCR analysis and sequencing

Of the 275 PCR reactions (11 gene amplifications x 25 isolates), there was 88% (242/275) agreement between the PCR and hybridization results (Table 4). Genes in the remaining 12% of discordant results tested positive with PCR, but were found to be negative (signal) with the hybridization experiments; all PCR reactions amplified a product of the predicted size for the corresponding genes (Table 2). To further verify the discordant results, PCR amplicons of *rhuM*, *spi4H* and *prgK* genes for *S.* Miami strain 180, *S.* Senftenberg strain 606, and *Salmonella* strain 801, respectively, and representative amplicons of *sugR*, *iacP*, *avr*, *invH* and *sopE* genes were sequenced and found to match the predicted genes based on NCBI's Blast search results.

Table 3. Prevalence of virulence-associated genes in 24 *Salmonella* isolates from food/food animal environment

Gene	Region	S. Typhimurium ATCC14028	S. Heidelberg 12	S. Heidelberg 580	S. Heidelberg 585	S. Heidelberg 589	*Salmonella* 801	*Salmonella* 802	*Salmonella* 804	*Salmonella* 805	*Salmonella* 806	S. Muenster 524	S. Muenster 807	S. Senftenberg 601	S. Senftenberg 606	S. Anatum 528	S. Anatum 181	S. Worthington 614	S. Enteritidis 1	S. Miami 180	S. Tallahassee 182	S. Newport 187	S. Inverness 188	S. Manhattan 193	S. Florida 197	S. Give 219
avrA	SPI-1		+	+	-	+	-	-	-	-	-	-	-	+	-	+	+	+	+	-	+	-	+	+	-	+
hilA	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
iacP	SPI-1	-	+	+	+	+	+	+	-	+	+	-	+	+	-	-	+	+	+	+	-	+	+	-	-	+
invA	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
invB	SPI-1	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+
invC	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
invE	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+
invF	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
invG	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
invH	SPI-1		-	+	-	-	-	-	-	-	-	+	+	+	-	+	+	+	+	+	-	+	+	-	+	+
orgA	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
prgH	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
prgI	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
prgJ	SPI-1	+	+	+	+	+	-	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
prgK	SPI-1		+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sicA	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sicP	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sipA	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sipB	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sipC	SPI-1	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sipD	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sirA	SPI-1	+	-	-	-	-	+	+	+	+	+	-	-	+	+	+	+	+	-	+	-	+	+	+	-	+
sopE	SPI-1	-	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-	+	+	+	+	+	-	-	+	+
spaO	SPI-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
spaP	SPI-1	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	-	+	+	+	+
spaQ	SPI-1	+	+	+	+	+	+	+	+	-	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+
ORF 242	SPI-2	+	+	+	-	-	-	-	-	-	-	+	+	+	-	+	+	+	+		+	+	+	-	+	+
ORF 245	SPI-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ORF 319	SPI-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ORF 408	SPI-2	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+
ORF 70	SPI-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ttrA	SPI-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ttrB	SPI-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ttrC	SPI-2	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+
ttrR	SPI-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ttrS	SPI-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cigR	SPI-3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
fidL	SPI-3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

marT	**SPI-3**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
mgtB	**SPI-3**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
mgtC	**SPI-3**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
misL	**SPI-3**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rhuM	**SPI-3**	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-
rmbA	**SPI-3**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
slsA	**SPI-3**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sugR	**SPI-3**	+	+	+	+	+	+	+	+	-	-	-	-	+	-	-	-	+	+	-	-	-	-	-	-
spi4A	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
spi4B	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
spi4C	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+
spi4E	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
spi4F	**SPI-4**	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
spi4H	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+
spi4I	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+
spi4J	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+
spi4K	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+
spi4L	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+
spi4M	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+
spi4N	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+
spi4O	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
spi4P	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
spi4Q	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+	+	+
spiD	**SPI-4**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
pipA	**SPI-5**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+
pipB	**SPI-5**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
pipC	**SPI-5**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
pipD	**SPI-5**	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sopB	**SPI-5**	+	+	+	+	+	+	+	+	-	+	-	+	+	-	+	+	+	+	-	+	-	+	+	+
purR		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
yihT		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Gray box indicates gene presence; white box indicates gene absence; and black indicates uncertain existence of the gene based on threshold cutoff used for data normalization.

Table 4. Comparison of microarray hybridization and PCR amplification data

Isolate	Virulence-associated genes										
	purR	*rmbA*	*rhuM*	*sugR*	*spi4H*	*ttrB*	*iacP*	*avrA*	*prgK*	*InvH*	*sopE*
S. Enteritidis 1	p	p	p	a	p	p	p	p	p	p	p
S. Heidelberg 12	p	p	p	p	p	p	p	p	p	p	p
S. Typhimurium ATCC14028	p	p	p	p	p	p	p	p	p	p	p
S. Miami 180	p	p	p	a	p	p	p	p	p	p	p
S. Anatum 181	p	p	a	a	p	p	p	p	p	p	p
S. Tallahassee 182	p	p	a	a	p	p	p	p	p	p	p
S. Newport 187	p	p	a	a	p	p	p	p	p	p	p
S. Inverness 188	p	p	a	a	p	p	p	p	p	p	p
S. Manhattan 193	p	p	a	a	p	p	p	p	p	p	p
S. Florida 197	p	p	a	a	p	p	p	a	p	p	p
S. Give 219	p	p	a	a	p	p	p	p	p	p	p
S. Muenster 524	p	p	a	a	p	p	p	a	p	p	p
S. Anatum 528	p	p	a	a	p	p	p	p	p	p	p
S. Heidelberg 580	p	p	p	p	p	p	p	p	p	p	p
S. Heidelberg 585	p	p	p	p	p	p	p	p	p	p	p
S. Heidelberg 589	p	p	p	p	p	p	p	p	p	p	p
S. Senftenberg 601	p	p	a	p	p	p	p	p	p	p	p
S. Senftenberg 606	p	p	a	p	p	p	p	p	p	p	p
S. Worthington 614	p	p	a	p	p	p	p	p	p	p	p
Untypeable *Salmonella* 801	p	p	p	p	p	p	p	a	p	p	p
Untypeable *Salmonella* 802	p	p	p	p	p	p	p	a	p	p	p
Untypeable *Salmonella* 804	p	p	p	p	p	p	p	a	p	p	p
Untypeable *Salmonella* 805	p	p	p	a	p	p	p	a	p	p	p
Untypeable *Salmonella* 806	p	p	p	a	p	p	p	a	p	p	p
S. Muenster 807	p	p	a	a	p	p	p	a	p	p	p

a: absent; p: present; Highlighted cells represent the discordant results between the microarray analysis and PCR assay. These were positive by PCR and negative by microarray.

Discussion

Although 97% of the genome sequence is identical among different *Salmonella* serovars [22]; comparative genomics using microarray have revealed conserved and variable gene components associated with fimbriae, pathogenicity and phage elements [23,24]. Virulence genes and plasmids can be used as biomarkers for detection of *Salmonella* serotypes, such as Typhimurium and Newport [25]. This study examined the profiles of 69 virulence-associated genes within *Salmonella* serovars from different environmental sources. Most genes in the spotted array were located in SPI-1 to SPI-3, and encode proteins responsible for secretion and translocation of *Salmonella* proteins in the host cells and intracellular survival and replication [7,9], while the other SPI (SPI-2, SPI-4 and SPI-5)-associated genes encode effector proteins that facilitate intracellular survival of *Salmonella* in the host cells, T1SS toxins and survival of these bacteria in macrophages [8, 26, 27]. This platform adds to list of other studies wherein microarray technology has been used for diagnostic characterization of disease-causing potential of bacterial pathogens [7,13,16,18].

The virulence genes associated with SPI-1 encode effector proteins that disrupt cytoskeletal and bacterial cell barriers resulting in *Salmonella* invasion of the host gastrointestinal epithelium [27]. The *sirA* and *invH* genes were absent from *S.* Heidelberg isolates (with the exception of *S.* Heidelberg strain 580 for *invH* gene), while *avrA* and *invH* were absent from the untypeable *Salmonella* isolates. However, a recent study has shown that SPI-1 deficient *S.* Senftenberg can cause human

enteropathogenic infections [26], indicating that SPI-1 associated genes are not essential to cause human gastroenteritis in this *Salmonella* serotype. In our study, all SPI-1 related genes in *S.* Senftenberg strain 601 were present; however, the *avrA, iacP*, *invE* and *invH* were absent from *S.* Senftenberg strain 606 (Table 3), suggesting that other virulence factors may contribute to the pathogenicity of this isolate. Several other virulence genes from SPI-2 (*ORF242*, *ORF408* and *ttrC*), SPI-3 (*rhuM* and *sugR*), SPI-4 (*spi*-class) and SPI-5 (*sopB*) were concurrently found to be absent from *S.* Senftenberg strain 606 as opposed to *S.* Senftenberg strain 601, suggesting the uniqueness of this strain.

Of the genes analyzed, 58% were found in all the isolates from the various serovars and sources, which potentially indicate that they may serve as a core set of virulence-associated genes in *Salmonella enterica.* Outside this core group, there was diversity in the virulence factor profiles among the isolates: only two isolates (*S.* Senftenberg strain 601 and *S.* Worthington strain 614) shared a common profile and both strains were isolated from the same sampled flock. On the other hand, the two isolates of *S.* Senftenberg (601 and 606) displayed variability for 13 virulence genes (Table 3), indicating that when multiple isolates of *Salmonella* serotypes were analyzed, differences in gene content could be detected in all of them. Additional strains of *S.* Senftenberg should be tested to evaluate if these gene differences were unique to these isolates. There were no unique pattern differences in the virulence gene profiles between farm/egg house and the outbreak strains. *S.* Enteritidis strain 1 and *S.* Heidelberg strain 12, isolated from the same egg house, exhibited nearly identical virulence-associated gene profiles, despite belonging to different serovars (Table 3).

The *avrA, iacP, invH, ORF242, rhuM, sirA or sugR* genes were absent from several *Salmonella* isolates. *S.* Heidelberg strains 585, 589, 606, *S.* Manhattan strain 193, and unidentifiable *Salmonella* strains 801, 802, 804, 805 and 806 were found not to possess the *ORF242* and *invH* genes. *ORF242*, located in SPI-2, encodes a protein similar to the AraC-like family of transcriptional regulators [28]. There was no evidence of a relationship between *ORF242* and *invH* in SPI-1, a lipoprotein required for *invG* localization to the outer membrane [29].

A subset of the target genes were selected for PCR analysis to specifically validate the microarray data demonstrating weaker signals. Nearly 88% accordance was observed between the microarray and PCR data. The discrepancies resulted from a target being detected by PCR and not meeting the positive detection threshold using the microarray probes. Over half (19/33) of the discrepancies were observed for *iacP* or *invH* genes and were probably caused by the stringency of hybridization conditions or sequence variability in the probe binding site. Future refinement of the array will need to address the problems with these two probes.

Although whole genome arrays can generate a lot of information, they are not always cost effective and require expensive software for data analysis. To support such analysis, ArrayTrack™ has been recently been upgraded to manage and analyze the genetic profiling data related to bacterial foodborne pathogens [30]. ArrayTrack™ libraries have been populated with bioinformatics data from the public domains related to bacterial pathogen species. Data processing and visualization tools have been enhanced with customized options to facilitate analysis of genetic profiling microarray data. Specifically, three new functions have been developed and are particularly effective for analysis of the microarray data: flag-based HCA, a flag concordance (FC) heat map, and flag indicators in the mixed scatter plot. These functions can be relevant and effective for the identification and characterization of bacterial pathogens through microarray genetic profiling data [30].

One of the benefits of this spotted array technology is the ability to rapidly and simultaneously analyze multiple genes on a single platform. Since much of the *Salmonella enterica* genome is invariant, we have chosen a unique set of virulence genes that would represent *Salmonella enterica* pathogenicity. The chip utilized in this study has the added benefit that it contains eight identical arrays, enabling simultaneous analysis of four samples in duplicate. This design results in a per sample cost reduction and an increase in the sample throughput over single arrays. Such a platform, along with the free analysis software, could be particularly useful for laboratories with limited resources. When this improved microarray chip design was compared to a microarray containing a similar but single set of probes [9], the multi-sample array functioned identically. Other microarray platforms previously reported have evaluated only a select list of virulence genes [16,25].

In summary, our study highlights a simple and effective method for a single-step screening of multiple virulence-associated genes for multiple

Salmonella enterica isolates that may render these bacteria pathogenic in a food and/or food animal environment. The ArrayTrack™ free software for microarray data analysis could be used by researchers in developing countries. Such a detection system can be easily modified and adapted to include additional probes for newly described virulence-associated genes such as those encoding for regulation-effector (*phoP*/*phoQ*), fimbriae (*safC*, *sefA*, *stbD* and *stcC*), or phage-associated genes (*sseI*, *gtgA* and *STM4210*) [12, 31]. High density screening of *Salmonella* isolates for genetic elements, such as pathogenicity islands, plasmids, and phages can provide better insights into the mechanisms of acquisition of these virulence factors by a particular strain.

Acknowledgements

We thank Drs. Mark Walderhaug from the USFDA Center for Food Safety and Applied Nutrition (CFSAN), Brett Kenney from West Virginia University, Doris Farmer from the USFDA Office of Regulatory Affairs (ORA), and Virginia L. Miller from Washington University for providing the *Salmonella* test and control strains. We thank Dr. Ronald Lorentzen from CFSAN for the travel fund from the USFDA Scientific Collaboration Program between CFSAN and the National Center for Toxicological Research (NCTR). We are grateful to Drs. James J. Chen, John Sutherland, Chris A. Elkins, Huizhong Chen, and Nathan Twaddle for critical reading of this manuscript. Wen Zou and Jing Han acknowledge support of their fellowships from the Oak Ridge Institute for Science and Education, administered through an interagency agreement between the US Department of Energy and the US Food and Drug Administration.

Disclaimer

The use of the trade names is for identification purposes only and does not imply endorsement by the Food and Drug Administration or by the US Department of Health and Human Services. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Food and Drug Administration. The authors have no conflicting financial interests.

References

1. Coburn B, Grassl GA, Finlay BB (2007) *Salmonella*, the host and disease: a brief review. Immunol Cell Biol 85: 112-118.
2. van Asten AJ, van Dijk JE (2005) Distribution of "classic" virulence factors among *Salmonella* spp. FEMS Immunol Med Microbiol 44: 251-259.
3. Fierer J, Guiney DG (2001) Diverse virulence traits underlying different clinical outcomes of *Salmonella* infection. J Clin Invest 2001 107: 775-780.
4. Huehn S, La Ragione RM, Anjum M, Saunders M, Woodward MJ, Bunge C, Helmuth R, Hauser E, Guerra B, Beutlich J, Brisabois A, Peters T, Svensson L, Madajczak G, Litrup E, Imre A, Herra-Leon S, mevius D, Newell DG, Malorny B (2010) Virulotyping and antimicrobial resistance typing of *Salmonella enterica* serovars relevant to human health in Europe. Foodborne Pathog Dis 7: 523-535.
5. Hayward RD, Koronakis V (2002) Direct modulation of the host cell cytoskeleton by *Salmonella* actin-binding proteins. Trends Cell Biol 12: 15-20.
6. Muller AJ, Hoffmann C, Galle M, Van Den Broeke A, Heikenwalder M, Falter L, Misselwitz B, Kremer M, Beyaert R, Hardt WD (2009) The *S.* Typhimurium effector SopE induces caspase-1 activation in stromal cells to initiate gut inflammation. Cell Host Microbe 6: 125-136.
7. Litrup E, Torpdahl M, Malorny B, Huehn S, Helms M, Christensen H, Nielsen EM (2010) DNA microarray analysis of *Salmonella* serotype Typhimurium strains causing different symptoms of disease. BMC Microbiol 10: 96.
8. Schmidt H, Hensel M (2004) Pathogenicity islands in bacterial pathogenesis. Clin Microbiol Rev 17: 14-56.
9. Courtney S, Mossoba ME, Hammack TS, Keys C, Al-Khaldi SF (2006) Using PCR amplification to increase the confidence level of *Salmonella* Typhimurium DNA microarray chip hybridization. Mol Cell Probes 20: 163-171.
10. Rotger R, Casadesus J (1999) The virulence plasmids of *Salmonella*. Int Microbiol 2: 177-184.
11. Collinson SK, Liu SL, Clouthier SC, Banser PA, Doran JL, Sanderson KE, Kay WW (1996) The location of four fimbrin-encoding genes *agfA*, *fimA*, *sefA* and *sefD* on the *Salmonella* Enteritidis and/or *S.* Typhimurium *XbaI-BlnI* genomic restriction maps. Gene 169: 75-80.
12. Malorny B, Bunge C, Guerra B, Prietz S, Helmuth R (2007) Molecular characterisation of *Salmonella* strains by an oligonucleotide multiprobe microarray. Mol Cell Probes 21: 56-65.
13. Kato-Maeda M, Gao Q, Small PM (2001) Microarray analysis of pathogens and their interaction with hosts. Cell Microbiol 3: 713-719.
14. Al-Khaldi SF, Mossoba MM (2004) Gene and bacterial identification using high-throughput technologies: genomics, proteomics, and phonemics. Nutrition 20: 32-38.
15. Lindsey RL, Frye JG, Fedorka-Cray PJ, Welch TJ, Meinersmann RJ (2010) An oligonucleotide microarray to characterize multidrug resistant plasmids. J Microbiol Methods 81: 96-100.
16. Chizhikov V, Rasooly A, Chumakov K, Levy DD (2001) Microarray analysis of microbial virulence factors. Appl Environ Microbiol 67: 3258-3263.
17. Bryant PA, Venter D, Robins-Browne R, Curtis N (2004) Chips with everything: DNA microarrays in infectious diseases. Lancet Infect Dis 4: 100-111.
18. Peterson G, Bai J, Nagaraja TG, Narayanan S (2010) Diagnostic microarray for human and animal bacterial diseases and their virulence and antimicrobial resistance genes. J Microbiol Methods 80: 223-230.
19. Garaizar J, Rementeria A, Porwollik S (2006) DNA microarray technology: a new tool for the epidemiological typing of bacterial pathogens? FEMS Immunology & Medical Microbiology 47: 178-189.
20. Nayak R, Stewart-King T (2008) Molecular epidemiological analysis and microbial source tracking of *Salmonella enterica* serovars in a preharvest turkey production environment. Foodborne Pathog Dis 5: 115-126.

21. Chen H, Li XL, Blum DL, Ximenes EA, Ljungdahl LG (2003) CelF of *Orpinomyces* PC-2 has an intron and encodes a cellulase (CelF) containing a carbohydrate-binding module. Appl Biochem Biotechnol 105-108: 775-785.
22. Scaria J, Palaniappan RU, Chiu D, Phan JA, Ponnala L, McDonough P, Grohn YT, Porwollik S, McClelland M, Chiou CS, Chu C, Chang YF (2008) Microarray for molecular typing of *Salmonella enterica* serovars. Mol Cell Probes 22: 238-243.
23. Anjum MF, Marooney C, Fookes M, Baker S, Dougan G, Ivens A, Woodward MJ (2005) Identification of core and variable components of the *Salmonella enterica* subspecies I genome by microarray. Infect Immun 73: 7894-7905.
24. Porwollik S, Boyd EF, Choy C, Cheng P, Florea L, Proctor E, McClelland M (2004) Characterization of *Salmonella enterica* subspecies I genovars by use of microarrays. 186: 5883-5898.
25. Chen S, Zhao S, McDermott PF, Schroeder CM, White DG, Meng J (2005) A DNA microarray for identification of virulence and antimicrobial resistance genes in *Salmonella* serovars and *Escherichia coli*. Mol Cell Probes 19: 195-201.
26. Hu Q, Coburn B, Deng W, Li Y, Shi X, Lan Q, Wang B, Coombes BK, Finlay BB (2008) *Salmonella enterica* serovar Senftenberg human clinical isolates lacking SPI-1. J Clin Microbiol 46: 1330-1336.
27. Hansen-Wester I, Hensel M (2001) *Salmonella* pathogenicity islands encoding type III secretion systems. Microbes Infect 3: 549-559.
28. Hensel M (2000) *Salmonella* pathogenicity island 2. Mol Microbiol 36: 1015-1023.
29. Crago AM, Koronakis V (1998) *Salmonella* InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization. Mol Microbiol 30: 47-56.
30. Fang H, Xu J, Ding D, Jackson SA, Patel IR, Frye JG, Zou W, Nayak R, Foley SL, Chen J, Su Z, Ye Y, Turner S, Harris S, Zhou G, Cerniglia CE, Tong W (2010) An FDA bioinformatics tool for microbial genomics research on molecular characterization of bacterial foodborne pathogens using microarrays. BMC Bioinformatics (In press).
31. Shin D, Lee EJ, Huang H, Groisman EA (2006) A positive feedback loop promotes transcription surge that jump-starts *Salmonella* virulence circuit. Science 314: 1607-1609.

Corresponding author
Rajesh Nayak
Division of Microbiology
National Center for Toxicological Research
Food and Drug Administration
Jefferson, AR 72079
Telephone: 870-543-7482; Fax 870-543-7307
Email: Rajesh.Nayak@fda.hhs.gov

Conflict of interests: No conflict of interests is declared.