AEM
Journals.ASM.org

Development of a DNA Microarray for Molecular Identification of All 46 *Salmonella* O Serogroups

Dan Guo,[a,b] Bin Liu,[a,b] Fenxia Liu,[a,b] Boyang Cao,[a,b,d,e] Min Chen,[c] Xiyan Hao,[a,b] Lu Feng,[a,b,d,e] Lei Wang[a,b,d,e]

TEDA School of Biological Sciences and Biotechnology, Nankai University, TEDA, Tianjin, People's Republic of China[a]; The Key Laboratory of Molecular Microbiology and Technology of the Ministry of Education, College of Life Sciences, Nankai University, Tianjin, People's Republic of China[b]; Shanghai Municipal Center For Disease Control and Prevention, Shanghai, People's Republic of China[c]; Tianjin Key Laboratory of Microbial Functional Genomics, TEDA, Tianjin, People's Republic of China[d]; Tianjin Research Center for Functional Genomics and Biochip, TEDA, Tianjin, People's Republic of China[e]

***Salmonella* is a major cause of food-borne disease in many countries. Serotype determination of *Salmonella* is important for disease assessment, infection control, and epidemiological surveillance. In this study, a microarray system that targets the O antigen-specific genes was developed for simultaneously detecting and identifying all 46 *Salmonella* O serogroups. Of these, 40 serogroups can be confidently identified, and the remaining 6, in three pairs (serogroups O67 and B, E1 and E4, and A and D1), need to be further distinguished from each other using PCR methods or conventional serotyping methods. The microarray was shown to be highly specific when evaluated against 293 *Salmonella* strains, 186 *Shigella* strains, representative *Escherichia coli* strains, and 10 strains of other bacterial species. The assay correctly identified 288 (98%) of the *Salmonella* strains. The detection sensitivity was determined to be 50 ng genomic DNA per sample. By testing simulated samples in a tomato background, 2 to 8 CFU per gram inoculated could be detected after enrichment. This newly developed microarray assay is the first molecular protocol that can be used for the comprehensive detection and identification of all 46 *Salmonella* O serogroups. Compared to the traditional serogrouping method, the microarray provides a reliable, high-throughput, and sensitive approach that can be used for rapid identification of multiple *Salmonella* O serogroups simultaneously.**

Salmonella is a facultative anaerobic Gram-negative bacterium belonging to the family *Enterobacteriaceae* (14). The genus *Salmonella* is composed of two species, *Salmonella enterica* and *Salmonella bongori*, seven subgroups, and more than 2,600 defined serotypes (29). *Salmonella* is widely distributed in nature, and it is transmitted via contaminated food, including meat, poultry, eggs, dairy products, and fresh produce, such as tomato and lettuce, thereby gaining entry into almost every aspect of the human food chain (2, 15, 24). It is recognized as the leading cause of food-borne outbreaks and infections in many countries, causing human illnesses such as typhoid fever, paratyphoid fever, and other salmonelloses (10, 25, 35). From May to November 2010, a multistate outbreak of human infection across the United States was attributed to the consumption of eggs contaminated with *Salmonella enterica* serovar Enteritidis, which resulted in 1,939 reported illnesses (30). It is estimated that there are 1.3 million cases of salmonellosis, 15,000 hospitalizations, and 400 deaths annually in the United States (9).

Salmonella isolates are serotyped according to the White-Kauffmann-Le Minor scheme on the basis of surface antigen identification of O (somatic) and H (flagellar) antigenic epitopes, permitting the characterization of 46 O serogroups and 119 H antigens and, thereby, more than 2,500 different serotypes. The White-Kauffmann-Le Minor scheme is utilized by public health organizations worldwide and is considered the gold standard for the determination of *Salmonella* serotypes (29, 32). The O antigen is the outermost component of the lipopolysaccharide (LPS), and it is extremely polymorphic due to variations in the types of sugars present, arrangements, and linkages, and it contributes major antigenic diversity to the cell surface. The genes required for O antigen biosynthesis are organized mainly in a large regulon on the chromosome, which is located between the *galF* and *gnd* genes (39). The O antigen variation among the 46 *Salmonella* O serogroups is mainly due to the extensive genetic diversity present within their respective O antigen gene clusters. Typically, three different types of genes are present within O antigen gene clusters: (i) nucleotide sugar synthesis genes that encode enzymes involved in the synthesis of the sugars that form the O subunit; (ii) sugar transferase genes that assemble the sugar substituents into the O subunit; and (iii) O unit-processing genes that encode proteins involved in the processing and assembly steps to build the O antigen from the O subunit, such as the O antigen flippase (*wzx*) and O antigen polymerase (*wzy*) (7). While the sequences of sugar synthesis genes within the O antigen gene cluster are relatively conserved, the processing genes *wzx* and *wzy* are highly variable and are considered to be specific to each O serogroup (18).

Both the virulence and host range of *Salmonella* isolates are serotype specific (5). Therefore, the rapid and accurate determination of *Salmonella* serotype is essential for human disease surveillance and outbreak control of this important food-borne pathogen (38). The highly variable nature of the O antigen provides the basis for serotyping. Despite its widespread utility, the traditional serotyping method has deficiencies. The method is carried out by assessing agglutination reactions using antisera raised against the O-standard reference strains. It requires well-trained technicians and high-quality antisera, both of which may be diffi-

Received 21 January 2013 **Accepted** 18 March 2013
Published ahead of print 22 March 2013
Address correspondence to Lei Wang, wanglei@nankai.edu.cn.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.00225-13.

doi:10.1128/AEM.00225-13

cult to obtain consistently and are very costly in resource-limited settings (19). Furthermore, it is time-consuming and tedious, as well as subjective in interpretation due to cross-reactivity. In recent years, molecular typing methods that target the O antigen-specific genes have been developed, and these have become an attractive alternative to traditional approaches for O serogroup determination (3, 7, 22, 37). Fitzgerald et al. developed an O group-specific Bio-Plex array to detect six serogroups of *Salmonella* in the United States (6). However, these methods only target a very small subset of the *Salmonella* serogroups, which are inadequate for comprehensive and reliable detection of *Salmonella*. DNA microarrays allow a large number of specific DNA sequences to be detected simultaneously, and it has been shown to be useful as a rapid and sensitive pathogen detection strategy that overcomes the disadvantages of conventional PCR and multiplex PCR methods.

In this study, a DNA microarray assay was developed to identify and detect all 46 *Salmonella* O serogroups. PCR primers and probes were designed based on the specific DNA markers of the O antigen gene clusters, which were retrieved either from the GenBank database or from our in-house databases. The specificity and sensitivity assessments using pure cultures and simulated food samples demonstrated that this new DNA microarray assay is specific and reliable, and it could serve as an effective alternative to traditional serotyping approaches.

MATERIALS AND METHODS

Bacterial strains examined and genomic DNA extraction. The bacterial strains used in this study, including 290 *Salmonella* strains, 186 representative strains of all *Shigella* and *Escherichia coli* serotypes, and 10 strains from other bacterial species, are listed in Table 1. All strains were inoculated into Luria-Bertani (LB) medium and incubated overnight at 37°C. Genomic DNA was prepared using a phenol-chloroform method as described previously (1).

Oligonucleotide primers and probes. The oligonucleotide primers and probes used in this study are listed in Table S1 in the supplemental material, and these were designed based on the GenBank sequences and our in-house database information of O antigen gene clusters for all 46 *Salmonella* O serogroups (B. Liu, Y. A. Knirel, L. Feng, A. V. Perepelov, S. N. Senchenkova, P. R. Reeves, and L. Wang, unpublished data). Two primers and three to four probes were designed against the targeted genes using Primer Premier 5.0 (Premier Boost International, Palo Alto, CA) and OligoArray 2.0 software, respectively (31). The primer pair used to amplify 16S rRNA genes has been reported previously (17). A probe based on the region of 16S rRNA that is conserved for all bacteria was used as the positive control. To enhance hybridization yield, each probe was 5′-amino modified and contained 15 poly(T) oligonucleotides at the 5′ end. A probe containing 40 poly(T) oligonucleotides was used as the negative control, and a 3′-Cy3-labeled probe was used as the positional reference and printing control.

Multiplex PCR. Multiplex PCR, used to amplify the target genes, was carried out in five groups (see Table S1 in the supplemental material). Each multiplex PCR was performed in a 30-μl reaction mixture containing 100 ng of DNA; 1× PCR buffer (50 mM KCl, 10 mM Tris-HCl [pH 8.3]); 2.5 mM $MgCl_2$; 0.16 mM (each) dATP, dCTP, dGTP, and dTTP; 0.05 to 0.2 μM respective primers based on each of targeted genes; 0.06 μM (each) two primers based on the 16S rRNA gene; and 2.5 U *Taq* DNA polymerase (TaKaRa Biotechnology, Dalian, China). The PCR was performed with an initial denaturation step of 95°C for 5 min, followed by 30 cycles at 95°C of 30 s, 55°C for 45 s, and 72°C for 1 min, with a final elongation step of 72°C for a 5-min elongation step. PCR products were purified using the Microcon centrifugal filter devices kit (Millipore Corporation, MA). Two microliters of each PCR product was run on an agarose gel to confirm the presence of appropriate amplicons (Fig. 1).

Target DNA labeling. PCR products were labeled using the PCR conditions described above, except that only reverse primers were used, and 0.3 μl of 25 nM Cy3-dUTP (Amersham Biosciences UK Ltd., Little Chalfont, England) was included. Five microliters of purified amplification product generated from the multiplex PCR described above was added as the template. Thermal cycling conditions were the same as those for the multiplex PCR described above. All labeled DNA was purified using the Microcon centrifugal filter device kit and then stored at −20°C in the dark.

Oligonucleotide probe sets and microarray construction. Probes were dissolved in 50% dimethyl sulfoxide (DMSO) to a final concentration of 1 μg/μl and then spotted onto aldehyde group-modified glass slides (CapitalBio Corporation, Beijing, China) using a SpotArray 72 instrument (Perkin-Elmer Corporation, CA). Each probe was spotted in duplicate to eliminate irregular data due to physical defects on the glass slides. Printed slides were dried for 24 h at room temperature in the dark and scanned at 532 nm to assess and confirm spotting quality. A schematic diagram of the respective probe positions is shown in Fig. 2A.

DNA microarray hybridization. Eight microliters of the labeled target DNA was mixed with an equal volume of preheated (50°C) hybridization buffer (25% formamide, 0.1% sodium dodecyl sulfate, 6× SSPE [1× SSPE consists of 0.18 M NaCl, 10 mM NaH_2PO_4, and 1 mM EDTA, pH 7.7]). The mixture was applied to a hybridization chamber and incubated at 50°C for 1.5 h in a water bath. After hybridization, the slide was washed with solution A (1× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate] and 0.1% SDS) for 3 min, followed by a wash with 0.05× SSC for 3 min and a final wash with 95% ethanol for 1.5 min. The slide was dried under a gentle airstream prior to scanning.

Simulated food samples. Fresh tomatoes were purchased from a local market, fully cleaned, and homogenized by a sterile blender. Bacterial cultures were serially diluted and quantified by the plate-counting method. Aliquots (100 μl) of each dilution were used to seed 25 g of tomato homogenate in 225 ml buffered peptone water (BPW). A total of 10 *Salmonella* serogroup strains were selected randomly and mixed with tomato samples, including serogroups E1, C1, B, D3, O61, O65′O52, O53, F, and M. The spiked tomato samples were incubated at 37°C overnight with shaking. After the enrichment, DNA was extracted from each sample (1 ml) and tested on the DNA microarray.

Fluorescence scanning and data analysis. The hybridized microarray was scanned using a GenePix personal 4100A microarray scanner (Axon Instruments, CA) with a laser beam at 532 nm, the photomultiplier tube (PMT) gain set to 650, and a pixel size of 5 μm. Two files were generated with GenePix Pro 6.0, one with the images saved as a TIF file and the other, for signal intensities, was saved as a GPR file. The signal-to-noise ratio was calculated for each spot using Bactarray Analyzer 1.0, developed in our laboratory, with the signal-to-noise ratio threshold set at 3.0. A serotype was confirmed and reported when the positive-control probes, the negative-control probes, and the printing control probes generated were correct and all of the probes of the given serogroup generated positive signals above the threshold.

RESULTS

Identification of target genes. The O antigen gene clusters of all 46 *Salmonella* serogroups were retrieved from GenBank and our in-house databases. Multiple-sequence alignments of O antigen genes were carried out to select specific genes for different O serogroups. The highly divergent O-unit-processing genes *wzx* and *wzy* were selected as target genes for all of the *Salmonella* O serogroups, with the exception of serogroup O54 and serogroups A and D1, for which the glycosyltransferase gene *wbbE* and the O antigen synthesis gene *prt* (CDP-paratose synthase) were selected, respectively.

Primer screening. Specific primers were designed based on the

TABLE 1 *Salmonella* strains and other bacterial species used in this study

Genus/species and serogroup	Serotype	No. of strains from each source	Total no.
Salmonella ($n = 293$)			
A(O2)	Paratyphi A, Nitra, Kiel, Koessen	3,[a] 2[b]	5
B(O4)	Hessarek, Drogana, Essen, Kingston, II, Derby, Hato, Agona, Chester, Stanley, Saintpaul, Huettwilen, Banana, Coeln, Albert, Ball, Haduna, Clackamas, Brezany, Kiambu, Indiana	35,[a] 2[b]	37
C1(O6,7)	Choleraesuis, IIIb, Sanjuan, Montevideo, II, Livingstone, Oranienburg, Birkenhead, VI	12[a]	12
C2(O8)	Muenchen, Glostrup, Newport, Tallahassee, Bovismorbificans, Litchfield, Kottbus, Takoradi	8,[a] 2[b]	10
D1(O9)	Dublin, II	2,[a] 2[b]	4
D2(O9,46)	Strasbourg, Marylebone, Bergedorf	5[a]	5
D3(O9,46,27)	II	5[a]	5
E1(O3,10)	Anatum, Westhampton, II	4[a]	4
E4(O1,3,19)	Senftenberg	1[a]	1
F(O11)	II, VI, IIIb, IIIa, Marseille, Gallen, Luciana, Aberdeen	11,[a] 5[b]	16
G(O13)	Durham, Raus, II, IIIa, IIIb, Havana, Poona	10[a]	10
H(O6,14)	Carrau, II, IIIb, Soahanina	5[a]	5
I(O16)	Hannover, Brazil, Hvittingfoss, II, IV	5[a]	5
J(O17)	Jangwani, IIIb, IV, IIIa, Bonames	8[a]	8
K(O18)	Cerro, IIIa, IIIb, II, Blukwa, IV	5,[a] 3[b]	8
L(O21)	Assen, Ghana, II, IIIa, IIIb	6[a]	6
M(O28)	Dakar, II, IIIb, Brisbane, Solna	6[a]	6
N(O30)	Urbana, Sternschanze, Overvecht, II	4[a]	4
O(O35)	Adelaide, Umhlatazana, Monschaui, II	6[a]	6
P(O38)	Lansing, IV, IIIa, IIIb	5[a]	5
Q(O39)	Champaign, II, Wandsworth	3[a]	3
R(O40)	Karamoja, Shikmonah, II, IIIa, IIIb, IV, V	8[a]	8
S(O41)	Burundi, Waycross, II, IV, VI	11[a]	11
T(O42)	Faji, Ursenbach, II, IIIa	7[a]	7
U(O43)	Berkeley, II, IIIb, IV	5[a]	5
V(O44)	Niakhar, Niarembe, Tiergarten, II, IIIa, V	8[a]	8
W(O45)	Dugbe, Meekatharra, Karachi, IIIa, VI, II	8[a]	8
X(O47)	II, Bootle, Sya	4[a]	4
Y(O48)	II, IIIa, IIIb, IV, V, VI, Buckeye, Dahlem	13[a]	13
Z(O50)	II, IIIa, IIIb, IV, VI	13[a]	13
O51	Dan, II, IIIa, IIIb, Karaya	7[a]	7
O52	Utrecht, Ord, II	4[a]	4
O53	II, IIIa, Leda	6[a]	6
O54	Uccle	1[a]	1
O55	II	2[a]	2
O56	II	1[a]	1
O57	II, IIIb, Antonio, Maryland	6[a]	6
O58	II	1[a]	1
O59	II, IIIa, IIIb	4[a]	4
O60	II, IIIb	4[a]	4
O61	II, IIIb	3[a]	3
O62	IIIa	2[a]	2
O63	IIIa	2[a]	2
O65	IIIb	1[a]	1
O66	V	3[a]	3
O67	Crossness	1[a]	1
Other ($n = 196$)			
Shigella and *E. coli*		165,[a] 20,[c] 1[f]	186
Pseudomonas aeruginosa		1[d]	1
Staphylococcus aureus		1[e]	1
Klebsiella pneumoniae		1[e]	1
Yersinia enterocolitica		1[a]	1
Enterococcus faecalis		1[d]	1
Enterococcus faecium		1[f]	1
Enterococcus cloacae		1[d]	1
Vibrio parahaemolyticus		1[b]	1
Vibrio cholerae		1[a]	1
Citrobacter freundii		1[g]	1

[a] Institute of Medical and Veterinary Science, Adelaide, Australia.
[b] Center for Disease Control and Prevention, Shanghai, China.
[c] Federal Institute for Risk Assessment, Berlin, Germany.
[d] Institute of Microbiology, Chinese Academy of Sciences, China.
[e] National Center for Medical Culture Collection, China.
[f] Chinese Center for Disease Control and Prevention.
[g] Tianjin Entry-Exit Inspection and Quarantine Bureau.

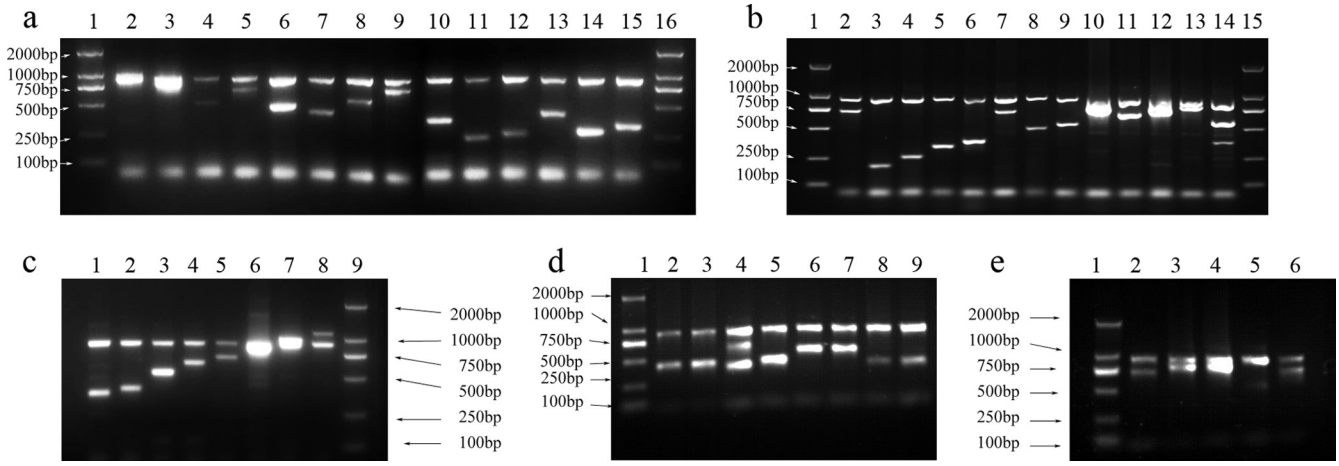

FIG 1 Agarose gel electrophoresis of multiplex PCR products for all 46 *Salmonella* serogroups. (a) Group A. Lanes 1 and 16, DL2000 DNA markers; lanes 2 to 15, DNA templates from *Salmonella* serogroups F, G, H, I, J, K, L, M, S, R, Q, P, O, and N, respectively. (b) Group B. Lanes 1 and 15, DL2000 DNA markers; lanes 2 to 14, DNA templates from *Salmonella* serogroups O57, O55, O54, O53, O52, O51, Z, Y, X, W, V, U, and T, respectively. (c) Group C. Lane 9, DL2000 DNA marker; lanes 1 to 8, DNA templates from *Salmonella* serogroups O66, O65, O63, O62, O61, O60, O59, and O58, respectively. (d) Group D. Lane 1, DL2000 DNA marker; lanes 2 to 9, DNA templates from *Salmonella* serogroups A, D1, D3, D2, B, O67, E1, and E4. (e) Group E. Lane 1, DL2000 DNA marker; lanes 2 to 6, the DNA templates from *Salmonella* serogroups C1, E1, E4, O56, and C2.

selected target genes for each serogroup. These primers were screened against the representative strains from each of the 46 *Salmonella* O serogroups with the capacity to generate single PCR products of the anticipated sizes of the corresponding strains, while negative results were generated consistently from the strains belonging to other serogroups or other species. The identifications based on PCR products were validated with all the serogroups except for three pairs of serogroups, A/D1, B/O67, and E1/E4. Each of these three pairs generated the same PCR product from the same primers due to the high levels of similarities between their O antigen gene clusters, and a further serological test was needed for their differentiation. Nonspecific PCR products were detected from the serogroups D2 and D3 with the primer pair of serogroups A/D1 and also from the serogroups E1/E4 with the primer pair of serogroup D2 (Table 2). The serogroup O54 strain used in our study was subtyped as serotype Uccle (O3, 54) and produced positive results with both O54 and E1/E4 primers. Since the serogroups D2, D3, E1/E4, and O54 have their own serogroup-specific probes, these serogroups could still be correctly identified. Therefore, 43 specific primer pairs were selected from the 81 primer pairs initially designed for the multiplex PCR amplification and labeling (see Table S1 in the supplemental material).

Multiplex PCR to amplify the target genes. Multiplex PCR was used to amplify and label the target genes in five groups. Group A targeted serogroups F, G, H, I, J, K, L, M, N, O, P, Q, R, and S; group B targeted serogroups T42, U43, V44, W45, X47, Y48, Z50, O51, O52, O53, O54, 055, and O57; group C targeted serogroups O58, O59, O60, O61, O62, O63, O65, and O66; group D targeted serogroups B/O67, A/D1, D2, and D3; and group E targeted serogroups C1, C2, O56, and E1/E4. We initially used the same concentration of 0.2 μM for all primers; however, several serogroups failed to generate expected hybridization signals under these conditions, which was probably due to the competition or interference between different primer pairs in the multiplex PCR. Therefore, the primer concentrations were adjusted in the range of 0.03 to 0.58 μM, and these were evaluated as the optimized amplification conditions (see Table S1 in the supplemental material). For each target strain, at least two bands were generated by the multiplex PCR, one for the 16S rRNA primer pair and the other for the O serogroup-specific primer pair. PCR products ranging in size from 205 to 1,132 bp were amplified (Fig. 1).

Specificity of the probes. The DNA microarray was tested against 148 *Salmonella* strains representing all 46 O serogroups, 186 *Shigella* and *E. coli* strains, and 10 strains from other bacterial species (Table 1). From the 183 oligonucleotide probes initially designed, 133 probes were selected for use, including 130 specific probes, one positive-control probe, one negative-control probe, and one printing and reference control probe (see Table S1 in the supplemental material). All representative *Salmonella* strains tested against the probes on the microarray hybridized consistently to their corresponding probes and could be successfully differentiated into the correct serogroups. The isolates belonging to *Shigella* and *E. coli* and the other bacterial species reacted only to the 16S rRNA gene probe, and none of them bound to the serogroup-specific probes on the microarray. The microarray hybridization patterns observed for 20 representative *Salmonella* O serogroups are shown in Fig. 2B.

Sensitivity of detection with DNA. To determine the sensitivity of the assay, serial dilutions of genomic DNA from strains representing each of the 46 serogroups were prepared in volumes of 1, 10, and 50 to 100 ng and used as the templates for multiplex PCR. The positive signals were obtained with DNA levels of 10 ng or above, while the negative results or very weak fluorescence signals were detected at lower DNA amounts. However, 50 ng DNA provided consistent identification results; thus, this amount was chosen as the base amount for the microarray. At this concentration, all 46 *Salmonella* serogroups could be detected accurately. Therefore, the sensitivity of the assay was determined as 50 ng DNA.

Blind test. The specificity and sensitivity of the designed microarray system was further tested using a double-blind approach. A total of 142 strains were randomly selected (Table 1) and used to hybridize to the microarrays without any prior knowledge of their identity, and agglutination tests with specific antisera were per-

A

1	OA-Cy3	OA-2801	OA-2802	OA-2803	OA-532	OA-2795	OA-2796	OA-2797	WL-4006
2	OA-2779	OA-2780	OA-2781	OA-2782	OA-2783	OA-2784	OA-2785	OA-2786	OA-2787
3	OA-2788	OA-2789	OA-2790	OA-3114	OA-3115	OA-3116	OA-2798	OA-2799	OA-2800
4	OA-2809	OA-3200	OA-3201	OA-2812	OA-2813	OA-2814	OA-2872	OA-2873	OA-2874
5	OA-3118	OA-3119	OA-3208	OA-3122	OA-532	OA-2804	OA-2805	OA-2806	50%DMSO
6	OA-2831	OA-2832	OA-2833	OA-2838	OA-2839	OA-2840	OA-2849	OA-2850	OA-2851
7	OA-2856	OA-2857	OA-2858	OA-2859	OA-2860	OA-2861	OA-2866	OA-2867	OA-3099
8	OA-2822	OA-2823	OA-2824	OA-2825	OA-2826	OA-2827	OA-2829	OA-2830	OA-3097
9	OA-2818	OA-2819	OA-2820	50%DMSO	OA-532	OA-2815	OA-2816	OA-2817	50%DMSO
10	OA-2845	OA-2846	OA-2848	OA-2852	OA-2854	OA-2855	OA-2875	OA-2876	OA-2877
11	OA-2882	OA-2883	OA-2884	OA-3204	OA-3011	OA-3012	OA-3021	OA-3022	OA-3023
12	OA-3025	OA-3026	OA-3027	OA-3014	OA-3015	OA-3016	OA-3018	OA-3019	OA-3020
13	OA-2835	OA-2836	OA-2837	50%DMSO	OA-532	OA-2842	OA-2843	OA-2844	50%DMSO
14	OA-2862	OA-2863	OA-2865	OA-2878	OA-2880	OA-2881	OA-3029	OA-3030	OA-3031
15	OA-2886	OA-2887	OA-2888	OA-3102	OA-3103	OA-3104	OA-3107	OA-3108	OA-3109
16	OA-3205	OA-3206	OA-3112	50%DMSO	50%DMSO	OA-3197	OA-3198	OA-3199	OA-Cy3

B

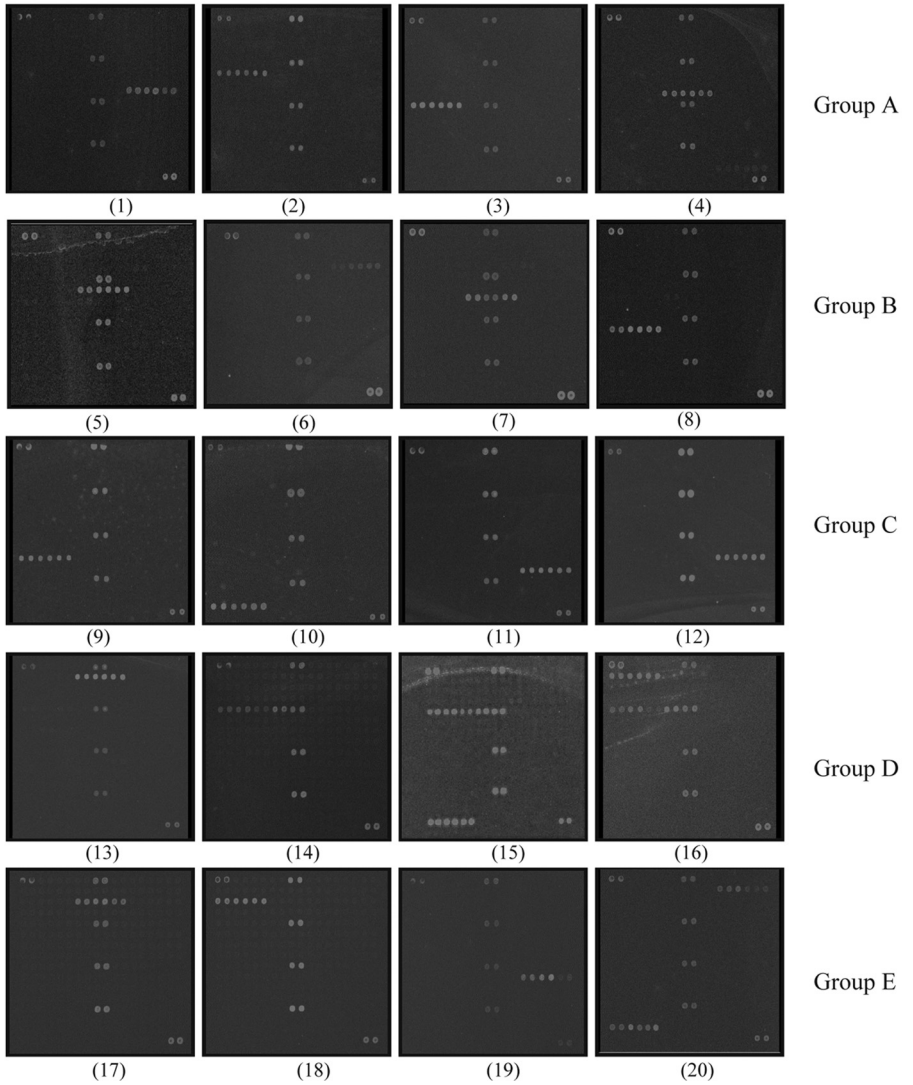

FIG 2 (A) Probe positions on the microarray slide. OA-532 is the positive-control probe based on the bacterial 16S rRNA gene. WL-4006 is the negative-control probe. OA-Cy3 is the positional reference and printing control probe. The rest of the probes are specific probes for the target strains. (B) Microarray hybridization patterns for representative *Salmonella* serogroup strains. Pattern 1, *Salmonella* serogroup P; 2, *Salmonella* serogroup S; 3, *Salmonella* serogroup M; 4, *Salmonella* serogroup O; 5, *Salmonella* serogroup U; 6, *Salmonella* serogroup O55; 7, *Salmonella* serogroup O51; 8, *Salmonella* serogroup W; 9, *Salmonella* serogroup O58; 10, *Salmonella* serogroup O59; 11, *Salmonella* serogroup 62; 12, *Salmonella* serogroup O63; 13, *Salmonella* serogroup B; 14, *Salmonella* serogroup A; 15, *Salmonella* serogroup D2; 16, *Salmonella* serogroup D3; 17, *Salmonella* serogroup C1; 18, *Salmonella* serogroup C2; 19, *Salmonella* serogroup O56; and 20, *Salmonella* serogroup E1.

TABLE 2 Molecular characterization of six closely related serogroups

Serogroup	Positive probes by microarray		Serogroup by microarray determination
	Serogroup specific	Nonspecific	
A	OA-3118, OA-3119, OA-3208, OA-3122		A or D1
D1	OA-3118, OA-3119, OA-3208, OA-3122		A or D1
D2	OA-3205, OA-3206, OA-3112	OA-3118, OA-3119, OA-3208, OA-3122	D2
D3	OA-2779, OA-2780, OA-2781	OA-3118, OA-3119, OA-3208, OA-3122	D3
E1	OA-2785, OA-2786, OA-2787	OA-3205, OA-3206, OA-3112	E1 or E4
E4	OA-2785, OA-2786, OA-2787	OA-3205, OA-3206, OA-3112	E1 or E4

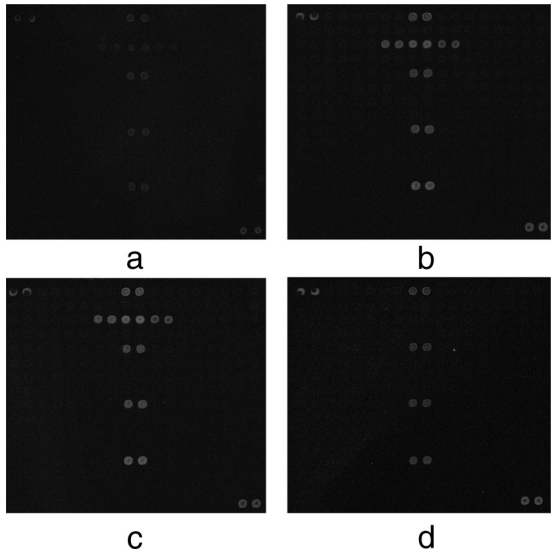

FIG 3 Microarray hybridization patterns representative of a *Salmonella* serogroup from DNA extracted from spiked tomato samples. (a) *Salmonella* serogroup B, 2 CFU/g; (b) *Salmonella* serogroup B, 11 CFU/g; (c) *Salmonella* serogroup B, 10^2 CFU/g; (d) *Salmonella* serogroup B, unspiked tomato sample.

formed separately. All of the detection results from the microarrays were consistent with those obtained by the conventional serotyping methods with five exceptions, three for serogroup O28 and two for serogroup O59 (Table 3). To evaluate the simultaneous detection efficacy of multiple strains, mixed cultures containing two *Salmonella* O serogroups strains (i.e., serogroups G and O50, Q and O57, O61 and O66, D1 and D3, or C1 and C2) were prepared and tested with the microarray. All of the samples gave the correct results. The consistency of the double-blinded test with the conventional serotyping method was 96.48%.

Test of the simulated food samples. Artificially contaminated tomato samples were evaluated using the microarray. A total of 10 *Salmonella* O serogroup strains, including serogroups E1, C1, B, D3, O61, O65′O52, O53, F, and M, were cultured and mixed with tomato samples. After overnight enrichment, DNA was extracted from the samples and tested using the microarray assay. As few as 2 to 8 CFU (initial inoculum) of *Salmonella* was detected in 1 g tomato sample, and all serotypes inoculated were successfully identified. The unspiked tomato sample failed to give a positive signal. Representative hybridization patterns using DNA extracted from tomato samples are shown in Fig. 3.

DISCUSSION

In this present study, we describe a microarray method for the determination of *Salmonella* O serogroups based on the O serogroup-specific genes. This is the first report to describe a molecular method for the comprehensive detection and identification of all 46 *Salmonella* O serogroups. Compared to other molecular-based detection methods, such as PCR and DNA sequencing, microarrays are more applicable for high throughput, are highly specific in combination with PCR, and are more efficient than DNA sequencing. PCR is rapid and easy to process, but it is hard to differentiate similar amplicon sizes from different serotypes. The DNA sequencing is very accurate; however, it requires more expensive equipment, has higher testing costs, and takes a lot of time and work for library preparation and sequencing (34).

In this study, the O unit-processing genes *wzx* and *wzy* were selected to be the target genes for the majority of *Salmonella* serogroups. In the cases where other genes were selected, *Salmonella* O54, for which the transferase gene *wbbE* was selected as a specific gene (13), has its O antigen gene cluster on a small mobilizable plasmid which lacks *wzx* and *wzy* genes, and it usually has a chromosomally encoded O antigen which is coexpressed with the O54 factor (12). The O54 strain used in our study also produced positive results with E1/E4 primers, as it is subtyped as serotype Uccle (O3, 54) and expresses additional O3 (E) factor. The O antigen gene clusters of serogroups A and D1 shared high levels of homology to serogroups B, D2, D3, E1, and E4. The *wzx* genes had 94 to 99% identities among serogroups A, D1, D2, and D3, while the *wzy* genes of serogroups A, D1, and B are identical (4, 26). Therefore, the O antigen synthesis gene *prt* was used for molecular differentiation of serogroups A and D1 from other serogroups (22). The serogroups D2 and D3 also produced positive results with the primer pair and probes for serogroup A/D1, as they have identical *prt* genes (4) and were distinguished by their own serogroup-specific primer pair and probes (Table 2). The *wzy* genes of serogroups E1/E4 shared 98% identities with serogroup D2; therefore, E1/E4 also produced positive results with the primer pair and probes for D2 (40). The serogroups E1/E4 were distinguished from D2 by their own serogroup-specific primer pair and probes (Table 2).

TABLE 3 Properties of *Salmonella* serogroups O28 and O59

Serogroup	Serotype	Antigenic formula	Microarray determination
O28	Dakar	28:a:1,6	Positive
O28	Dakar	28:a:1,6	Positive
O28	IIIb	28:z10:z	Positive
O28	Brisbane	28:z;e,n,z_{15}	Negative
O28	Solna	28:a:1,5	Negative
O28	II	28:a:e,n,x	Negative
O59	II	59:k:(z)	Positive
O59	II	59:k:z_{65}	Positive
O59	IIIa	59:z_4,z_{23}:-	Negative
O59	IIIb	59:c:e,n,x,z_{15}	Negative

The microarray assay provided serogroup designations comparable to those of traditional methods in the vast majority of cases. However, this method alone was unable to differentiate three closely related serogroup pairs: serogroups E1 and E4, serogroups A and D1, and serogroups O67 and B. The O antigen gene cluster of serogroup E1 shares great similarity to that of serogroup E4 (39). The O antigen gene cluster of serogroup A is identical to that of serogroup D1 except for a frameshift mutation in the *tyv* gene (21), and there is not a serogroup-specific probe that can distinguish them (8, 22). Molecular markers for the differentiation of those three pairs are currently unavailable, and further serological tests are needed for the differentiation of these serogroups. Serogroup O67 is very rare and appears to be a variant of serogroup B (6, 18), and the region between *galF* and *gnd* in *Salmonella* serogroup O67 is the same as that in the serogroup B O antigen gene cluster. However, there is no cross-reaction for group B and O67 antisera, and their structural analysis revealed that the O67 gene cluster is not located between *galF* and *gnd*. The draft genome of the *Salmonella* O67 type strain was obtained by our laboratory recently, and a potential O67 O antigen gene cluster was found elsewhere in the genome, which is consistent with its O antigen chemical structure (Liu et al., unpublished). Based on this gene cluster, we designed an additional PCR primer pair targeting the *wejV* gene (see Table S2 in the supplemental material). PCR assays showed that this primer pair was specific to serogroup O67, as none of the serogroup B strains tested gave positive results. In future work, probes based on the O67-specific gene can be incorporated into the microarrays to distinguish between the serogroups O67 and B.

The unexpected negative results were obtained with 5 of the 290 *Salmonella* isolates tested, including three serogroup O28 strains (six in total) and two serogroup O59 strains (four in total) (Table 3). These might be due to variations of *Salmonella* O antigen clusters within a serogroup, as such cases were reported previously (6). The two serogroup O59 strains that failed to be identified by the microarray belong to subspecies IIIa and IIIb serogroup O59. It has been reported that O59 in subspecies IIIb had different O-polysaccharide structures than O59 in subspecies II, although both of their O antigens include a common disaccharide, which could provide a common epitope responsible for their serological relatedness (28). Since O antigen gene cluster sequence information for those two subspecies are not available, the specific primers and probes designed based on the O antigen gene cluster of subspecies II serogroup O59 might not be suitable for subspecies IIIa and IIIb. We are currently in the process of sequencing the O antigen clusters of serogroup O59 subspecies IIIa and IIIb to reveal any sequence variations that can be used for the identification of these two subspecies. *Salmonella* serogroup O28 was originally divided into three subfactors ($O28_1$, $O28_2$, and $O28_3$) without structural differences being ascribed (16). The *wzx* and *wzy* genes from the O antigen gene clusters of serotype Dakar ($O28_1$ and $O28_3$) and serotype Pomona ($O28_1$ and $O28_2$) share 51 and 48% identity, respectively (11). The specific primers and probes for O28 were targeted on the *wzy* gene of *S. enterica* serotype Dakar in our microarray. An additional PCR primer pair was designed against the *wzy* gene of *S. enterica* serotype Pomona (see Table S2 in the supplemental material). The three O28 strains initially giving negative results were positive when tested using this additional primer pair (data not shown). Thus, the extra primer pair could be incorporated into the microarray in further studies to increase the power of the assay.

Salmonella and *E. coli* are known to be evolutionarily related (27). There are several pairs of *Salmonella* and *E. coli* strains that are known to have identical or closely related O antigen organizations, such as serotype O35 and *E. coli* O111, serotype O50 and *E. coli* O55, and serotype O30 and *E. coli* O157 (33, 36). Nonetheless, their gene clusters at the nucleotide level are divergent, and the average identities of *wzx* and *wzy* are 70 and 63%, respectively, which is enough difference to discriminate them from each other. Our microarray was tested against the 186 *Shigella* and *E. coli* representative strains to validate the specificity of the new assay, and each of these strains produced a negative signal against the *Salmonella*-specific probes.

In conclusion, this study presents a new multiplex PCR-based microarray assay for the comprehensive detection and identification of all 46 *Salmonella* O serogroups. This new method provides an accurate and reliable approach for differentiating *Salmonella* at the serogroup level and contributes significantly to large-scale epidemiology studies, and it could be employed to monitor local, regional, and national trends in human salmonellosis. Although the microarrays developed here allow the analysis of all *Salmonella* O serogroups simultaneously, serotyping of *Salmonella* based on flagellar antigens (H antigens) still needs to be done using the conventional method. Currently, the sequences of flagellin alleles representing 67 out of the 114 known *Salmonella* flagellar antigenic types have been reported (23), with more to come. In future work, it will be worthwhile to incorporate flagellar antigen-specific probes into the microarrays for more comprehensive analysis and practical uses.

ACKNOWLEDGMENTS

This work was supported by the National Key Program for Infectious Diseases of China (2013ZX10004216-001-001), the National 863 Program of China (2012AA020103 and 2011AA100901-2), the National 973 Program of China (2011CB504900), National Natural Science Foundation of China (NSFC) Program (31170094, 31030002, and 31270003), the Tianjin Research Program of Application Foundation and Advanced Technology (10JCYBJC10000), the Research Fund for the Doctoral Program of Higher Education of China (20090031120023), the Fundamental Research Fund for the Central Universities (65020121 and 65020061), and the Public Health Key Disciplines in Shanghai Health Microbiology (12GWZX0801).

REFERENCES

1. **Bastin DA, Reeves PR.** 1995. Sequence and analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111. Gene **164**:17–23.
2. **Behravesh CB, Blaney D, Medus C, Bidol SA, Phan Q, Soliva S, Daly ER, Smith K, Miller B, Taylor T, Nguyen T, Perry C, Hill TA, Fogg N, Kleiza A, Moorhead D, Al-Khaldi S, Braden C, Lynch MF.** 2012. Multistate outbreak of *Salmonella* serotype Typhimurium infections associated with consumption of restaurant tomatoes, U. S. A., 2006: hypothesis generation through case exposures in multiple restaurant clusters. Epidemiol. Infect. **140**:2053–2061.
3. **Brown PK, Romana LK, Reeves PR.** 1992. Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar Muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B. Mol. Microbiol. **6**:1385–1394.
4. **Curd H, Liu D, Reeves PR.** 1998. Relationships among the O-antigen gene clusters of *Salmonella enterica* groups B, D1, D2, and D3. J. Bacteriol. **180**:1002–1007.
5. **Fierer J, Guiney DG.** 2001. Diverse virulence traits underlying different clinical outcomes of *Salmonella* infection. J. Clin. Investig. **107**:775–780.
6. **Fitzgerald C, Collins M, van Duyne S, Mikoleit M, Brown T, Fields P.**

2007. Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups. J. Clin. Microbiol. **45**:3323–3334.
7. **Fitzgerald C, Sherwood R, Gheesling LL, Brenner FW, Fields PI.** 2003. Molecular analysis of the *rfb* O antigen gene cluster of *Salmonella enterica* serogroup O:6,14 and development of a serogroup-specific PCR assay. Appl. Environ. Microbiol. **69**:6099–6105.
8. **Franklin K, Lingohr EJ, Yoshida C, Anjum M, Bodrossy L, Clark CG, Kropinski AM, Karmali MA.** 2011. Rapid genoserotyping tool for classification of *Salmonella* serovars. J. Clin. Microbiol. **49**:2954–2965.
9. **Hardnett FP, Hoekstra RM, Kennedy M, Charles L, Angulo FJ.** 2004. Epidemiologic issues in study design and data analysis related to FoodNet activities. Clin. Infect. Dis. **38**(Suppl. 3):S121–S126.
10. **Heissenhuber A, Hautmann W, Arenz S, Kugler R, Kleih W, Ludwig S, Wildner M.** 2005. Accumulated occurrence of illnesses with *Salmonella* Enteritidis in hospitals and nursing homes in the district Oberallgaeu, Bavaria, in July 2004. Gesundheitswesen **67**:845–852.
11. **Hu B, Perepelov AV, Liu B, Shevelev SD, Guo D, Senchenkova SN, Shashkov AS, Feng L, Knirel YA, Wang L.** 2010. Structural and genetic evidence for the close relationship between *Escherichia coli* O71 and *Salmonella enterica* O28 O-antigens. FEMS Immunol. Med. Microbiol. **59**: 161–169.
12. **Keenleyside WJ, Perry M, Maclean L, Poppe C, Whitfield C.** 1994. A plasmid-encoded $rfb_{O:54}$ gene cluster is required for biosynthesis of the O:54 antigen in *Salmonella enterica* serovar Borreze. Mol. Microbiol. **11**: 437–448.
13. **Keenleyside WJ, Whitfield C.** 1996. A novel pathway for O-polysaccharide biosynthesis in *Salmonella enterica* serovar Borreze. J. Biol. Chem. **271**:28581–28592.
14. **Kim S, Frye JG, Hu JX, Fedorka-Cray PJ, Gautom R, Boyle DS.** 2006. Multiplex PCR-based method for identification of common clinical serotypes of *Salmonella enterica* subsp. *enterica*. J. Clin. Microbiol. **44**:3608–3615.
15. **Kuhn KG, Sorensen G, Torpdahl M, Kjeldsen MK, Jensen T, Gubbels S, Bjerager GO, Wingstrand A, Porsbo LJ, Ethelberg S.** 2012. A long-lasting outbreak of *Salmonella* Typhimurium U323 associated with several pork products, Denmark, 2010. Epidemiol. Infect. **26**:1–9.
16. **Kumirska J, Szafranek J, Czerwicka M, Paszkiewicz M, Dziadziuszko H, Kunikowska D, Stepnowski P.** 2007. The structure of the O-polysaccharide isolated from the lipopolysaccharide of *Salmonella* Dakar (serogroup O:28). Carbohydr. Res. **342**:2138–2143.
17. **Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR.** 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. U. S. A. **82**:6955–6959.
18. **Li Q, Reeves PR.** 2000. Genetic variation of dTDP-L-rhamnose pathway genes in *Salmonella enterica*. Microbiology **146**(Pt 9):2291–2307.
19. **Lim BK, Thong KL.** 2009. Application of PCR-based serogrouping of selected *Salmonella* serotypes in Malaysia. J. Infect. Dev. Ctries. **3**:420–428.
20. **Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN, Reeves PR, Wang L.** 2008. Structural diversity in *Salmonella* O antigens and its genetic basis. FEMS Microbiol. Rev. **32**:627–653. (Erratum, **34**:606, 2010).
21. **Liu D, Verma NK, Romana LK, Reeves PR.** 1991. Relationships among the *rfb* regions of *Salmonella* serovars A, B, and D. J. Bacteriol. **173**:4814–4819.
22. **Luk JM, Kongmuang U, Reeves PR, Lindberg AA.** 1993. Selective amplification of abequose and paratose synthase genes (*rfb*) by polymerase chain reaction for identification of *Salmonella* major serogroups (A, B, C2, and D). J. Clin. Microbiol. **31**:2118–2123.
23. **McQuiston JR, Parrenas R, Ortiz-Rivera M, Gheesling L, Brenner F, Fields PI.** 2004. Sequencing and comparative analysis of flagellin genes *fliC*, *fljB*, and *flpA* from *Salmonella*. J. Clin. Microbiol. **42**:1923–1932.
24. **Mertens E, Kreher H, Rabsch W, Bornhofen B, Alpers K, Burckhardt F.** 2012. Severe infections caused by *Salmonella* Enteritidis PT8/7 linked to a private barbecue. Epidemiol. Infect. **19**:1–7.
25. **Muhlenberg W.** 1993. Epidemiologic public health conclusions from observations of illnesses caused by *Salmonella* Enteritidis. Gesundheitswesen **55**:21–27.
26. **Naide Y, Nikaido H, Maekelae PH, Wilkinson RG, Stocker BA.** 1965. Semirough strains of *Salmonella*. Proc. Natl. Acad. Sci. U. S. A. **53**:147–153.
27. **Ochman H, Wilson AC.** 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26**:74–86.
28. **Perepelov AV, Liu B, Senchenkova SN, Shashkov AS, Guo D, Feng L, Knirel YA, Wang L.** 2011. Structures of the O-polysaccharides of *Salmonella enterica* O59 and *Escherichia coli* O15. Carbohydr. Res. **346**:381–383.
29. **Popoff MY, Bockemuhl J, Gheesling LL.** 2003. Supplement 2001 (no. 45) to the Kauffmann-White scheme. Res. Microbiol. **154**:173–174.
30. **Centers for Disease Control and Prevention.** 2010. Investigation update: multistate outbreak of human *Salmonella* Enteritidis infections associated with shell eggs. CDC, Atlanta, GA.
31. **Rouillard JM, Zuker M, Gulari E.** 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. Nucleic Acids Res. **31**:3057–3062.
32. **WHO Collaborating Center for Reference and Research on Salmonella.** 2007. Antigenic formulae of the *Salmonella* serovars, 9th ed. Institut Pasteur, Paris, France.
33. **Samuel G, Hogbin JP, Wang L, Reeves PR.** 2004. Relationships of the *Escherichia coli* O157, O111, and O55 O-antigen gene clusters with those of *Salmonella enterica* and *Citrobacter freundii*, which express identical O antigens. J. Bacteriol. **186**:6536–6543.
34. **Shallom SJ, Tae H, Sarmento L, Preston D, McIver L, Franck C, Dickerman A, Adams LG, Garner HR.** 2012. Comparison of genome diversity of *Brucella* spp. field isolates using Universal Bio-signature detection array and whole genome sequencing reveals limitations of current diagnostic methods. Gene **509**:142–148.
35. **Somily AM, Adam MH, Gad El Rab MO, Morshed MG, Shakoor Z.** 2011. Detection of *Salmonella typhi* agglutinins in sera of patients with other febrile illnesses and healthy individuals. Ann. Afr. Med. **10**:41–44.
36. **Wang L, Reeves PR.** 2000. The *Escherichia coli* O111 and *Salmonella enterica* O35 gene clusters: gene clusters encoding the same colitose-containing O antigen are highly conserved. J. Bacteriol. **182**:5256–5261.
37. **Wang L, Romana LK, Reeves PR.** 1992. Molecular analysis of a *Salmonella enterica* group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism. Genetics **130**:429–443.
38. **Wollin R.** 2007. A study of invasiveness of different *Salmonella* serovars based on analysis of the Enter-net database. Euro Surveill. **12**:E070927.3.
39. **Xiang SH, Haase AM, Reeves PR.** 1993. Variation of the *rfb* gene clusters in *Salmonella enterica*. J. Bacteriol. **175**:4877–4884.
40. **Xiang SH, Hobbs M, Reeves PR.** 1994. Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for its origin from an insertion sequence-mediated recombination event between group E and D1 strains. J. Bacteriol. **176**:4357–4365.