International Journal of Food Microbiology 309 (2019) 108332

Contents lists available at ScienceDirect

International Journal of Food Microbiology

journal homepage: www.elsevier.com/locate/ijfoodmicro

Developing a novel molecular serotyping system based on capsular polysaccharide synthesis gene clusters of *Vibrio parahaemolyticus*

Yu Pang[a], Xi Guo[a,b,c], Xiao Tian[a], Fenxia Liu[a,b,c], Lu Wang[a], Junli Wu[a,b,c], Si Zhang[a], Shujie Li[a], Bin Liu[a,b,c,*]

[a] TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, TEDA, Tianjin, PR China
[b] Key Laboratory of Molecular Microbiology & Technology, Ministry of Education, Tianjin Economic-Technological Development Area, Tianjin, PR China
[c] Center for Microbial Functional Genomics and Detection Technology, Ministry of Education, Nankai University, 23 Hongda Street, TEDA, Tianjin, PR China

ARTICLE INFO

Keywords:
Vibrio parahaemolyticus
Capsular polysaccharide gene cluster
Molecular K-serotyping
Whole genome sequencing
Epidemiological surveillance

ABSTRACT

Vibrio parahaemolyticus is a major food-borne pathogen. *V. parahaemolyticus* infections are associated with various serotypes; to date, 71 K-serogroups of *V. parahaemolyticus* have been determined based on capsular polysaccharide (CPS) diversity. In this study, the capsular polysaccharide gene clusters (CPSgcs) of 55 K-serogroups were identified by whole-genome sequencing and analysis. These CPSgcs exhibit a high level of genetic diversity. A microsphere-based suspension array (MSA) was established for the detection and identification of 55 *V. parahaemolyticus* K-serogroups based on CPSgc-specific genes. To evaluate our array, a double-blind test with 120 clinical isolates was carried out. In addition, an *in silico* K-serotyping system was established based on *V. parahaemolyticus* CPSgc-specific genes. This system was then used to examine 845 publicly available *V. parahaemolyticus* genomes; the results demonstrated that 813 isolates belong to one of 43 K-serogroups. Taken together, these results demonstrate that the molecular system developed in this study is suitable for rapid serotyping of *V. parahaemolyticus* isolates from environmental and clinical samples. In addition, the system could be applied to epidemiological investigations of this important food-borne pathogen.

1. Introduction

Vibrio parahaemolyticus is a gram-negative, halophilic, rod-shaped bacterium (Baffone et al., 2006), which naturally inhabits estuarine and inshore marine water and is also part of the natural flora of shellfish. *V. parahaemolyticus* was first discovered in Japan in 1950, as the cause of a food poisoning outbreak that affected 272 individuals (Fujino et al., 1953). Subsequently, *V. parahaemolyticus* has been reported in food-borne outbreaks worldwide including China, Japan, Korea, and the United States of America (L.H. Lee et al., 2018; Ward and Bej, 2006, Wong et al., 2000, Yu et al., 2011). Since the discovery of *V. parahaemolyticus*, it has been recognized as a human pathogen that causes several clinical diseases (Morris and Black, 1985; Yeung and Boor, 2004). To date, *V. parahaemolyticus* is regarded as the leading cause of food poisoning around the world (Blake et al., 1980; Jiang et al., 2018; Y. Lee et al., 2018; Pan et al., 1997; Wong et al., 2000; Wu et al., 2018).

Owing to the diversity of surface polysaccharides, serotyping is considered the "gold standard" of methods for identifying pathogenic isolates in the environment and clinical samples, as well as for epidemiological monitoring and tracking (Wang et al., 2010). Moreover, *V. parahaemolyticus* serotypes have been identified as a distinct trait of clinical specimens (Han et al., 2016). *V. parahaemolyticus* is typically classified into 13 O-serogroups and 71 K-serogroups based on its O and K antigens (Iguchi et al., 1995). Unlike *Vibrio cholerae*, for which only two serogroups (serogroups O1 and O139) are associated with epidemic diseases, multiple serotypes of *V. parahaemolyticus* have been reported to be virulent and cause infection (Bhuiyan et al., 2002) including O1:K25, O1:K41, O1:K56, O3:K6, O3:K58, O3:K68, O4:K8, and O10:K60 (Bhuiyan et al., 2002, Chowdhury et al., 2004a; Jones et al., 2012, Laohaprertthisan et al., 2003, Nair et al., 2007, Ueno et al., 2016). However, traditional serotyping assays using commercially available antisera targeting *V. parahaemolyticus* strains are limited because of their sensitivity and specificity (Bogdanovich et al., 2003). Molecular methods based on PCR and microarrays targeting serotype-specific genes have been shown to be more specific and sensitive for the identification of bacterial serotypes compared with traditional methods (Liu et al., 2008).

Abbreviations: CPS, capsular polysaccharide; CPSgc, CPS gene cluster; MSA, microsphere-based suspension array; MFI, median fluorescence intensities; HG, homology group; TMAC, tetramethylammonium chloride; EDC, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide

* Corresponding author at: TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, TEDA, Tianjin, PR China.
E-mail address: liubin1981@nankai.edu.cn (B. Liu).

https://doi.org/10.1016/j.ijfoodmicro.2019.108332
Received 8 February 2019; Received in revised form 26 August 2019; Accepted 31 August 2019
Available online 01 September 2019

In general, capsular polysaccharide (CPS) represents K-antigens comprising hydrophilic high molecular-weight polymers that form a dense coat outside of the bacterial cell. The CPS synthesis genes are always located within a cluster in the genome. The genes within the CPS gene cluster (CPSgc) are normally classified into three main categories: sugar synthesis genes, sugar transferase genes, and processing genes for CPS polymerization and translocation (Liu et al., 2014). The role of the Wzx/Wzy pathway in CPS synthesis and translocation has been extensively studied in *Escherichia coli* and *Klebsiella* spp. (Pan et al., 2015; Whitfield, 2006). The synthesis of group 1 and group 4 CPSs in *E. coli* has been shown to be mediated by a Wzx/Wzy-dependent pathway. Synthesis of the capsular repeat unit is initiated by the transfer of a sugar phosphate from an NDP-sugar to undecaprenyl phosphate (UndP); sugar chain elongation is then sequentially catalyzed by specific sugar transferases, thus forming an Und-PP linked unit (Drummelsmith and Whitfield, 1999). Newly synthesized Und-PP linked units are then flipped across the inner membrane into the periplasmic face (in a process requiring Wzx), where they are polymerized by the polymerase protein, Wzy. Subsequently, the channel (Wza), together with Wzb and Wzc, which regulate the polymerization and transport processes, transport the polymer to the bacterial surface. Wzx belongs to the polysaccharide transport (PST) family; the general feature of this protein is the presence of 12 transmembrane segments (TMS) (Islam and Lam, 2014). Wzx has a preference for its cognate O- or K- unit and Wzx sequence variability is thought to be required to accommodate the wide range of available O- or K- unit structures (Marolda et al., 2010). The Wzy polymerase, which normally possesses 10–14 TMS, also exhibits relatively low sequence conservation between or within strains. However, to date, the molecular mechanism underlying the substrate specificity of Wzx and Wzy is yet to be fully determined, as no tertiary structures have been elucidated for these two proteins. It has been reported that the region between the *gmhD* and *rjg* genes constitutes the genetic determinant of CPS in *V. parahaemolyticus* (Chen et al., 2007; Guvener and McCarter, 2003). Previously, we developed a multiplex PCR system based on an O-serogroup-specific gene for the detection and identification of 13 strains belonging to various *V. parahaemolyticus* O-serogroups (Chen et al., 2012). However, only the genetic features of K6 and K8 CPS have been deciphered (Chen et al., 2007; Li et al., 2017). The development of a molecular method for K-serotyping of *V. parahaemolyticus* has been hindered by the limited information available regarding the CPSgcs of *V. parahaemolyticus*.

The microsphere-based suspension array (MSA) platform (liquid chip platform) developed by the Luminex Corporation offers a molecular diagnostic platform for simultaneous, high-throughput, and multiplex detection of up to 100 targets in nucleic acid studies (Chen et al., 2016). The MSA platform provides an attractive approach to molecular diagnostics and high-throughput detection. This method has been approved by the United States of America's FDA for clinical diagnosis and has been widely applied in many fields (Guo et al., 2018; Liu et al., 2011; Qian et al., 2018; Silbereisen et al., 2015). In this study, we obtained 55 *V. parahaemolyticus* CPSgcs by whole-genome sequencing. Next, we developed a molecular K-serotyping system based on the MSA platform using CPSgc-specific genes, establishing an efficient and genome-wide system for serotyping *V. parahaemolyticus*. In addition, we downloaded 845 *V. parahaemolyticus* genome sequences from GenBank and subjected them to the genome-based serotyping system. The results demonstrated that 813 sequences were sorted into 43 K-serogroups. This study provides alternatives for rapid and simple K-serotyping of *V. parahaemolyticus* using both bioinformatic and experimental techniques. This method could potentially be further developed for clinical identification and epidemiological surveillance of *V. parahaemolyticus*.

2. Materials and methods

2.1. Bacterial strains and genomic DNA extraction

Fifty-five *V. parahaemolyticus* strains belonging to different K-serogroups were used in this study (Table 1). All *V. parahaemolyticus* strains were grown in LB medium, with 3.5% salinity (Matsumoto et al., 2000), and cultured with shaking at 37 °C overnight. The K-serogroup of the 55 *V. parahaemolyticus* strains was determined by agglutination tests with specific antisera according to the manufacturer's instructions (Denka-Seiken Ltd., Tokyo, Japan). The genomic DNA of the 55 strains was extracted using a DNA Extraction Kit (Sangon Biotech, Shanghai, China).

Table 1
V. parahaemolyticus strains used for sequencing.

Lab collection no.	Serogroup	Accession no.	Lab collection No.	Serogroup	Accession no.
G2929[a]	K1	MK455076	G3590[b]	K51	MK473658
G2941[a]	K4	MK455077	G2863[c]	K52	MK473659
G2932[a]	K5	MK455078	G3562[b]	K53	MK473660
G2871[a]	K6	MK455079	G3563[b]	K54	MK482084
G2927[a]	K7	MK455080	G2910[a]	K55	MK482085
G2874[a]	K8	MK455081	G2881[a]	K56	MK482086
G2939[a]	K9	MK455082	G3586[b]	K57	MK482087
G3490[b]	K11	MK455083	G3565[b]	K59	MK482088
G2877[a]	K12	MK455084	G3566[b]	K60	MK482089
G2944[a]	K13	MK455085	G2880[a]	K63	MK482090
G2872[a]	K15	MK455086	G3588[b]	K64	MK482091
G2854[c]	K17	MK482100	G3568[b]	K65	MK482092
G2888[a]	K18	MK482101	G3569[b]	K66	MK482093
G2857[c]	K19	MK463646	G3570[b]	K67	MK482094
G3508[b]	K20	MK463647	G2940[a]	K68	MK482095
G3494[b]	K22	MK463648	G3571[b]	K69	MK482096
G3495[b]	K23	MK463649	G3572[b]	K70	MK482097
G3550[b]	K24	MK482098			
G2943[a]	K25	MK463650			
G3499[b]	K30	MK463651			
G3500[b]	K31	MK473657			
G3501[b]	K32	MK473656			
G2865[a]	K33	MK473654			
G3503[b]	K34	MK473655			
G3551[b]	K36	MK473653			
G3552[b]	K37	MK473652			
G2873[a]	K38	MK482099			
G3554[b]	K39	MK473651			
G3555[b]	K40	MK473650			
G2922[a]	K41	MK473649			
G3556[b]	K42	MK473648			
G3557[b]	K43	MK473647			
G3581[b]	K44	MK473646			
G3582[b]	K45	MK473645			
G3583[b]	K46	MK473644			
G3558[b]	K47	MK473643			
G2884[a]	K48	MK473642			
G2879[a]	K49	MK473641			

[a] Department of Microbiology, Soochow University, Taipei, Taiwan 111, China.
[b] Center for Disease Control and Prevention, Shanghai, China.
[c] Bioresources Collection and Research Center (BCRC), Taiwan, China.

2.2. Sequencing and annotation

Each genome was sequenced using a Solexa Genome Analyzer IIx system (Illumina, Little Chesterford, UK) with a coverage depth approaching 100-fold. The obtained reads were *de novo* assembled using Velvet Optimizer v2.2. BLAST and PSI-BLAST were used for alignment with genes and proteins from the GenBank database (www.ncbi.nlm.nih.gov/genbank) and the PFAM protein motif database (pfam.sanger.ac.uk). The TMHMM v2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/) program was used to predict transmembrane domains within protein sequences.

2.3. Construction of single-nucleotide polymorphisms (SNP) phylogenetic tree

Ortherfinder (Emms and Kelly, 2015) was performed to identify the conserved genomic single core genes shared by genomes. These

conserved genomic regions were extracted, aligned with MAFFT (Katoh and Standley, 2013) and concatenated. Then the SNPs were extracted, SNPs due to recombination were removed by ClonalFrameML (Katoh and Standley, 2013), and SNPs due to mutation were used to construct the phylogenetic tree by PhyML (Guindon et al., 2010) using the maximum-likelihood algorithm.

2.4. Homology group (HG) categorization

The predicted proteins were clustered into HGs using the OrthoMCL v2.0 program (http://orthomcl.org/common/downloads/software/v2.0/). The amino acid identity in each HG was ≥50%. The same name was assigned to proteins within the same HG. Hypothetical proteins with uncertain functions in CPS synthesis were named according to the HG number (Table S1 in Supplementary material).

2.5. Primer and probe design

The primers were designed based on specific *wzy* and *wzx* genes using Primer Premier v5.0 software (Premier Biosoft International, Palo Alto, CA, USA). Biotin was used to label the 5′ end of the reverse primer. For MSA, a C12 modification was synthesized at the 5′ end of each probe (Table 2).

2.6. Multiplex PCR

The target sequences of the CPSgcs were amplified by multiplex PCR. Multiplex PCR amplification was conducted in a 25 μL volume consisting of 1× PCR buffer, 100 ng genomic DNA, 20 μM of each deoxynucleoside triphosphate, 1.5 units of Q5 high-fidelity DNA polymerase (New England BioLabs), and 0.5 μM of the forward and reverse primer. The reaction parameters were as follows: hot start at 95 °C for 10 min; 35 cycles at 95 °C for 30 s, 55 °C for 45 s, and 72 °C for 1 min; and a final extension at 72 °C for 5 min. The size of each amplicon was analyzed by agarose gel electrophoresis and images were captured using a gel imaging system (Tanon 4200, Tanon, Shanghai, China).

2.7. Development of the MSA platform

As described above (Qian et al., 2018), probes containing a 5′ end C12 modifier (AuGCT, Beijing, China) were bonded to microspheres. Briefly, 50 μL of microspheres (Bio-Rad Laboratories, CA, USA) were centrifuged in a centrifuge tube at 12,000*g* for 2 min, the supernatant was discarded, and the microspheres were suspended in 10 μL of 0.1 MMES buffer (pH 4.5). Next, 2 μL of the probes and 2 μL of fresh 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC; 10 mg/mL) were added to the centrifuge tube; the tube was thoroughly mixed and incubated in the dark for 30 min. The microspheres were then washed twice with 0.02% Tween 20 (Sigma-Aldrich, St. Louis, MO, USA) and 0.1% SDS (Sigma-Aldrich). The coupled microspheres were resuspended in 20 μL of TE buffer (pH 8.0; Sigma-Aldrich) and the products were stored in the dark at 4 °C.

The microsphere mixture was prepared by adding 5 μL of each microsphere-probe to a centrifuge tube containing 475 μL of 1.5× tetramethylammonium chloride (TMAC) and mixed by vortexing. Next, 33 μL of the microsphere mixture, 5 μL of the PCR amplicons, and 12 μL of TE were used as the hybridization system. The hybridization reactions were denatured for 5 min at 94 °C and then hybridized for 30 min at 53 °C. The hybridization reaction products were then washed and centrifuged twice at 12,000*g* for 2 min with 100 μL of 1× TMAC. The products were resuspended in 80 μL of 1× TMAC containing 10 ng/mL of streptavidin labeled with R-phycoerythrin (Molecular Probes, Eugene, OR, USA) and incubated for 15 min at 50 °C. The products were analyzed based on fluorescence intensity using a Bio-Plex platform (Bio-Rad) and the data were analyzed using Bio-Plex Manager 4.1 software. The data are presented as median fluorescence intensity (MFI), which represents the total fluorescence intensity minus the background intensity. The specific fluorescence signal was at least 3-fold higher than the nonspecific signals. TE buffer was used as the negative control in this study.

To determine the sensitivity of the MSA method, 10-fold serial dilutions of *V. parahaemolyticus* genomic DNA (100, 10, 1, and 0.1 ng) were prepared and used as templates for MSA detection.

2.8. Double blind test

A total of 120 clinical isolates were obtained from the Shanghai City Disease Control and Prevention Center. These isolates were extracted from the stool specimens of patients affected by *V. parahaemolyticus*. The K-serogroups of the 120 clinical isolates were determined by agglutination tests with antisera according to the manufacturer's instructions (Denka-Seiken Ltd., Tokyo, Japan). Nine polyvalent antisera were used to divide these isolates into nine groups; each group contained seven K-serogroups. In each group, the corresponding monovalent antiserum was used to determine the specific K-serogroup of each isolate. The K70 antiserum was not included in the nine polyvalent antisera groups, but was used as a single monovalent antiserum to test each isolate. Genomic DNA was extracted from the 120 clinical isolates using a DNA Extraction Kit (Sangon Biotech).

2.9. Construction of a molecular serotyping program

We constructed a python script to predict *V. parahaemolyticus* K-serogroups using genomic data from 845 *V. parahaemolyticus* genome sequences obtained from the database and the complete genomes of 23 isolates. A molecular K-serotyping database was set up based on the K-serogroup singular genes *wzx* and *wzy*; all specific genes of the *V. parahaemolyticus* isolates belonging to the 55 K-serogroups were stored in the database. To perform a BLASTn search using the K-serogroup-specific gene database, the minimum cutoff rate identified was 98% and the length was > 95%. The script output of the BLAST analysis listed the most highly matched genes and the identity between the genes detected in the tested genome and the K-serogroup specific genes. The script output predicted K-serogroup based on the most highly matched gene.

2.10. Accession numbers of CPSgc nucleotide sequences

The nucleotide sequences of the 55 *V. parahaemolyticus* CPSgcs are available from the GenBank database. The accession numbers are listed in Table 1.

3. Results

3.1. General features of 55 V. parahaemolyticus CPSgcs

The genomes of the 55 *V. parahaemolyticus* strains belonging to different K-serogroups were sequenced. The CPSgc is located between *gmhD* and *rjg* in the genome of type strain RIMD 2210633 (Chen et al., 2007). The CPSgc sequences were extracted from the 55 genomes. Of these, 45 CPSgcs mapped between *gmhD* and *rjg*. In the CPSgc of K18, K31, K38, and K60, *gmhD* was found at the 5′ end, but *rjg* was absent at the 3′ end. In the CPSgc of K12, K47, K52, K65, and K70, *rjg* was found at the 3′ end, but *gmhD* was absent at the 5′ end. In the CPSgc of K41, neither *gmhD* nor *rjg* were found (Fig. 1). The length of these CPSgcs ranged from 16,359 to 46,011 bp and 1375 predicted coding sequences were identified in the 55 CPSgcs.

The conserved gene set *yjbH-yjbG-yjbF* was located at the 5′ end in all CPSgcs except for K12, K41, K47, K52, K65, and K70. The genes *yjbH-yjbG-yjbF* plus *yjbE* are always expressed together and may function in extracellular polysaccharide assembly in *E. coli* (Ferrieres et al., 2007). However, their precise functions in *V. parahaemolyticus* CPS

Table 2
Primers and probes used in *V. parahaemolyticus* molecular typing.

Serogroup	Target gene	Forward/reverse primers (5′-3′)	Product size (bp)	Probes
K4	*wzy*	wl-71037:ATTTACTTTGGGTATGGG wl-71038: CAAATTCATTGCCTCTGT	313	OA-5056:NH2-(C12)-TTGTGGATGGAATGGCTG
K36	*wzy*	wl-71039:AATGAATGGACTTAGGCA wl-71040: CATCAAAGGGAGAACAAAT	142	OA-5057:NH2-(C12)-ACTTGTCTGAACAGCGTCG
K53	*wzy*	wl-69612:TTGTATCTGTCCTTTAGCAT wl-69613:CACTTAGTAGGTAGTCGTTA	241	OA-4798:NH2-(C12)-ACTAGTGGGGCTAAGTCA
K55	*wzy*	wl-69614:TTAGACTTTGTCCTTTTCGC wl-69615:ACCAAAGCACTCATGTGAAA	194	OA-4799:NH2-(C12)-GTACGTTTCGCCAAGGTA
K68	*wzy*	wl-69616: AATAATCCTTTGGCTAGTTT wl-69617:CCCGATGCATATAGGACAGA	204	OA-4800:NH2-(C12)-CTACTTCAATGGGTGGGA
K19	*wzy*	wl-69659: TTGATGTGGGTTCTTTGT wl-69660:TATAGAAGGCCGATATTG	164	OA-4808:NH2-(C12)-GTGAATGTCATTACGGGCAA
K56/K57	*wzy*	wl-69661:GCTGCTCTGTTTCTCATG wl-69662:TCCTCAACTTTGCCTTCT	233	OA-4809:NH2-(C12)-TTTTGGCGTTCCGTTTCAC
K39	*wzy*	wl-73233:AGCGACTAGAGCTGGTAT wl-73234:CAAATCTTGTAGAGGGAC	208	OA-6061:NH2-C12-TTATTCCAGCATACGGGC
K59	*wzy*	wl-73237:CATTAGCCGCTGAAGGAT wl-73238:CTATTTTGTTGCATTTAC	266	OA-5785:NH2-(C12)-TCTTAATTGTACCAACGTT
K9	*wzy*	wl-69673:CCGTTATTGTTATTTTGTC wl-69674:CTGTAAACAGAGCCATAAG	214	OA-4815:NH2-(C12)-AAAGTTGTGGTCATCACGCTGTT
K8	*wzy*	wl-73241:AATCTAGCACTTAATCGAAAGC wl-73242:TCCGACACTGCACCCATA	356	OA-6067:NH2-C12-TTCCGACTTGTTATACAGCATC
K44	*wzy*	wl-73243:CAGCGATGTTCTGGTCTA wl-73244:GTTCCAAGCAAGGGATTT	329	OA-4817:NH2-(C12)-AACAAAGCCAATTCCGTGG
K48/K47	*wzy*	wl-69681:AGACTTATTTGGATTACG wl-69682:AAGAACATAAAGCACAGT	215	OA-4819:NH2-(C12)-AAGCCGAATGGTTTAGTTGC
K25	*wzy*	wl-73249:ACCATTGTTATTAGGGAC wl-73250:TTACATAAAACCACCACC	350	OA-4818:NH2-(C12)-GGGGTTCTCCTTATTTAGCGA
K54	*wzy*	wl-69685:GACGCTAGTATAGAATCAC wl-69686:AATAGCCAACCATAAGTG	253	OA-4821:NH2-(C12)-ACCCAATTAAGCGGTGTTATG
K32	*wzy*	wl-69687:TGGTGGTAATGCTCAAAC wl-69688:TTACAAAGCAGAGCAAGG	227	OA-4822:NH2-(C12)-AATTTGCTGCAATCGCTGTG
K37	*wzy*	wl-69691:CCTATACGCGGAGATGTT wl-69692:CTAATCCAACCGTAACAAA	195	OA-5426:NH2-(C12)-TCTGACTCATACTCAACTCCTC
K45	*wzy*	wl-69693:AAAGTTCCGCCTAATAAT wl-69694:GGGTATCCCAACAGTGCC	184	OA-6049:NH2-C12-CAGTAATACTCATATCTGCTGGAC
K6	*wzy*	wl-71049:TTGGTTTGATCGGACTTT wl-71050:AGGGCTTACTCCTTCACC	113	OA-5063:NH2-(C12)-AGCATATTAGAACTTGCA
K34	*wzy*	wl-71051:CATTTATTGGGATAGCAAG wl-71052:TGAACTCAGGCCAGAAGC	248	OA-5064:NH2-(C12)-CTTGAGACAGCATCTAGC
K7	*wzy*	wl-71683:TTCTACTATGGCGATGGT wl-71684:CTATACCGAACAATAGGCC	115	OA-4829:NH2-(C12)-GGGCGGTTAGACTTCTTGAGG
K66	*wzy*	wl-69705:GCAACAGGATTAAAGCTGGGA wl-69706:AGGGTAAGATAGCGATAGCCA	206	OA-4831:NH2-(C12)-GGCAACTATGCAGGATAC
K64	*wzy*	wl-69709:GTACAATGTTAGACAGCATA wl-69710:TAGCAATAATCCATCTGT	185	OA-4833:NH2-(C12)-TGTTTGCAGGGCTCTTTCAC
K17	*wzy*	wl-71679:GATTCTATACAGGCGAACT wl-71680:ATAATTGCCCTATCTACG	176	OA-5059:NH2-(C12)-CTTCGAGTATAACTTCCCT
K41	*wzy*	wl-71681:ACCAGTATTTGAACCATT wl-71682:ACGGAAGAGCTAATGACT	168	OA-5060:NH2-(C12)-GAGGTCTATCTAACTCAATT
K43	*wzy*	wl-69703:AAGGAAACTGGAAGAATA wl-69704:CCACAGAACCAAGTAGAT	145	OA-5789:NH2-(C12)-TATTTTGAACCGTTATTT
K42	*wzy*	wl-69707:GCTTTTATAGGTTATTTCAT wl-69708:AAAACTGCCTCTTCACTT	242	OA-4832:NH2-(C12)-ATTCGCAATAAGCACTGTTCG
K46	*wzy*	wl-73217:AAGGTGCTGGAGTGTTTG wl-73218:TTGATGCGAGACCGTAAT	459	OA-4834:NH2-(C12)-AGGGCACATAAACACTATTTCG
K18	*wzy*	wl-71019:CTTGCATAGCGATACTTC wl-71020:CAAGCCTAGAATAGCAGA	272	OA-5048:NH2-(C12)-CTTATGATTTGGGGTTGT
K24	*wzy*	wl-71021:TTGGTCGTCGATTCTCGT wl-71022:GTAAAATAGAGCGCCTGA	127	OA-5780:NH2-(C12)-GCCATAAGTATGCTTTTGAA
K65	*wzy*	wl-69596:TAATGGACTATGGGGATG wl-69597:AAGAAGAGCATAACTCGC	214	OA-4790:NH2-(C12)-ATCTGGCGTTTCGGCTGTAT
K13	*wzy*	wl-69598:GTGCAATAATTCTGGTTT wl-69599:CTAGTGGGCTATACTTCC	261	OA-4791:NH2-(C12)-ACTGGAAACGATTGGCGAGA
K38	*wzy*	wl-69600:CCTTGTATTTTCTAGCTG wl-69601:AAAGTGAATAGACAATGC	155	OA-4792:NH2-(C12)-TGGTCCGCAGCTTCAACAGT
K1	*wzy*	wl-69602:TATCGTGAAATAATGGGA wl-69603:GAAATACGGTTAAAGGAG	139	OA-5050:NH2-(C12)-AACTCAGTATTTATTGGGGC
K20	*wzy*	wl-69604:GTTTCAAACAAACCCTAAT wl-69605:TAAAGGATCAAAGGAACTA	174	OA-4794:NH2-(C12)-GTTTGTATCATTCCGTTGCTTTGTG
K69	*wzy*	wl-69572:TTGTTCGTCCCTAGAAAG wl-69573:CGTTGGGATAACATACATC	258	OA-4778:NH2-(C12)-TTGCTTTGATTCGCTTCC
K40	*wzy*	wl-69576:TTCCTTGTATTTATGCCTTTCG wl-69577:TTCGGCTACATTGAATCAGC	119	OA-4780:NH2-(C12)-TTTTATCGTCTCGCAGTCGTG

(*continued on next page*)

Table 2 (*continued*)

Serogroup	Target gene	Forward/reverse primers (5′-3′)	Product size (bp)	Probes
K63	*wzy*	wl-71023:ATTAACCGGCGTATTCTA wl-71024:ATATCCCTCCCGATGTAG	289	OA-5051:NH2-(C12)-TTGATTACGCAAAAGATACC
K30	*wzy*	wl-69584:AACTTTCCTTACCTGCTT wl-69585:CTGATGATAATACTCCAA	297	OA-6046:NH2-C12-TACTTTAGGTTTTGGATCTGGCTA
K15	*wzy*	wl-71027:AGCCTGGCTATGCGTCC wl-71028:CAGCATGAAGAACTGGGT	109	OA-4785:NH2-(C12)-AAAGACATTGGCACTGGCTGGA
K70	*wzy*	wl-71031:GTGCGGATATGCGTTTAG wl-71032:GCGTATAGCCAACTTCCT	173	OA-4787:NH2-(C12)-AGCGAAGTGGGAACTGATTAT
K11	*wzy*	wl-71025:GTCAGATGGTATCACTATG wl-71026:GTACTTATCTTAACCGAAC	102	OA-5775:NH2-(C12)-GGCTTGTTGGGGTTTATA
K23	*wzy*	wl-69582:TCTGAATTAGGACGGTTTA wl-69583:AGAATGGCTTGTAGGGTT	203	OA-4783:NH2-(C12)-TCGTATTAACCCAGTCTATGG
K33	*wzy*	wl-73251:ATGGCTTACCCAAATGAT wl-73252:TTAATCAACGCTGCTAAA	389	OA-4786:NH2-(C12)-TCTGTGCGATAAAGATAGAGCC
K67	*wzy*	wl-73255:TTTAACTAAAACTTGGTTGG wl-73256:CCTTTAATAAATAGCGATGA	219	OA-4777:NH2-(C12)-CTTCCGTTCTCACAAGTAAGTCAT
K5	*wzy*	wl-74071:GCGGGCTCGAAATGATAGTC wl-74072:ACTCCCTTTTCAGGAGCAAAA	286	OA-6432:NH2-(C12)-CGGTCCGTGGTCCTCATATATAG
K12	*wzy*	wl-71445:TTTTCTTTGATGGGAACG wl-71446:ATTGGAGCGGTAGGGAGC	146	OA-5793:NH2-(C12)-AACCCCATGCCAAGTAGA
K31	*wzy*	wl-72623:GATTTCTACTCGCAAGAT wl-72624:AGCCGATCACTTAACCTAT	266	OA-5796:NH2-(C12)-TCCCAGATATGTATAGAAGAT
K51	*wzy*	wl-71451:GTTCCTTGTAAGATTCCCTGTT wl-71452:CAGCGACCATTTGTAGGC	134	OA-5311:NH2-(C12)-TGGCTTATTAGTGTCGTGTT
K60	*wzx*	wl-71455:TGACGCAGGTTCATACAA wl-71456:TACCCTTCAGGAAATAGG	267	OA-5798:NH2-(C12)-TGCTCTTAAGGTGCCATT
K22	*wzx*	wl-71687:TCGGTATGCCGTTAGTAC wl-71688:TATTAGTGGGATCGCTTT	223	OA-5800:NH2-(C12)-AAGGAAGTCATGCGTTTAA
K49	*wzy*	wl-73229:GCTTACCTGCTGTAGCGA wl-73230:TCGACGGAACACCAATAT	339	OA-6060:NH2-C12-TATAAGTCTTCATTTAATTTGGAT
K52	*wzx*	wl-73223:TTGGTGGTAGATTGATGG wl-73224:TCCGAAAGCAGATAAAGA	374	OA-5312:NH2-(C12)-GCTCAGACTGGATTATACTC

synthesis remain unknown. In our study, we found that *yjbH*, *yjbG*, and *yjbF* formed a conserved gene set at the 5′ end of each CPSgc. In addition to *yjbH-yjbG-yjbF*, *wbfE*, *wbfF*, and *wzz*, which play a role in CPS assembly (Chen et al., 2007), were observed in most CPSgcs. In K15, *wbfE*, *wbfF*, and *wzz* were absent in the CPSgc; instead, *wza* was identified in this region. Furthermore, we found the conserved *ugd-galU* gene set at the 3′ of all CPSgcs.

The *yjbF-ugd* set is responsible for the high diversity of CPSgcs. The *rmlABCD* genes, which encode proteins responsible for dTDP-L-rhamnose synthesis (Wang et al., 2007), were found in 12 CPSgcs. These four genes were usually found together; however, the CPSgc of K32 contained only *rmlA*, the CPSgc of K18 contained only *rmlC*, the CPSgcs of K1 and K38 contained only *rmlD*, and the CPSgc of K56 and K57 contained only *rmlAB*. *manCBA*, which encode proteins responsible for GDP-Man synthesis (Perepelov et al., 2015), were found in 21 CPSgcs. Seventeen CPSgcs were found to contain *fcl* and *gmd*, the products of which are involved in GDP-L-Fuc synthesis (Kneidinger et al., 2001); *ugd*, the product of which is responsible for UDP-D-GlcA synthesis (Grangeasse et al., 2003); *galU*, responsible for the formation of UDP-D-Glc (Bonofiglio et al., 2005); and *galE*, responsible for UDP-D-Gal synthesis (Samuel and Reeves, 2003), were also present in most CPSgcs. *wecB*, the product of which catalyzes the formation of UDP-D-ManNAc (Campbell et al., 2000), was found in the CPSgcs of K15, K17, K39, K51, and K69. *wecC*, which encodes a protein that catalyzes the formation of UDP-D-GalNAc (Cunneen et al., 2013), was found in the CPSgcs of K44 and K69. *fnlABC*, the products of which are responsible for the formation of UDP-L-FucNAc (Liu et al., 2008), were found in the CPSgcs of K30, K56 and K57. *rmlB*, *rmlA*, *fcf1*, and *fcf2*, which encode proteins involved in D-Fuc*f* synthesis (Feng et al., 2004), were found in the CPSgc of K56 and K57. Six genes (*pseB*, *pseC*, *pseF*, *pseG*, *pseH*, and *pseI*), thought to encode proteins involved in CMP-Pse5Ac7Ac biosynthesis (Schoenhofen et al., 2006), were found in the CPSgc of K60. *glf*, the product of which catalyzes the conversion of UDP-D-Gal to UDP-D-Gal*f* (Clarke and Whitfield, 1992), was present in the CPSgcs of K46, K52, K56, K57 and K60.

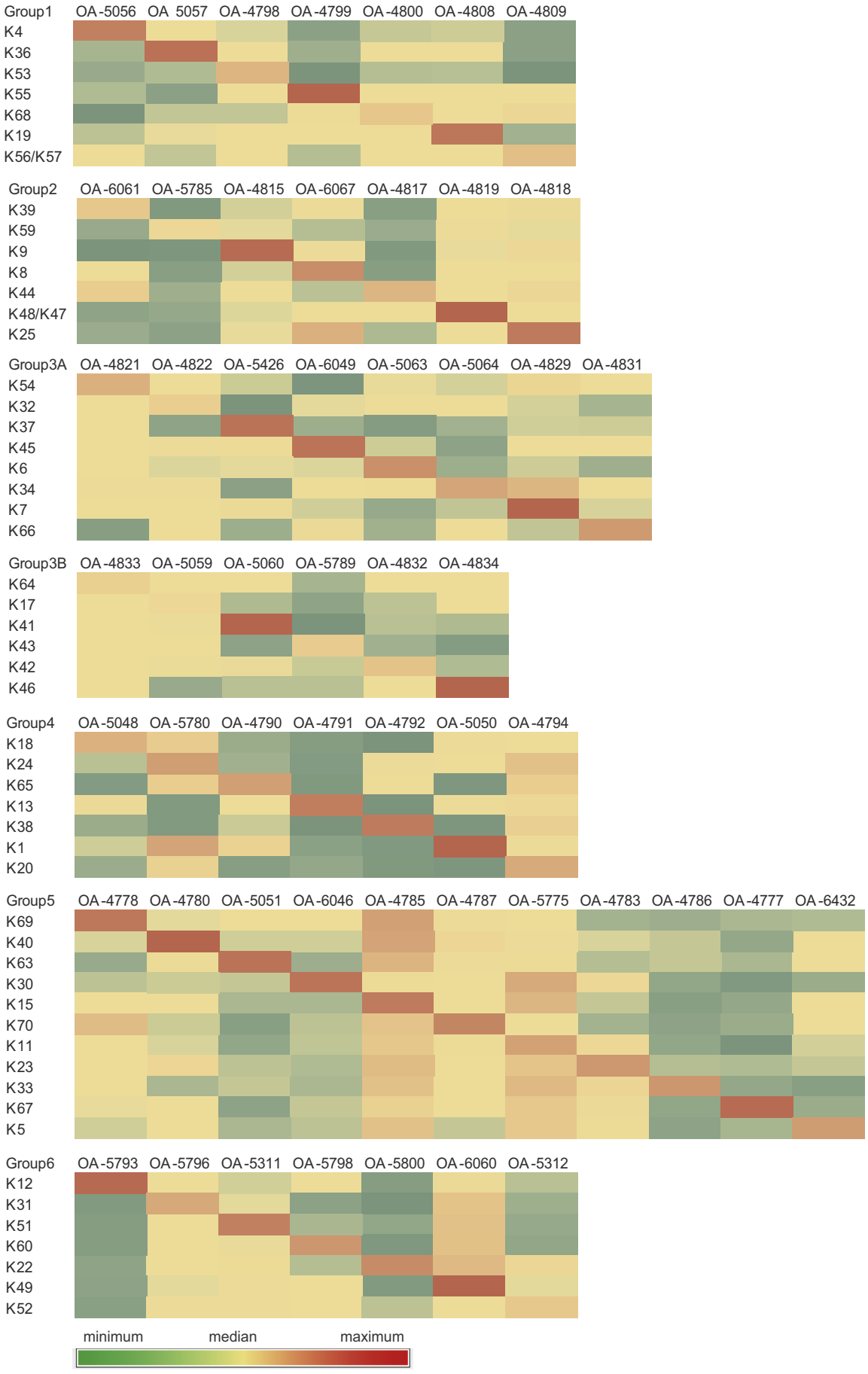
All CPSgcs contained the *wzx* and *wzy* genes, suggesting that the CPS is synthesized *via* the Wzx/Wzy pathway. The CPSgc of K38 only contained *wzx*; this atypical feature has been reported in *Klebsiella* spp. K29 and K50 (Pan et al., 2015). The CPSgcs of K8 and K45 contained only *wzy*; this atypical feature has been reported in *Klebsiella* spp. K11 and K34 (Pan et al., 2015). Sixteen CPSgcs carried genes encoding acetyltransferases and three CPSgcs contained genes encoding pyruvyltransferases. These genes may be involved in modification of CPS synthesis. The number of glycosyltransferases (GTs) in each CPSgc ranged from three to eleven. An initial GT gene (*wecA* or *wcaJ*), the product of which has been shown to transfer galactose-1-phosphate to UndP to initiate CPS synthesis (Liu et al., 1993; Patel et al., 2012), was present in 28 CPSgcs (Fig. 1 and Table S3 in Supplementary material).

A phylogenetic tree was constructed using genome SNP profile of 55 *V. parahaemolyticus*. The result showed that most of our strains exhibited a great genetic diversity and did not belong to the same clone (Fig. S2). However, we found K56 and K57 were closely related. In consistent, our result showed that the CPSgcs of them were almost identical (Fig. 1).

3.2. Distribution of CPS genes into HGs

The Tribe-MCL program was used to assemble 1375 predicted proteins into 106 HGs (≥2 members in each HG; Table S1 in Supplementary material). The products of *wzz*, *wbfE*, *wbfF*, *fcl*, *galU*, *galE*, *gmd*, *ugd*, *manA*, *manB*, and *manC* were each classified into a single HG, demonstrating that these proteins were conserved in different K-serogroups. In contrast, the GTs Wzy and Wzx were clustered into different HGs, suggesting they were diverse in various K-serogroups. Wzy was classified into 41 HGs and Wzx was categorized into 14 HGs. Intriguingly, the CPS modification proteins (pyruvyltransferases and acetyltransferases) were clustered in various HGs (two and 12 HGs, respectively).

Fig. 1. CPS biosynthesis gene clusters of 55 *V. parahaemolyticus* strains.
The arrows represent transcription direction and gene location in the CPSgcs. The different colors represent different types of gene functions.



(caption on next page)

Fig. 2. Heatmap representing the K-serogroups detected based on the MSA platform.
Each rectangle represents a fluorescence signal from mixed microspheres that have combined with a template PCR amplicon.

3.3. Development of an MSA-based molecular serotyping system

Molecular serotyping systems based on the CPS processing genes *wzy* and *wzx* have been widely employed for detecting gram-positive and negative bacteria (Okura et al., 2013; Qian et al., 2018). In this study, all primers and probes for implementing molecular K-serotyping of *V. parahaemolyticus* were designed based on *wzy*, apart from three serogroups, K22, K52, and K60, for which the primers and probes were designed based on the *wzx* gene (Table 2). These chosen genes were subjected to bioinformatic analysis; however, no sequence similarity was found between any two sequences following pairwise alignment.

To evaluate the specificity of the molecular serotyping approach based on the MSA system, we used 64 *V. parahaemolyticus* strains and other bacteria including *E. coli* (n = 3), *Salmonella* spp. (n = 3), *Vibrio cholerae* (n = 3), *Shigella* spp. (n = 2), *Klebsiella* spp. (n = 2), *Citrobacter freundii* (n = 2), and *Enterobacter cloacae* (n = 1). The multiplex PCR was divided into six groups. Non-specific products were not detected and each K-serogroup only produced an independent specific band of the anticipated size (Fig. S1 in Supplementary material). The subsequent MSA analysis of each group demonstrated that every single CPSgc-specific probe correctly identified the corresponding K-serogroup. Signals corresponding to other K-serogroups or other bacteria were not detected (Fig. 2 and Fig. S2 in Supplementary material).

We conducted a double-blind test using 120 clinical isolates from the Shanghai City Disease Control and Prevention Center; 31 different serogroups were detected in these isolates. Seven isolates were designated as K56/K57 serogroup; four isolates were designated as K4, K7, K41, K63, and K68 serogroups; three isolates were designated as K5, K6, K9, K13, K15, K18, and K49 serogroups; and the remaining 18 serogroups contained one or two isolates. In addition, traditional antisera agglutination tests were performed to evaluate the double-blind test MSA analysis results. Our results showed that most of the clinical isolates were classified in the same K-serogroup by either method, however some of them were untypeable by antisera agglutination test (Table 3). For isolates which were untypeable by antisera agglutination test, the MSA primers were used to amplify the target sequences. Each PCR amplicon was confirmed by ABI sequencing, showing consistent with the results generated by the MSA analysis. This indicated that our MSA assay was reliable and could gave higher resolution compared to conventional method. The sensitivity of the MSA molecular serotyping system was tested using 10-fold serial dilutions of genomic DNA (100, 10, 1, and 0.1 ng) derived from the 55 *V. parahaemolyticus* strains. The results showed that the sensitivity of this system for K-serogroup detection using genomic DNA is 1 ng.

Table 3
Comparison of MSA analysis and traditional antisera agglutination tests in K-serotyping of *V. parahaemolyticus*.

K-serogroup	MSA analysis	Antisera test	Inconsistent
K56/K57	7	7	0
K4	4	4	0
K7	4	3	1
K41	4	2	2
K63	4	3	1
K68	4	2	2
K5	3	2	1
K6	3	1	2
K9	3	3	0
K13	3	2	1
K15	3	3	0
K18	3	1	2
K49	3	2	1
K11	2	1	1
K19	2	2	0
K24	2	2	0
K33	2	1	1
K36	2	2	0
K38	2	1	1
K40	2	1	1
K52	2	1	1
K54	2	2	0
K55	2	1	1
K8	1	1	0
K17	1	0	1
K30	1	1	0
K32	1	1	0
K37	1	0	1
K39	1	1	0
K42	1	0	1
K43	1	1	0

3.4. In silico K-serotyping using V. parahaemolyticus genome data

Subsequently, we developed a program to classify the K-serogroups using the genomic data of *V. parahaemolyticus*. The database covered the 55 K-serogroups of *V. parahaemolyticus* analyzed in this study, consisting of the target genes that were employed in the MSA molecular serotyping scheme (Table 2). Fifty-two K-serogroups were identified using the *wzy* gene and the K22, K52, and K60 serogroups were identified using the *wzx* gene. We used 845 publicly available *V. parahaemolyticus* genomes from GenBank to design and evaluate the computerized serotyping program. Of the 845 *V. parahaemolyticus* genomes, 813 were assigned to 43 K-serogroups using this method with a threshold of 98% and a minimum length of 95%. Of the isolates, 191 strains were reported as belonging to 40 K-serogroups based on a conventional antisera method (Table S2 in Supplementary material) (Dong et al., 2017; Jensen et al., 2013; Makino et al., 2003). The serogroups identified for these 191 isolates using our *in silico* method were consistent with the results based on traditional serotyping, confirming the reliability of our method. Of the 813 isolates, 295 (36%) strains belonged to the K17 serogroup, followed by the K63 serogroup (15%) and K6 serogroup (11%). These results indicated that the K17 and K63 serogroups were the two major serogroups in the *V. parahaemolyticus* strains in GenBank. The K6 serogroup, which was reported as the most prevalent from 1995 to 1999 in Taiwan (Chiou et al., 2000), was the third major serogroup. The remaining 32 unallocated *V. parahaemolyticus* genomes possibly belong to the other 16 serogroups that were not covered by our method.

4. Discussion

Many methods have been developed for *V. parahaemolyticus* molecular typing such as pulsed-field gel electrophoresis (PFGE), multi-locus sequence typing (MLST), and whole-genome sequencing (WGS)-based approaches. PFGE provides a highly visual sense of global genomic monitoring and is performed using standard methods varying little for different organisms; this method has been frequently applied to determine the relatedness of *V. parahaemolyticus* clones (Ellingsen et al., 2008; Tsai et al., 2013) according to the standard PulseNet protocol (http://www.pulsenetinternational.org). The traditional MLST assay, which is based on several housekeeping genes and provides the phylogeny relationship of strains, has also been utilized and found suitable for subtyping *V. parahaemolyticus* strains (N.R. Chowdhury et al. 2004; Gonzalez-Escalona et al., 2017; Yu et al., 2011). Recently, with the decreasing cost of DNA sequencing and the development of user-friendly tools, the core genome MLST (cgMLST) method, which provides higher resolution and discriminatory power, has been suggested for global *V. parahaemolyticus* epidemiological studies (Gonzalez-

Escalona et al., 2017), showing ideal potential for use in outbreak investigations.

Serotyping based on the variability of O- or K-antigens has been in development since the 1940s and can provide important pathotype information for outbreak investigations, risk management, and epidemiological surveillance. *V. parahaemolyticus* infections and pandemics are associated with pathogenic strains of numerous specific serotypes (Han et al., 2017; Li et al., 2017). Therefore, the serotyping assay (combined with other molecular methods if necessary) should still be performed during the routine detection of *V. parahaemolyticus*, and not be replaced in the short-term. Moreover, as conventional serotyping is complicated, laborious, and time-consuming, our molecular serotyping method using the MSA platform developed in this study provides a novel tool for rapid, reliable, and cost-effective detection and serotyping of *V. parahaemolyticus*.

Although the CPS structures of *V. parahaemolyticus* remain unclear and need to be determined to support our study, our work, as a whole, elucidates the genetic basis of *V. parahaemolyticus* CPS and may contribute to understanding the evolution of this important food-borne pathogen. In addition, capsule antigens have been shown to confer enhanced colonization (Shifrin et al., 2008), environmental persistence (Gibson et al., 2006), and pathogenesis (Caboni et al., 2015). Strains that caused *V. parahaemolyticus* pandemics have been reported to be associated with different O- and K-serogroups (Chen et al., 2010). However, the mechanism of certain serotypes accounting for epidemics and pandemics remains largely unknown. In this study, there were no differences in the treatment of the different *V. parahaemolyticus* serotypes in clinical practice. Thus, understanding the precise mechanisms of different *V. parahaemolyticus* serotypes that cause infections could improve the clinical treatment of various serotype strains and increase the efficacy of clinical therapies. From this perspective, our CPSgc data provide the basis for the study of *V. parahaemolyticus* adaption, persistence, and pathogenesis.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijfoodmicro.2019.108332.

Acknowledgements

BL conceived the project. YP, XT and FL prepared the strain samples, preformed the sequence analyses, and developed the molecular serotyping system. YP, XG and LW conducted the bioinformatics analyses. SZ and SL developed the multiplexed Luminex-based array. XG, JW and XT performed the double-blind test. BL, YP and XG prepared the manuscript. All authors read and approved the final manuscript.

This work was supported by National Key Programs for Infectious Diseases of China (2017ZX10303405-001 and 2017ZX10104002-001-006), and National Natural Science Foundation of China (NSFC) Program (31820103002, 31770144, 81772148 and 81871624).

References

Baffone, W., Tarsi, R., Pane, L., Campana, R., Repetto, B., Mariottini, G.L., Pruzzo, C., 2006. Detection of free-living and plankton-bound *vibrios* in coastal waters of the Adriatic Sea (Italy) and study of their pathogenicity-associated properties. Environ. Microbiol. 8 (7), 1299–1305.

Bhuiyan, N.A., Ansaruzzaman, M., Kamruzzaman, M., Alam, K., Chowdhury, N.R., Nishibuchi, M., Faruque, S.M., Sack, D.A., Takeda, Y., Nair, G.B., 2002. Prevalence of the pandemic genotype of *Vibrio parahaemolyticus* in Dhaka, Bangladesh, and significance of its distribution across different serotypes. J. Clin. Microbiol. 40 (1), 284–286.

Blake, P.A., Weaver, R.E., Hollis, D.G., 1980. Diseases of humans (other than *cholera*) caused by *vibrios*. Annu. Rev. Microbiol. 34, 341–367.

Bogdanovich, T., Carniel, E., Fukushima, H., Skurnik, M., 2003. Use of O-antigen gene cluster-specific PCRs for the identification and O-genotyping of *Yersinia pseudotuberculosis* and *Yersinia pestis*. J. Clin. Microbiol. 41 (11), 5103–5112.

Bonofiglio, L., Garcia, E., Mollerach, M., 2005. Biochemical characterization of the pneumococcal glucose 1-phosphate uridylyltransferase (GalU) essential for capsule biosynthesis. Curr. Microbiol. 51 (4), 217–221.

Caboni, M., Pedron, T., Rossi, O., Goulding, D., Pickard, D., Citiulo, F., MacLennan, C.A., Dougan, G., Thomson, N.R., Saul, A., Sansonetti, P.J., Gerke, C., 2015. An O antigen capsule modulates bacterial pathogenesis in *Shigella sonnei*. PLoS Pathog. 11 (3), e1004749.

Campbell, R.E., Mosimann, S.C., Tanner, M.E., Strynadka, N.C., 2000. The structure of UDP-N-acetylglucosamine 2-epimerase reveals homology to phosphoglycosyl transferases. Biochemistry 39 (49), 14993–15001.

Chen, Y., Bystricky, P., Adeyeye, J., Panigrahi, P., Ali, A., Johnson, J.A., Bush, C.A., Morris Jr., J.G., Stine, O.C., 2007. The capsule polysaccharide structure and biogenesis for non-O1 *Vibrio cholerae* NRT36S: genes are embedded in the LPS region. BMC Microbiol. 7, 20.

Chen, Y., Dai, J., Morris Jr., J.G., Johnson, J.A., 2010. Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio parahaemolyticus* O3:K6. BMC Microbiol. 10, 274.

Chen, M., Guo, D., Wong, H.C., Zhang, X., Liu, F., Chen, H., Chen, M., Liu, B., Wang, L., Wu, F., Feng, L., 2012. Development of O-serogroup specific PCR assay for detection and identification of *Vibrio parahaemolyticus*. Int. J. Food Microbiol. 159 (2), 122–129.

Chen, Y.S., Li, H.R., Zhang, W., Hua, Z.D., Lin, X.H., Lin, M.Q., Huang, W.S., Huang, L.P., Yu, X.L., Xu, N.L., Lin, M., Xie, B.S., Shen, X.N., Xie, J.F., Wang, Y., Huang, M., Wu, Y.A., Hu, X.L., 2016. Development of a bead-based suspension array for the detection of pathogens in acute respiratory tract infections. Exp. Biol. Med. (Maywood) 241 (14), 1551–1558.

Chiou, C.S., Hsu, S.Y., Chiu, S.I., Wang, T.K., Chao, C.S., 2000. *Vibrio parahaemolyticus* serovar O3:K6 as cause of unusually high incidence of food-borne disease outbreaks in Taiwan from 1996 to 1999. J. Clin. Microbiol. 38 (12), 4621–4625.

Chowdhury, A., Ishibashi, M., Thiem, V.D., Tuyet, D.T., Tung, T.V., Chien, B.T., Seidlein Lv, L., Canh, D.G., Clemens, J., Trach, D.D., Nishibuchi, M., 2004a. Emergence and serovar transition of *Vibrio parahaemolyticus* pandemic strains isolated during a diarrhea outbreak in Vietnam between 1997 and 1999. Microbiol. Immunol. 48 (4), 319–327.

Chowdhury, N.R., Stine, O.C., Morris, J.G., Nair, G.B., 2004b. Assessment of evolution of pandemic *Vibrio parahaemolyticus* by multilocus sequence typing. J. Clin. Microbiol. 42 (3), 1280–1282.

Clarke, B.R., Whitfield, C., 1992. Molecular cloning of the rfb region of *Klebsiella pneumoniae* serotype O1:K20: the rfb gene cluster is responsible for synthesis of the D-galactan I O polysaccharide. J. Bacteriol. 174 (14), 4614–4621.

Cunneen, M.M., Liu, B., Wang, L., Reeves, P.R., 2013. Biosynthesis of UDP-GlcNAc, UndPP-GlcNAc and UDP-GlcNAcA involves three easily distinguished 4-epimerase enzymes, Gne, Gnu and GnaB. PLoS One 8 (6), e67646.

Dong, X., Bi, D., Wang, H., Zou, P., Xie, G., Wan, X., Yang, Q., Zhu, Y., Chen, M., Guo, C., Liu, Z., Wang, W., Huang, J., 2017. pirAB(vp)-bearing *Vibrio parahaemolyticus* and *Vibrio campbellii* pathogens isolated from the same AHPND-affected pond possess highly similar pathogenic plasmids. Front. Microbiol. 8, 1859.

Drummelsmith, J., Whitfield, C., 1999. Gene products required for surface expression of the capsular form of the group 1 K antigen in *Escherichia coli* (O9a:K30). Mol. Microbiol. 31 (5), 1321–1332.

Ellingsen, A.B., Jorgensen, H., Wagley, S., Monshaugen, M., Rorvik, L.M., 2008. Genetic diversity among Norwegian *Vibrio parahaemolyticus*. J. Appl. Microbiol. 105 (6), 2195–2202.

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157.

Feng, L., Senchenkova, S.N., Yang, J., Shashkov, A.S., Tao, J., Guo, H., Cheng, J., Ren, Y., Knirel, Y.A., Reeves, P.R., Wang, L., 2004. Synthesis of the heteropolysaccharide O antigen of *Escherichia coli* O52 requires an ABC transporter: structural and genetic evidence. J. Bacteriol. 186 (14), 4510–4519.

Ferrieres, L., Aslam, S.N., Cooper, R.M., Clarke, D.J., 2007. The *yjbEFGH* locus in *Escherichia coli* K-12 is an operon encoding proteins involved in exopolysaccharide production. Microbiology 153 (Pt 4), 1070–1080.

Fujino, T., Okuno, Y., Nakada, D., Aoyama, A., Fukai, K., Mukai, T., Ueho, T., 1953. On the bacteriological examination of shirasu-food poisoning. Med. J. Osaka Univ. 4 (2/3), 299–304.

Gibson, D.L., White, A.P., Snyder, S.D., Martin, S., Heiss, C., Azadi, P., Surette, M., Kay, W.W., 2006. *Salmonella* produces an O-antigen capsule regulated by AgfD and important for environmental persistence. J. Bacteriol. 188 (22), 7722–7730.

Gonzalez-Escalona, N., Jolley, K.A., Reed, E., Martinez-Urtaza, J., 2017. Defining a core genome multilocus sequence typing scheme for the global epidemiology of *Vibrio parahaemolyticus*. J. Clin. Microbiol. 55 (6), 1682–1697.

Grangeasse, C., Obadia, B., Mijakovic, I., Deutscher, J., Cozzone, A.J., Doublet, P., 2003. Autophosphorylation of the *Escherichia coli* protein kinase Wzc regulates tyrosine phosphorylation of Ugd, a UDP-glucose dehydrogenase. J. Biol. Chem. 278 (41), 39323–39329.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59 (3), 307–321.

Guo, X., Wang, M., Wang, L., Wang, Y., Chen, T., Wu, P., Chen, M., Liu, B., Feng, L., 2018. Establishment of a molecular serotyping scheme and a multiplexed Luminex-based array for *Enterobacter aerogenes*. Front. Microbiol. 9, 501.

Guvener, Z.T., McCarter, L.L., 2003. Multiple regulators control capsular polysaccharide production in *Vibrio parahaemolyticus*. J. Bacteriol. 185 (18), 5431–5441.

Han, C., Tang, H., Ren, C., Zhu, X., Han, D., 2016. Sero-prevalence and genetic diversity of pandemic *V. parahaemolyticus* strains occurring at a global scale. Front. Microbiol. 7, 567.

Han, D., Yu, F., Tang, H., Ren, C., Wu, C., Zhang, P., Han, C., 2017. Spreading of pandemic *Vibrio parahaemolyticus* O3:K6 and its serovariants: a re-analysis of strains isolated from multiple studies. Front. Cell. Infect. Microbiol. 7, 188.

Iguchi, T., Kondo, S., Hisatsune, K., 1995. *Vibrio parahaemolyticus* O serotypes from O1 to

O13 all produce R-type lipopolysaccharide: SDS-PAGE and compositional sugar analysis. FEMS Microbiol. Lett. 130 (2–3), 287–292.
Islam, S.T., Lam, J.S., 2014. Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. Can. J. Microbiol. 60 (11), 697–716.
Jensen, R.V., Depasquale, S.M., Harbolick, E.A., Hong, T., Kernell, A.L., Kruchko, D.H., Modise, T., Smith, C.E., McCarter, L.L., Stevens, A.M., 2013. Complete genome sequence of prepandemic *Vibrio parahaemolyticus* BB22OP. Genome Announc. 1 (1).
Jiang, Y., Chu, Y., Xie, G., Li, F., Wang, L., Huang, J., Zhai, Y., Yao, L., 2018. Antimicrobial resistance, virulence and genetic relationship of *Vibrio parahaemolyticus* in seafood from coasts of Bohai Sea and Yellow Sea, China. Int. J. Food Microbiol. 290, 116–124.
Jones, J.L., Ludeke, C.H., Bowers, J.C., Garrett, N., Fischer, M., Parsons, M.B., Bopp, C.A., DePaola, A., 2012. Biochemical, serological, and virulence characterization of clinical and oyster *Vibrio parahaemolyticus* isolates. J. Clin. Microbiol. 50 (7), 2343–2352.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780.
Kneidinger, B., Graninger, M., Adam, G., Puchberger, M., Kosma, P., Zayni, S., Messner, P., 2001. Identification of two GDP-6-deoxy-D-lyxo-4-hexulose reductases synthesizing GDP-D-rhamnose in Aneurinibacillus thermoaerophilus L420-91T. J. Biol. Chem. 276 (8), 5577–5583.
Laohaprertthisan, V., Chowdhury, A., Kongmuang, U., Kalnauwakul, S., Ishibashi, M., Matsumoto, C., Nishibuchi, M., 2003. Prevalence and serodiversity of the pandemic clone among the clinical strains of *Vibrio parahaemolyticus* isolated in southern Thailand. Epidemiol. Infect. 130 (3), 395–406.
Lee, L.H., Ab Mutalib, N.S., Law, J.W., Wong, S.H., Letchumanan, V., 2018a. Discovery on antibiotic resistance patterns of *Vibrio parahaemolyticus* in Selangor reveals carbapenemase producing *Vibrio parahaemolyticus* in marine and freshwater fish. Front. Microbiol. 9, 2513.
Lee, Y., Choi, Y., Lee, S., Lee, H., Kim, S., Ha, J., Lee, J., Oh, H., Kim, Y., Yoon, Y., 2018b. Occurrence of pathogenic *Vibrio parahaemolyticus* in seafood distribution channels and their antibiotic resistance profiles in S. Korea. Lett. Appl. Microbiol. 68 (2), 128–133.
Li, B., Yang, X., Tan, H., Ke, B., He, D., Ke, C., Zhang, Y., 2017. *Vibrio parahaemolyticus* O4:K8 forms a potential predominant clone in southern China as detected by whole-genome sequence analysis. Int. J. Food Microbiol. 244, 90–95.
Liu, D., Haase, A.M., Lindqvist, L., Lindberg, A.A., Reeves, P.R., 1993. Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica*: identification and characterization of transferase genes of groups B, C2, and E1. J. Bacteriol. 175 (11), 3408–3413.
Liu, B., Knirel, Y.A., Feng, L., Perepelov, A.V., Senchenkova, S.N., Wang, Q., Reeves, P.R., Wang, L., 2008. Structure and genetics of *Shigella* O antigens. FEMS Microbiol. Rev. 32 (4), 627–653.
Liu, J., Kibiki, G., Maro, V., Maro, A., Kumburu, H., Swai, N., Taniuchi, M., Gratz, J., Toney, D., Kang, G., Houpt, E., 2011. Multiplex reverse transcription PCR Luminex assay for detection and quantitation of viral agents of gastroenteritis. J. Clin. Virol. 50 (4), 308–313.
Liu, B., Knirel, Y.A., Feng, L., Perepelov, A.V., Senchenkova, S.N., Reeves, P.R., Wang, L., 2014. Structural diversity in *Salmonella* O antigens and its genetic basis. FEMS Microbiol. Rev. 38 (1), 56–89.
Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., Iijima, Y., Najima, M., Nakano, M., Yamashita, A., Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H., Hattori, M., Iida, T., 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V cholerae*. Lancet 361 (9359), 743–749.
Marolda, C.L., Li, B., Lung, M., Yang, M., Hanuszkiewicz, A., Rosales, A.R., Valvano, M.A., 2010. Membrane topology and identification of critical amino acid residues in the Wzx O-antigen translocase from *Escherichia coli* O157:H4. J. Bacteriol. 192 (23), 6160–6171.
Matsumoto, C., Okuda, J., Ishibashi, M., Iwanaga, M., Garg, P., Rammamurthy, T., Wong, H.C., Depaola, A., Kim, Y.B., Albert, M.J., Nishibuchi, M., 2000. Pandemic spread of an O3:K6 clone of *Vibrio parahaemolyticus* and emergence of related strains evidenced by arbitrarily primed PCR and *toxRS* sequence analyses. J. Clin. Microbiol. 38 (2), 578–585.
Morris Jr., J.G., Black, R.E., 1985. *Cholera* and other *vibrioses* in the United States. N. Engl. J. Med. 312 (6), 343–350.
Nair, G.B., Ramamurthy, T., Bhattacharya, S.K., Dutta, B., Takeda, Y., Sack, D.A., 2007. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. Clin. Microbiol. Rev. 20 (1), 39–48.
Okura, M., Takamatsu, D., Maruyama, F., Nozawa, T., Nakagawa, I., Osaki, M., Sekizaki, T., Gottschalk, M., Kumagai, Y., Hamada, S., 2013. Genetic analysis of capsular polysaccharide synthesis gene clusters from all serotypes of *Streptococcus suis*: potential mechanisms for generation of capsular variation. Appl. Environ. Microbiol. 79 (8), 2796–2806.
Pan, T.M., Wang, T.K., Lee, C.L., Chien, S.W., Horng, C.B., 1997. Food-borne disease outbreaks due to bacteria in Taiwan, 1986 to 1995. J. Clin. Microbiol. 35 (5), 1260–1262.
Pan, Y.J., Lin, T.L., Chen, C.T., Chen, Y.Y., Hsieh, P.F., Hsu, C.R., Wu, M.C., Wang, J.T., 2015. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of *Klebsiella* spp. Sci. Rep. 5, 15573.
Patel, K.B., Toh, E., Fernandez, X.B., Hanuszkiewicz, A., Hardy, G.G., Brun, Y.V., Bernards, M.A., Valvano, M.A., 2012. Functional characterization of UDP-glucose:undecaprenyl-phosphate glucose-1-phosphate transferases of *Escherichia coli* and *Caulobacter crescentus*. J. Bacteriol. 194 (10), 2646–2657.
Perepelov, A.V., Shashkov, A.S., Guo, X., Filatov, A.V., Weintraub, A., Widmalm, G., Knirel, Y.A., 2015. Structure and genetics of the O-antigen of *Escherichia coli* O169 related to the O-antigen of *Shigella boydii* type 6. Carbohydr. Res. 414, 46–50.
Qian, C., Du, Y., Li, H., Wu, P., Wang, L., Wei, Y., Cao, H., Yin, Z., Zhang, Y., Zhu, Y., Guo, X., Liu, B., 2018. Development of rapid and simple experimental and in silico serotyping systems for *Citrobacter*. Future Microbiol. 13, 1511–1522.
Samuel, G., Reeves, P., 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr. Res. 338 (23), 2503–2519.
Schoenhofen, I.C., McNally, D.J., Brisson, J.R., Logan, S.M., 2006. Elucidation of the CMP-pseudaminic acid pathway in *Helicobacter pylori*: synthesis from UDP-N-acetylglucosamine by a single enzymatic reaction. Glycobiology 16 (9), 8c–14c.
Shifrin, Y., Peleg, A., Ilan, O., Nadler, C., Kobi, S., Baruch, K., Yerushalmi, G., Berdichevsky, T., Altuvia, S., Elgrably-Weiss, M., Abe, C., Knutton, S., Sasakawa, C., Ritchie, J.M., Waldor, M.K., Rosenshine, I., 2008. Transient shielding of intimin and the type III secretion system of enterohemorrhagic and enteropathogenic *Escherichia coli* by a group 4 capsule. J. Bacteriol. 190 (14), 5063–5074.
Silbereisen, A., Tamborrini, M., Wittwer, M., Schurch, N., Pluschke, G., 2015. Development of a bead-based Luminex assay using lipopolysaccharide specific monoclonal antibodies to detect biological threats from *Brucella* species. BMC Microbiol. 15, 198.
Tsai, S.E., Jong, K.J., Tey, Y.H., Yu, W.T., Chiou, C.S., Lee, Y.S., Wong, H.C., 2013. Molecular characterization of clinical and environmental *Vibrio parahaemolyticus* isolates in Taiwan. Int. J. Food Microbiol. 165 (1), 18–26.
Ueno, H., Tomari, K., Kikuchi, K., Kobori, S., Miyazaki, M., 2016. The first report of *Vibrio parahaemolyticus* strain O10:K60 in Japan, a new combination of O and K serotypes isolated from a patient with gastroenteritis. Jpn. J. Infect. Dis. 69 (1), 28–32.
Wang, Y., Xu, Y., Perepelov, A.V., Qi, Y., Knirel, Y.A., Wang, L., Feng, L., 2007. Biochemical characterization of dTDP-D-Qui4N and dTDP-D-Qui4NAc biosynthetic pathways in *Shigella dysenteriae* type 7 and Escherichia coli O7. J. Bacteriol. 189 (23), 8626–8635.
Wang, Q., Ruan, X., Wei, D., Hu, Z., Wu, L., Yu, T., Feng, L., Wang, L., 2010. Development of a serogroup-specific multiplex PCR assay to detect a set of *Escherichia coli* serogroups based on the identification of their O-antigen gene clusters. Mol. Cell. Probes 24 (5), 286–290.
Ward, L.N., Bej, A.K., 2006. Detection of *Vibrio parahaemolyticus* in shellfish by use of multiplexed real-time PCR with TaqMan fluorescent probes. Appl. Environ. Microbiol. 72 (3), 2031–2042.
Whitfield, C., 2006. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. Annu. Rev. Biochem. 75, 39–68.
Wong, H.-C., Liu, S.-H., Wang, T.-K., Lee, C.-L., Chiou, C.-S., Liu, D.-P., Nishibuchi, M., Lee, B.-K., 2000. Characteristics of *Vibrio parahaemolyticus* O3: K6 from Asia. Appl. Environ. Microbiol. 66 (9), 3981–3986.
Wu, G., Yuan, Q., Wang, L., Zhao, J., Chu, Z., Zhuang, M., Zhang, Y., Wang, K., Xiao, P., Liu, Y., Du, Z., 2018. Epidemiology of foodborne disease outbreaks from 2011 to 2016 in Shandong Province, China. Medicine (Baltimore) 97 (45), e13142.
Yeung, P.S., Boor, K.J., 2004. Epidemiology, pathogenesis, and prevention of foodborne *Vibrio parahaemolyticus* infections. Foodborne Pathog. Dis. 1 (2), 74–88.
Yu, Y., Hu, W., Wu, B., Zhang, P., Chen, J., Wang, S., Fang, W., 2011. *Vibrio parahaemolyticus* isolates from southeastern Chinese coast are genetically diverse with circulation of clonal complex 3 strains since 2002. Foodborne Pathog. Dis. 8 (11), 1169–1176.