A Thesis for the Degree of Master of Science

Genome and Transcriptome Analyses of *Vibrio parahaemolyticus* FORC_022

장염비브리오 FORC_022에 대한 유전체, 전사체 분석

August, 2016

Byungho Lee

Department of Agricultural Biotechnology
College of Agriculture and Life Sciences
Seoul National University

석사학위논문

Genome and Transcriptome Analyses of *Vibrio parahaemolyticus* FORC_022

지도교수 최 상 호

이 논문을 석사학위논문으로 제출함

2016 년 8 월

서울대학교 대학원

농생명공학부

이 병 호

이병호의 석사학위논문을 인준함

2016 년 8 월

위원장 서 진 호 (인)

부위원장 최 상 호 (인)

위원 유 상 렬 (인)

Abstract

Vibrio parahaemolyticus is a Gram negative, curved rod-shaped bacterium that causes gastrointestinal illness associated with the consumption of seafood. To characterize *V. parahaemolyticus* that have caused outbreaks in South Korea, ten strains of *V. parahaemolyticus* were obtained from Korean Ministry of Food and Drug Safety. Virulence gene-specific PCR screening and lactate dehydrogenase (LDH) release assay were employed to evaluate the virulence of these strains. Among strains, the *V. parahaemolyticus* H8 strain, which was isolated from soy sauce marinated crab, showed a higher level of cytotoxicity than the positive control used in the LDH release assay. Therefore, the H8 strain was selected for genome sequencing and was designated as FORC_022. The FORC_022 genome consists of two circular chromosomes and a plasmid. The genome was predicted to have 4,858 open reading frames, 133 tRNA genes, and 37 rRNA genes. The FORC_022 strain did not include major virulence factor such as thermostable direct hemolysin (*tdh*), and TDH-related hemolysin (*trh*), but it had many genes related to virulence and several *V. parahaemolyticus* islands. Phylogenetic tree analysis of *Vibrio* 16S rRNA sequence showed that all the *V. parahaemolyticus* strains formed a group, which included the FORC_022 strain as well. Also, average nucleotide identity (ANI) analysis using nine complete *V. parahaemolyticu*s genome sequences available from NCBI revealed that FORC_022 was closely related to the CDC_K4557 strain (98.55%). Comparative genomic analysis of FORC 022 with CDC_K4557 revealed that FORC 022 carried an additional genomic

region encoding virulence factors like Zona occludens toxin and type III secretion factors. To identify differentially expressed genes of *V. parahaemolyticus* FORC_022 exposed to crab, RNA-sequencing was performed. Transcriptome analysis showed that the gene related to amino acid transport and metabolism, lipid transport and metabolism were down-regulated and the gene related to carbohydrate transport and metabolism, translation, ribosomal structure were up-regulated. Also, some virulence factors like type III secretion system, Tad locus, thermolabile hemolysin were up-regulated. The results suggested the FORC_022 strain may be pathogenic and there is high risk of infection when human eat crab contaminated strain FORC_022.

Key words: *V. parahaemolyticus*, Genome sequence, Food-borne pathogen, Comparative analysis, Virulence factor; *V. parahaemolyticus* island, Transcriptomics

Student Number: 2014-21909

Contents

List of Figures

List of Tables

I. INTRODUCTION

Vibrio parahaemolyticus is curved rod-shape, Gram-negative halophilic bacterium with a single polar flagellum. It inhabits marine and estuarine environments. The bacterium is leading cause of seafood-borne gastroenteritis, infection route is usually associated with the consumption of raw or undercooked seafood. The symptoms include watery diarrhea, vomiting, nausea, abdominal cramps, septicemia and even death (Joseph et al., 1982). In South Korea, *V. parahaemolyticus* has been one of top cause of food-borne disease outbreaks. It caused about 10% (175 of 1794) of all outbreaks from 2007 to 2012 and infected 2972 patients (Moon *et al*, 2014). Therefore, determination of the genomic information of various *V. parahaemolyticus* strains is important to understand the pathogenesis of these organisms.

The swimming crab *Portunus trituberculatus* is one of the most consumed aquatic animals in South Korea. Many data showed high amounts of the average annual catch and consumption of swimming crab in South Korea. (Statistics Korea, 2015 and Moon et al., 2009). Also, some outbreaks caused by *Vibrio* species in crab were reported. So, analyzing the transcriptomic changes of *V. parahaemolyticus* when it is exposed to crab, provides important insight into the understanding of survival and pathogenesis of the *V. parahaemolyticus* in humans as well as in crabs.

Recently, the advent of high throughput sequencing technologies make bacterial whole-genome sequencing faster and cheaper. Bacterial genome sequences can now be generated in many labs, in a matter of hours or days.

Such developments in DNA-sequencing technologies are significantly affect the diagnosis and monitoring of all pathogen. It can be used in many parts like identifying the species of an isolate, analyzing bacterial properties (virulence and resistance to antibiotics), monitoring the emergence and spread of bacterial pathogens (Didelot *et al*., 2012). RNA-sequencing (RNA-seq) techniques also have advanced and it allowed precise examination of transcriptome at a whole genome level, in particular physiological condition. Also, increased sensitivity of RNA-seq enabled the detection and quantification of RNA expressed at very low level (Pinto et al. 2011).

While the pathogenesis of *V. parahaemolyticus* has been studied in physiological and molecular levels, characterization and pathogenesis studies of the genomes of *V. parahaemolyticus* are insufficient. Also, only few *V. parahaemolyticus* were completely sequenced. In this study, *V. parahaemolyticus* FORC_022 was isolated soy sauce marinated crab contaminated with *V. parahaemolyticus* in South Korea, and its genome was sequenced and analyzed. Also, transcriptional analysis were performed to analyze gene expression pattern when it exposed to swimming crab. The results will improve epidemiological investigations and further understand the survival and pathogenesis of *V. parahaemolyticus* FORC_022.

II. MATERIALS AND METHODS

Strains, culture media

The strains used in this study are listed in Table 1. *Vibrio parahaemolyticus* strains cultivated aerobically at 30°C in modified Luria-Bertani medium (LB) supplemented with 1% (w/v) NaCl for 12 h (Sakazaki, 1983).

DNA extraction and virulence gene specific PCR

Genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA), according to the manufacturer's protocol. The virulence genes (*toxR*, *colla*, *groEL*) ORF was PCR-amplified from genomic DNA of *V. parahaemolyticus* strains using pairs of primers (VP_toxR_F and VP_toxR_R, VP_colla_F and VP colla_R, VP_gro_F and VP_gro_R). The amplified DNA fragments were resolved on a 1% agarose gel. The location of genes was visualized by using Geldoc^{TM} EZ Imager (Bio-Rad, Richmond, CA)

Cytotoxicity test

Cytotoxicity was evaluated by measuring the cytoplasmic lactate dehydrogenase (LDH) activity that is release from the human cell-lines by damage of plasma membranes. LDH release assay were performed using INT-407 (ATCC CCL-6) human intestinal epithelial cells. The preparation of the INT-407 cells and infection with the bacterial cultures were performed in a 96-well tissue culture plate (Nunc, Roskilde, Denmark) as previously described (Park *et al*, 2006). The cytotoxicity was then determined by measuring the

activity of lactate dehydrogenase (LDH) released in the supernatant using a Cytotoxicity Detection Kit (Roche, Mannheim, Germany), and expressed using the total LDH activity released from the cells completely lysed by 1% Triton X-100 as 100%.

Transmission electron microscope

The *V. parahaemolyticus* cells were negatively stained with uranyl acetate (UA) for one minute and were observed using TEM JEM-2100 (JEOL, Tokyo, Japan) at 200 kV. The FORC_022 strain was a curved, rod-shaped bacterium that was 1.5–2.0 μm in length and 0.6–0.8 μm in width with a single polar flagellum (Fig 1).

Genome sequencing and annotation

The whole genome sequencing and assembly were conducted at the ChunLab Genome sequencing was conducted at ChunLab Inc. (Seoul, South Korea) using hybrid sequencing technology with Illumina MiSeq (Illumina, San Diego, CA, USA) and PacBio RS II platform (Pacific Biosciences, Menlo Park, CA, USA) sequencers according to the manufacturers' protocols. Raw sequences obtained from PacBio RS II were assembled by PacBio SMRT Analysis ver. 2.3.0 software (Pacific Biosciences), and raw sequences from Illumina MiSeq were assembled by CLC Genomics Workbench ver. 7.5.1 (CLC bio, Aarhus, Denmark). The hybrid assembly of generating contigs from both systems was performed using the CodonCode Aligner (CodonCode, Co., Dedham, MA, USA).

Prediction of open reading frames (ORFs) and their annotation were performed using GeneMarkS program (Besemer *et al*, 2001)and the Rapid Annotation using Subsystem Technology (RAST) server (Aziz e*t al*. 2008) . Prediction of ribosome binding sites (RBS) were conducted using RBSfinder (J. Craig Venter Institute, Rockville, MD, USA). Subsequent predictions of the functions of ORFs and their conserved protein domains were carried out using InterProScan5 and COG-based WebMGA programs (Jones *et al*., 2014, Wu *et al*., 2011).

The putative virulence factors of the *V. parahaemolyticus* FORC_022 strain were predicted and characterized using BLAST search in the Virulence Factor Database (VFDB; http://www.mgc.ac.cn/VFs/main.htm). The circular genome maps of the FORC_022 strain were drawn using GenVision program (DNASTAR, Madision, WI, USA).

Comparative genome analysis

The strain was identified as *V. parahaemolyticus* using 16S rRNA sequencing and its identity was confirmed with comparative phylogenetic tree analysis using MEGA6 (Tamura et al., 2013) with various 16S rRNA sequences of the genus *Vibrio*. ANI phylogenetic tree analysis of complete genome sequences of *V. parahaemolyticus* (CDC_K4557, RIMD 2210633, BB22OP, UCM-V493, FDA_R31, FORC_004, FORC_006 and FORC_008) was conducted using the JSpecies program (Richter et al., 2009) to show the evolutionary relationship among *V. parahaemolyticus* strains. The comparative genome analysis was conducted with Artemis Comparison Tool (ACT) (Carver *et al*., 2005). Also,

V. parahaemolyticus islands (VPaIs) identification in FORC_022 was identified by BLAST analysis.

RNA isolation

V. parahaemolyticus FORC_022 was grown to mid-log phase (A_{600} of 0.8) and incubated in VFMG (*Vibrio fisheri* minimal medium containing glycerol) medium with the presence or absence of crab for 4hr. The culture were filtrated with syringe, sterilized gauze and vacuum filter with Whatman no. 1 filter paper (Whatman International Ltd, Maidstone, England). Subsequently, the culture transferred to 50 ml tubes (SPL, Kyungki, Korea) and centrifuged at 5,000 xg and 4˚C for 10 minutes. After centrifugation, pellet was resuspended with using 0.5 ml of cold-DEPC treated PBS and the solutions were quickly mixed with 1 ml of RNAprotectⓇ Bacteria Reagent (QIAGEN, Valencia, CA). Total RNAs were isolated using miRNeasy Mini Kit (QIAGEN) according to the manufacturer's procedure. Contaminated DNAs were digested by TURBO DNase (AMbion, Austin, TX) and then RNAs were cleaned up using RNeasy MinElute Cleanup kit (QIAGEN). The quality of total RNAs was verified using Agilent 2100 Bioanalyzer and Agilent RNA 6000 Nano reagents (Agilent Technologies, Waldbronn, Germany) by Chunlab (Seoul, South Korea).

RNA sequencing and data analysis

The procedures for a strand-specific cDNA library construction and RNA-sequencing were conducted by Chunlab. Briefly, mRNA was selectively

enriched by depleting ribosomal RNAs by using Ribo-Zero™ rRNA Removal kit (Epicentre, Madison, WI). Enriched mRNA was subjected to the cDNA library construction using TruSeq Stranded mRNA Sample Preparation kit (Illumina, San Diego, CA) following manufacturer's instruction. The quality of cDNA libraries was evaluated as described above for the quality verification of total RNA, except that Agilent DNA 1000 Reagents (Agilent Technologies) was used. Strand-specific paired-ended 100 nucleotide reads from each cDNA library were obtained using HiSeq2500 (Illumina). For biological replication, two libraries were constructed and sequenced from RNAs isolated from two independent filtered culture supernatants of *V. parahaemolyticus* FORC_022.

The reads obtained from RNA-sequencing were mapped to the *V. parahaemolyticus* FORC_022 reference genome (GenBank™ accession numbers CP013248, CP013249 and CP013250, www. ncbi.nlm.nih.gov) using CLC Genomics Workbench 5.5.1 (CLC Bio). The relative transcript abundance was measured by reads per kilobase of transcript per million mapped sequence reads (RPKM) (Mortazavi *et al*., 2008). The fold changes of RPKM values and their significance were assigned and the genes with 2 or greater fold change with P-values < 0.01 were considered to be differentially expressed in samples.

RNA purification and transcript analysis

Total cellular RNAs from the *V. parahaemolyticus* FORC_022 were isolated using RNeasy Mini kit (QIAGEN), and cDNA was synthesized using iScript™ cDNA Synthesis Kit (Bio-Rad). Real-time PCR amplification of the cDNA was performed by using the Chromo 4 real-time PCR detection system (Bio-Rad)

with pair of specific primers listed in Table 2 as described previously (Lim and Choi, 2014). Relative expression levels of the specific transcripts were calculated by using the 16S rRNA expression level as the internal reference for normalization.

Table 1. Bacterial strains used in this study

Strain	Relevant characteristics	Reference or source
Bacterial strains		
V. parahaemolyticus		
FORC_022	Environmental strain	Jeollanam-do Institute of Health and Environment
E1	Environmental strain	Gyeongnam Institute of Health & Environment
E2	Environmental strain	Gyeongnam Institute of Health & Environment
E3	Environmental strain	Gyeongnam Institute of Health & Environment
E5	Environmental strain	Busan Metropolitan City Institute of Health & Environment
E6	Environmental strain	Busan Metropolitan City Institute of Health & Environment
E7	Environmental strain	Busan Metropolitan City Institute of Health & Environment
E8	Environmental strain	Busan Metropolitan City Institute of Health & Environment
E9	Environmental strain	Busan Metropolitan City Institute of Health & Environment
F1	Environmental strain	Busan Metropolitan City Institute of Health & Environment

Table 2. Oligonucleotides used in this study

Oligonucleotide	Sequence (5'→3')	Use
For virulence gene specific PCR screening[a]		
VP_toxR_F	TGTACTGTTGAACGCCTAA	PCR analysis of *toxR*
VP_toxR_R	CACGTTCTCATACGAGTG	
VP_Colla_F	GAAAGTTGAACATCATCAGCACGA	PCR analysis of *Colla*
VP_Colla_R	GGTCAGAATCAAACGCCG	
VP_gro_F	AGGTCAGGCTAAGCGCGTAAGC	PCR analysis of *groEL*
VP_gro_R	GTCACCGTATTCACCCGTCGCT	
For qRT-PCR[b]		
FORC22_0119_F	GGCAGACCAACAAAAAGCGA	qRT-PCR of FORC22_0119
FORC22_0119_R	TTGGGTACACGCTGTTGTCT	
FORC22_0125_F	TTGAGAAAAAGCGGGGGTGT	qRT-PCR of FORC22_0125
FORC22_0125_R	CGTGGTTGCACACGAATCAG	
FORC22_0728_F	CAGCGAACATTGAAGTGGGC	qRT-PCR of FORC22_0728
FORC22_0728_R	GGTCGAAGAGGTTGAACGGT	
FORC22_0931_F	AATTTGTTGGCCCATCTCTG	qRT-PCR of FORC22_0931
FORC22_0931_R	CCGCATCAAGTTTCTTCGTT	
FORC22_1399_F	TGGTGATACGCAAGAAGACG	qRT-PCR of FORC22_1399
FORC22_1399_R	TTGGGGTACGTTCTTCGATT	
FORC22_1642_F	CGCTAAAAGAAGCGGGTGTG	qRT-PCR of FORC22_1642
FORC22_1642_R	CAAAGGTCACGATTGCACCG	

FORC22_1677_F	GTTGCTCGGAGATTTTGCTGG	qRT-PCR of FORC22_1677
FORC22_1677_R	GCTTGTTCTTTCTGAGCGCC	
FORC22_1684_F	TCTCGATTGGGCTTTCGCTT	qRT-PCR of FORC22_1684
FORC22_1684_R	AACGCACTTTGGTCGGTTTG	
FORC22_2280_F	AATCCCTGCAACTTCTGGCA	qRT-PCR of FORC22_2280
FORC22_2280_R	CAAGACTACCAAATGCGCGG	
FORC22_2298_F	AAACCGTAAAGACGGCCCAT	qRT-PCR of FORC22_2298
FORC22_2298_R	GCCTGAACAGCTCCAGTGAA	
FORC22_2737_F	CGCCGTGTTTGTCCAGATTG	qRT-PCR of FORC22_2737
FORC22_2737_R	ACAACCTGCTCCTTTGTGGA	
FORC22_3287_F	ACGCAAAACTGGCGAAGAAC	qRT-PCR of FORC22_3287
FORC22_3287_R	CGCTTTCACTTCTGGAACGC	
FORC22_3661_F	GCGGATAACGGCAAGTTAAA	qRT-PCR of FORC22_3661
FORC22_3661_R	ATACCATGCCCACTTTACGG	
FORC22_3789_F	GATGCCGAGGGTGCTCTTAG	qRT-PCR of FORC22_3789
FORC22_3789_R	GCGTCTGGCTGTAGTTGAGA	
FORC22_3797_F	GGCAGAAGAGGGGATTAGCG	qRT-PCR of FORC22_3797
FORC22_3797_R	AAAAGCCCTGTCGATGCTGA	
FORC22_3991_F	CGGCTTTCCATCAACTGAAT	qRT-PCR of FORC22_3991
FORC22_3991_R	CATTGAGGACTCGTCGCTTT	

FORC22_4409_F	qRT-PCR of FORC22_4409
FORC22_4409_R	
FORC22_4411_F	qRT-PCR of FORC22_4411
FORC22_4411_R	
FORC22_4415_F	qRT-PCR of FORC22_4415
FORC22_4415_R	

The oligonucleotides were designed using the [a]*V. parahaemolyticus* RIMD2210633 genome sequence (GenBank™ accession number BA000031, BA000032) and [b]*V. parahaemolyticus* FORC_022 genome sequence (GenBank™ accession number CP13248, CP13249)

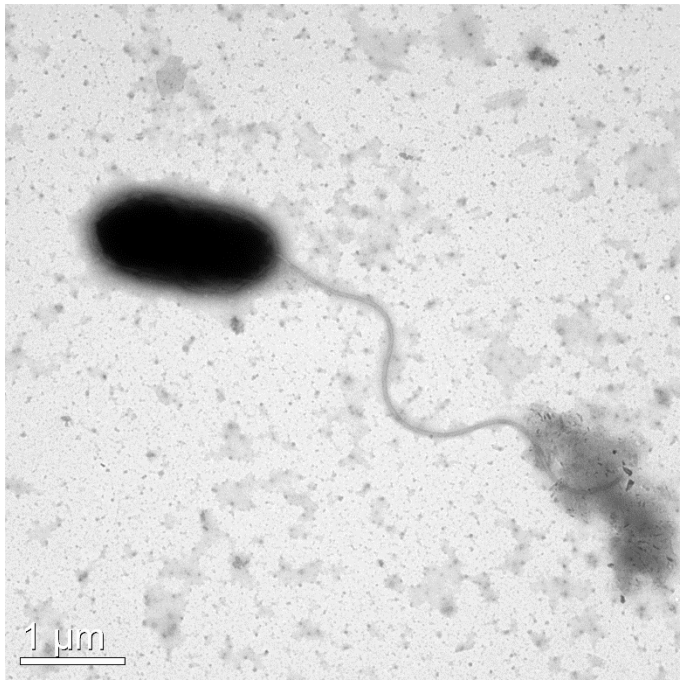

Fig 1. Transmission electron micrograph (TEM) image of *V. parahaemolyticus* FORC_022. The cells were negatively stained with 2.0% uranyl acetate (UA), for one minute. It was observed using TEM JEM-2100 (JEOL, Tokyo, Japan) at 200 kV.

III. RESULTS

Virulence gene-specific PCR screening

To select virulent strain, three virulence specific gene of *V. parahaemolyticus* were analyzed using PCR. The size of *toxR, Colla, groEL* is 503 bp, 272 bp, and 510 bp, respectively. Three genes were detected in all of strains. (Fig. 2)

Cytotoxicity analysis for ten strains of *V. parahaemolyticus*

To selection of virulent strain, LDH release assay was performed using the INT-407 cells. The host cells were infected with each strain and the activities of LDH released from the INT-407 cells were compared at various MOIs for 2 hours and 3 hours. LDH- releasing activity of the H8 strains higher than other strains (Fig. 3).

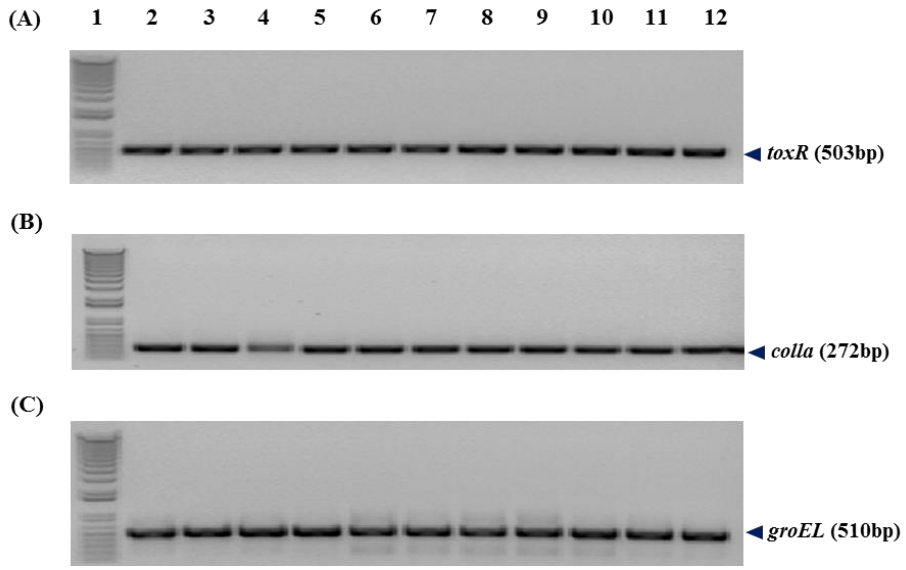

Fig. 2. PCR analysis for genes of *toxR*, *colla* and *groEL* in *V. parahaemolyticus*. In order to perform primary selection, three virulence specific gene of *V. parahaemolyticus*, *toxR* (A), *colla* (B) and *groEL* (C) were analyzed using PCR. *toxR*, trans-membrane protein involved in regulating virulence-associated gene; *colla*, gene encoding putative collagenase; *groEL*, heat shock protein. Molecular weight marker, lane 1; control, *V. parahaemolyticus* KCTC2471, lane 2; the E1 strain, lane 3; the E2 strain, lane 4; the E3 strain, lane 5; the E5 strain, lane 6; the E6 strain, lane 7; the E7 strain, lane 8; the E8 strain, lane 9; the E9 strain, lane 10; the F1 strain, lane 11; the H8 strain, lane 12.

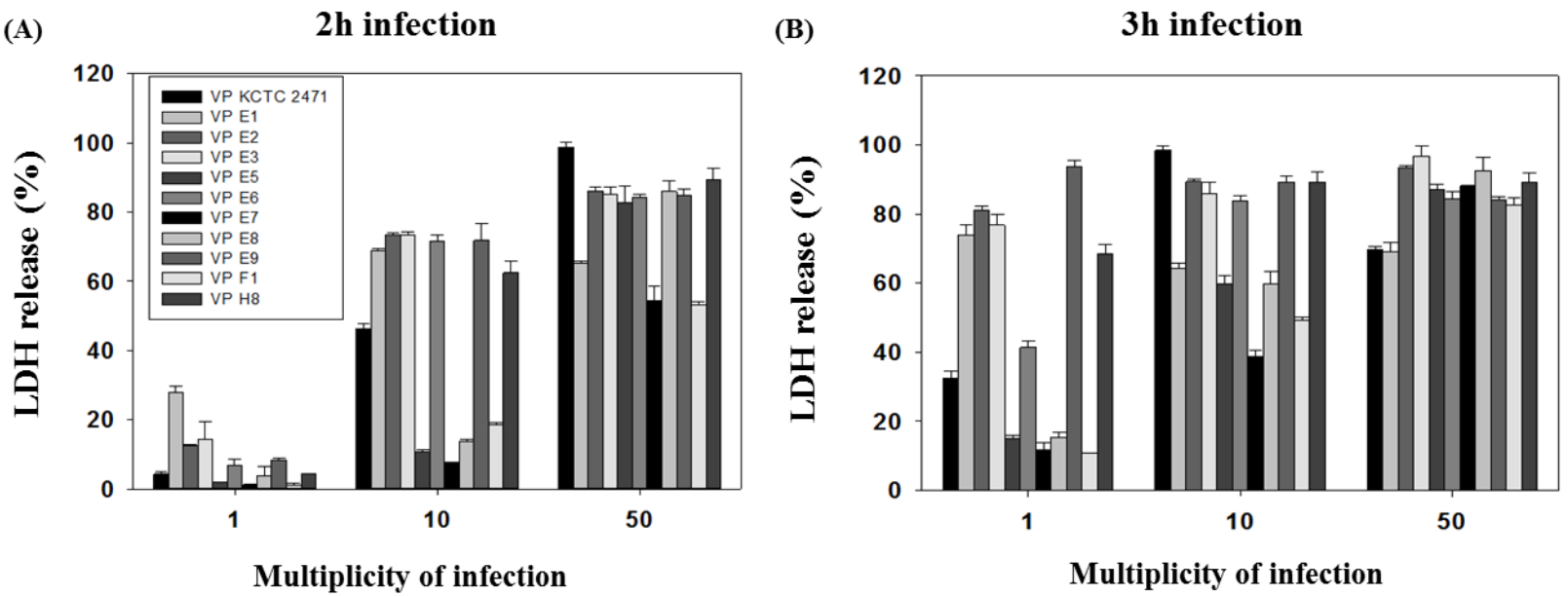

Fig. 3. Cytotoxicity analysis for 10 strains of *V. parahaemolyti*cus. INT-407 cells were infected with the *V. parahaemolyticus* strains at various MOIs for (A) 2h and (B) 3h. The cytotoxicity was determined by an LDH release assay and expressed using the total LDH released from the cells completely lysed by 2% Triton X-100 as 100%. Error bars represent the standard errors of the means (SEM).

Genome properties of *Vibrio parahaemolyticus* FORC_022

The hybrid platforms of MiSeq and PacBio with a total fold coverage of 326.62 presented that the complete genome of *V. parahaemolyticus* FORC_022 consists of two circular DNA chromosomes and a plasmid (Table 3)

Chromosome I consists of 3,397,828-bp with a GC content of 45.20% containing 3,066 predicted ORFs, 119 tRNA genes, and 34 rRNA genes. Among the ORFs, 2,423 ORFs (79.02%) were predicted to be functional and 643 ORFs (17.71%) to encode hypothetical proteins. Chromosome II consists of 1,879,989 bp with a GC content of 45.37% containing 1,685 predicted ORFs, 14 tRNA genes, and 3 rRNA genes. Among the ORFs, 1,322 ORFs (78.46%) were predicted to be functional and another 363 ORFs (21.54%) to encode hypothetical proteins. The plasmid, pFORC22.1, consists of 101,597 bp with a GC content of 44.49% containing 107 predicted ORFs. 41 ORFs (38.32%) were predicted to be functional and another 66 ORFs (61.68%) to encode hypothetical proteins. (Table 4). Through the bioinformatics analysis of these chromosomes and a plasmid, the circular genome maps were drawn (Fig. 4).

Table 3. Summary of *V. parahaemolyticus* FORC_022 genome

Property	Term
Finishing quality	Finished
Libraries used	Illumina 300 base pair paired-end library PacBio SMRTbell™ library (> 10 kb) for draft assembly
Sequencing platforms	Illumina MiSeq, PacBio RS II
Assemblers	CLCbio CLC Genomics Workbench 7.5.1 PacBio SMRT Analysis 2.3.0
Gene calling method	RAST ver. 2.0 (Glimmer 3), GeneMarkS
Average genome coverage	326.62x
Chromosome length (bp)	5,379,414 (PacBio and Illumina Miseq) 3,397,828 (Chromosome I) 1,879,989 (Chromosome II) 101,597 (Plasmid)
Contigs no.	3
Scaffolds no.	3
N50	3,397,828
Locus Tag	FORC22
Genbank ID	CP013248, CP013249, CP013250
Genbank Date of Release	2016-11-05
BIOPROJECT	PRJNA301198
Source Material Identifier	FORC_022
Project relevance	Agricultural

Table 4. The chromosomal properties of *V. parahaemolyticus* FORC_022

Sample	DNA	GenBank Accession #	Genome size (bp)	Protein coding genes	Annotated genes	Hypothetical proteins	tRNA	rRNA
Vibrio parahaemolyticus FORC022	Chromosome I	CP013248	3,397,828	3,066	2423	643	119	34
	Chromosome II	CP013249	1,879,989	1,685	1322	363	14	3
	pFORC22.1	CP013250	101,597	107	41	66	0	0

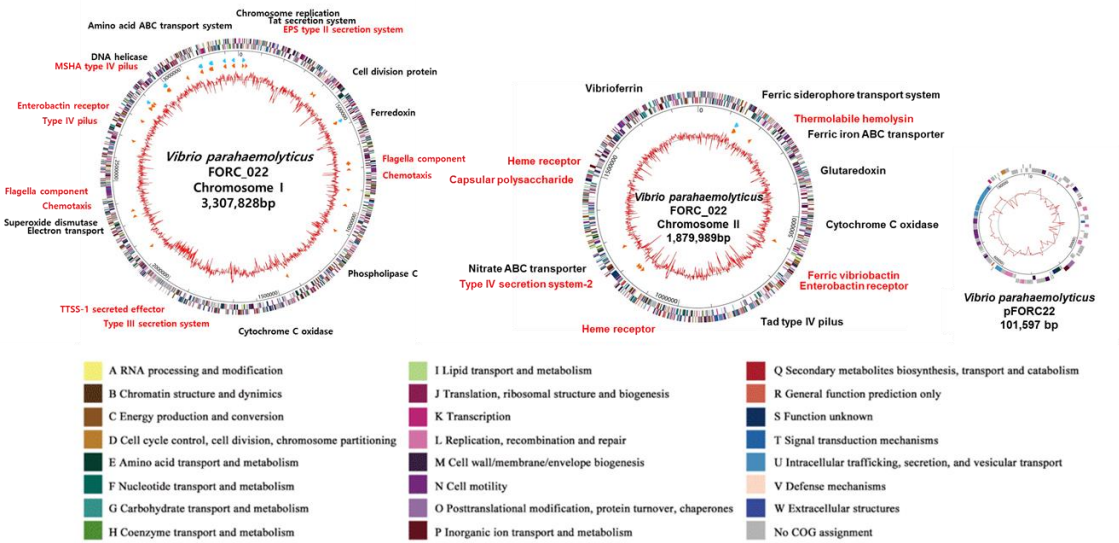

Fig. 4. Genome map of *Vibrio parahaemolyticus* FORC_022. The outer circle indicates the locations of all annotated ORFs, and the inner circle with the red peaks indicates GC content. Between these circles, sky blue arrows indicate the rRNA operons and orange arrows indicate the tRNAs. All annotated ORFs were colored differently according to the COG assignments.

Comparative phylogenetic tree analysis

To reveal the evolutionary relationship of *V. parahaemolyticus* FORC_022 with other *V. parahaemolyticus*, 16S rRNA sequence-based phylogenetic tree analysis and whole genome-based average nucleotide identity (ANI) analysis were conducted. Phylogenetic tree analysis of *Vibrio* 16S rRNA sequence showed that all the *V. parahaemolyticus* strains formed a group, which included the FORC_022 strain as well (Fig. 5). Also, average nucleotide identity (ANI) analysis using nine complete *V. parahaemolyticu*s genome sequences revealed that FORC_022 was closely related to clinically isolated strains, *V. parahaemolyticus* CDC_K4557 with the ANI value of 98.55 (Fig. 6).

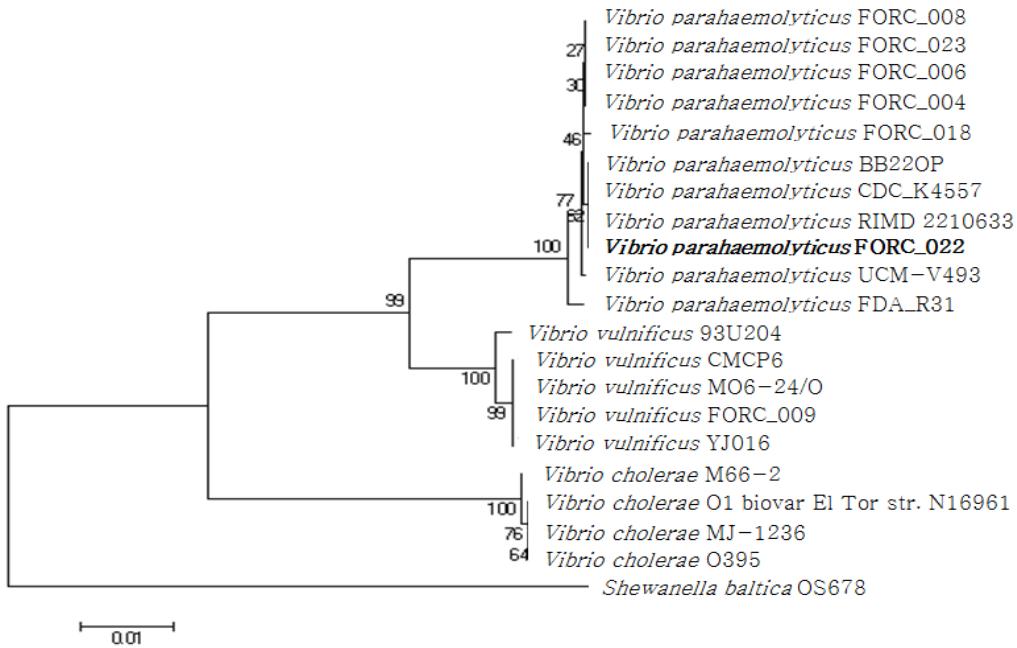

Fig. 5. The phylogenetic tree of strain FORC_022 with closely related *Vibrio* species aligned by using ClustalW. Phylogenetic trees were constructed from the aligned sequences using the neighbor-joining method with 1,000 bootstrap replicates via MEGA6 software. *Shewanella baltica* OS678 was used as outgroup.

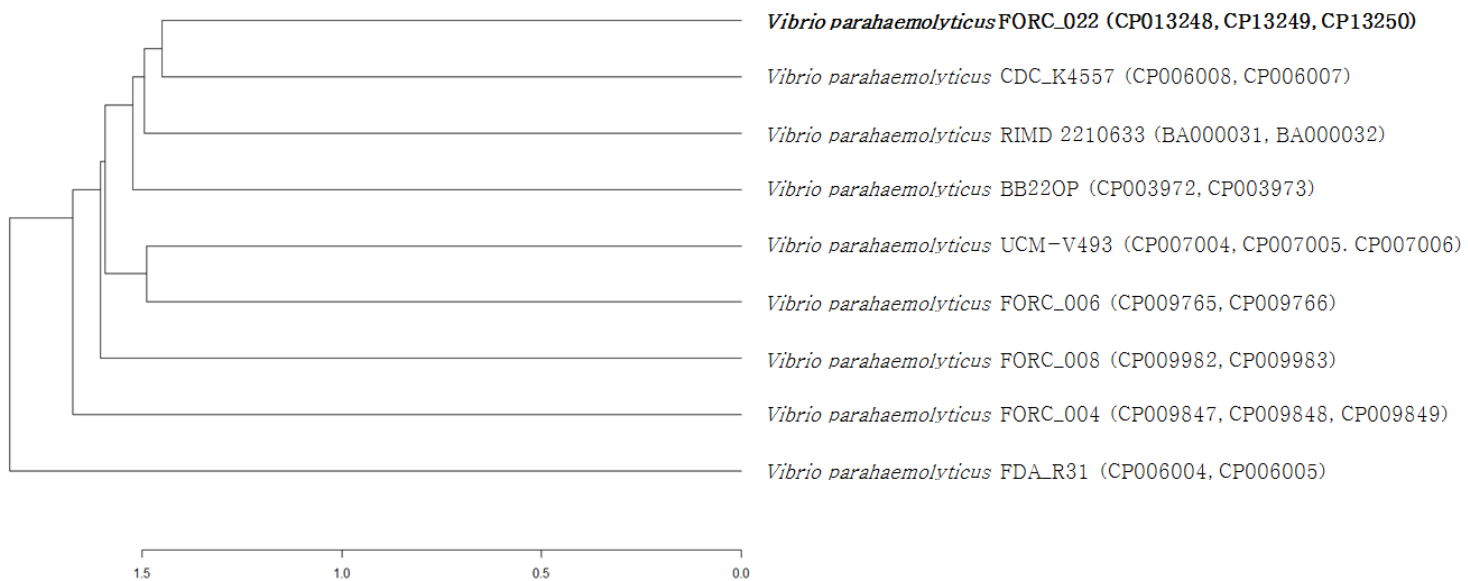

Fig. 6. The genome tree of *V. parahaemolyticus* strains constructed using the ANI values. The ANI values were calculated using JSpecies by comparing whole genome sequences between *V. parahaemolyticus* strains, which were fragmented into 1020 bp, based on BLAST algorism. The tree was constructed using R program.

Virulence factors of *V. parahaemolyticus* FORC_022

Virulence factors of FORC_022 are listed in Table 5. FORC_022 had genes related to adherence, anti-phagocytosis, secretion system, and toxin. In particular, type III secretion system was involved in the cytotoxicity of the *V. parahaemolyticus* to eukaryotic cells (Park et al., 2004, Ono et al., 2006). Also, FORC_022 had genes responsible for iron uptake, chemotaxis and motility. However, the genome of FORC_022 did not include major virulence genes such as thermostable direct hemolysin (*tdh*) and TDH-related hemolysin (*trh*).

Table 5. Virulence factors of *V. parahaemolyticus* FORC_022

Virulence factor	Annotation	Chromosome	Location	Function
Adherence				
mshH, mshI, mshJ, mshK, mshL, mshM, mshN, mshE, mshG, mshF, mshB, mshA, mshC, mshD, mshO mshP, mshQ	Mannose-sensitive hemagglutinin (MSHA type IV pilus)	Chromosome I	2953746~2966634 (FORC22_2728~FORC22_2744)	Hemagglutinin activity
pilA, pilB, pilC, pilD	Type IV pilus	Chromosome I	2767326~2771646 (FORC22_2563~FORC22_2566)	Adhesion, motility
Antiphagocytosis				
cpsA, cpsB, cpsC, cpsD, cpsE, cpsF, cpsG, cpsH, cpsI, cpsJ	Capsular polysaccharide	Chromosome II	1479903~1491496 (FORC22_4411~FORC22_4421)	Protection
Chemotaxis and motility				
flaC, flaA, flgL, flgK, flgJ, flgI, flgH, flgG, flgF, flgE, flgD, flgC, flgB, flgA, flgM, flgN, flgP, flgO, flgT	Flagella	Chromosome I	800920~804968 (FORC22_0717~FORC22_0723) 807219~819746 (FORC22_0726~FORC22_0736) 820418~8215554 (FORC22_0739)	Flagella component
flhB, fliR, fliQ, fliP, fliO, fliN, filM, fliL, fliK, fliJ, fliI, fliH, fliG, fliF, fliE, flrC, flrA, fliS, flaI, fliD, flaG, flaB, flaD, flaE, fliA, flhG, flhF, flhA	Flagella	Chromosome I	2446674~2477357 (FORC22_2279~FORC22_2307)	Flagella component
cheR, cheV	Chemotaxis protein	Chromosome I	805141~806906 (FORC22_0723~FORC0724)	Chemotaxis
cheW, cheB, cheA, cheZ, cheY	Chemotaxis protein	Chromosome I	2439615~2440109 (FORC22_2272) 2442023~2446640 (FORC22_2275~FORC22_2278)	Chemotaxis

motA, motB, motY, motX	Flagellar motor protein	Chromosome I	717038~718762 (FORC22_0646~FORC22_0647) 2336730~2337611 (FORC22_2179) 3084307~3084942 (FORC22_2841)	Motility
Iron uptake				
vctA	Enterobactin receptors	Chromosome II	654353~656359 (FORC22_3685)	Ferric ion uptake
irgA	Enterobactin receptors	Chromosome I	2848859~2850817 (FORC22_2639)	Ferric ion uptake
hutA, hutR	Heme receptors	Chromosome II	1051751~1053832 (FORC22_4014) 1568680~1570818 (FORC22_4477)	Iron, heme uptake
vctP, vctD, vctG, vctC	Periplasmic binding protein-dependent ABC transport systems	Chromosome II	649040~652607 (FORC22_3679~FORC22_3682)	Ferric ion uptake
Quorum sensing				
luxS	Autoinducer-2	Chromosome I	2779697~2780215(FORC22_2577)	Autoinducer production
cqsA	CAI-1 autoinducer synthase	Chromosome II	724842~726023 (FORC22_3731)	Autoinducer production
Secretion system				
epsN, epsM, espL, epsK, epsJ, epsI, epsH, epsG, epsF, epsE, gspD, epsC	EPS type II secretion system	Chromosome I	134810~146130 (FORC22_0115~FORC22_0126)	Secretion system
vopD, vopB, vcrH, vcrV, vcrG, vcrR, vcrD, vscY, vscX, sycN, tyeA, vopN, vscN, vscO, vscP, vscQ, vscR, vscS, vscT, vscU, vscL, vscK, vscJ, vscI, vscH, vscG, vscF, vscD, vscC, vscB, vscA, virF, virG, vxsC	type III secretion protein	Chromosome I	1819885~1857460 (FORC22_1641~FORC22_1686)	Secretion system
vopS, vopR, vopQ	TTSS-1 secreted effectors	Chromosome I	1840471~1841949 (FORC22_1665) 1842645~1843622 (FORC22_1668) 1844327~1845490 (FORC22_1671)	Secretion system
Toxin				
tlh	Thermolabile hemolysin	Chromosome II	229639~230895 (FORC_3287)	pathogenesis

Comparative genomics with *V. parahaemolyticus* CDC_K4557 and VPaIs identification

Comparative genomic analysis of *V. parahaemolyticus* FORC_022 and its closest strain CDC_K4557 was conducted. Different genomic regions (ranging from FORC22_1561 to FORC22_1576; 1,719,645-1,726,760 bp in chromosome I, ranging from FORC22_3757 to FORC22_3831; 751,784 – 833,321 bp in chromosome II) were detected. The FORC_022 genome may contain additional virulence factors like accessory cholera enterotoxin (FORC22_1570), zona occludens toxin (FORC22_1571 and Tad locus (FORC22_3784 – FORC22_3797) compared with the CDC_K4557 (Fig. 7).

VPIs identification was conducted by comparing with *V. parahaemolyticus* RIMD 2210633. Four regions, VPaI-1 (FORC22_0350 – FORC22_0352, FORC22_0358 – FORC22_0360, FORC22_0362, FORC22_0365 – FORC22_0368, FORC22_0372 – FORC22_0373), FORC23 2245), VPaI-3 (FORC22_1007 – FORC22_1009, FORC22_1034 – FORC22_1041, VPaI-6 (FORC22_ 4366) and VPaI-7 (FORC22_4405 – FORC22_4406), were present in the FORC_022 genome (Table 6).

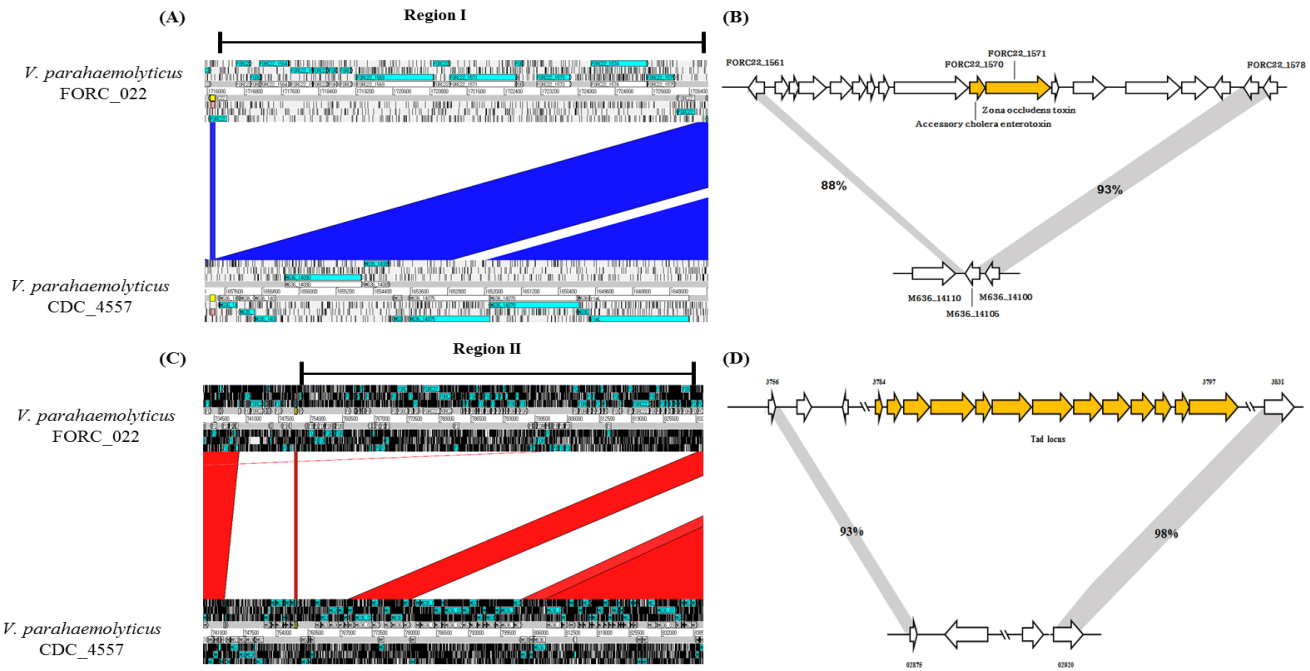

Fig. 7. Comparative genome analysis of the closely related strains, FORC_022 and CDC_K4557. (A) Region I included Zona occludens toxin region (B) Gene map of Region I (C) Region II included Tight adhesion locus region (D) Gene map of Region II

Table 6. Genomic islands of *V. parahaemolyticus* FORC_022.

ORFs of RIMD2210633	Annotation	Location on FORC_022	Detected gene number (Identity %)	Chromosome
Genomic Islands (GIs)				
VP0380 - VP0403	*V. parahaemolyticus* island-1 (VPaI-1)	367725-370261, 375474-378670, 380436-381926, 384740-390666, 394100-395834 (FORC22_0350-FORC22_0352, FORC22_0358-FORC0360, FORC22_0362, FORC22_0365-FORC22_0368, FORC22_0372-FORC22_0373)	VP0380-VP0382, VP0384-VP0386, VP0388, VP0395-VP0400, VP0402-VP0403 (98%)	Chromosome I
VP0635 - VP0643	*V. parahaemolyticus* island-2 (VPaI-2)	Absent		Chromosome I
VP1071 - VP1094	*V. parahaemolyticus* island-3 (VPaI-3)	1122265-1123693 and 1143337-1154849 (FORC22_1007-1009 and FORC22_1034-1041)	VP1073-VP1075, and VP1086-VP1094 (99%)	Chromosome I
VP2131 - VP2144	*V. parahaemolyticus* island-4 (VPaI-4)	Absent		Chromosome I
VP2900 - VP2910	*V. parahaemolyticus* island-5 (VPaI-5)	Absent		Chromosome I
VPA1253 - VPA1270	*V. parahaemolyticus* island-6 (VPaI-6)	1437093-143761 (FORC22_4366)	VPA1253 (99%)	Chromosome II
VPA1312 - VPA1398	*V. parahaemolyticus* island-7 (VPaI-7)	1472577=1473493 (FORC22_4405 - FORC22_4406)	VPA1397-VPA1398 (99%)	Chromosome II

Identification of differentially expressed genes under exposed to crab

The transcriptome analysis was used to compare the transcriptional profiles of the control, which was not expose to crab and treatment, which was expose to crab for 4h. Differentially expressed genes following the exposure to crab were identified. Average RPKM values from the biological duplicate samples were used to represent the expression level of each gene. The volcano plot showed that genes are differentially expressed with significance (*p*-value<0.01, 2 fold threshold) (Fig. 8). A total of 1283 genes were identified to be differentially expressed when FORC_022 exposed to crab; 650 up-regulated and 633 down-regulated. The differentially expressed genes (*p*- value <0.01, 2 fold threshold) were clustered into functionally related groups using WebMGA server (http://weizhong-lab.ucsd.edu/metagenomic-analysis/) for *V. parahaemolyticus* FORC_022 genome, which showed that the genes with various function were differentially expressed (Fig. 9). Also, the heat map shows the expressed virulence related genes are shown in *P* value ≤ 0.01 (Fig. 10). This results show that EPS type II secretion system, Type III secretion system, MSHA type IV pilus, thermolabile hemolysin, capsular polysaccharide, and heme receptors are up-regulated (Table 9).

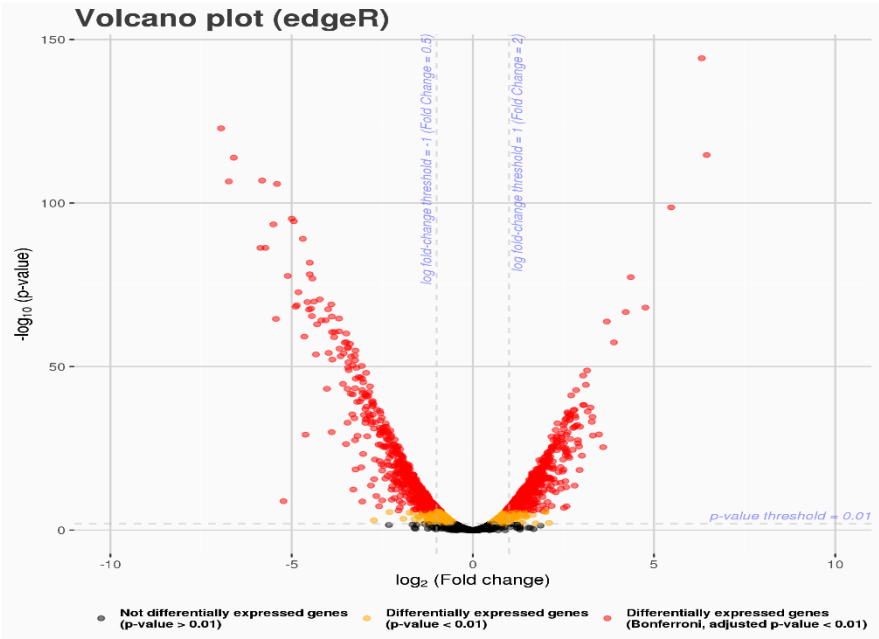

Fig. 8. Transcriptome comparisons of the RNA-seq samples. Volcano-plots of genes differentially expressed between the cells exposed and non – exposed to crab were generated. Number on the X- and Y-axis represent the fold change (log_2) and p-value (log_{10}). Red dots represent the differentially expressed genes.

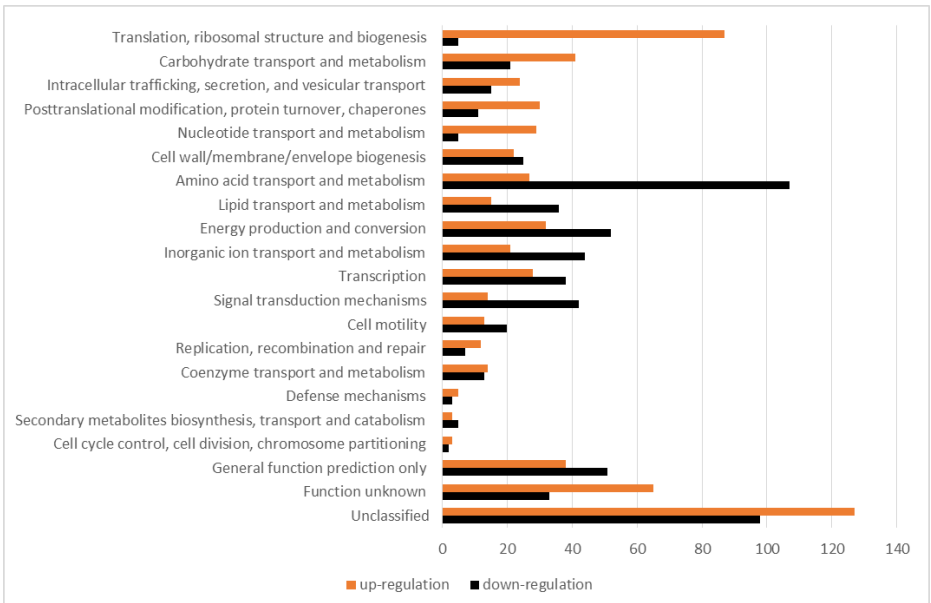

Fig. 9. Functional categorization of genes differentially expressed under exposed to crab. Gene with expression ratio of ≥2 on the basis of the RNA sequencing results were considered to be differentially expressed under exposed to crab. Functional categories (COG) are based on the database for the *V. parahaemolyticus* FORC_022 genome. Genes up-regulated (orange) and down-regulated (black) under exposed to crab were represented.

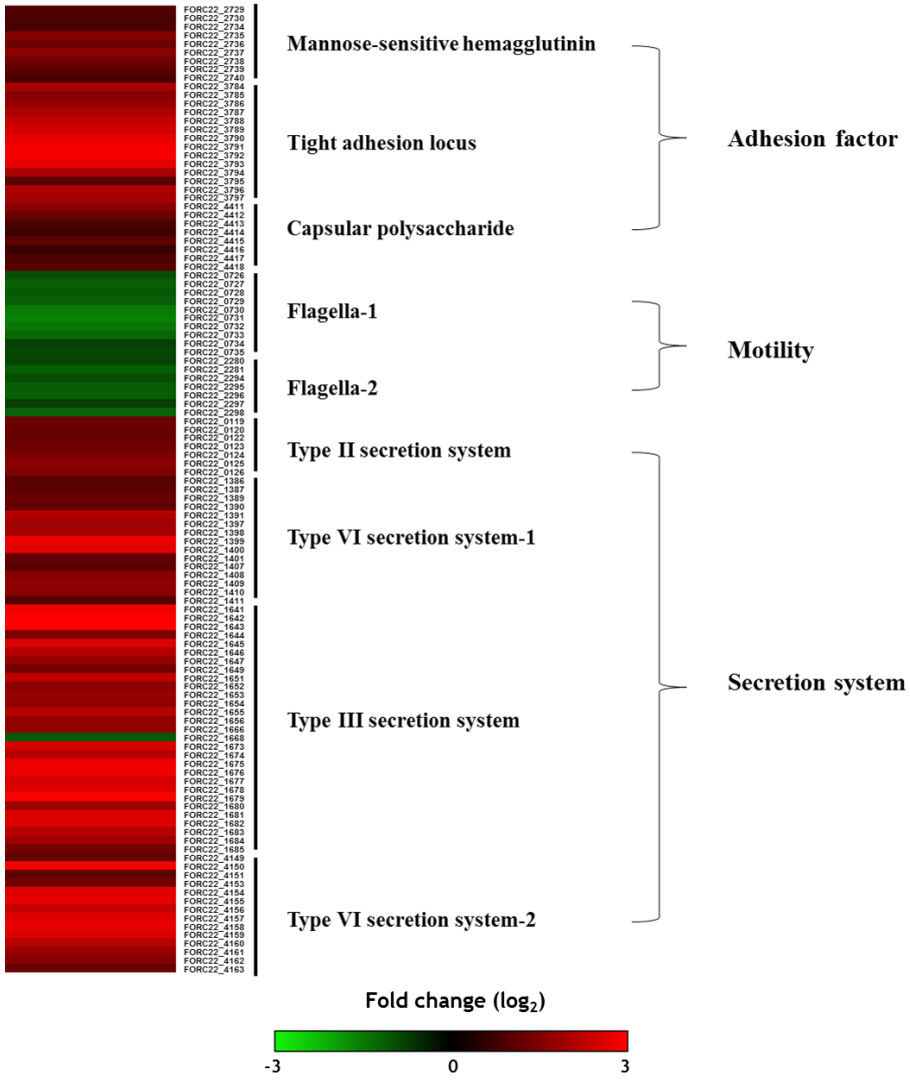

Fig. 10. Transcriptome of virulence genes under exposed to crab (*P*-value ≤ 0.01). Heat map was obtained using the Gitools.

Table 7. Transciptional change of virulence genes when FORC_022 strain exposed to crab

Locus tag	Product	Fold change	*P*-value[a]
Type III secretion system			
FORC22_1641	Type III secretion host injection and negative regulator protein (YopD)	7.19	0
FORC22_1642	Type III secretion host injection protein (YopB)	8.93	0
FORC22_1643	Type III secretion chaperone protein for YopD (SycD)	10.00	0
FORC22_1644	type III secretion cytoplasmic LcrG inhibitor LcrV	2.77	0
FORC22_1645	Type III secretion cytoplasmic plug protein LcrG	5.86	0.00071
FORC22_1646	Type III secretion low calcium response protein LcrR	4.28	0.00001
FORC22_1647	Type III secretion inner membrane channel protein	3.42	0
FORC22_1649	Type III secretion protein SctX	2.53	0.00660
FORC22_1651	Type III secretion outermembrane negative regulator of secretion TyeA	4.63	0.00006
FORC22_1652	Type III secretion outermembrane contact sensing protein	7.19	0.00001
FORC22_1673	Type III secretion cytoplasmic protein YscL	5.82	0
FORC22_1674	Type III secretion cytoplasmic protein YscK	4.42	0
FORC22_1675	Type III secretion bridge between inner and outermembrane lipoprotein	6.95	0
FORC22_1676	Type III secretion cytoplasmic protein YscI	7.18	0
FORC22_1677	Type III secretion effector protein YopR,	5.70	0
FORC22_1678	Type III secretion spans bacterial envelope protein YscG	6.03	0
FORC22_1679	Type III secretion cytoplasmic protein YscF	7.67	0
FORC22_1680	Type III secretion protein YscE	3.60	0.00055
FORC22_1681	Type III secretion inner membrane protein	5.88	0
FORC22_1682	Type III secretion outermembrane pore forming protein	6.06	0
FORC22_1683	Type III secretion chaperone protein for YopN (SycN,YscB)	4.48	0
FORC22_1684	Type III secretion negative regulator LscZ	3.69	0
FORC22_1685	type III secretion regulator ExsA	2.59	0
Thermolabile hemolysin (TLH)			
FORC22_3287	Thermolabile hemolysin precursor	3.99	0
Tight adhesion locus			
FORC22_3784	Flp pilus assembly protein, pilin Flp	3.90	0
FORC22_3785	Type IV prepilin peptidase TadV/CpaA	3.00	0

FORC22_3786	Flp pilus assembly protein RcpC/CpaB	3.34	0
FORC22_3787	Type II/IV secretion system secretin RcpA/CpaC	4.10	0
FORC22_3788	hypothetical protein	5.01	0
FORC22_3789	Type II/IV secretion system ATPase TadZ/CpaE,	5.60	0
FORC22_3790	Type II/IV secretion system ATP hydrolase TadA/VirB11/CpaF	6.52	0
FORC22_3791	Flp pilus assembly protein TadB	7.34	0
FORC22_3792	Type II/IV secretion system protein TadC	7.58	0
FORC22_3793	Flp pilus assembly protein TadD, contains TPR repeat	6.51	0
FORC22_3794	Flp pilus assembly membrane protein TadE	3.96	0
FORC22_3795	Flp pilus assembly surface protein TadF, ATP/GTP-binding motif	2.08	0
FORC22_3796	Protein TadG, associated with Flp pilus assembly	4.00	0
FORC22_3797	Outer membrane lipoprotein precursor, OmpA family	3.81	0

[a]The *P*-value less than six decimal places were denoted as zero.

IV. Discussion

In this study, virulence strain was selected for whole-genome sequencing. Ten strains of *V. parahaemolyticus* were obtained from Korean Ministry of Food and Drug Safety (Table 1). To evaluate the virulence of these strains, virulence gene-specific PCR screening and LDH release assay were conducted. Among the strains, *V. parahaemolyticus* H8, which was isolated from a soy sauce marinated crab by Jeollanam-do of Health & Environment in South Korea, showed a higher level of cytotoxicity than the positive control used in the LDH release assay (Fig. 2). Therefore, the H8 strain was selected, designated as FORC_022, and the complete genome sequence was obtained. To characterize of FORC_022 strain, comparative analysis of the complete genome were performed. Also, transcriptome analysis were employed to elucidate transcriptional response of the strain exposed to crab.

Previous studies, major virulence factors of *V. parahaemolyticus* are thermostable direct hemolysin (TDH), which induces beta-type hemolysin and clinical strains of *V. parahaemolyticus* almost have this gene (Miyamoto *et al*., 1969 and Kishishita *et al*., 1992). The genome of FORC_022 do not include the *tdh* and *trh* genes. Nonetheless, the cytotoxicity of FORC_022 strain showed the stronger than that of the positive control KCTC 2471 in LDH release assays and it indicated the FORC_022 strain may have other unknown virulence factors. Through the BLAST search against the VFDB, the virulence factors of the FORC_022 strain were predicted. Several secretion system were detected

in the genome sequence of FORC_022 (Type II secretion system: FORC22_0115-FORC22_0126 in chromosome I; Type III secretion system: FORC22_FORC22_1641 – FORC22_1686 in chromosome I; Type VI secretion system: FORC22_4148 – FORC22_4165 in chromosome II). Type II secretion system required for secretion of toxin that contribute to virulence in *V. cholerae* (cholera toxin), *P. aeruginosa* (exotoxin A) (Sandkvist, 2001) and type III secretion system was involved in the cytotoxicity of the *V. parahaemolyticus* to eukaryotic cells (Park *et al*., 2004 and Ono *et al*., 2006). Type VI secretion system deliver effectors can have either virulence activities against eukaryotic cells or antibacterial activities (Salomon and Kim, 2015). Because these secretion system associated virulence, the FORC_022 strain may have the potential to cause disease. The verification of these secretion systems in causing disease will be conducted in further study.

Comparative analysis with closest strain CDC_K4557 showed the FORC_022 strain had different regions compared with the CDC_K4557, clinical isolate from the stool of a patient in Louisiana in 2007. (Fig. 6). Region I (FORC22_1562 – FORC22_1576) contain accessory cholera enterotoxin and zona occludens toxin gene. Accessory cholera enterotoxin has the activity of a classic enterotoxin in an in vivo model, it contributes to intestinal secretion and diarrhea by stimulating Ca^{2+}-dependent Cl^-/HCO_3^- symporters (Trucksis *et al*., 2000). Zona occludens toxin modulate intestinal tight junction and leads to intestinal secretion (Fasano *et al*., 1997). Region II (FORC22_3757 – FORC22_3830) contain tight adhesion (Tad) locus. Tad locus was involved in pathogenicity of bacteria like *Pasteurella multocida*, *Yersinia ruckeri*

(Fernandez and Guijarro, 2004 and Fuller *et al*., 2000). These results showed that the FORC_022 strain may have additional virulence factors compared with CDC_K4557 strain

V. parahaemolyticus islands (VPaIs) identification in FORC_022 was performed by BLAST analysis (Table 5). Four regions, VPaI-1, VPaI-3, VPaI-6 and VPaI-7 were present in the FORC_022 genome. VPaI-1 encoded a type 1 restriction modification gene cluster and VPaI-6 encoded putative virulence genes. They may involve in protection and virulence (Hurley *et al*., 2006). The FORC_022 genome contain many putative virulence factor, suggesting that the FORC_022 strain may be pathogenic although this strain was isolated from environment.

Transcriptome analysis showed that many genes were differentially expressed when FORC_022 exposed to crab; 650 up-regulated and 633 down-regulated (Fig. 7). Among these gene, the gene related to amino acid transport and metabolism, lipid transport and metabolism were down-regulated and the gene related to carbohydrate transport and metabolism, translation, ribosomal structure were up-regulated (Fig. 8). Also, expression of genes related to adhesion factor, secretion system increased. Since these factors are involved in adherence to epithelial cell or pathogenicity(Sandkvist, 2001, Park *et al*., 2004, Ono *et al*., 2006, Fernandez and Guijarro, 2004, Fuller *et al*., 2000, Hurley *et al*., 2006), increased expression of the factors may be attributed to strong possibility of infection.

In conclusion, *V. parahaemolyticus* FORC_022 showed a high level of cytotoxicity although it is environmental strain and its genome does not include

two major hemolysin, thermostable direct hemolysin and TDH-related hemolysin. Other virulence factors like Tad locus, type III secretion system, a number of VPaI regions in FORC_022 may influence cytotoxicity. Also, transcriptional analysis showed the expression of these factors increased under exposed to model food, crab. All data shows *V. parahaemolyticus* FORC_022 may be pathogenic and there is high risk of infection when human eat crab contaminated strain FORC_022.

V. Reference

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., ... & Meyer, F. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, *9*(1), 75.

Besemer, J., Lomsadze, A., & Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic acids research*, *29*(12), 2607-2618.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., & Parkhill, J. (2005). ACT: the Artemis comparison tool. *Bioinformatics*, *21*(16), 3422-3423.

Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E., & Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics*, *13*(9), 601-612.

Fasano, A., Uzzau, S. E. R. G. I. O., Fiore, C. A. R. A., & Margaretten, K. L. A. R. A. (1997). The enterotoxic effect of zonula occludens toxin on rabbit small intestine involves the paracellular pathway. *Gastroenterology*, *112*(3), 839-846.

Fernandez, L., Marquez, I., & Guijarro, J. A. (2004). Identification of specific in vivo-induced (ivi) genes in *Yersinia ruckeri* and analysis of ruckerbactin, a catecholate siderophore iron acquisition system. *Applied and environmental microbiology*, *70*(9), 5199-5207.

Fuller, T. E., Kennedy, M. J., & Lowery, D. E. (2000). Identification of *Pasteurella multocida* virulence genes in a septicemic mouse model using signature-tagged mutagenesis. *Microbial pathogenesis*, *29*(1), 25-38.

Gao, W. W., Liu, H., Liu, C., Lu, Y., Gu, Q. F., Sun, S. F., & WU, C. F. (2011). Risk analysis on *Vibrio Parahaemolyticus* in *Portunus Trituberculatus*. *Journal of Environmental & Occupational Medicine*, *28*(414), e418.

Hurley, C. C., Quirke, A., Reen, F. J., & Boyd, E. F. (2006). Four genomic islands that mark post-1995 pandemic *Vibrio parahaemolyticus* isolates. *BMC genomics*, *7*(1), 1.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... & Pesseat, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236-1240.

Joseph, S. W., Colwell, R. R., & Kaper, J. B. (1982). *Vibrio parahaemolyticus* and related halophilic vibrios. *CRC Critical Reviews in Microbiology*, *10*(1), 77-124.

Kishishita, M., Matsuoka, N., Kumagai, K., Yamasaki, S., Takeda, Y., & Nishibuchi, M. (1992). Sequence variation in the thermostable direct hemolysin-related hemolysin (trh) gene of *Vibrio parahaemolyticus*. *Applied and Environmental Microbiology*, *58*(8), 2449-2457.

Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., ... & Kubota, Y. (2003). Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *The Lancet*, *361*(9359), 743-749.

Miyamoto, Y., Kato, T., Obara, Y., Akiyama, S., Takizawa, K., & Yamai, S. (1969). In vitro hemolytic characteristic of *Vibrio parahaemolyticus*: its close correlation with human pathogenicity. *Journal of Bacteriology*, *100*(2), 1147.

Moon, H. B., Kim, H. S., Choi, M., Yu, J., & Choi, H. G. (2009). Human health risk of polychlorinated biphenyls and organochlorine pesticides resulting from seafood consumption in South Korea, 2005–2007. *Food and chemical toxicology*, *47*(8), 1819-1825.

Moon, S., Sohn, I. W., Hong, Y., Lee, H., Park, J. H., Kwon, G. Y., ... & Youn, S. K. (2014). Emerging pathogens and vehicles of food-and water-borne disease outbreaks in Korea, 2007–2012. *Osong public health and research perspectives*, *5*(1), 34-39.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, *5*(7), 621-628.

Ono, T., Park, K. S., Ueta, M., Iida, T., & Honda, T. (2006). Identification of proteins secreted via *Vibrio parahaemolyticus* type III secretion system 1. *Infection and immunity*, *74*(2), 1032-1042.

Park, K. S., Ono, T., Rokuda, M., Jang, M. H., Okada, K., Iida, T., & Honda, T. (2004). Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*. *Infection and immunity*, *72*(11), 6659-6665.

Park, N. Y., Lee, J. H., Kim, M. W., Jeong, H. G., Lee, B. C., Kim, T. S., & Choi, S. H. (2006). Identification of the *Vibrio vulnificus wbpP* gene and evaluation of its role in virulence. *Infection and immunity*, *74*(1), 721-728.

Pinto, A. C., Melo-Barbosa, H. P., Miyoshi, A., Silva, A., & Azevedo, V. (2011). Application of RNA-seq to reveal the transcript profile in bacteria. *Genet Mol Res*, *10*(3), 1707-18.

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences*, *106*(45), 19126-19131.

Sakazaki, R. (1983). Vibrio parahaemolyticus as a food-spoilage organism. *Economic microbiology*.

Salomon, D., & Orth, K. (2015). Type VI secretion system. *Current Biology*, *25*(7), R265-R266.

Sandkvist, M. (2001). Type II secretion and pathogenesis. *Infection and immunity*, *69*(6), 3523-3535.

Statistics Korea. (2015). 2015 first half production Fishing Survey.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, mst197.

Trucksis, M., Conn, T. L., Wasserman, S. S., & Sears, C. L. (2000). *Vibrio cholerae* ACE stimulates Ca2+-dependent Cl−/HCO3− secretion in T84 cells in vitro. *American Journal of Physiology-Cell Physiology*, *279*(3), C567-C577.

Wu, S., Zhu, Z., Fu, L., Niu, B., & Li, W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC genomics*, *12*(1), 1.

VI. 국문초록

장염비브리오균은 식중독을 일으키는 주요 원인균으로, 해산물을 날 것으로 섭취할 경우에 감염될 수 있다. 국내에서 분리된 장염비브리오균들 중 높은 세포독성을 가진 균주를 선발하여, FORC_022로 명명하고 유전체 분석을 진행하였다. FORC_022의 유전체는 2개의 염색체와 하나의 플라스미드(plasmid)로 구성되어 있으며, 4,858개의 전사 해석틀 (open reading frame)과 133개의 운반RNA, 37개의 리보솜RNA가 존재한다고 예측된다. FORC_022 균주는 주요한 독성인자인 thermostable direct hemolysin 유전자를 가지고 있지 않았지만, 이외에 다른 독성인자들을 가지고 있었다. 비브리오 속의 16S 리보솜RNA의 계통 분석을 통해 FORC_022 균주가 장염비브리오균에 속한다는 것을 다시 확인하였으며, 9개의 장염비브리오균 유전체를 이용하여 진행한 average nucleotide identity (ANI) 분석을 통하여 FORC_022 균주가 CDC_K4557 균주와 계통적으로 가장 가깝다는 것을 밝혀냈다. FORC_022 균주와 CDC_K4557균주 간의 유전체 비교 분석을 통하여 FORC_022가 Zona occludens toxin and type III secretion factors와 같은 독성인자들을 추가적으로 가지고 있다는 것을 확인하였다. 모델 식품인 꽃게에 노출되었을 때, FORC_022의 유전자 발현양상의 변화를 확인하기 RNA-sequencing 기법을 이용하였

다. 단백질 수송 및 대사, 지방 수송 및 대사와 관련된 유전자들은 발현이 감소한 반면, 탄수화물 수송 및 대사, 단백질 번역과 관련된 유전자들의 발현은 증가하였다. 이외에도 type III secretion system, Tad locus, thermolabile hemolysin와 같은 독성인자들의 발현 또한 증가했다. 종합적으로, 장염비브리오균 FORC_022는 여러 독성인자를 가지고 있어 발병의 위험을 가지고 있으며, 꽃게와 같은 식품에 접촉하였을 때, 독성인자들을 발현시키기 때문에 FORC_022 균주에 오염된 식품을 섭취할 경우 감염의 위험성이 높아질 것이라고 예측할 수 있다.

주요어 : 장염비브리오균, 유전체 분석, 비교유전체학, 전사체 분석