Molecular virulence characterization of *Vibrio parahaemolyticus*

DISSERTATION

zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Department Chemie
Hamburg School of Food Science
der Universität Hamburg, Deutschland

With the aim of achieving a doctoral degree
at the Faculty of Mathematics, Informatics and Natural Sciences
Department of Chemistry
Hamburg School of Food Science
of the University of Hamburg, Germany

In partial fulfillment at the
Division of Seafood Science and Technology, U.S. Food and Drug Administration,
1 Iberville Dr, Dauphin Island, AL 36528, USA

Submitted by
Catharina Hermy Mariechen Lüdeke
from
Hamburg

Hamburg 2017

The present work was performed in the period between November 2011 till April 2015 under the supervision of Prof. Dr. Markus Fischer, Dr. Capt. William Burkhardt and Dr. Jessica Jones at the U.S. Food and Drug Administration, Division of Seafood Science and Technology, Dauphin Island, USA in collaboration with the University of Hamburg, Hamburg, Germany.

Dissertation Committee:	Prof. Dr. Markus Fischer
	Prof. Dr. Bernward Bisping
Disputation:	06.10.2017
Disputation committee:	Prof. Dr. Markus Fischer
	Prof. Dr. José Broekaert
	Dr. Angelika Paschke-Kratzin
Date of approval for publication:	06.10.2017

For my Family

I. Table of contents

II. List of tables

III. List of figures

IV. List of abbreviations

APW	Alkaline peptone water
BAM	FDA Bacteriological Analytical Manual
BPI	Bactericidal permeability-increasing proteins
CC	Clonal complex
CD	Cluster of differentiation
CDC	Centers for Disease Control and Prevention
CDS	Coding sequences
CE	Capillary electrophoresis
CLB	Cell lysis buffer
COG	Cluster of Orthologous Groups
CPS	Capsular polysaccharides
CSB	Cell suspension buffer
DABCYL	4-(dimethylaminoazo)benzene-4-carboxylic acid
DBG	De bruijn graph
DGREA	Direct genome restriction enzyme analysis
DLV	Double-locus variant
DNA	Desoxyribonucleic acid
Eth-Ac	Ethanol-acetic acid
FAM	6-carboxyfluorescein
FDA	Food and Drug Administration
FRET	Fluorescence resonance energy transfer
GPI	Glycosylphosphatidylinositol
HB	Hybridization buffer
HEX	Hexacholoro-6-carboxyfluorescein
HRM	High Resolution Melt

ISR-1	Intergenic spacer region-1
KAP	Kanagawa phenomenon detection
KP	Kanagawa phenomenon
kSNP	Kmer single nucleotide polymorphism
Kut	K untypeable
LB	Luria bertani
LBP	Lipid binding protein
LDH	Lactate dehydrogenase
LPS	Lipopolysaccharide
LS	Lysis solution
LS-BSR	Large scale BLAST score ratio
OLC	Overlap-layout-consensus
ONPG	O-nitrophenyl- β-D-galactopyranoside
MAM	Multivalent adhesion molecule
ME	Minimum evolution
ML	Maximum likelihood
MLST	Multi-locus sequence typing
MLVA	Multiple-locus variable-tandem repeat analysis
NaCl	Sodium chloride
NGS	Next generation sequencing
NJ	Neighbor-joining
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis
PGAAP	Prokaryotic Genomes Automatic Annotation Pipeline
PGN	Peptidoglycans

RAxML	Randomized Axelerated Maximum Likelihood
ROS	Reactive oxygen species
RPLA	Reverse passive latex agglutination
SDS	Sodium dodecyl sulfate
SSC	Standard saline citrate solution
SLV	Single locus variant
SNP	Single nucleotide polymorphism
ST	Sequence type
T3SS	Type-III-Secretion System
TAE buffer	Tris-Acetate-EDTA buffer
TAMRA	6-carboxytetramethylrhodamine
TBE buffer	Tris-Borate EDTA buffer
TCBS	Thiosulfate-citrate-bile salts-sucrose agar
TDH	Thermostable direct hemolysin
tdh	Thermostable direct hemolysin gene
TE buffer	Tris-EDTA buffer
TET	Tetrachloro-6-carboxy- fluorescein
TLH	Thermolabile hemolysin
tlh	Thermolabile hemolysin gene
TLR4/MD-2	Toll-like receptor 4 with lymphocyte antigen 96 complex
TRH	TDH-related hemolysin
trh	*tdh*-related hemolysin gene
TSA	Tryptic soy agar
TSA sheep	Tryptic soy agar with 10% sheep blood
TSB	Tryptic soy broth
TSB glycerol	Tryptic soy broth with 30% glycerol
UPGMA	Unweighted pair group method using arithmetic averages

WGS	Whole genome sequencing
WHO	World Health Organization

1. Summary

The main objective of this thesis was to characterize a diverse set of clinical and shellfish, primarily oyster, *Vibrio parahaemolyticus* isolates with the intention of identifying potential factors contributing to the virulence mechanism of this organism, especially by identification of factors that allow differentiation of clinical and shellfish and/or cytotoxic and non-cytotoxic isolates. The reason for this research is that reports have questioned the reliability of existing pathogenicity markers, thermostable direct hemolysin (*tdh*), *tdh*-related hemolysin (*trh*) and the type-III-secretion system (T3SS) genes, to fully explain *V. parahaemolyticus* virulence.

Initially, the isolate panel was examined for biochemical, serological, and virulence gene (using the established *tdh*, *trh* and T3SS) profiles. This part of the thesis was the first report on the presence of the type-III-secretion system 2β (T3SS2β) in tdh^+/trh^+ isolates, which was previously noted as absent in isolates of this genotype. Moreover, this study section emphasized that pathogenicity is more complex than previously thought as 25% of the clinical isolates were tdh^-/trh^- and did not possess T3SS2, which is generally indicative of a very low virulence potential. Serotype was the characteristic least shared by clinical and oyster isolates, as 17 of the 35 serotypes were found only in clinical isolates. The variety of serotypes was eventually determined to be too broad for use as an indicator of pathogenicity, in later sections of the study.

Throughout this thesis, subtyping as well as phylogenetic methods such as pulsed-field gel electrophoresis (PFGE), intergenic spacer region (ISR-1) analysis, direct genome restriction enzyme analysis (DGREA), multi-locus sequence typing (MLST), and high resolution melt-multiple-locus variable-tandem repeat analysis (HRM-MLVA), all functioning on the DNA level, were applied to the isolate panel. All these methods have been regularly utilized in epidemiological and evolutionary studies and these capabilities were again demonstrated in this thesis, with the exception of DGREA. DGREA has high discriminatory power that could be of use in phylogenetic studies. However, almost 20% of the isolates in this study were

untypeable by this method and, thus, it was determined to not be a reliable method for subtyping *V. parahaemolyticus*. Notably, ISR-1 was able to separate clinical and oyster isolates, indicating likely differences in the virulence potential. However, by comparing the results to further in-depth analysis using single nucleotide polymorphisms (SNPs) of the whole genome sequences (WGS), no correlation to that analysis' clustering was observed. PFGE had a high discriminatory power as well; diverse isolates produced unique pattern combinations. Additionally, isolates tended to cluster by their serotype using PFGE, which supports the use of serotyping as a basic screening method. Furthermore, a combination of PFGE, MLST, and HRM-MLVA presented to be a promising tool for outbreak or evolutionary investigations. MLST by itself could define identity and phylogenetic relationships of *V. parahaemolyticus* isolates; however, the HRM-MLVA method was able to differentiate between isolates within a PFGE cluster or of identical ST. By utilizing WGS, the MLST examination was expanded to an *in silico* analysis. A diverse set of 59 STs was found within 132 isolates. Although MLST is a widely used method to distinguish evolutionary relationships and outbreak source tracking, it is not a method of virulence indication.

Due to the unexpected challenges from automated annotations and/or assemblies of WGS sequence data, it is currently difficult to identify meaningful genetic differences on the species level in *V. parahaemolyticus*. However, based on the kSNP matrix differences in phylogeny between clinical and oyster isolates were observed. This clustering can hypothetically serve as a basis for the search of potential new pathogenicity factors. Furthermore, a cytotoxicity assay was developed targeting HeLa and Caco-2 cells. Based on the statistical analysis, Caco-2 cells seem to be the more suitable cell line for virulence investigation. After correlation of the cytotoxicity assay results with the genotype of each isolate, cytotoxicity appears partially associated with *tdh* and T3SS2. However, the results of these assays emphasized once more that *tdh*, *trh,* and T3SS are likely not definitive markers of strain virulence. The results of the KAP-reversed passive latex agglutination (RPLA) kit, indicating TDH production, further strengthens this conclusion. Isolates of the serotype O4:K12 and O4:Kut, which are associated with an outbreak strain from 2012 and 2013, seem

to possess higher virulence potential, whereas isolate serotype does not otherwise correlate with cytotoxicity. This is another indication that while the serotype is a valuable screening tool, it alone may not be predictive of strain virulence.

Taking the phenotype into consideration, a differentiation of cytotoxic and non-cytotoxic isolates could be made based on the phylogenetic analysis of concatenated sequences of a subset of the *V. parahaemolyticus* isolate panel from this study. Even though no specific genes could be found in all of the cytotoxic isolates that were not present in the non-cytotoxic isolates and vice versa, groups of genes were correlated to phenotypic results.

V. parahaemolyticus is a very diverse organism, complicating investigations into its pathogenicity mechanism. This research contributed a large amount of subtyping and phylogenetic data to the vibrio community. Comparing *V. parahaemolyticus* isolates from all coastlines of the United States, as well as from patients, by phenotype and genotype has, up to day, not been reported. The genes of the functional groups related to the cytotoxic phenotype lend insight into additional virulence factors. Further identified SNPs in cytotoxic strains add to understanding the pathogenic potential of *V. parahaemolyticus*. WGS and the establishment of a cytotoxicity assay contribute to the existing catalogue of subtyping and phylogenetic methods for *V. parahaemolyticus*.

2. Zusammenfassung

Ziel dieser Arbeit war die Charakterisierung einer Sammlung von verschiedenen *Vibrio parahaemolyticus* Isolaten, welche sowohl von Patienten als auch von Schalentieren, insbesondere Austern, stammten. Diese Arbeit sollte neue Erkenntnisse zum Virulenzmechanismus von *V. parahaemolyticus* liefern. Am Virulenzmechanismus beteiligte Virulenzfaktoren sollten eine Differenzierung von Patienten- und Schalentierisolaten und von zytotoxischen und nicht-zytotoxischen Isolaten ermöglichen. Grundlage dieser Forschungsarbeit waren Veröffentlichungen, die den Einfluss der existierenden Pathogenitätsmarker *tdh*, *trh* und des T3SS als einzige einflussnehmende Faktoren im *V. parahaemolyticus* Virulenzmechanismus in Frage stellten.

Zu Beginn der Arbeit wurde für die jeweiligen vorhandenen Stämme ein biochemisches und serologisches Profil sowie ein Virulenzgenprofil der bislang bekannten Pathogenitätsfaktoren *tdh*, *trh* und T3SS erstellt. Dadurch konnte zum ersten Mal die Anwesenheit des Virulenzfaktors T3SS2β in *tdh*$^+$/*trh*$^+$ Isolaten nachgewiesen werden. Von T3SS2β wurde bislang angenommen, dass dieser nicht in Isolaten dieses Genotyps vorkommt. Diese Untersuchungen zeigten deutlich, dass die Pathogenität viel komplexer ist als angenommen. Aus diesen Untersuchungen ging auch hervor, dass 25% der klinischen *tdh*$^-$/*trh*$^-$ Isolate kein T3SS2 besaßen und somit automatisch dadurch kein Virulenzpotential aufzeigten.

Siebzehn von insgesamt 35 verschiedene Serotypen wurden ausschließlich in klinischen Isolaten und nicht auch in Austernisolaten gefunden. Die Vielfalt an unterschiedlichen Serotypen war möglicherweise zu groß, um als Pathogenitätsindikator zu dienen. Im weiteren Verlauf dieser Forschungsarbeit wurden Typisierungsmethoden wie PFGE (*pulsed-field gel electrophoresis*), ISR- 1 (*intergenic spacer region analysis*), DGREA (*direct genome restriction enzyme analysis*), MLST (*multi-locus sequence typing*) und HRM-MLVA (*high resolution melt-multiple-locus variable-tandem repeat analysis*) angewandt. Alle Methoden verwenden DNA als Target und wurden bereits erfolgreich für epidemiologische Experimente

und für phylogenetische Charakterisierungen eingesetzt. Die Anwendbarkeit dieser Methoden (mit Ausnahme von DGREA) konnte somit erneut mit dieser Arbeit belegt werden. Mit Hilfe von DGREA können Bakterienisolate grundsätzlich differenziert werden, jedoch waren bei der vorliegenden Studie fast 20% der Isolate mittels dieser Methode nicht typisierbar. Demnach ist DGREA keine verlässliche Methode für die Typisierung von *V. parahaemolyticus* Isolaten. Mit Hilfe von ISR-1 hingegen konnten Patienten- und Schalentierisolate unterschieden werden. Nachdem diese Ergebnisse zusätzlich mit einer detaillierten SNP-Analyse der genomischen Sequenzen verglichen wurden, konnten jedoch keine Zusammenhänge der Ergebnisse festgestellt werden. Auch PFGE führt zu einer starken Differenzierung; diverse Isolate produzierten spezifische Banden-Muster im Agarosegel. Zusätzlich gruppierten sich die Isolate basierend auf ihrem Serotyp, was die Verwendung einer Serotypisierung als eine Voruntersuchung ermöglicht. Darüber hinaus zeigte eine Kombination aus PFGE, MLST und HRM-MLVA ein großes Potential für Ausbruchs- oder Evolutionsuntersuchungen. Mithilfe der modifizierten HRM-MLVA-Methode konnten Isolate mit identischem PFGE Muster oder ST unterschieden werden. MLST allein konnte bereits die Identität und die phylogenetischen Beziehungen der *V. parahaemolyticus* Isolate definieren. Durch die Einbeziehung von *Whole-Genome*-Sequenzen konnte die MLST-Analyse zudem auf eine *in silico* Analyse ausgeweitet werden. Es wurden 61 Sequenztypen identifiziert. Auch wenn MLST vermehrt als Methode für Ausbruch- und Evolutionsuntersuchungen angewendet wird, ist es keine geeignete Methode für Virulenzindikatoren. Aufgrund fehlerhafter automatisierter Annotierungen und/oder Assemblierung von *Whole-Genome*-Sequenzdaten ist es zurzeit nicht oder nur schwer möglich, genetische Unterschiede auf dem Spezieslevel in *V. parahaemolyticus* zu erkennen. Anhand der kSNP Matrix konnten jedoch Unterschiede zwischen Patienten- und Schalentierisolaten beobachtet werden. Diese Cluster-Bildung kann als hilfreiche Basis bei der Suche nach einem potenziellen neuen Pathogenitätsfaktor dienen. Des Weiteren wurden Zytotoxizitätsassays für die Zelllinien HeLa und Caco-2 entwickelt. Nach statistischer Analyse erwiesen sich die Caco-2 Zellen als die geeignete Zelllinie für

Virulenzuntersuchungen. Anhand der Ergebnisse des Zytotoxizitätsassays und dem Genotyp jedes Isolates ergab sich eine teilweise Assoziierung von Zytotoxizität und *tdh* sowie dem T3SS2. Allerdings verdeutlichen die Ergebnisse dieser Zellassays erneut, dass *tdh*, *trh* und das T3SS möglicherweise nicht die einzigen Marker für die Virulenz eines Bakterienstammes sind. Auch die Ergebnisse des KAP-RPLA (Kanagawa phenomenon detection kit by reversed passive latex agglutination) Kits zur Bestimmung der TDH (thermostable direct hemolysin) Produktion festigten diese Schlussfolgerung. Isolate mit dem Serotypen O4:K12 und O4:Kut, welche mit einem Ausbruchstamm in Verbindung gebracht werden, scheinen dagegen ein höheres Virulenzpotential zu besitzen. Im Allgemeinen korreliert der Serotyp, hingegen nicht mit den Zytotoxizitätswerten. Dies ist ein weiterer Indikator dafür, dass der Serotyp ein wichtiges Screeningwerkzeug darstellt, jedoch allein kein Faktor für die Virulenz eines Stammes ist. Unter Beachtung des Phänotypes konnte eine Differenzierung von zytotoxischen und nichtzytotoxischen Isolaten (basierend auf der phylogenetischen Analyse der verknüpften Sequenzen) beobachtet werden. Obwohl keine spezifischen Gene in zytotoxischen Isolaten gefunden wurden, die nicht auch in nicht-zytotoxischen Isolaten vorhanden waren, konnten Gengruppen mit dem Phänotyp korreliert werden. Zusammenfassend stellt die Heterogenität von *V. parahaemolyticus* eine besondere Herausforderung bei Untersuchungen zum zugrundeliegenden Pathogenitätsmechanismus dar.

Ein Vergleich von Geno- und Phänotyp einer so großen Menge an Isolaten von klinischen Proben und Umweltproben aller Küstenregionen der Vereinigten Staaten, wurde bis heute noch nicht erstellt. Die Gene der funktionellen Gruppen des zytotoxischen Phänotyps könnten als zusätzliche Virulenzmarker fungieren. Des Weiteren könnten die identifizierten SNPs in zytotoxischen Stämmen zur Aufklärung des pathogenen Potenzials von *V. parahaemolyticus* beitragen. Whole Genome Sequenzierung und die Entwicklung des Zytotoxizitätsassays erweitern zudem den existierenden Katalog an Typisierungs- und Phänotypisierungsmethoden für *V. parahaemolyticus*.

3. Background

3.1. Bivalve molluscan shellfish

Bivalve molluscan shellfish, or the class of Bivalvia, are animals with two shell valves hinged by an elastic ligament, such as mussels, oysters, and clams. They belong to the phylum Mollusca, which are soft bodied animals surrounded by a protective shell (Gosling, 2008). Bivalve shellfish can be found in a range of aquatic habitats, saltwater and freshwater ecosystems (Potasman et al., 2002). Oysters belong to the family of *Ostreidae*. There are twelve different species of living oysters; three of these species are of most interest to this thesis. *Crassostrea virginica* is the Eastern Oyster native to the Atlantic and Gulf Coast of the United States. *Crassostrea gigas* has its natural habitat in the Pacific coast of Asia, which leads to its colloquial name Pacific or Japanese Oyster. They also occur in the United States. Additively, the oyster native to Europe is of the species *Ostrea edulis* and occurs across the whole European coastline (Galtstoff, 1964). However, through growth in tray cultures in the state of Maine, the European oyster is also present in the United States (Menzel, 1991). The natural habitat of oysters is in the nutrient-rich, shallow coastal waters (Menzel, 1991). They grow attached to rocks, underwater structures, or buried in sediment. These oyster populations are connected to each other in shell-to-shell forming oyster beds, banks or reefs (Galtstoff, 1964).

Another important member of bivalve molluscan shellfish is the hard clam. Hard clams belong to the family of *Veneridae*. *Mercenaria mercenaria* is one of the four clams mainly harvested in the United States: hard clams, ocean quahogs (*Artica islandica*), softshell clams (*Mya arenaria*), and surf clams (*Spisula solidissima*) (Eversole, 1987). Hard clams mainly occur on the Eastern coast of the United States. They grow usually in groups or beds in sand or muddy sand sediments of coastal areas (Kraeuter and Castagna, 2001). In general, clams prefer a salty environment with an optimal salinity from 24 – 28 ppt (Eversole, 1987).

Bivalve molluscs are filter feeders and therefore pass large volumes of water across their gills to obtain oxygen and food. Many microorganisms ingested via this feeding survive the digestive process (Ward and Hackney, 1991).

3.2. Foodborne bacterial infection

Foodborne bacterial infections are generally caused by invasive microorganisms entering the human body and spreading from the inside (Krämer, 2007). Foodborne pathogenicity can be subdivided into three forms: foodborne infection, foodborne intoxication, and foodborne toxicoinfection (Bhunia, 2007). By entering the human body, living microorganisms can cause a foodborne infection. Infections are caused when bacteria colonize the intestine and cross the intestinal barrier after intake of food or water harboring the organism(s). From there, microorganisms can cause local tissue damage concomitant with inflammation or spread to extraintestinal sites (Krämer, 2007). In case of foodborne intoxication, the bacteria produce toxins in food which contacts the gastrointestinal tract by food consumption. The toxins can cause inflammation or reach other organs or tissues in the human body (Krämer, 2007). Other bacteria can colonize the intestinal surface and produce exotoxins after consumption of food (foodborne toxicoinfection), which can either damage local cells or tissues, or enter the blood stream and cause disease there (Bhunia, 2007).

Across the world, approximately 600 million cases of foodborne illness were reported in 2010 (WHO, 2015). According to CDC, each year one in six Americans gets sick by consuming contaminated foods and beverages corresponding to 48 million illnesses, 128,000 hospitalizations, and 3,000 deaths (Johnson et al., 2014). The leading cause of foodborne illnesses is norovirus, followed by *Listeria*, *Salmonella,* and Shiga toxin-producing *Escherichia coli* (STEC) O157 (Johnson et al., 2014). In Europe, *Campylobacter* leads the list of foodborne illnesses ahead of *Salmonella*, *Listeria,* and verocytotoxigenic *E. coli* (VTEC) infections. A total of 55,453 cases from foodborne outbreaks were reported in the European Union in 2012 with 5,118 hospitalizations and 41 deaths (European Food Safety

Authority and Control). For public health support and prevention of foodborne infections, risk analysis is a necessity. Microbiological risk analysis is a process consisting of three components: Risk assessment, risk management, and risk communication, with the overall goal to ensure public health protection. Risk Assessment is defined as a scientifically based process consisting of hazard identification, hazard characterization, exposure assessment, and risk characterization (World Health Organisation, 2009). Studies screen for the potential presence of bacterial pathogens in the environment and foods and correlate that information to probability-of-illness models for evaluation of risk (Dickinson et al., 2013; Ebel and Williams, 2015). Microbial risk assessment can serve as a useful tool for the management of risk related to bacterial pathogens. To prevent microbial contamination the World Health Organization (WHO), Food and Drug Administration (FDA), Centers for Disease Control and Prevention (CDC) and other public health institutions have communicated to the public the risks of not adhering to safe food preparation. It is of great importance to keep surfaces, as well as equipment used for food preparation, and hands, clean and sanitized. Food preparation should always include safe water and washing fruits and vegetables prior to handling. Raw foods such as raw meats, poultry, and seafood should be separated from other food items and require thorough cooking to ensure safety. Cooked and perishable foods need proper refrigeration, according to their shelf life (WHO, 2006).

The quality of coastal and estuarine waters based on environmental conditions and pollution has a big impact on the level of pathogenic bacteria. Most shellfish-borne diseases are caused by enteric viruses, pathogenic vibrios, as well as fecal-borne bacteria. Bivalve molluscan shellfish, as filter-feeding organisms which can accumulate pathogens from the contaminated estuarine waters, may present a health risk when consumed raw or only lightly cooked (Malham et al., 2014). Across the world, infectious disease outbreaks with bivalve molluscs as a food vehicle have increased (Potasman et al., 2002). Therefore, microbial risk assessment can serve as a useful tool for the risk management of marine pathogens.

3.3. Genus Vibrio

Vibrios are species of the genus *Vibrio* belonging to the family of *Vibrionaceae*. They are gram-negative, facultatively anaerobic, mesophilic, rod-shaped bacteria (West, 1989). Their natural habitat is in marine and estuarine environments across the globe. For growth, they prefer waters with increased salinity and temperatures between 10˚C and 30˚C (Murray et al., 2002). Vibrios can also appear in and/or on marine organisms, i.e. corals, fish, molluscan shellfish, seagrass, sponges, shrimp, and zooplankton (Thompson et al., 2004).

Members of the genus *Vibrio* total more than 100 species; twelve of them are documented as pathogenic to humans: *Vibrio cholerae*, *Vibrio vulnificus*, *V. parahaemolyticus*, *Vibrio alginolyticus*, *Grimontia hollisae* (formerly *V. hollisae*), *Vibrio fluvialis*, *Vibrio furnissii*, *Photobacterium damsela* (formerly *V. damsela*), *Vibrio mimicus*, *Vibrio metschnikovii*, *Vibrio cincinnatiensis* and *Vibrio carchariae* (Murray et al., 2002). The clinical conditions of a *Vibrio* infection, also referred to as vibriosis, vary in severity and can range from mild cases of gastroenteritis to septicemia and invasive skin and soft tissue infections (Janda et al., 2015). However, *Vibrio* species have been isolated from numerous other anatomic sites such as ear, eye, gallbladder, sinuses, and peritoneal fluid (Janda et al., 2015).

Diseases with symptoms comparable to a *V. cholerae* infection can be tracked back to 400 BC. The Italian physician Filippo Pacini discovered the first *Vibrio* species, *V. cholerae*, while studying outbreaks of this disease in 1854 in Florence. At the same time the English medical doctor, John Snow, discovered the source of *V. cholerae* infections in polluted drinking water during investigation of the epidemiology of cholera in several cities of England after the infectious disease had killed tens of thousands of people (Barua, 1992). Almost 30 years later, Robert Koch was able to obtain pure cultures of *V. cholerae* on gelatin plates after examining several outbreaks in Egypt and India. Thereby, Koch and his team discovered the typical comma shape and motility of the microorganism (Brock, 1999). During an outbreak of cholera in 1893 in Hamburg, Germany, Koch proposed that water supply systems should include filtration of drinking water in order to remove the infection-causing bacteria. Concurrently, Koch discovered that vibrios were ubiquitous in aquatic environments and

some *Vibrio* species were non-pathogenic to humans (Brock, 1999). The first nonpathogenic *Vibrio* species such as *V. fischeri*, *V. splendidus*, and *Photobacterium phosphoreum* were isolated from the aquatic environment by the Dutch microbiologist Martinus Beijerinck in the late 1880s (Brock, 1999).

In Germany, non-cholera *Vibrio* species are detected in the North and Baltic Sea, especially during the warmer summers (Breidenbach and Frank, 2012). Although *Vibrio* infections rarely occur in Germany, since strains usually lack pathogenic potential and/or a specific notification process for non-cholera *Vibrio* infections is lacking, due to climate change and warming of the marine environment the presence of *Vibrio* species might increase and, therefore, the risk of infection (Huehn et al., 2014).

In the United States, at least 8000 *Vibrio* infections occur annually; about 75% originate from food sources (Dechet et al., 2008). Based on epidemiological data, foodborne vibriosis is increasing in the United States (Crim et al., 2015). Current data suggests that various conditions associated with *Vibrio* infections are either misdiagnosed or not considered as part of the diagnosis during initial presentation with accompanying symptoms. Estimated vibriosis is much higher than current laboratory-confirmed cases (Dechet et al., 2008; Newton et al., 2012). Furthermore, rising sea temperatures have been correlated to higher levels of *Vibrio* species along the east coast of the United States (Banakar et al., 2011).

3.3.1. *Vibrio parahaemolyticus*

In 1950 a food poisoning outbreak occurred from consumption of shirasu, small half-dried sardines, in Osaka, Japan. From outbreak samples a new bacterium was isolated, which was named *Pasteurella parahaemolytica*. A couple of years later this name was changed to *Vibrio parahaemolyticus* due to its similar motility of *Vibrio cholerae* and halophilic characteristics (Shinoda, 2011). As a species of the genus *Vibrio, V. parahaemolyticus* is a gram-negative, non-spore forming, rod-shaped curved bacterium and its motility is attributed to polar and lateral flagella (Garrity et al., 2007). The bacterium's natural habitat is costal and estuarine

environment and, therefore, can be present in bivalve molluscan shellfish (Kueh and Chan, 1985).

Through consumption of *V. parahaemolyticus*-containing raw or undercooked seafood, especially molluscan shellfish, this bacterium can cause acute gastroenteritis including symptoms such as diarrhea, headache, vomiting, nausea, abdominal cramps, and a low fever (Powell, 1999). Also, open wounds or cuts, which contact water harboring the organism, or injuries from cutting fish or seafood, can become infected by *V. parahaemolyticus*. These wound infections can cause subdermal damage and necrosis of the tissue. Both wound infections as well as gastroenteritis, can lead to septicemia. Immunocompromised patients are at a higher risk of developing severe illness due to infection with *V. parahaemolyticus*, including possible death (Breidenbach and Frank, 2012).

The annual estimate of *V. parahaemolyticus* infections in the United States is 35,000 (Scallan et al., 2011). The organism can multiple quickly in oysters in warmer temperatures. Therefore, even low levels of *V. parahaemolyticus* at harvest can be of risk upon exposure to elevated temperatures (Depaola et al., 2000). In Japan, *V. parahaemolyticus* infections have been a dominant cause of foodborne infections since the 1960s. However, from 1999 to 2001 the Ministry of Health, Labour, and Welfare (MHLW) of Japan established regulations for seafood safety from the production stage to the consumer and lead to a decrease of *V. parahaemolyticus* infections (Morris, 2014). In Europe, *V. parahaemolyticus* infections are less frequent. Sporadic outbreaks occurred in Spain and France in the last twenty years (Martinez-Urtaza et al., 2005; Su and Liu, 2007).

3.3.2. Virulence factors

3.3.2.1. Thermostable direct hemolysin (TDH)/TDH-related hemolysin

In 1969 a thermolabile hemolysin (TLH) was discovered in *V. parahaemolyticus*. All *V. parahaemolyticus* isolates produce TLH, which can be inactivated by heating to 60˚C (Fujino et al., 1969; Taniguchi et al., 1986). Furthermore, clinical *V. parahaemolyticus* isolates can

initiate β-type hemolysis on Wagatsuma agar, a special high-salted blood agar. This occurrence is called Kanagawa phenomenon (KP). KP is caused by the thermostable direct hemolysin (TDH) (Sakazaki et al., 1968; Sakurai et al., 1973). The KP test was used as an indicator for virulence, since most environmental samples were lacking this β-hemolysin activity (Sakazaki et al., 1968; Nishibuchi and Kaper, 1995). In 1988 *V. parahaemolyticus* isolates from patients with diarrhea were confirmed as KP- and TDH-negative; those strains produced a TDH-related hemolysin (TRH) (Honda et al., 1988) (Figure 1). A study of a clinical strain TH3766 revealed the simultaneous production of TDH and TRH (Xu et al., 1994). Since, TDH and TRH have been considered the major virulence markers/toxins of *V. parahaemolyticus*; clinical strains predominantly possess TDH, TRH, or both proteins.

TDH cannot be inactivated by heating at temperatures between 80 and 100˚C (Takeda et al, 1974). TDH is a tetramer in solution and forms a central pore. This pore is responsible for membrane disruption (Yanagihara et al., 2010). Moreover, TDH functions as an amyloid toxin. It can change its confirmation in lipid vesicles and can therefore damage membranes. It remains unknown, how this affects TDH hemolytic activity (Fukui et al., 2005). In comparison of the membrane toxicity of TDH and TRH, they both share the same hemolytic activity, whereas TRH has less amyloidogenicity. Therefore, the membrane disruption is more likely caused by the tetrameric pore than by amyloidogenicity (Yanagihara et al., 2010). The lipid composition of TDH's target cells plasma membrane consists of cholesterol- and sphingolipid-enriched microdomains, also known as "lipid rafts". Cholesterol and sphingomyelin are associated with cytotoxicity of TDH, while the hemolytic activity is independent from lipid rafts. The reduction of the lipid raft content interferes with cytotoxicity (Matsuda et al., 2010).

TDH and TRH are encoded by the *tdh* and *trh* genes, respectively (Makino et al., 2003). The sequence of *Vibrio parahaemolyticus* RIMD2210633 has two copies of the *tdh* gene: *vpa1314* (*tdhA*) and *vpa 1378* (*tdhS*) (Iida and Yamamoto, 1990; Honda et al., 1991). The *trh* gene is divided into two sequence variants, *trh1* and *trh2* (Kishishita et al., 1992). Comparative analysis showed *tdh* and *trh* have about 70% homology to each other

(Nishibuchi et al., 1989). A previous study reported that TDH induces activation of the NLRP3 inflammasome in tissue culture. Pore-forming toxins have been responsible for this phenomenon in other bacteria and can cause caspase-1 activation (Higa et al., 2013). However, the deletion of both *tdh* genes had no effect on cytotoxicity and enterotoxicity in cell culture studies (Park et al., 2004a; Lynch et al., 2005). Additionally, some clinical isolates do not possess the *tdh* or *trh* gene. Thus, it has been suggested *tdh* and *trh* have limited influence on cytotoxicity (Lynch et al., 2005). This could indicate the existence of additional virulence markers.

3.3.2.2. Type-III-Secretion-System

Sequencing of the *V. parahaemolyticus* strain RIMD2210633 revealed the presence of two sets of genes, referred to as the Type-III-Secretion-Systems 1 and 2 (T3SS1 and T3SS2). The role of the T3SS is to recognize and secret proteins into the host cell via needle-like machinery (Galan and Wolf-Watz, 2006). The homologues of the T3SS1 genes are also present in other *Vibrio* species such as *V. harveyi* and *V. alginolyticus* (Park et al., 2004b). The T3SS1 is present in all *V. parahaemolyticus* isolates, clinical and environmental, and plays an important role in survival in the environment (Zhang and Orth, 2013). Furthermore, the T3SS2 presents two variants: T3SS2α and T3SS2β. T3SS2α corresponds with *tdh*$^+$/*trh*$^-$ isolates and T3SS2β with *tdh*$^-$/*trh*$^+$ isolates (Park et al., 2004b).

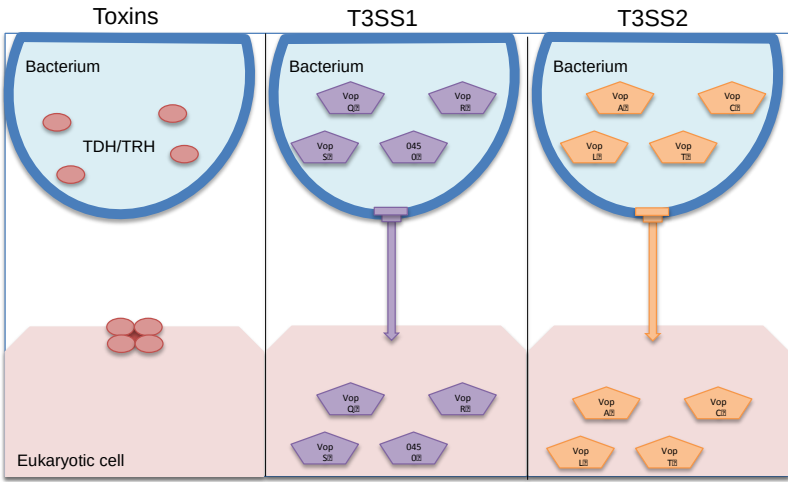


Figure 1: Illustration of the three major virulence factors (Broberg et al., 2011)

The T3SS1 is present in all *V. parahaemolyticus* isolates, clinical and environmental, and thus the question was raised how T3SS1 contributes to virulence (Park et al., 2004b; Okada et al., 2010). According to previous studies, the T3SS1 is involved in cytotoxicity to HeLa cells (Park et al., 2004b; Ono et al., 2006) (Figure 1). T3SS1 includes effector proteins, which can induce cell lysis and release of nutrients. The three recently identified effectors are VopQ, VopS, and VPA0450 (Broberg et al., 2011). HeLa cells were infected with these effectors of the T3SS1, which led to autophagy, followed by rounding of the host cell due to changes in the cytoskeleton (Broberg et al., 2011). Hence, cellular contents were released out of the host cell. VopQ (encoded by VP1680) can induce a P13-kinase independent autophagy (Burdette et al., 2009); VopS (VP1686) is responsible for rounding cells by targeting the actin cytoskeleton and mediating AMP addition to GTPases (Yarbrough et al., 2009). A recent study revealed that the effectors VopQ and VopS also induce autophagy and Cdc42, a cell division control protein, inactivation, respectively. Associated with those processes the inflammasome NLRC4 was activated (Higa et al., 2013). However, it is still unclear, how effectors of the T3SS1 work together to induce cytotoxicity, or in the overall virulence mechanism (Burdette et al., 2008; Burdette et al., 2009).

The T3SS2 is mostly associated with enterotoxicity (Figure 1). Animal studies demonstrated that T3SS2 effectors cause fluid accumulation and inflammation in the intestine and diarrhea. Many effectors of the T3SS2 also manipulate signaling pathways and actin cytoskeleton organization by targeting Mitogen-Activated Protein Kinase (MAPK), small Rho GTPases, and F-actin (Ham and Orth, 2012). Responsible for the activation of GTPase is the effector VopC, which is encoded by the gene VPA1321 and functions as a deamidase. Therefore GTPase is able to alter the actin cytoskeleton allowing *V. parahaemolyticus* to enter a non-phagocytic host cell (Zhang et al., 2012). Another effector putting stress on the actin cytoskeleton is VopL. VopL is encoded by the gene VPA1370 and targets actin as a nucleate. By disturbing actin, VopL serves as a nucleation factor for the actin filament polymerization (Liverman et al., 2007). The effector VopA/VopP, encoded by the gene *VPA1346*, acetylates serine and threonine residues of MAPKs. Thus, the kinase inactivation

is enabled (Trosky et al., 2004; Trosky et al., 2007). Another effector being studied is VopT. VopT is encoded by the VPA1327 gene and contains an ADP-ribosyltransferase. This enzyme modifies the G-protein Ras which eventually induces cytotoxicity; the entire mechanism is still not fully understood (Ham and Orth, 2012). As the presence of TDH had no influence on fluid accumulation, the presence of T3SS2 in the rabbit ileal loop test was examined and revealed an important impact on enterotoxicity (Hiyoshi et al., 2010). The important effector for enterotoxicity in the rabbit ileal loop was determined to be VopV (encoded by VPA*1357*). By binding of VopV to F-actin, these actin polymers accumulated *in vivo*. How enterotoxicity is affected by this phenomenon of F-actin polymer accumulation still remains unknown (Hiyoshi et al., 2011). A recent study showed environmental strains of *V. parahaemolyticus* possess T3SS2, but lack *tdh* and *trh* (Caburlotto et al., 2010).

3.3.2.3. Other virulence factors

V. parahaemolyticus can possess additional contributing virulence factors. The organism synthesizes three major surface antigens: O antigens (Lipopolysaccharide [LPS]), capsular K antigens (capsular polysaccharides [CPS]) and H antigens (flagellar antigens). The CPS types correlate with the ability to adhere to intestinal cells. Additionally, they can induce mucosal and systemic immune responses (Hsieh et al., 2003). The LPS consists of lipid A and core oligosaccharides as the determinant of the O-serotype (Han and Chai, 1992; Iguchi et al., 1995). This additional layer of lipids functions as an extra defense barrier (Neyen and Lemaitre, 2016). The serotyping scheme for *V. parahaemolyticus* includes 11 different O antigens and 71 different K types (Nair et al., 2007). The serotypes O3:K6, O4:K68, O4:K12, O1:K25, O1:K41, and O1:K untypeable (Kut) are considered the pandemic serogroup of *V. parahaemolyticus* (Chowdhury et al., 2000; Laohaprertthisan et al., 2003; Chowdhury et al., 2004a).

Beyond the layer of LPS the cell wall, a double layer of DAP-like petidoglycans (PGN), is located and thereby protected from the immune response receptors of the host (Neyen and Lemaitre, 2016). In vertebrates the extracellular detection of LPS is controlled by soluble lipid

binding proteins (LBP) and bactericidal permeability-increasing proteins (BPI), which collect LPS from serum and deliver it to glycosylphosphatidylinositol (GPI)-anchored cluster of differentiation (CD) 14. Hereby it activates the toll-like receptor 4 with lymphocyte antigen 96 (TLR4/MD-2) complex, which triggers cytosolic sensor caspase 11 of the host (Neyen and Lemaitre, 2016).

V. parahaemolyticus has two different types of flagella (Figure 2). Flagella are a common factor in pathogenesis of infectious bacteria. In this case one, the polar flagellum, is used for swimming, while the second one, the lateral flagellum, is expressed on the swarming cell type (Broberg et al., 2011). The lateral flagellum is expressed in highly viscous media or on a surface. The regulation of lateral flagella is based on environmental factors. Using those flagella, *V. parahaemolyticus* is able to adhere to a host cell and to form biofilms (Park et al., 2005; Merino et al., 2006).

Many bacteria possess an outer membrane molecule, the multivalent adhesion molecule (MAM), involved in the initial binding to the host cell and is important for the activation and delivery of virulence factors. MAM7 is specific to gram-negative bacteria; therefore it is present in *V. parahaemolyticus* as well (Krachler et al., 2011; Krachler and Orth, 2011) (Figure 2). MAM7 is responsible for the initial attachment of the pathogen to the host cell and could, therefore, be an important trigger for early infection. The two initial receptors of MAM7, fibronectin and phosphatitic acid, have been identified. Fibronectin is the initial receptor with high affinity but weak binding; thus phosphatidic acid as a secondary receptor results in a high binding (Krachler et al., 2012). During later phases other adhesion factors probably play a more strain-specific role and strengthen the attachment. Additionally,

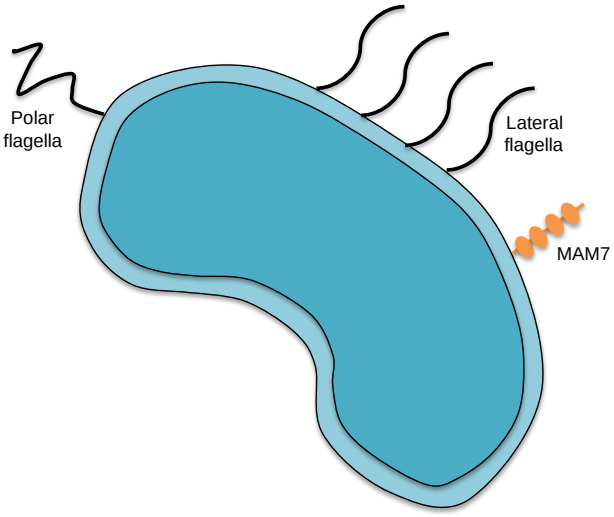


Figure 2: Illustration of other virulence factors of *V. parahaemolyticus* (following (Broberg et al., 2011)

MAM7 inhibits infection by other pathogens by blocking the host cell receptor. However, the exact cycle of adhesion is still widely unexplored and needs further investigation (Krachler and Orth, 2011).

4. Research objectives

The FDA has, as part of its mission, the goal to reduce the risk of foodborne illnesses at all points in the food chain. As an approach, FDA utilizes risk assessment principles at each step as food moves from growers and producers to the consumer. Therefore, it is of great importance to be able to detect bacterial presence. Concerning *V. parahaemolyticus,* microbiological and molecular methods are in place to detect this bacterium in food or water.

Not every *V. parahaemolyticus* strain is capable of infecting humans. For current risk analysis purposes, the pathogenicity of *V. parahaemolyticus* is based on the presence of two hemolysin genes, *tdh* and *trh*. However, recent studies have shown that *tdh* and *trh* are not present in about 30% of clinical isolates. Based on cell culture and animal experiments, these two hemolysins had a minimal effect on cytotoxicity and/or enterotoxicity. The question is therefore raised, whether other pathogenicity factor/s exist and how they influence the virulence potential of *V. parahaemolyticus*. For risk assessment purposes, a more definitive determination of the reliability of existing markers and/or identification of new markers would refine the evaluation of potential risk.

V. parahaemolyticus is a very diverse organism which complicates the search for novel pathogenicity markers. This research project focused on the characterization of a diverse set of *V. parahaemolyticus* isolates. These strains were isolated from clinical and shellfish (oyster, clam) samples. By applying first fingerprinting and sequence based subtyping methods, these isolates are going to be characterized and put into related clusters. Following, next-generation sequencing, as well as cell culture experiments for phenotypic profiling, add to the techniques utilized to evaluate these isolates. Connecting the fields of genomics and phenotyping to compare clinical versus oyster isolates could lend insight into the virulence potential of *V. parahaemolyticus* and might prove the hypothesis of the need of an additional pathogenicity marker.

5. Results

The first part of this chapter describes the characterization of the environmental and clinical strains utilized in this work. Furthermore, this section focuses on examining the diversity and relations between the *V. parahaemolyticus* strains by using several subtyping and phylogenetic methods. The second part intensifies the search for a virulence marker. WGS covered the genomic part of this thesis and, is correlated with cell culture studies, which provides detail to the phenotype of the strains. In conclusion, combining the puzzle of subtyping, genomics, and phenotypic analysis lends insight into the relatedness of *V. parahaemolyticus* isolates.

5.1. Biochemical, serological, and virulence characterization of clinical and oyster *Vibrio parahaemolyticus* isolates

In the first section of this chapter the 144 isolates of this thesis were characterized with basic molecular methods such as real-time polymerase chain reaction (PCR) specific for the genes *tlh*, *tdh*, *trh* and conventional PCR for the T3SSs. Additionally, all isolates were serotyped and were subjected to biochemical tests via API20 test strips. Twenty-eight of 67 (41.8%) oyster isolates and 42 of 77 (54.6%) clinical isolates were identified by the API 20E test as *V. parahaemolyticus*. Most frequently, the misidentified isolates gave API codes for *V. vulnificus* or *Aeromonas hydrophila*; but identifications of *V. fluvialis*, *V. cholerae*, and *V. mimicus* were also made. Table 1 lists the biochemical properties of the clinical and oyster isolates. All *V. parahaemolyticus* isolates were positive for oxidase, indole, and glucose fermentation. Only two of the oyster isolates (3%) were sucrose positive, and one clinical isolate (1.3%) was VP positive, both unusual traits for this organism. *V. parahaemolyticus* is generally considered to be o-nitrophenyl-β-D-galactopyranoside (ONPG) negative (no β-galactosidase production) (Bryant et al., 1986; Kaysner and Depaola, 2004), but 51% of the isolates were positive. This one test was

responsible for the majority (>90%) of the misidentifications by the API 20E system, indicating that the isolates otherwise produced biochemical profiles typical of *V. parahaemolyticus* (Bryant et al., 1986). No traits distinguishing between the biochemical profiles of the clinical and oyster isolates were observed.

Table 1: Biochemical properties of clinical and oyster *V. parahaemolyticus* strains examined in this chapter *

	Clinical Isolates		Oyster Isolates	
β-galactosidase	V	46%	V	57%
Arginine dihydrolase	-	99%	-	99%
Lysine decarboxylase	+	96%	+	100%
Ornithine decarboxylase	+	84%	+	100%
Citrate utilization	-	81%	-	93%
H_2S production	-	100%	-	99%
Urease	V	58%	V	73%
Tryptophane deaminase	-	100%	-	100%
Indole production	+	100%	+	100%
Voges Proskauer (acetoin production)	-	99%	-	100%
Gelatinase	+	96%	+	100%
Glucose fermentation	+	100%	+	100%
Mannitol fermentation	+	100%	+	99%
Inositol fermentation	-	99%	-	100%
Sorbitol fermentation	-	97%	-	100%
Rhamnose fermentation	-	99%	-	100%
Saccharose fermentation	-	100%	-	97%
Melibiose fermentation	-	100%	-	90%
Amygdalin fermentation	+	81%	+	91%
Arabinose fermentation	+	92%	+	93%
Cytochrome oxidase	+	100%	+	100%
Sucrose utilization	-	100%	-	97%

*Symbols: +, positive trait for greater than 80% of isolates (exact percentage given in adjacent column); -, negative trait for at least 80% of isolates (exact percentage given in adjacent column); V, variable trait (percentage of positive isolates given in adjacent column).

Among the 144 isolates tested, 35 serotypes were identified (Table 2). There were representatives of all but three (O types 2, 7, and 9) of the known O types, but nearly half (49.3%) of the isolates were untypeable for the K antigen. Fifteen isolates (10.4%) had unique serotypes within this study. Only nine of the serotypes were shared by clinical and oyster isolates (O1:Kut, O1:K20, O3:Kut, O4:Kut, O4:K8, O4:K9, O5:Kut,

O10:Kut, and O11:Kut). Thirteen serotypes were found only in clinical isolates (O1:K33, O1:K56, O3:K39, O3:K56, O4:K4, O4:K13, O4:K53, O4:K63, O5:K17, O5:K30, O5:K47, O6:K18, and O8:K41), and nine serotypes were found only in oyster isolates (O1:K43, O3:K5, O4:K10, O4:K34, O4:K37, O4:K42, O6:Kut, O8:Kut, and O8:K70). O1:Kut was the dominant serotype overall, with 18 isolates (12.5%) from clinical and oyster sources, geographically distributed. O1:Kut is one of the serogroups associated with pandemic *V. parahaemolyticus* strains (Chowdhury et al., 2000; Chowdhury et al., 2004b); of the remaining pandemic serotypes, four O3:K6, one O4:K68, and one O1:K25 isolate were identified among the clinical isolates but none of the oyster isolates. The prevalence of pandemic serotypes was higher among clinical isolates (22.1%) than among oyster isolates (10.4%), but the difference was only marginally significant (P, < 0.10; 95% confidence interval, -1.6%, 25%).

Table 2: Distribution of serotypes for all 144 *V. parahaemolyticus* isolates based on isolation source and location .

Sero-type	Clinical isolates			Oyster isolates		
	n	Source (n)	Reporting state (n)	n	Harvest month (n)	Harvest state (n)
O1:Kut	11	Stool (9), Wound (1), Other (1)	WA (4), MA (2), LA (1), NV (1), OK (1), OR (1), NY (1)	7	Mar. (3), May (2), Jul. (2)	FL (4), ME (2), LA (1)
O1:K20	2	Stool (2)	VA (2)	6	Nov. (3)	SC (3)
O1:K25	1	Wound (1)	LA (1)	0	---	---
O1:K33	2	Stool (1), Blood (1)	LA (1), VA (1)	0	---	---
O1:K43	0	---	---	2	Jul. (1)	ME (1)
O1:K56	1	Stool (1)	SD (1)	0	---	---
O3:Kut	3	Stool (1), Wound (1), Other (1)	MS (2), LA (1)	9	Jul. (6), Jan. (1), May (1), Nov. (1)	WA (6), LA (1), SC (1), TX (1)
O3:K5	0	---	---	18	Jul. (9)	Canada (9)
O3:K6	4	Stool (4)	GA (1), NY (2), TX (1)	0	---	---
O3:K39	1	Wound (1)	LA (1)	0	---	---
O3:K56	1	Stool (1)	MD (1)	0	---	---
O4:Kut	6	Stool (6)	WA (5), NY (1)	2	Apr. (1), Jul. (1)	FL (1), ME (1)
O4:K4	1	Stool (1)	HI (1)	0	---	---
O4:K8	2	Stool (2)	GA (1), WA (1)	3	Mar. (1), Apr. (1), Jun. (1)	AL (1), LA (1), NJ (1)
O4:K9	1	Stool (1)	GA (1)	2	Apr. (1)	FL (1)
O4:K10	0	---	--	2	Mar. (1)	LA (1)
O4:K12	11	Stool (11)	WA (3), IA (2), IN (1), MD (1), NY (1), NV (1), OK (1), PA (1)	0	---	---
O4:K13	1	Stool (1)	GA (1)	0	---	---

Table 2: *continued*

Sero-type	Clinical isolates			Oyster isolates		
	n	Source (n)	Reporting state (n)	n	Harvest month (n)	Harvest state (n)
O4:K34	0	---	---	2	Aug. (1)	VA (1)
O4:K37	0	---	---	2	Oct. (1)	FL (1)
O4:K42	0	---	---	2	Oct. (1)	FL (1)
O4:K53	3	Stool (3)	MA (2), NY (1)	0	---	---
O4:K63	2	Stool (2)	AK (1), IN (1)	0	---	---
O4:K68	1	Stool (1)	GA (1)	0	---	---
O5:Kut	8	Other (5), Stool (3)	TX (3), HI (2), VA (2), ME (1)	8	Nov. (4), Jul. (3), May (1)	FL (4), ME (2), TX (1), WA (1)
O5:K17	3	Other (2), Stool (1)	VA (2), HI (1)	0	---	---
O5:K30	1	Wound (1)	MD (1)	0	---	---
O5:K47	1	Stool (1)	MD (1)	0	---	---
O6:Kut	0	---	---	1	Aug. (1)	FL (1)
O6:K18	2	Other (2)	HI (1), NC (1)	0	---	---
O8:Kut	0	---	---	4	Aug. (3), Jul. (1)	VA (3), AL (1)
O8:K41	2	Stool (2)	VA (2)	0	---	---
O8:K70	0	---	---	2	Aug. (2)	FL (2)
O10:Kut	4	Stool (3), Other (1)	NY (4)	8	Mar. (4), Jul. (2), May (1), Oct. (1)	TX (4), FL (1), LA (1), ME (1), WA (1)
O11:Kut	2	Stool (2)	NY (1), GA (1)	6	Oct. (3), Jul. (2), May (1)	FL (4), Canada (2)

Concerning the hemolysin gene profile, all 144 isolates tested were positive for the presence of the *tlh* gene, confirming their identity as *V. parahaemolyticus*. For this study, oyster isolates were preferentially selected for the presence of *tdh* and/or *trh*; however, all clinical isolates submitted to the CDC in 2007 were included. Twenty-one of 77 (27%) clinical isolates and 14 of 67 (21%) oyster isolates were negative for both *tdh* and *trh* (Table 3). Thirty-five (45.5%) clinical and 23 (34.3%) oyster isolates contained both the *tdh* and *trh* genes. Nine (11.7%) clinical and two (3.0%) oyster isolates were tdh^+/trh^-; 12 (15.6%) clinical and 28 (41.8%) oyster isolates were tdh^-/trh^+. The most common virulence genotype (45%) among clinical *V. parahaemolyticus* isolates was positivity for both *tdh* and *trh*; all but one of these strains were isolated from stool specimens. All nine of the tdh^+/trh^- clinical isolates were also isolated from stool specimens. Of the 12 tdh^-/trh^+ isolates, 8 (66.7%) were isolated from stool specimens. The 25 (37.3%) oyster isolates that were positive for *tdh* (whether positive or negative for *trh*) were from market oysters harvested between March and November from Gulf and Mid-Atlantic states. In contrast, the majority (22 of 28 [78.6%]) of tdh^-/trh^+ isolates came from oysters harvested from the

North Atlantic or Pacific Northwest during July.

Table 3: Distribution of hemolysin genotypes for all 144 *V. parahaemolyticus* isolates based on isolation source and location

Genotype	Clinical Isolates			Oyster Isolates		
	N	Isolation Source (n)	Reporting State (n)	N	Harvest Month (n)	Harvest State (n)
tdh$^+$, *trh*$^-$	9	Stool (9)	GA (3), MA (2), NY (2), TX (1), WA (1)	2	Oct. (2)	FL (2)
tdh$^-$, *trh*$^+$	12	Stool (8), Other (3), Wound (1)	WA (4), VA (2), LA (1), NY (1), ME (1), MD (1), TX (1), OK (1)	28	Jul. (22), Mar. (4), Jan. (1), May (1)	Canada (9), WA (7), ME (6), TX (6)
tdh$^+$, *trh*$^+$	35	Stool (34), Other (1)	WA (8), NY (7), VA (3), GA (3), IA (2), IN (2), MA (2), MD (2), NV (2), AK (1), OK (1), PA (1), SD (1),	23	Nov. (6), Aug. (5), Mar. (4), May (3), Apr. (2), Jun. (1), Jul. (1), Oct. (1),	FL (11), LA (3), SC (3), VA (3), AL (2), NJ (1)
tdh$^-$, *trh*$^-$	21	Stool (8), Other (8) Wound (4), Blood (1)	HI (5), LA (4), VA (4), MS (2), TX (2), MD (1), NC (1), NY (1), OK (1)	14	Jul. (4), May (2), Aug. (2), Oct. (2), Nov. (2), Mar. (1), Apr. (1),	FL (6), Canada (2), LA (2), ME (1), SC (1), VA (1), WA (1)

In the testing of isolates for T3SS1 genes by multiplex PCR, 17 isolates were missing at least one of the four genes (representative strains are shown in chapter 11.1.1 in Figure 25). These 17 isolates were retested for amplification of each of the missing genes by simplex PCR, and seven were found to contain the genes not amplified by the multiplex PCR (see chapter 11.1.1 Figure 26). The remaining ten isolates demonstrated weak amplification by simplex PCR, sometimes with a stronger product at a size other than that expected. Similar problems were observed in the application of the T3SS2α multiplex PCR, including weak amplification in many isolates (see chapter chapter 11.1.1 in Figure 28). In the two *tdh*$^+$/*trh*$^-$ oyster isolates, only three of the four genes were strongly amplified, even by simplex PCR (see chapter 11.1.1 Figure 27). Additionally, three oyster isolates showed strong amplification of VPA1335 (*vscS2*) when tested by simplex PCR. All but three of the clinical isolates (96%) contained all four T3SS1 genes (Table 4). Two (5.7%) clinical isolates positiv for both *tdh* and *trh* did not amplify with primers for VP1694 (*vscF*), and one (8.3%) *tdh*$^-$/*trh*$^+$

isolate was negative for VP1686. Only 60 of 67 (90%) oyster isolates amplified all four T3SS1 genes. Two isolates positive for both *tdh* and *trh*, one *tdh*$^-$/*trh*$^+$ isolate, and two isolates negative for both *tdh* and *trh* did not amplify VP1686; one *tdh*$^-$/*trh*$^+$ isolate and one isolate negative for both *tdh* and *trh* did not amplify VP1694 (*vscF*). All nine of the *tdh*$^+$/*trh*$^-$ clinical isolates contained all four of the T3SS2α genes. The two *tdh*$^+$/*trh*$^-$ oyster strains did not amplify VPA1362 (*vopB2*) but were positive for the other three genes. The observed difference in the prevalence of VPA1362 (*vopB2*) among clinical versus oyster *tdh*$^+$/*trh*$^-$ isolates was statistically significant (P, 0.018). Additionally, one oyster isolate was positive for both *tdh* and *trh*, one negative for *tdh* and positive for *trh*, and one negative for both *tdh* and *trh* amplified VAP1335 (*vscC2*) of the T3SS2α system. Overall, the observed association between the *tdh*$^+$/*trh*$^-$ genotype and the presence of T3SS2α system genes was statistically significant (P <0.0001).

All four of the T3SS2β genes were detected in all 12 clinical and 28 oyster isolates that were *trh*$^+$/*tdh*$^-$. Additionally all 35 clinical isolates and 22 of 23 (95.7%) oyster isolates positive for both *tdh* and *trh* amplified all four T3SS2β genes tested. The one remaining oyster isolate positive for both *tdh* and *trh* amplified all but *vopC*. Overall, the observed association between the *tdh*$^+$/*trh*$^+$ and *tdh*$^-$/*trh*$^+$ genotypes and the presence of T3SS2β system genes was statistically significant (P <0.0001).

Table 4: Distribution of T3SS genes for all 144 *V. parahaemolyticus* isolates based on isolation source and hemolysin genotype

		Clinical Isolates (n=77)				Oyster Isolates (n=67)			
		tdh$^+$/*trh*$^+$ (n=35)	*tdh*$^+$/*trh*$^-$ (n=9)	*tdh*$^-$/*trh*$^+$ (n=12)	*tdh*$^-$/*trh*$^-$ (n=21)	*tdh*$^+$/*trh*$^+$ (n=23)	*tdh*$^+$/*trh*$^-$ (n=2)	*tdh*$^-$/*trh*$^+$ (n=28)	*tdh*$^-$/*trh*$^-$ (n=14)
T3SS1	all 4 genes present	33	9	11	21	21	2	26	11
	VP1670 (*vscP*)	35	9	12	21	23	2	28	14
	VP1686 (putative)	35	9	11	21	21	2	27	12
	VP1689 (*vscK*)	35	9	12	21	23	2	28	14
	VP1694 (*vscF*)	33	9	12	21	23	2	27	13
T3SS2α	all 4 genes present	0	9	0	0	0	0	0	0
	VPA1362 (*vopB2*)	0	9	0	0	0	0	0	0
	VPA1339 (*vscC2*)	0	9	0	0	0	2	0	0
	VPA1335 (*vscS2*)	0	9	0	0	1	2	1	1
	VPA1327 (*vopT*)	0	9	0	0	0	2	0	0
T3SS2β	all 4 genes present	35	0	12	0	22	0	28	0
	vscC2	35	0	12	0	23	0	28	0
	vopB2	35	0	12	0	23	0	28	0
	vopC	35	0	12	0	22	0	28	0
	vscS2	35	0	12	0	23	0	28	0

5.2. Suitability of the molecular subtyping methods intergenic spacer region, direct genome restriction analysis, and pulsed-field gel electrophoresis for clinical and environmental *Vibrio parahaemolyticus* isolates

The objective of this section was to identify phylogenetic relationships that reveal differences between oyster and clinical strains, which could reveal insights into *V. parahaemolyticus* strain diversity, and eventually virulence. Concurrently, the suitability of intergenic spacer region typing (ISR-1), direct genome restriction enzyme analysis (DGREA), and pulsed-field gel electrophoresis (PFGE) for subtyping *V. parahaemolyticus* was evaluated.

5.2.1. ISR-1

A representative gel from ISR-1 analysis is shown in Figure 3. These fingerprints were analyzed using bands between 300 and 800 bp as described previously (Gonzalez-Escalona et al., 2006). All of the 144 isolates produced useable patterns with ISR-1, resulting in 23 patterns, and a discriminatory index of 0.8665 (Table 5). The *V. parahaemolyticus* isolates clustered into four clusters (see Table 6 and chapter 11.1.2, Figure 29). The largest cluster, 1, contained 63 oyster and 64 clinical isolates with 33 different serotypes and all combinations of virulence genotypes.

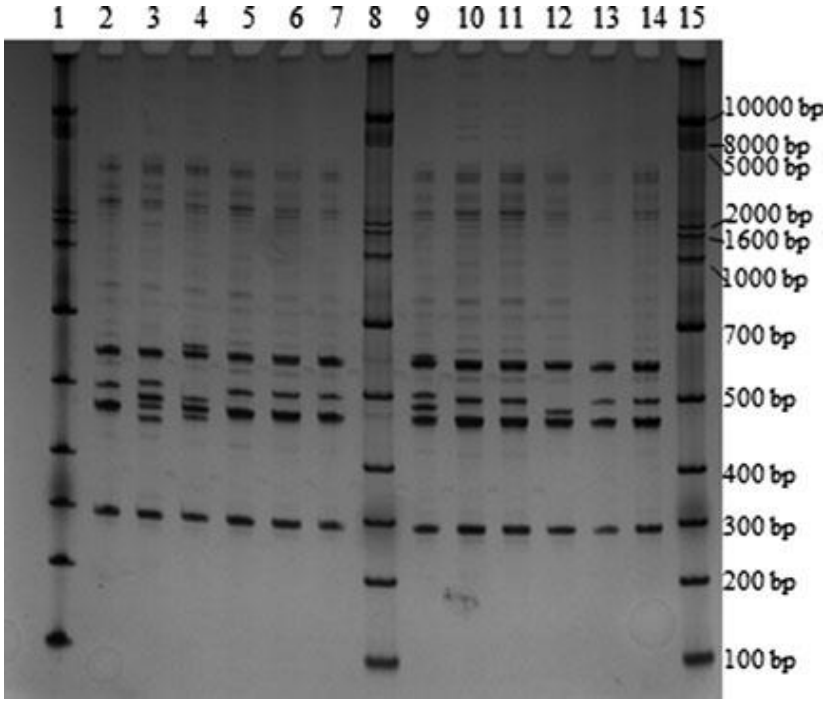


Figure 3: Representative gel image of intergenic spacer region-1 analysis after heteroduplex resolution. Lane 1, ladder; lane 2, CDC_K5328; lane 3, CDC_K5330; lane 4, CDC_K5331; lane 5, CDC_K5345_1; lane 6, CDC_K5345_2; lane 7, CDC_ K5346, lane 8, ladder; lane 9, CDC_K5428; lane 10, CDC_K5429; lane 11, CDC_K5433; lane 12, CDC_K5435; lane 13, CDC_K5436; lane 14, CDC_K5437; lane 15, ladder (Ludeke et al., 2014).

Table 5: Discriminatory indices of *V. parahaemolyticus* subtyping methods

Method		Number of patterns	Discriminatory Index (DI)	% Typeable
ISR	300-800bp	23	0.8665	100
	50-2500bp	130	0.9986	100
DGREA	50-1500bp	116	0.9993	82.6
	50-2500bp	122	0.9995	86.8
PFGE	Combined	142	0.9998	100
	*Sfi*I	93	0.9894	100
	*Not*I	99	0.9910	100

Due to a large number of products outside of the previously described analysis scale, bands between 50 and 2500 bp were also analyzed. In this analysis scale, 130 patterns were acquired (chapter 11.1.2, Figure 30), providing a discriminatory index of 0.9986 (Table 5). The isolates formed eleven clusters (Table 6). Remarkably, clusters 1 and 2 contained mostly (39/40; 97.5%) oyster isolates, while clusters 5–11 harbored mostly (29/34; 85.3%) clinical isolates, regardless of serotype or virulence genotype. The largest cluster, 3, contained a mix of oyster (n = 21) and clinical (n = 41) isolates, with strains subclustering based on oyster versus clinical origin.

Table 6: Clinical and environmental strains utilized in this study and their subtyping pattern groups

Isolate ID	Source of isolate	Sero-type	*tdh*	*trh*	ISR-1		DGREA		PFGE		
					300-800bp	50-2500bp	50-1500bp	50-2500bp	Com	*Sfi*I	*Not*I
CDC_K5010-1	Clinical	O1:Kut	+	-	1	1	11	11	78	29	11
FDA_R2	Oyster	O3:Kut	-	+	1	1	33	38	1	21	19
FDA_R10	Oyster	O1:Kut	+	+	1	1	21	30	5	6	10
FDA_R12	Oyster	O4:K8	+	+	1	1	12	16	21	6	22
FDA_R47	Oyster	O4:K8	+	+	1	1	12	21	21	6	22
FDA_R16	Oyster	O4:K9	+	+	1	1	15	28	22	6	22
FDA_R17	Oyster	O4:Kut	-	-	1	1	13	28	53	2	37
FDA_R21	Oyster	O5:Kut	-	+	1	1	40	37	9	4	5
FDA_R29	Oyster	O11:Kut	-	-	1	1	8	29	64	10	18
FDA_R30	Oyster	O1:Kut	+	+	1	1	18	30	59	26	6
FDA_R31	Oyster	O1:Kut	+	+	1	1	18	30	59	26	6
FDA_R32	Oyster	O10:Kut	+	+	1	1	18	30	48	6	1
FDA_R33	Oyster	O3:Kut	-	+	1	1	38	40	75	27	10
FDA_R26	Oyster	O4:K8	+	+	1	1	12	21	21	6	22
FDA_R51	Oyster	O8:Kut	+	+	1	1	-	-	20	15	10
FDA_R52	Oyster	O3:Kut	-	+	1	1	14	21	74	30	14
FDA_R53	Oyster	O3:Kut	-	+	1	1	14	22	74	30	14
FDA_R54	Oyster	O3:Kut	-	+	1	1	14	22	74	30	14
FDA_R55	Oyster	O3:Kut	-	+	1	1	-	-	74	30	14
FDA_R56	Oyster	O3:Kut	-	+	1	1	14	22	74	30	14
FDA_R60	Oyster	O10:Kut	-	+	1	1	22	29	42	3	8
FDA_R61	Oyster	O1:K43	-	+	1	1	19	35	26	10	13
FDA_R62	Oyster	O1:Kut	-	+	1	1	18	23	26	10	13
FDA_R63	Oyster	O4:Kut	-	+	1	1	16	27	68	9	34
FDA_R65	Oyster	O5:Kut	-	+	1	1	-	7	63	23	40
FDA_R42	Oyster	O10:Kut	-	+	1	1	24	29	65	10	19
FDA_R94	Oyster	O3:K5	-	+	1	1	7	43	15	28	1
FDA_R95	Oyster	O3:K5	-	+	1	1	28	22	15	28	1
FDA_R96	Oyster	O11:Kut	-	-	1	1	34	22	50	15	9

Table 6: *continued*

Isolate ID	Source of isolate	Sero-type	*tdh*	*trh*	ISR-1		DGREA		PFGE		
					300-800bp	50-2500bp	50-1500bp	50-2500bp	Com	*Sfi*I	*Not*I
FDA_R98	Oyster	O3:K5	-	+	1	1	28	22	15	28	1
FDA_R86	Oyster	O6:Kut	-	-	1	1	30	14	69	24	20
FDA_R87	Oyster	O8:K70	+	+	1	1	-	47	72	5	31
FDA_R88	Oyster	O8:K70	+	+	1	1	-	47	72	5	31
FDA_R75	Oyster	O8:Kut	+	+	1	1	27	36	20	15	10
FDA_R76	Oyster	O8:Kut	+	+	1	1	10	9	20	15	10
FDA_R77	Oyster	O8:Kut	+	+	1	1	10	9	20	15	10
FDA_R57	Oyster	O3:Kut	-	-	3	1	11	23	12	16	16
FDA_R59	Oyster	O5:Kut	-	-	3	1	9	22	25	4	1
FDA_R74	Oyster	O4:K34	-	-	3	1	9	16	2	20	41
FDA_R97	Oyster	O3:K5	-	+	4	2	-	-	17	4	1
FDA_R149	Oyster	O1:Kut	+	+	1	3	21	30	6	6	10
FDA_R150	Oyster	O1:Kut	+	+	1	3	21	30	5	6	10
FDA_R45	Oyster	O5:Kut	-	+	1	3	40	32	73	19	21
FDA_R99	Oyster	O3:K5	-	+	1	3	28	22	15	28	1
FDA_R100	Oyster	O3:K5	-	+	1	3	28	43	15	28	1
FDA_R108	Oyster	O3:K5	-	+	1	3	-	-	15	28	1
FDA_R109	Oyster	O3:K5	-	+	1	3	28	22	15	28	1
FDA_R110	Oyster	O3:K5	-	+	1	3	28	22	15	28	1
FDA_R111	Oyster	O11:Kut	-	-	1	3	29	24	49	15	9
FDA_R129	Oyster	O11:Kut	-	-	1	3	26	23	83	35	42
FDA_R130	Oyster	O4:K37	+	-	1	3	36	27	81	9	39
FDA_R131	Oyster	O10:Kuk	+	+	1	3	10	9	54	1	1
FDA_R125	Oyster	O11:Kut	+	-	1	3	-	-	31	10	35
FDA_R126	Oyster	O4:K42	-	-	1	3	13	10	16	16	29
FDA_R144	Oyster	O5:Kut	+	+	1	3	15	19	32	12	10
FDA_R145	Oyster	O5:Kut	+	+	1	3	15	19	32	12	10
FDA_R146	Oyster	O5:Kut	+	+	1	3	15	19	32	12	10
FDA_R135	Oyster	O3:Kut	-	-	1	3	20	22	38	9	23
FDA_R136	Oyster	O1:K20	+	+	1	3	26	23	55	17	1
FDA_R137	Oyster	O1:K20	+	+	1	3	10	23	55	17	1
FDA_R138	Oyster	O1:K20	+	+	1	3	10	23	55	17	1
CDC_K5009-1	Clinical	O4:K53	+	+	1	3	33	1	10	13	27
CDC_K5009-2	Clinical	O4:K53	+	+	1	3	33	1	10	13	27
CDC_K4762	Clinical	O5:K17	-	-	1	3	32	6	24	12	14
CDC_K4858	Clinical	O4:K4	-	-	1	3	14	15	46	15	12
CDC_K4764-1	Clinical	O8:K41	-	-	1	3	3	34	36	3	8
CDC_K4764-2	Clinical	O8:K41	-	-	1	3	23	31	82	3	8
CDC_K4842	Clinical	O5:K47	-	+	1	3	10	8	40	33	23
CDC_K4556-2	Clinical	O1:Kut	-	-	1	3	17	20	58	4	17
CDC_K4775	Clinical	O3:K6	+	-	1	3	35	19	80	29	11
CDC_K4859	Clinical	O6:K18	-	-	1	3	13	15	81	18	33

Table 6: *continued*

Isolate ID	Source of isolate	Sero-type	*tdh*	*trh*	ISR-1		DGREA		PFGE		
					300-800bp	50-1500bp	50-1500bp	50-2500bp	Com	*Sfi*I	*Not*I
CDC_K5073	Clinical	O3:K56	+	+	1	3	9	11	47	22	12
CDC_K4981	Clinical	O1:Kut	-	-	1	3	26	22	33	14	10
CDC_K5276	Clinical	O11:Kut	+	+	1	3	30	14	13	33	26
CDC_K5330	Clinical	O5:Kut	-	+	1	3	17	9	29	11	3
CDC_K5067	Clinical	O1:K56	+	+	1	3	10	23	56	17	1
CDC_K5059-1	Clinical	O5:Kut	-	-	1	3	31	20	27	10	4
CDC_K5125	Clinical	O3:Kut	-	-	1	3	37	18	37	20	1
CDC_K5512	Clinical	O4:K12	+	+	1	3	-	-	8	7	2
CDC_K5428	Clinical	O1:Kut	+	+	1	3	29	10	3	31	24
CDC_K5485	Clinical	O6:K18	-	-	1	3	17	11	4	10	32
CDC_K5280	Clinical	O4:K12	+	+	1	3	-	-	8	7	2
CDC_K5433	Clinical	O4:Kut	+	+	1	3	42	49	8	7	2
CDC_K5615	Clinical	O4:K53	+	+	1	3	5	13	11	13	27
CDC_K5345-1	Clinical	O4:K12	+	+	1	3	42	49	8	7	2
CDC_K5345-2	Clinical	O4:K12	+	+	1	3	42	49	8	7	2
CDC_K5457	Clinical	O4:Kut	+	+	1	3	-	-	8	7	2
CDC_K5429	Clinical	O4:K12	+	+	1	3	-	-	8	7	2
CDC_K5618	Clinical	O10:Kut	+	+	1	3	44	44	52	28	36
CDC_K5346	Clinical	O4:K12	+	+	1	3	42	49	8	7	2
CDC_K5620	Clinical	O10:Kut	+	+	1	3	-	44	52	28	36
CDC_K5437	Clinical	O4:Kut	+	+	1	3	-	-	8	7	2
CDC_K5528	Clinical	O4:K68	+	-	1	3	-	47	77	29	11
CDC_K5277	Clinical	O1:Kut	-	+	1	3	16	25	43	27	28
CDC_K5279	Clinical	O1:Kut	-	+	1	3	16	25	43	27	28
CDC_K5328	Clinical	O4:K12	+	+	1	3	42	49	8	7	2
CDC_K5456	Clinical	O4:Kut	+	+	1	3	-	-	8	7	2
CDC_K5579	Clinical	O4:K63	+	+	1	3	6	2	67	10	14
CDC_K4763	Clinical	O5:Kut	+	+	2	3	33	5	35	14	1
CDC_K4857-1	Clinical	O5:K17	-	-	2	3	11	12	45	34	2
CDC_K4857-2	Clinical	O5:Kut	-	-	2	3	21	6	45	34	2
CDC_K5331	Clinical	O4:K8	+	-	3	3	-	-	61	9	30
CDC_K5435	Clinical	O1:Kut	-	+	3	3	13	8	43	27	28
CDC_K5439	Clinical	O4:K8	+	-	3	3	-	-	62	9	29
CDC_K4588	Clinical	O5:Kut	-	+	1	4	13	15	70	23	1
CDC_K4558-1	Clinical	O3:K39	-	-	1	4	33	19	76	4	17
CDC_K4558-2	Clinical	O3:Kut	-	-	1	4	37	7	76	4	17
CDC_K5306	Clinical	O4:K9	+	+	1	4	15	28	22	6	22
FDA_R5	Oyster	O10:Kut	-	+	1	4	27	39	39	25	3
FDA_R6	Oyster	O10:Kut	-	+	1	4	27	37	39	25	3

Table 6: *continued*

Isolate ID	Source of isolate	Sero-type	*tdh*	*trh*	ISR-1		DGREA		PFGE		
					300-800bp	50-2500bp	50-1500bp	50-2500bp	Com	*Sfi*I	*Not*I
CDC_K5308	Clinical	O4:K63	+	+	1	5	39	4	66	10	14
CDC_K5126	Clinical	O3:Kut	-	-	1	5	8	17	41	8	22
CDC_K5324-1	Clinical	O1:K20	+	+	1	5	26	22	57	17	16
CDC_K5278	Clinical	O4:K12	+	+	1	5	-	-	8	7	2
CDC_K5323-1	Clinical	O5:K17	-	+	1	5	1	45	23	13	13
CDC_K5324-2	Clinical	O1:K20	+	+	3	5	10	22	57	17	16
CDC_K5438	Clinical	O1:Kut	-	+	3	5	13	25	43	27	28
CDC_K5323-2	Clinical	O5:Kut	-	+	3	5	2	46	23	13	13
CDC_K4636	Clinical	O10:Kut	+	+	1	6	41	42	52	28	36
CDC_K4637-1	Clinical	O3:K6	+	-	1	6	13	11	79	29	11
CDC_K5638	Clinical	O4:K12	+	+	1	7	43	48	8	7	2
FDA_R7	Oyster	O10:Kut	-	+	1	7	20	41	18	10	15
FDA_R8	Oyster	O10:Kut	-	+	1	7	20	41	19	10	15
FDA_R13	Oyster	O4:K10	-	-	1	7	-	-	60	26	4
FDA_R64	Oyster	O1:Kut	-	+	1	7	18	23	26	10	13
CDC_K5629	Clinical	O4:K13	+	+	1	7	43	48	8	7	2
CDC_K5701	Clinical	O1:Kut	-	+	3	7	16	26	44	27	28
CDC_K5621	Clinical	O1:Kut	-	+	3	7	16	26	43	27	28
CDC_K4638	Clinical	O10:Kut	-	+	1	8	25	11	51	22	26
CDC_K4637-2	Clinical	O3:K6	+	-	1	8	16	19	79	29	11
CDC_K4556-1	Clinical	O1:K25	-	-	1	8	38	33	58	4	17
CDC_K4557	Clinical	O1:K33	-	-	2	8	-	33	30	13	35
CDC_K4760	Clinical	O1:K33	-	-	1	9	-	-	28	20	38
CDC_K5010-2	Clinical	O1:Kut	+	-	1	9	11	19	78	29	11
CDC_K4639-1	Clinical	O4:K12	+	+	1	9	-	-	8	7	2
CDC_K5059-2	Clinical	O5:Kut	-	-	1	9	31	3	27	10	4
CDC_K5282	Clinical	O5:Kut	-	-	1	9	15	4	14	13	24
CDC_K5281	Clinical	O4:K12	+	+	1	9	-	-	8	7	2
CDC_K5582	Clinical	O11:Kut	+	+	1	9	4	11	13	33	26
CDC_K4639-2	Clinical	O4:Kut	+	+	1	10	-	-	8	7	2
CDC_K5058	Clinical	O3:K6	+	-	1	10	11	28	77	29	11
CDC_K5436	Clinical	O4:Kut	+	+	1	10	42	50	8	7	2
FDA_R143	Oyster	O5:Kut	-	-	1	10	12	19	34	10	1
CDC_K5635	Clinical	O5:K30	-	-	3	11	26	23	7	6	7

5.2.2. DGREA

A representative gel from DGREA is shown in Figure 4. These fingerprints were analyzed using bands between 50 and 1500 bp, as previously described (Gonzalez-Escalona et al., 2007b). Any isolate producing a pattern of fewer than four products was omitted from this analysis. Of the 144 isolates examined, 58 of 67 (86.6%) oyster and 61 of 77 (79.2%) clinical isolates produced an usable pattern. DGREA produced 116 unique patterns from the 119 typeable strains, resulting in a discriminatory index of 0.9993 (Table 5).

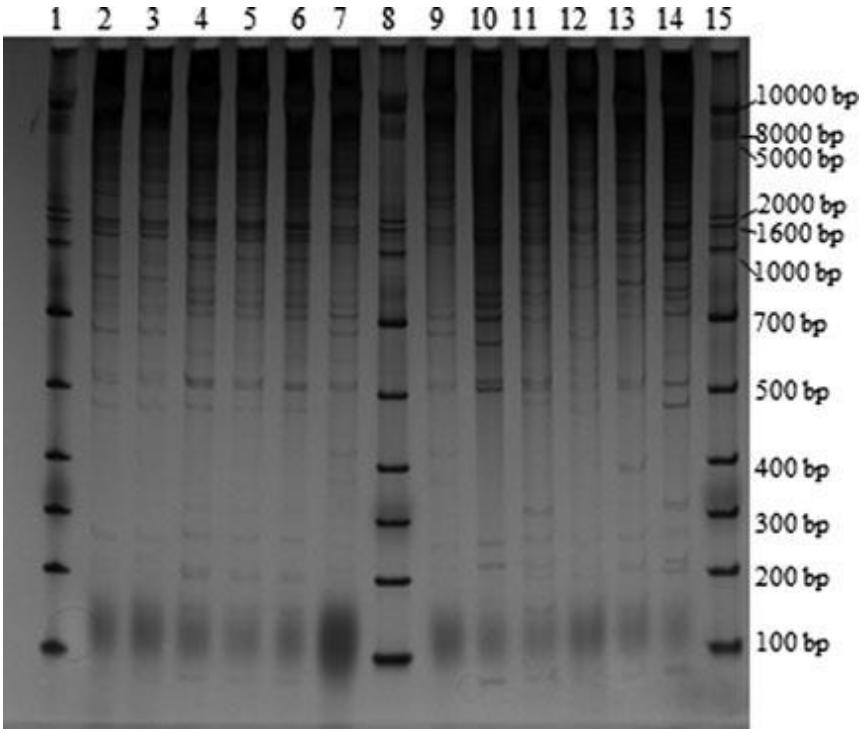


Figure 4: Representative gel image of direct genome restriction analysis product separation. Lane 1, ladder; lane 2, CDC_K5009_1; lane 3, CDC_K5009_2; lane 4, CDC_K5010_1; lane 5, CDC_K5010_2; lane 6, CDC_K5058_2; lane 7, CDC_K5059_1, lane 8, ladder; lane 9, CDC_K5059_2; lane 10, CDC_K5067; lane 11, CDC_K5073; lane 12, CDC_K5125; lane 13, CDC_K5126; lane 14, CDC_K5276; lane 15, ladder (Ludeke et al., 2014).

DGREA cluster analysis showed 44 individual clusters, 18 of which contained only a single isolate (Table 6). Cluster 28 contained only *trh*$^+$ oyster isolates with the serotype O3:K5. Clusters 42 and 43 contained eight *tdh*$^+$/*trh*$^+$ clinical isolates of serotypes O4:K12, O4:K13, and O4:Kut. Although no other clusters contained only one serotype or virulence genotype (other than the single isolate clusters), most strains of the same serotype were in the same cluster. Similar to the ISR-1 analysis, a second analysis range of bands

between 50 and 2500 bp was used. Sixty-one (91.0%) oyster and 64 (83.1%) clinical isolates produced acceptable fingerprint patterns in this analysis range, producing 122 unique patterns from 125 isolates (chapter 11.1.2, Figure 31). This resulted in a discriminatory index of 0.9995 (Table 5). The resultant dendrogram revealed 50 clusters (Table 6 and chapter 11.1.2, Figure 32). Clusters 48, 49, and 50 harbored only *tdh*$^+$/*trh*$^+$ clinical isolates of serotypes O4:K12, O4:K13, or O4:Kut. Additionally, cluster 25 contained three *tdh*$^-$/*trh*$^+$ O1:Kut clinical isolates. Remaining clusters contained a mix of clinical and oyster isolates with different serotypes and virulence genotypes. Only two sets of isolates shared a common pattern, regardless of the analysis range (Table 6): two *tdh*$^+$/*trh*$^+$ oyster isolates with serotype O5:Kut, and three *tdh*$^+$/*trh*$^+$ clinical isolates with serotype, O4:K12.

5.2.3. PFGE

A representative gel from PFGE is shown in Figure 5. DNA from the 144 *V. parahaemolyticus* isolates was successfully digested using both enzymes; however, DNA from isolate FDA_R129 (recovered from oysters) was uncut with *Sfi*I due to the restriction site being absent or inaccessible. All bands generated were used to determine the fingerprint pattern of that isolate, including the single band pattern produced by R129. The 144 isolates produced 93 patterns when digested with *Sfi*I and 99 patterns with *Not*I digestion (see chapter 11.1.2, Figure 33 and Figure 34). PFGE using *Sfi*I and *Not*I alone provided discriminatory powers of 0.9894 and 0.9910, respectively (Table 5). Combined analysis of both restriction patterns resulted in 142 unique pattern combinations (see chapter 11.1.2 Figure 35), providing a discriminatory index of 0.9998 (Table 5).

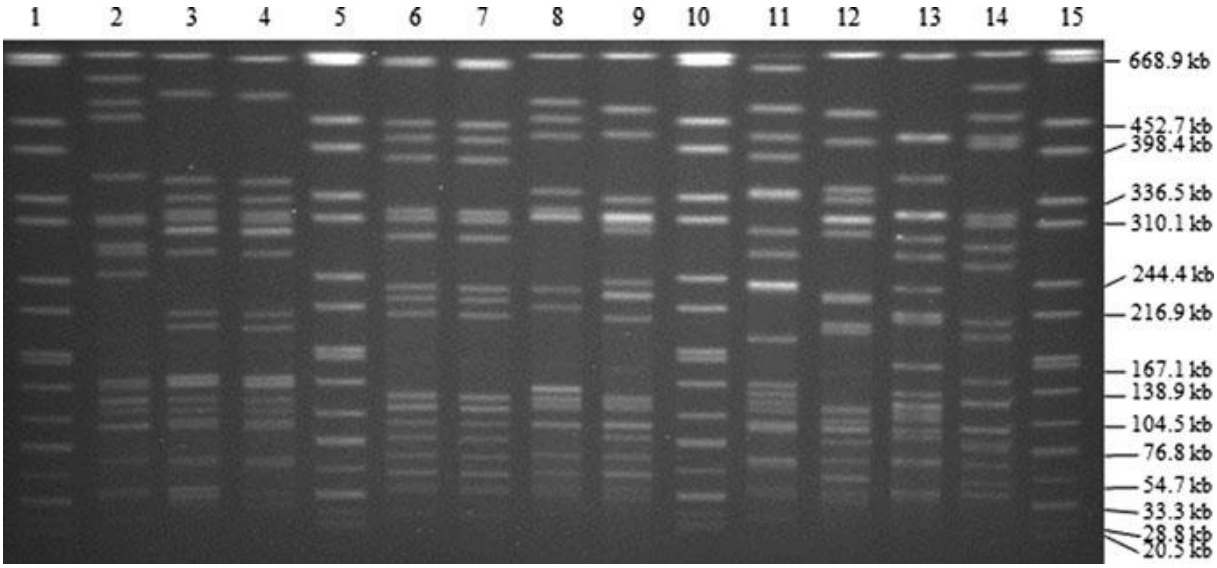


Figure 5: Representative gel image of pulsed-field gel electrophoresis product separation after digestion with SfiI . Lane 1, *Salmonella Braenderup* size standard; lane 2, FDA_R2; lane 3, FDA_R5; lane 4, FDA_R6; lane 5, *Salmonella Braenderup*; lane 6, FDA_R7; lane 7, FDA_R8, lane 8, FDA_R10; lane 9, FDA_R12; lane 10, *Salmonella Braenderup*; lane 11, FDA_R13; lane 12, FDA_R16; lane 13, FDA_R17; lane 14, FDA_R21; lane 15, *Salmonella Braenderup* (Ludeke et al., 2014).

Looking individually at patterns produced by each enzyme (*Sfi*I and *Not*I), isolates generally clustered based on serotypes (Table 6 and chapter 11.1.2, Figure 33 and Figure 34). When both enzyme patterns were analyzed together, the 144 isolates clustered into 83 groups (chapter 11.1.2, Figure 35), also primarily by serotype (Table 6). Cluster 8 was the largest and contained six clinical O4:Kut (100%) and nine of the eleven (81.8%) O4:K12 strains possessing both *tdh* and *trh*. Similarly, cluster 15 was comprised of eight of the nine (88.9%) tdh^-/trh^+ oyster isolates of the serotype O3:K5. The remaining clusters contained strains of multiple serotypes, virulence genotypes, and origins; however, isolates with the same serotypes grouped together within each cluster. Only two pairs of environmental isolates (FDA_R108/FDA_R99 and FDA_R76/FDA_R77) possessed indistinguishable patterns using combined enzyme analysis, and each pair was derived from a common oyster sample.

5.3. Examination of clinical and environmental *Vibrio parahaemolyticus* isolates by multi-locus sequence typing (MLST) and multiple-locus variable-number tandem-repeat analysis (MLVA)

The objective of this section was to identify phylogenetic relationships using multi-locus sequence typing (MLST) that reveal differences between oyster and clinical strains, which could reveal insights into *V. parahaemolyticus* strain diversity, and eventually virulence. Additionally, multiple-locus variable-tandem repeat analysis (MLVA) was used as a newly-developed High Resolution Melt (HRM) assay to differentiate between PFGE clusters and MLST sequence types (STs). Due to the labor-intensive MLST procedure a subset of the 144 *V. parahaemolyticus* isolates was used. These isolates were selected based to be representative of the sero- and genotype profiles.

5.3.1. MLST

From the 58 *V. parahaemolyticus* isolates analyzed, MLST analysis resulted in 41 different sequence types (ST) (Table 7). Four (6.9%) and one (1.7%) of the isolates were untypeable for *recA* and *pntA*, respectively; for these strains, no ST could be assigned. Twelve (42.9%) of the oyster and ten (40.0%) of the clinical isolates were a novel ST. The most frequently identified STs were ST36 (13.3%) and ST3 (10.0%) in clinical isolates and ST32 (10.7%), ST313 (7.1%), and ST676 (7.1%) in oyster isolates. All loci showed ratios of synonymous and non-synonymous substitutions (dN/dS) below 1 and, therefore, are under purifying selection, as expected for housekeeping genes.

Table 7: Isolates used in this study and their ST

Isolate ID	Source of isolate	Collection state	Serotype	*tdh*	*trh*	Allele types							ST
						dnaE	gyrB	recA	dtdS	pntA	pyrC	tnaA	
FDA_R2	Oyster	TX	O3:Kut	-	+	86	300	17	55	12	54	86	**729**
FDA_R5	Oyster	TX	O10:Kut	-	+	214	**329**	30	19	**165**	69	26	**730**
FDA_R10	Oyster	FL	O1:Kut	+	+	142	29	10	7	4	24	20	313
FDA_R12	Oyster	LA	O4:K8	+	+	20	25	15	6	7	11	4	32
FDA_R13	Oyster	LA	O4:K10	-	-	**241**	**330**	205	253	28	22	**188**	**732**
FDA_R16	Oyster	FL	O4:K9	+	+	20	25	15	13	7	11	5	34
FDA_R17	Oyster	FL	O4:Kut	-	-	14	30	49	11	49	11	13	536
FDA_R21	Oyster	TX	O5:Kut	-	+	9	21	15	13	4	10	26	12
FDA_R26	Oyster	NJ	O4:K8	+	+	20	25	15	6	7	11	4	32
FDA_R29	Oyster	FL	O11:Kut	-	-	**235**	22	25	273	164	254	20	**734**
FDA_R30	Oyster	FL	O1:Kut	+	+	17	16	UT	36	15	31	26	-
FDA_R45	Oyster	WA	O5:Kut	-	+	37	14	14	9	14	34	26	61
FDA_R47	Oyster	AL	O4:K8	+	+	20	25	15	6	7	11	4	32
FDA_R51	Oyster	AL	O8:Kut	+	+	60	106	31	72	66	62	65	676
FDA_R52	Oyster	WA	O3:Kut	-	+	4	13	11	38	18	46	23	**735**
FDA_R60	Oyster	ME	O10:Kut	-	+	63	**326**	**231**	13	48	120	24	**736**
FDA_R62	Oyster	ME	O1:Kut	-	+	31	327	UT	157	14	3	20	-
FDA_R74	Oyster	VA	O4:K34	-	-	26	58	53	19	28	9	26	108
FDA_R75	Oyster	VA	O8:Kut	+	+	60	106	31	72	66	62	65	676
FDA_R86	Oyster	FL	O6:Kut	-	-	45	**336**	143	7	171	**255**	36	**737**
FDA_R87	Oyster	FL	O8:K70	+	+	145	177	140	158	4	132	104	320
FDA_R94	Oyster	PEI (Canada)	O3:K5	-	+	47	328	UT	13	2	256	23	-
FDA_R125	Oyster	FL	O11:Kut	+	-	17	**331**	**235**	23	33	137	94	**739**
FDA_R126	Oyster	FL	O4:K42	-	-	36	285	25	250	26	227	26	**740**
FDA_R135	Oyster	SC	O3:Kut	-	-	26	16	41	224	31	32	23	**741**
FDA_R136	Oyster	SC	O1:K20	+	+	31	16	32	36	33	11	19	**775**
FDA_R143	Oyster	FL	O5:Kut	-	-	17	64	137	60	94	11	51	**743**
FDA_R149	Oyster	FL	O1:Kut	+	+	142	29	10	7	4	24	20	313
CDC_K4556_1	Clinical	LA	O1:K25	-	-	31	82	**236**	35	23	26	51	**744**
CDC_K4557	Clinical	LA	O1:K33	-	-	28	4	82	88	63	187	1	**799**
CDC_K4588	Clinical	ME	O5:Kut	-	+	56	16	**237**	8	33	59	20	**746**
CDC_K4857_1	Clinical	HI	O5:K17	-	-	35	43	38	21	31	35	37	79
CDC_K4858	Clinical	HI	O4:K4	-	-	27	84	127	139	54	124	37	283
CDC_K4981	Clinical	OK	O1:Kut	-	-	17	**327**	13	8	172	32	**181**	**748**
CDC_K5009_1	Clinical	MA	O4:K53	+	+	5	71	**238**	162	26	11	107	**749**
CDC_K5010_1	Clinical	MA	O1:Kut	+	-	3	4	19	4	29	4	22	3
CDC_K5058	Clinical	TX	O3:K6	+	-	3	4	19	4	29	4	22	3
CDC_K5067	Clinical	SD	O1:K56	+	+	31	16	13	36	33	11	19	**775**
CDC_K5073	Clinical	MD	O3:K56	+	+	17	57	52	**285**	44	28	36	**750**
CDC_K5125	Clinical	MS	O3:Kut	-	-	195	263	187	75	23	198	**190**	**772**
CDC_K5276	Clinical	NY	O11:Kut	+	+	222	128	21	69	46	236	12	631
CDC_K5278	Clinical	WA	O4:K12	+	+	21	15	1	23	23	21	16	36
CDC_K5282	Clinical	HI	O5:Kut	-	-	19	217	89	175	UT	62	51	-
CDC_K5306	Clinical	GA	O4:K9	+	+	20	25	15	13	7	11	5	34

Table 7: *continued*

Isolate ID	Source of isolate	Collection state	Serotype	*tdh*	*trh*	Allele types							ST
						dnaE	gyrB	recA	dtdS	pntA	pyrC	tnaA	
CDC_K5323_1	Clinical	VA	O5:K17	-	+	83	82	73	83	4	77	58	674
CDC_K5324_1	Clinical	VA	O1:K20	+	+	56	16	32	**286**	14	11	19	**752**
CDC_K5331	Clinical	GA	O4:K8	+	-	11	48	UT	48	26	48	26	-
CDC_K5345_1	Clinical	IA	O4:K12	+	+	21	15	1	23	23	21	16	36
CDC_K5428	Clinical	NV	O1:Kut	+	+	22	28	17	13	8	19	14	199
CDC_K5433	Clinical	WA	O4:Kut	+	+	21	15	1	23	23	21	16	36
CDC_K5436	Clinical	WA	O4:Kut	+	+	21	15	1	23	23	21	16	36
CDC_K5439	Clinical	WA	O4:K8	+	-	11	48	3	48	26	48	26	189
CDC_K5485	Clinical	NC	O6:K18	-	-	29	5	22	12	20	22	25	50
CDC_K5528	Clinical	GA	O4:K68	+	-	3	4	19	4	29	4	22	3
CDC_K5582	Clinical	GA	O11:Kut	+	+	222	128	21	69	46	236	12	631
CDC_K5618	Clinical	NY	O10:Kut	+	+	223	106	31	221	45	171	165	636
CDC_K5621	Clinical	NY	O1:Kut	-	+	39	9	27	39	3	37	30	65
CDC_K5635	Clinical	MD	O5:K30	-	-	158	131	31	**287**	128	43	**189**	**753**

*new ST in bold

eBURST analysis divided the 53 isolates for which a ST could be identified into 37 singletons and two groups: one single locus variant (SLV) and one double locus variant (DLV). No clonal complexes (CC) could be identified; demonstrating that none of the STs identified in this study share more than six alleles and, therefore, belong to different *V. parahaemolyticus* lineages.

A minimum evolution (ME) tree was constructed using the concatenated sequences of each allele (Figure 6). The isolates grouped into two main clusters, or lineages (I and II), with each lineage containing ST of clinical and oyster isolates. Isolates with the same ST generally had the same serotype; ST631 isolates possessed serotype O11:Kut, ST676 were serotype O8:Kut, ST36 were serotype O4:K12 or O4:Kut, and ST313 were serotype O1:Kut. However, the three ST3 isolates had all different serotypes.

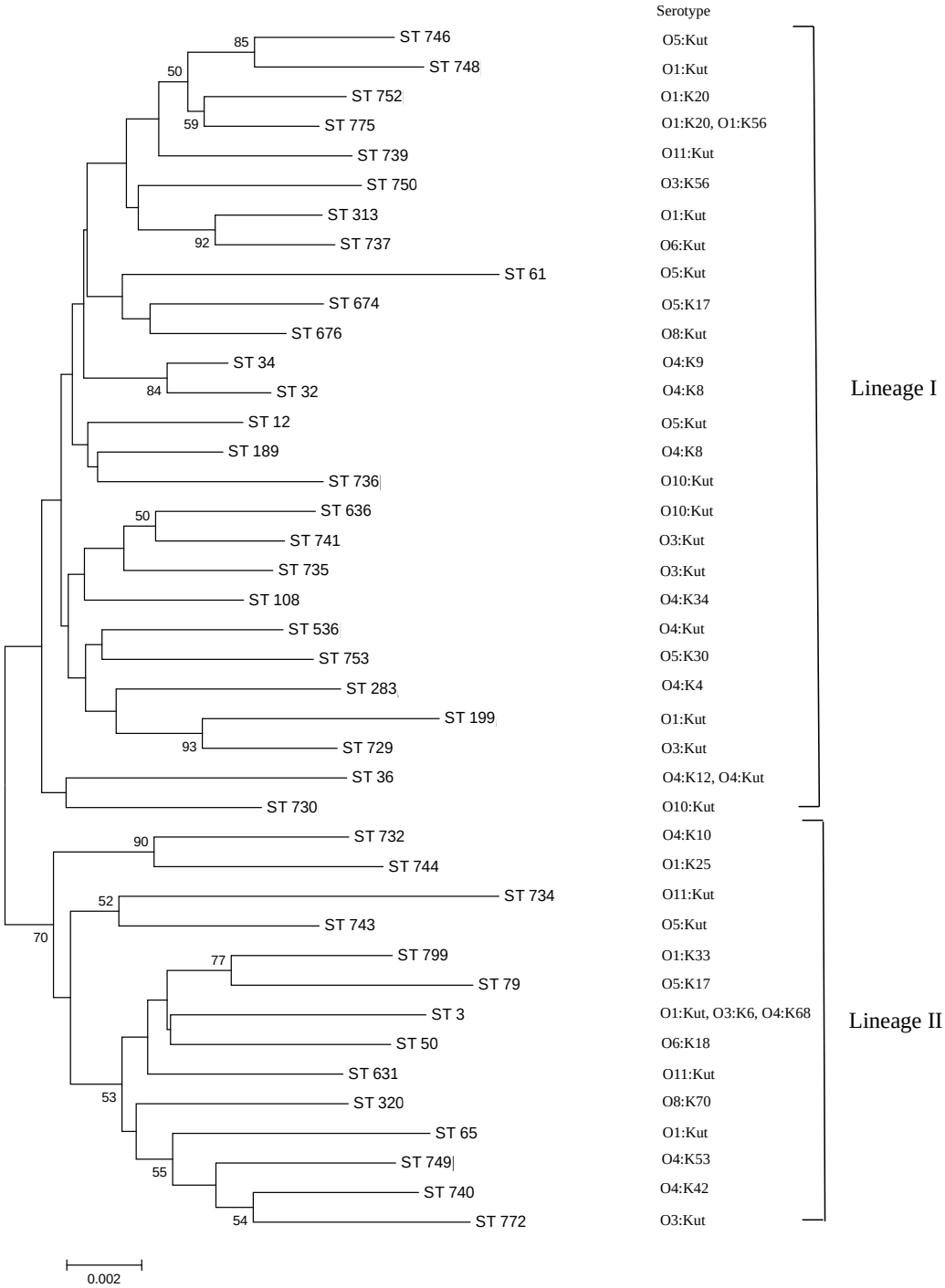


Figure 6: MLST minimum evolution tree of the 58 *V. parahaemolyticus* isolates (Ludeke et al., 2015).

5.3.2. MLVA

The 58 *V. parahaemolyticus* isolates from this study were further analyzed by MLVA with HRM analysis. Three multiplex PCRs covering twelve different loci were used. Each multiplex PCR generated reproducible melting curve profiles of select isolates (see chapter

11.1.3). In Table 8 are the percentages of isolates listed, from which each target sequence in the MLVA scheme was amplified. From the Multi A, VP2-03, VPTR7, or VP1-11 was not amplified in 10% and 14.3%, 46.6% and 39.3%, or 3.3% and 3.6% of clinical and oyster isolates, respectively; VPTR5 was amplified from all isolates. In Multi B, VP1-10 or VPTR8 was not present in 96.7% and 17.9%, 23.3% and 17.9% of clinical and oyster isolates, respectively; VPTR1 and VP1-17 was amplified in all isolates with the exception of 6.7% of clinical isolates not amplifying VPTR1. From Multi C, VPRT6 was not amplified in 83.3% and 85.7% of clinical and oyster isolates, respectively. All isolates amplified VPTR4, VPTR3, and VP2-07, with the exception of 3.6% of oyster isolates for VPTR4.

Table 8: Presence of individual MLVA genes in clinical and oyster isolates

		Clinical isolates (n=30)	Percentage	Oyster isolates (n=28)	Percentage
Multi A	VP2-03	27	90.0%	24	85.7%
	VPTR7	16	53.3%	17	60.7%
	VP1-11	29	96.7%	27	96.4%
	VPTR5	30	100.0%	28	100.0%
Multi B	VP1-10	1	3.3%	23	82.1%
	VPTR1	28	93.3%	28	100.0%
	VPTR8	23	76.7%	23	82.1%
	VP1-17	30	100.0%	28	100.0%
Multi C	VPTR4	30	100.0%	27	96.4%
	VPTR3	30	100.0%	28	100.0%
	VPTR6	5	16.7%	4	14.3%
	VP2-07	30	100.0%	28	100.0%

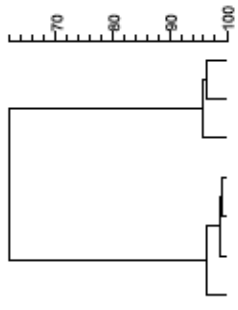
5.3.3. Comparison of MLST, MLVA, and PFGE

Based on the hypothesis MLVA can differentiate isolates with the same ST and PFGE pattern, these isolates' MLVA patterns were compared to the MLST data, as well as the PFGE results (Chapter 5.2.3). To compare these methods, dendrograms were built from the combined melting curves of the three MLVA multiplex PCRs and correlated to the PFGE cluster and ST of each isolate. MLVA allowed further differentiation of isolates with identical STs and PFGE clusters (Figure 7).

Key	Serotype	ST	PFGE pattern
FDA_R26	O4:K8	32	21
FDA_R47	O4:K8	32	21
FDA_R12	O4:K8	32	21
FDA_R10	O1:Kut	313	5
FDA_R86	O6:Kut	737	69
FDA_R143	O5:Kut	743	34
CDC_K4858	O4:K4	283	46
FDA_R136	O1:K20	775	55
FDA_R2	O3:Kut	729	1
CDC_K5067	O1:K56	775	56
CDC_K5073	O3:K56	750	47
CDC_K4588	O5:Kut	746	70
CDC_K5324-1	O1:K20	752	57
FDA_R125	O11:Kut	739	31
FDA_R29	O11:Kut	734	64
FDA_R16	O4:K9	34	22
FDA_R17	O4:Kut	536	53
CDC_K5276	O11:Kut	631	13
CDC_K5582	O11:Kut	631	13
CDC_K5618	O10:Kut	636	52
FDA_R126	O4:K42	740	16
FDA_R5	O10:Kut	730	39
FDA_R135	O3:Kut	741	38
FDA_R62	O1:Kut	-	26
FDA_R149	O1:Kut	313	6
FDA_R30	O1:Kut	-	59
FDA_R74	O4:K34	108	2
FDA_R45	O5:Kut	61	73
CDC_K4981	O1:Kut	748	33
CDC_K5621	O1:Kut	65	43
CDC_K4857-1	O5:K17	79	45
FDA_R13	O4:K10	732	60
CDC_K5331	O4:K8	-	61
CDC_K5439	O4:K8	189	62
FDA_R75	O8:Kut	676	20
FDA_R51	O8:Kut	676	20
CDC_K5010-1	O1:Kut	3	78
CDC_K5528	O4:K68	3	77
CDC_K5058	O3:K6	3	77
CDC_K5125	O3:Kut	772	37
CDC_K5306	O4:K9	34	22
CDC_K5282	O5:Kut	-	14
FDA_R60	O10:Kut	736	42
CDC_K5428	O1:Kut	199	3
CDC_K5278	O4:K12	36	8
CDC_K5345-1	O4:K12	36	8
CDC_K5433	O4:Kut	36	8
CDC_K5436	O4:Kut	36	8
FDA_R94	O3:K5	-	15
CDC_K4557	O1:K33	799	30
CDC_K5009-1	O4:K53	749	10
FDA_R52	O3:Kut	735	74
FDA_R87	O8:K70	320	72
CDC_K5323-1	O5:K17	674	23
FDA_R21	O5:Kut	12	9
CDC_K4556-1	O1:K25	744	58
CDC_K5635	O5:K30	753	7
CDC_K5485	O6:K18	50	4

Figure 7: Combined dendrogram of MLVA melting curves of the three multiplex PCRs built with BioNumerics software version 6.6. using Pearson correlation and the unweighted pair group method using arithmetic averages (UPGMA). Isolates originated from oysters starting with "FDA", isolates from clinical origin labeled with "CDC". The PFGE pattern designations are as previously reported (Ludeke et al., 2015).

Specifically, the isolates with ST3 and ST36 share the same PFGE cluster, but were distinguishable by MLVA melting curve profiles (Figure 8). The dendrogram with only ST3 and ST36 isolates showed ST-specific clusters, but separation within those clusters based on the combined melting curves of MLVA.



Key	Serotype	ST	PFGE pattern
CDC_K5010-1	O1:Kut	3	78
CDC_K5528	O4:K68	3	77
CDC_K5058	O3:K6	3	77
CDC_K5278	O4:K12	36	8
CDC_K5345-1	O4:K12	36	8
CDC_K5433	O4:Kut	36	8
CDC_K5436	O4:Kut	36	8

Figure 8: Dendrogram of MLVA melting curves of the three multiplex PCRs for the isolates carrying ST3 and ST36 built with BioNumerics software version 6.6. using Pearson correlation and UPGMA (Ludeke et al., 2015).

5.4. Complete genome sequences of a clinical and an environmental *Vibrio parahaemolyticus*

The objective of this section was to sequence a clinical and an environmental *V. parahaemolyticus* isolate and create, through assemblies, two closed whole genome sequences. In public databases not many closed genomes of *V. parahaemolyticus* have been deposited; therefore these sequences added beneficially to the publically available number of genomes and support the investigation of new pathogenicity factors.

The genomes were sequenced within the University of California at Davis 100K Pathogen Genome Project using the PacBio RSII sequencing platform (Pacific Biomarkers, Menlo Park, CA, USA). Through the annotation process, 4,771 and 4,937 genes for the clinical and oyster isolates, respectively, as well as 4,579 and 4,731 coding regions were identified. The presence or absence of the *tdh* and *trh* genes was confirmed in both isolates. Nucleotide sequence accession numbers are listed in 11.1.5 in Table 16.

5.5. Correlation of *Vibrio parahaemolyticus* cytotoxicity with the virulence markers, *tdh, trh,* T3SS2, and serotype

The objective of this section was to evaluate the cytotoxic potential of clinical and environmental *V. parahaemolyticus* isolates by applying cytotoxicity assays to Caco-2 and HeLa cells as well as to evaluate the utility of each cell line for virulence investigations.

It was hypothesized that Caco-2 would be a more suitable cell line, compared with HeLa, to study the potentially pathogenic effect of *V. parahaemolyticus* on humans as they are derived from colorectal carcinoma cells, and *V. parahaemolyticus* infects the digestive tract. Therefore, 75 shellfish and 89 clinical *V. parahaemolyticus* isolates were examined by lactate dehydrogenase (LDH)-releasing cytotoxicity assays using both HeLa and Caco-2 cells. In the HeLa assay, 45.3% of the shellfish isolates were cytotoxic, 24.0% semi-cytotoxic and 30.7% non-cytotoxic; for the clinical isolates 53.9% were cytotoxic, 21.4% were semi-cytotoxic, and 24.7% were non-cytotoxic (Figure 9). In contrast, 25.3% of the shellfish isolates were cytotoxic, 32.0% semi-cytotoxic, and 42.7% non-cytotoxic to Caco-2 cells. For the clinical isolates, 51.7% were cytotoxic, 20.2% were semi-cytotoxic, and 28.1% were non-cytotoxic (Figure 9). The cytotoxicity of clinical and shellfish isolates was significantly different in Caco-2 ($p=0.003$) but not HeLa cells ($p=0.534$).

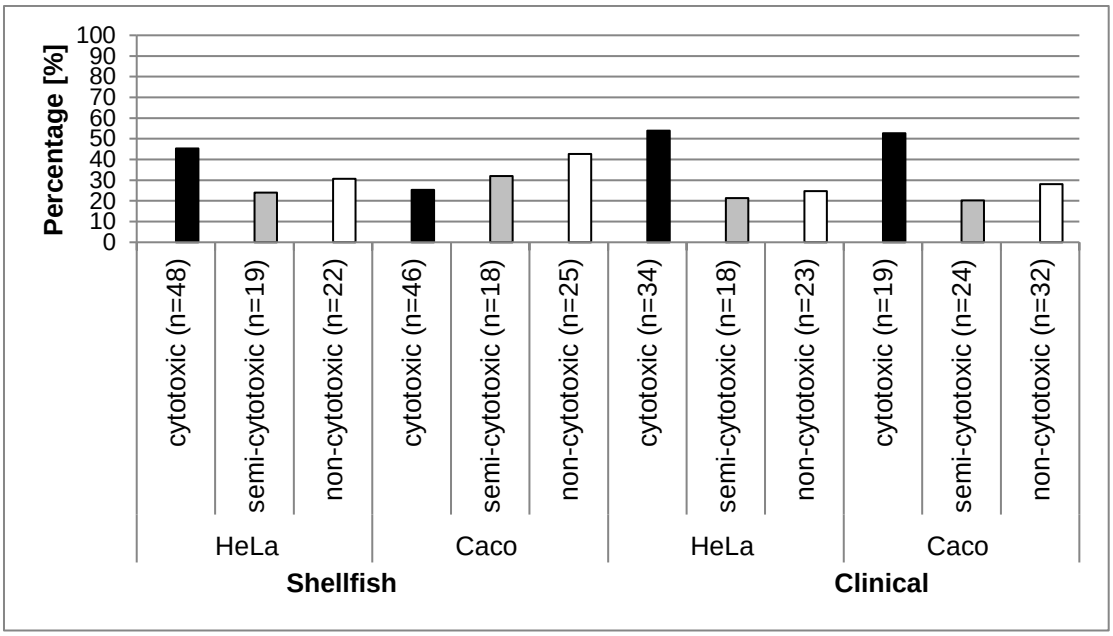


Figure 9: Percentage of isolates in each cytotoxicity category, based on isolate origin and infection target cell line; black: cytotoxic isolates, grey: semi-cytotoxic isolates, white: non-cytotoxic isolates.

Further investigations were aimed at evaluating associations between the existing virulence marker genes and cytotoxicity. It was assumed that *tdh*, *trh,* and T3SS2 are not the only indicators of strain cytotoxicity, as a proxy of virulence, and cytotoxicity is not completely determined by isolate origin (clinical or shellfish). To test these hypotheses, the correlation between the *tdh*, *trh,* and T3SS2 gene profile and the cytotoxicity response was evaluated both overall and by isolate type (clinical versus shellfish). Looking at HeLa cell response to the *tdh*$^+$/*trh*$^+$/T3SS2β$^+$ isolate subset, 42.3% of shellfish and 64.4% of clinical isolates were cytotoxic (Figure 10 and Figure 11). However, 47.6% of the *tdh*$^-$/*trh*$^-$/T3SS2$^-$ clinical isolates also showed cytotoxic potential towards HeLa cells. Similar results were seen in the Caco-2 cells; 52.4% of the clinical *tdh*$^-$/*trh*$^-$/T3SS2$^-$ isolates possessed cytotoxic potential. Overall, 56.3% and 47.9% of the total 71 *tdh*$^+$/*trh*$^+$/T3SS2β$^+$ isolates were cytotoxic for HeLa and Caco-2 cells, respectively. Also, of the 39 *tdh*$^-$/*trh*$^-$/T3SS2$^-$ isolates 43.6% were cytotoxic for each cell line. Of the eleven *tdh*$^+$/*trh*$^-$/T3SS2α$^-$ shellfish isolates 50.0% were cytotoxic in HeLa cells. In contrast, in Caco-2 cells 72.7% of these isolates were non-cytotoxic.

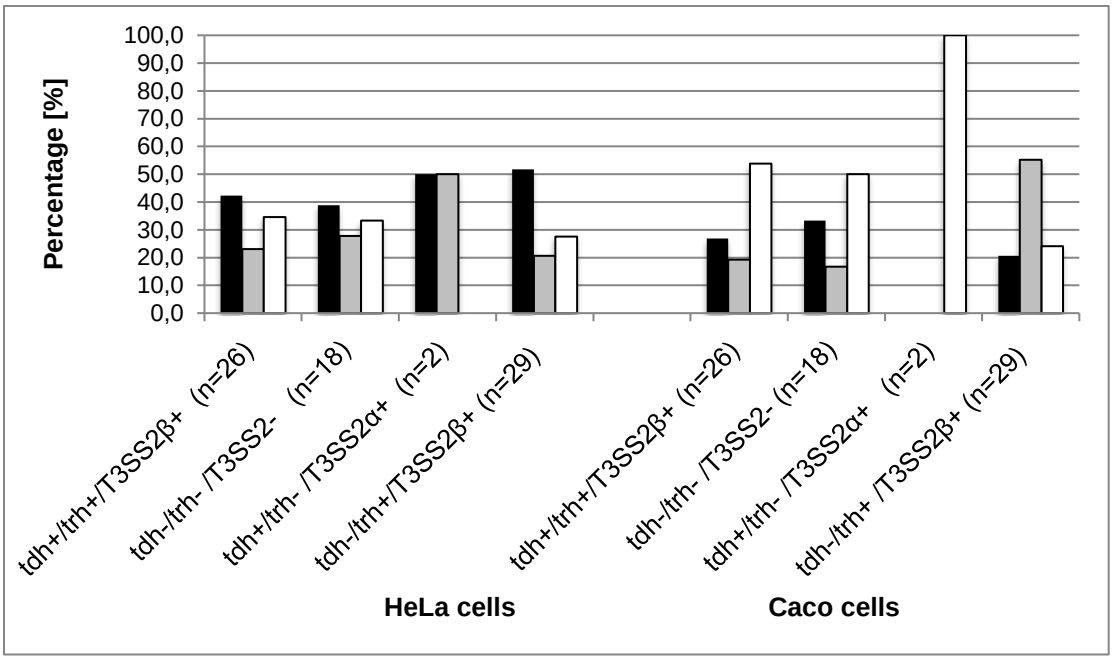


Figure 10: Percentage of shellfish isolates in each cytotoxicity category correlated to the virulence gene profile of *tdh, trh*, T3SS; black: cytotoxic isolates, grey: semi-cytotoxic isolates, white: non-cytotoxic isolates

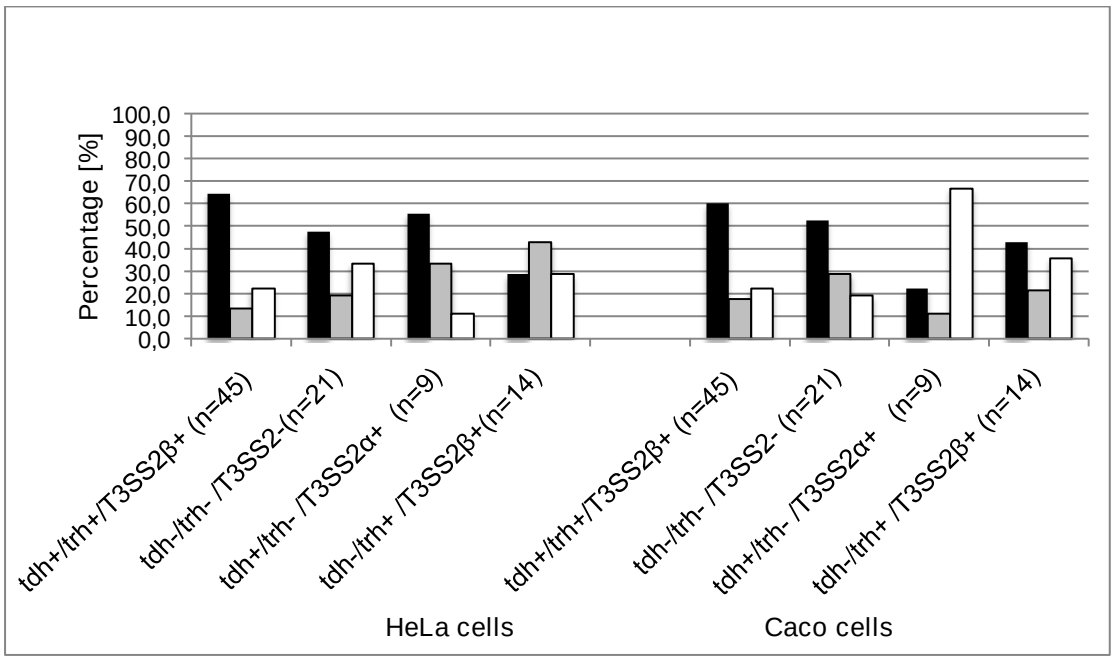


Figure 11: Percentage of clinical isolates in each cytotoxicity category correlated to the virulence gene profile of *tdh, trh*, T3SS; black: cytotoxic isolates, grey: semi-cytotoxic isolates, white: non-cytotoxic isolates

Looking at the presence of *tdh* and/or *trh* and the T3SS2, regardless of strain origin (shellfish or clinical), no significant difference was observed in cytotoxic potential between the isolates with or without individual virulence factors in the HeLa cell assay (chapter 11.1.5). However, a statistically significant difference was observed between the cytotoxic potential of isolates

with *tdh* (p=0.043) and T3SS2α (p=0.023) compared to those lacking the respective genes using the Caco-2 cell line (chapter 11.1.5). When examining clinical and shellfish isolates separately, cytotoxic potential is associated with T3SS2α in clinical (p=0.025), but not shellfish, isolates targeting Caco-2 cells. To confirm the observations of association between cytotoxic potential and *tdh* in clinical and shellfish isolates was not affected by gene expression, we applied a KAP-reverse passive latex agglutination (RPLA) kit to all *tdh* and/or *trh* positive isolates to test for TDH expression. Of the 26 *tdh*$^+$/*trh*$^+$ shellfish isolates, 25 (96.2%) showed agglutination with the TDH antibody, as well as both of the *tdh*$^+$/*trh*$^-$ isolates (100.0%), and none of the 29 *tdh*$^-$/*trh*$^+$ shellfish isolates. Twenty-one of the 45 clinical *tdh*$^+$/*trh*$^+$ isolates (46.7%) were positive for TDH agglutination, as were all 9 of the *tdh*$^+$/*trh*$^-$ isolates (100%), and none of the 14 *tdh*$^-$/*trh*$^+$ isolates produced TDH. Overall, comparing isolates producing TDH to those not producing TDH, no significant difference was observed between the cytotoxicity values from the HeLa cell line, however, a significant association was present in the Caco-2 cell line (p=0.041). Moreover, the significant correlation overall was largely attributable to the shellfish isolates in the Caco-2 cell assay where an association between TDH and cytotoxicity was most evident (p=0.019) (contingency tables in chapter 11.1.5).

With a particular interest in outbreak-related strains, we questioned if strain serotype is correlated with cytotoxicity. In 2012 and 2013, outbreaks occurred in the United States related to the serogroup O4:Kut and O4:K12 (Martinez-Urtaza et al., 2013b; Newton et al., 2014b); therefore, comparisons were focused on the effect of isolates carrying these serotypes as well as the serotypes of the pandemic group (O1:K25, O1:Kut, O3:K6, and O4:K68). Looking at the serotype profile, ≥90% of clinical O4:K12 and O4:Kut isolates were cytotoxic for both cell lines, with a significant positive association with cytotoxic potential in both cell assays (p<0.001 for each cell line, contingency tables in chapter 11.1.5). Of the pandemic serotypes, ~80% of the O1:Kut clinical isolates (n=11) were cytotoxic or semi-cytotoxic to both cell lines, and was one of the most common serotypes in this study.

However, of the shellfish isolates with serotypes O1:Kut (n=7) and O4:Kut (n=5), 50% and ~60% were cytotoxic, respectively. Further in the pandemic serogroup, the one O1:K25 isolate was semi-cytotoxic or cytotoxic, while the one O4:K68 isolate was semi-cytotoxic or non-cytotoxic in the HeLa or Caco cell assay, respectively. In this study, only 50% and 25% of O3:K6 isolates (n=4) were cytotoxic to HeLa and Caco-2 cells, respectively. These serotypes associated with the pandemic clade did not show a statistically significant association with cytotoxicity towards HeLa (p=0.205) or Caco-2 cells (p=0.100) (contingency tables in chapter 11.1.5).

5.6. Next-generation sequencing analysis of *V. parahaemolyticus* isolates

In this chapter, 144 draft sequences from clinical and oyster isolates from 2007 were completed. The sequences were generated using the Illumina HiSeq 2000 with PE100 plus index read in the University of California at Davis 100K Pathogen Genome Project. The sequences were analyzed with different phylogenetic methods such as *in silico* MLST and kSNP phylogeny.

5.6.1. Sequencing and assembly

Twelve sequences were excluded from analysis due to unusual genome sizes and the remaining 132 were further analyzed. The genome sizes varied between 4.8 and 5.3 Mb which has been described before for *V. parahaemolyticus* (Tiruvayipati et al., 2013), with an average coverage of 73x. The contigs for each isolate, coverage and genome sizes are summarized in the attachments (chapter 11.1.6, Table 50). The genome assemblies yielded between 76 and 297 contigs, with an average gene count of 4659.

5.6.2. *In silico* MLST

In this chapter, 61 STs were identified from the 132 isolates (including isolates used in chapter 5.3). The ST for each isolate is presented in the attachment 11.1.6 in Table 52. The three most frequent STs were ST36, ST1151 and ST3. Six isolates were untypeable by MLST due to untyepable *recA* loci, as well as *pntA*. The population snapshot of the eBurst analysis showed one SLV (ST189 and ST265) indicating these isolates share one allele type (Figure 12). No CC could be identified, as observed in the previous subset of these isolates examined, (chapter 5.3.1) indicating that none of the ST share more than six allele types (Ludeke et al., 2015). Some isolates of the same ST were isolated in different reporting states and were of different serotypes. Some STs were found in oyster and clinical isolates; i.e., one oyster and one clinical isolate were both of ST34; similarly, three isolates of oyster origin shared ST775 with one isolate of clinical origin. Since ST265 possesses the *recA* allele variant *recA*107 (Gonzalez-Escalona et al., 2015), it was removed from the MLST database and excluded from the ML tree analysis to gain a better resolution of the remaining STs relations. Since the new ST has not been published yet, the original assigned ST265 was kept in the results table. The ML tree resulted in three major clusters, divided into several subgroups (Figure 13).

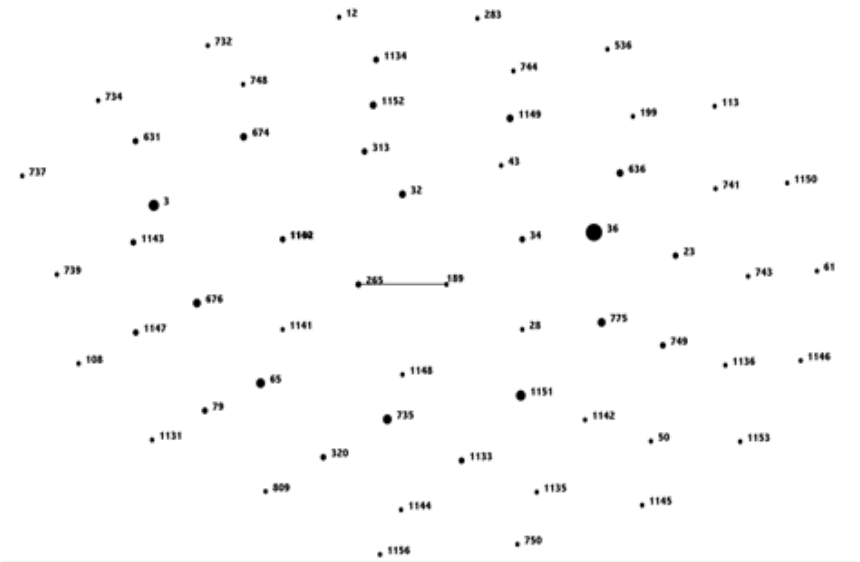


Figure 12: eBurst population snapshot of the 61 ST, the frequency of each ST is indicated by the font size, ST265 and ST189 are connected by a line as the SLV of the group.

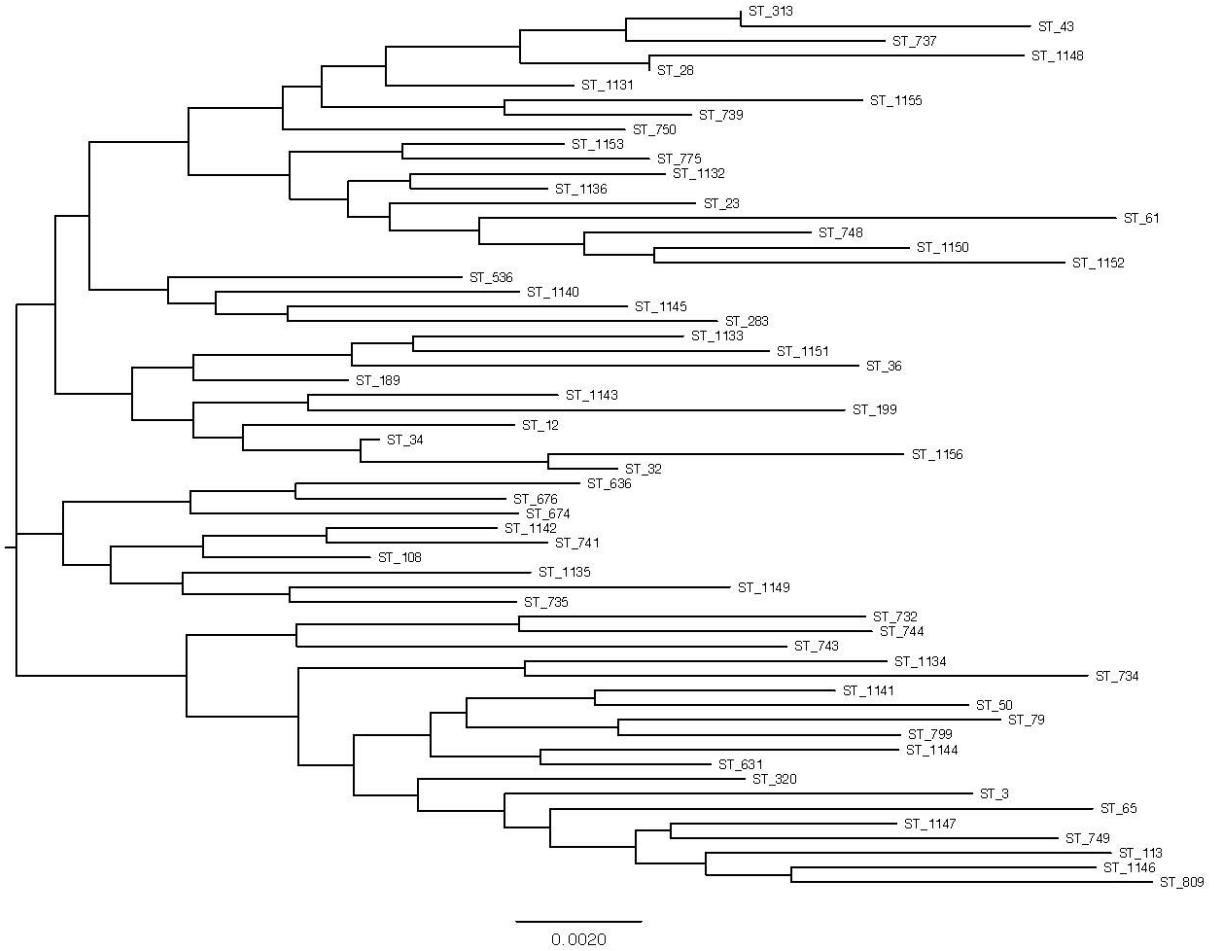


Figure 13: MLST maximum likelihood (ML) tree of the 60 ST of V. parahaemolyticus isolates. The tree was built with Mega software 6 using concatenated sequences by the kimura-2-model with 1000 bootstrap replicates. The scale represents the evolutionary distance.

The single nucleotide polymorphism (SNP) matrix included 129 taxa containing 110,813 SNPs, assuming that three isolates had similar SNP patterns to other isolates in the data set. The kSNP tree showed three clusters (A, B, and C). All clusters were further subdivided into two smaller sub-clusters with clusters B and C as the largest groups (Figure 13). As expected, isolates of the same ST grouped together throughout the tree. Compared to ML tree of the MLST analysis some of the sequences showed different relations to each other, as expected by the different methods in use. The distribution of the STs among the two clusters of the MLST ML tree are different from the shared SNP relations in the kSNP tree. Further, isolates of ST36 show immediate relations to ST1131 and ST1151 in kSNP; the

MLST ML tree has the STs ST536, ST32, ST34 and ST23 as related branches to ST36. The eBurst population snapshot supports the relations of each ST mentioned to the ST36 with the exemption of ST1131 and ST1151. In general, the kSNP tree presented a clustering based on the source of isolation of these samples. Cluster A branched out into two subgroups, A1 and A2. Thus, sub-cluster A1 contained a great number (44.1%; n=26 of 59) of oyster isolates carrying the same serotype as some clinical isolates; this relationship is demonstrated in the clustering. The sub-cluster A2 and further B1 only held a few isolates of both origins and from diverse locations. Sub-cluster B2 contained a large set (31.4%; n=22 of 70) of the clinical isolates differing by a small number of SNPs. This group held all O4:K12/O4:Kut and ST36 isolates. Several strongly supported lineages, C1 and C2, emerged from this clade. Within C1 various clinical isolates (37.1%; n=26 of 70) grouped together alongside a subgroup of oyster isolates (32.2%; n=19 of 59). In this cluster, though, some isolates of different origin were only different by a small set of SNPs, they shared the same O-serogroup or the same harvest state. The same pattern was observed in cluster C2: A small group of clinical and oyster isolates have their reporting state and O-serogroup in common, but differ by a larger number of SNPs.

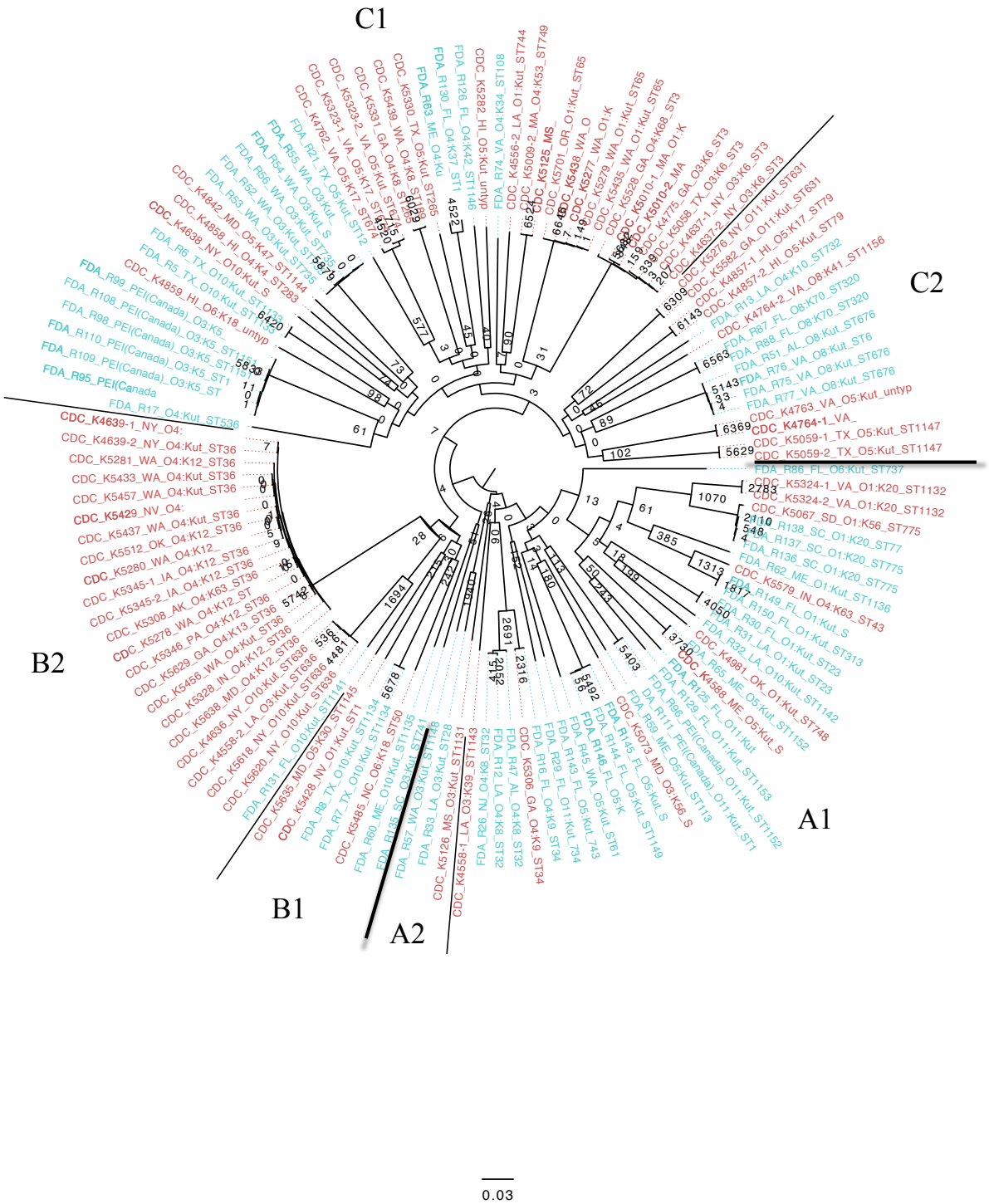


Figure 14: RAxML bootstrapped tree of kSNP matrix. Numbers on branches indicate the number of SNPs different in each isolate, visualized in FigTree v1.4.2., The isolates are labeled as follows: Isolate_ID_Harvest state_serotype_ST, the origin of each isolate is colour coded: Clinical isolates: red, oyster isolates: green. Clusters are labeled as follows: A1, A2, B1, B2, C1, and C2.

5.6.3. Presence/absence analysis of virulence related genes

While setting up a presence-absence matrix for the virulence genes *tdh* and *trh* the observation was made that the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (NPGAAP) did not annotate any of our isolates as possessing *tdh* or *trh*, in contrast to our real-time PCR results (Jones et al., 2012) (see chapter 11.1.6, Table 53). However, the NPGAAP annotations indicated a close correspondence between the presence of hemolysin activation protein (Table 53) and that of *tdh* and/or *trh* detected via real-time PCR. The hemolysin activation protein sequence was generally present in isolates with the real-time PCR genotypes of *tdh*$^+$/*trh*$^+$, *tdh*$^-$/*trh*$^+$, and *tdh*$^+$/*trh*$^-$, while it was generally not present in *tdh*$^-$/*trh*$^-$ isolates. Exceptions were five isolates *tdh*$^-$/*trh*$^+$ by real-time PCR in which the hemolysin activation protein was not found, and three *tdh*$^-$/*trh*$^-$ isolates with a hemolysin activation protein. Furthermore, when sequences from expressed proteins previously subjected to detailed biochemical investigation were used for comparison, *tdh*_2 [GI: 29611945, (Yanagihara et al., 2010)] and *trh* [GI:39748662, (Ohnishi et al., 2011)], the presence-absence matrix results concurred with the previous real-time PCR results, with the exception of four isolates (0.03%). The *V. parahaemolyticus*-specific gene, *tlh,* is present in every isolate (Taniguchi et al., 1986), a finding confirmed by real-time PCR (chapter 5.1), one-way BLAST analysis against selected genes downloaded from NCBI, and the Genbank files produced by Prokaryotic genomes automatic annotation pipeline (PGAAP). Taking into account that Interpro groups *tdh* and *trh* into one gene family, there was strong correspondence between real-time PCR results and Interproscan results. All strains, for which *tdh* or *trh* were detected via real-time PCR, were annotated as having a TDH family protein by Interproscan. The only discrepancies were three of the 132 isolates (0.02%) identified as positive for a TDH family protein by Interpro, but were negative for both genes by real-time PCR. When the assemblies from this study were blasted against the *V. parahaemolyticus* sequences from SwissProt, there was again a general correspondence between the presence of *tdh* or *trh* detected via Blast and detection by

means of real-time PCR, but the details indicate substantial inconsistency between the two approaches. Swiss-Prot *trh* only came up in the scan for isolates *tdh*$^+$/*trh*$^+$ by real-time PCR and two isolates *tdh*$^-$/*trh*$^-$ by real-time PCR. All isolates were negative in Swiss-Prot for *tdh*_1 with the exception of nine *tdh*$^+$/*trh*$^-$ real-time PCR isolates. Most isolates (*tdh*$^+$/*trh*$^+$, *tdh*$^-$/*trh*$^+$, and *tdh*$^+$/*trh*$^-$ by real-time PCR) were positive for *tdh*_2 in Swiss-Prot with only eight exceptions. Additionally, three isolates *tdh*$^-$/*trh*$^-$ by real-time PCR were positive for *tdh*_2 based on Blasting against Swiss-Prot.

5.7. Evaluation of next-generation sequencing data in relation to isolate cytotoxicity

As described in chapter 5.5 a cytotoxicity assay using Caco-2 cells for *V. parahaemolyticus* was developed. In this section, the cytotoxicity data was compared to the whole genome sequences of each isolate. In the Caco-2 cell assay, 56 of the isolates collected in 2006/2007 were cytotoxic, 36 isolates semi-cytotoxic and 52 isolates non-cytotoxic. To ease the analysis, a subpopulation of these isolates representative of the sero- and genotype profiles was used (list of used isolates in attachment 11.1.7 in Table 54 and Table 55). These isolates were also selected to have equal representation of cytotoxic, semi-cytotoxic, and non-cytotoxic strains.

5.7.1. Large scale BLAST score ratio (LS-BSR) analysis

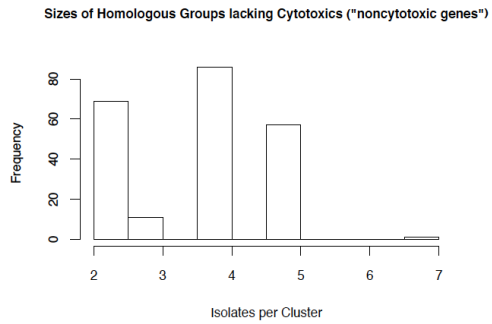


Figure 15: Sizes of homologous groups lacking cytotoxic isolates. Each bar represents a number of exclusively cytotoxic and in-between isolates in a cluster and the y-axis is how frequent a cluster with that number of strains occurs.

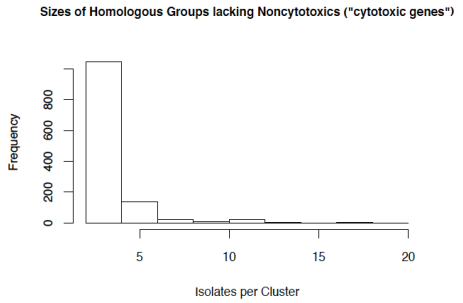


Figure 16: Sizes of homologous groups lacking non-cytotoxic isolates. Each bar represents a number of exclusively cytotoxic and in-between isolates in a cluster and the y-axis is how frequent a cluster with that number of strains occurs

Groups „lacking cytotoxics“ were considered to be isolates carrying non-cytotoxic genes, while the group „lacking non-cytotoxics“ was considered having only cytotoxic genes. The size distribution of the clusters that contain either no cytotoxic (Figure 15) or no non-cytotoxic (Figure 16) isolates is small. The histogram lacking non-cytotoxic isolates indicates that there are many genes shared exclusively by five or fewer cytotoxic strains. The largest number of strains in a single exclusively cytotoxic cluster is 7, far fewer than the 49 cytotoxic isolates that were included in the analysis. The histogram without cytotoxic isolates shows the same, five or fewer non-cytotoxic isolates represent the highest amount of shared genes. Overall, there were a total of 9650 clusters. The counts of the gene clusters per histogram were 1244 for all cytotoxic and semi-cytotoxic isolates and 244 for all non-cytotoxic and semi-cytotoxic isolates. Neither the cytotoxic nor the non-cytotoxic cluster are large enough to contain nearly the full set of either genes.

5.7.2. Functional categories of cluster of orthologous groups and average number of SNPs

Table 9 contains total counts and counts of cluster of orthologous groups (COGs) corresponding to COG IDs sorted by core, pan, cytotoxic, and non-cytotoxic.

Table 9: Total counts and counts of COGs sorted by their functional category

Functional category	core	pan	cytotoxic	non-cytotoxic
Function unknown	536	38	8	7
Energy production and conversion	167	1	1	0
Transcription	195	8	2	0
Defense mechanisms	43	2	2	0
Replication, recombination and repair	107	4	1	2
Translation, ribosomal structure and biogenesis	149	0	1	0
Inorganic ion transport and metabolism	185	2	0	0
Carbohydrate transport and metabolism	129	15	0	0
Cell wall/membrane/envelope biogenesis	165	8	0	1
Intracellular trafficking, secretion, and vesicular transport	76	3	2	1
Secondary metabolites biosynthesis, transport and catabolism	32	0	0	0
Amino acid transport and metabolism	216	2	1	1
Posttranslational modification, protein turnover, chaperones	116	0	1	0

Table 9: *continued*

Functional category	core	pan	cytotoxic	non-cytotoxic
Nucleotide transport and metabolism	75	0	0	0
Coenzyme transport and metabolism	115	0	2	0
General function prediction only	218	16	4	0
Signal transduction mechanisms	143	1	0	0
Cell cycle control, cell division, chromosome partitioning	33	0	1	0
Lipid transport and metabolism	73	1	0	0
Cell motility	46	0	0	0
RNA processing and modification	1	0	0	0
Total counts	2820	101	26	12

In the core genome approximately 20% of the COGs were classified as “function unknown”. Also, the pan genome contained 38% of the COG counts in the category of “function unknown”. While the core genome had a total count of 2820 COGs, this outbalanced the variable part of the genome with 101. A bar chart was created with the software R version 3.1.2. (Team, 2015) to visualize the functional groups containing potential association with cytotoxicity (Figure 17).

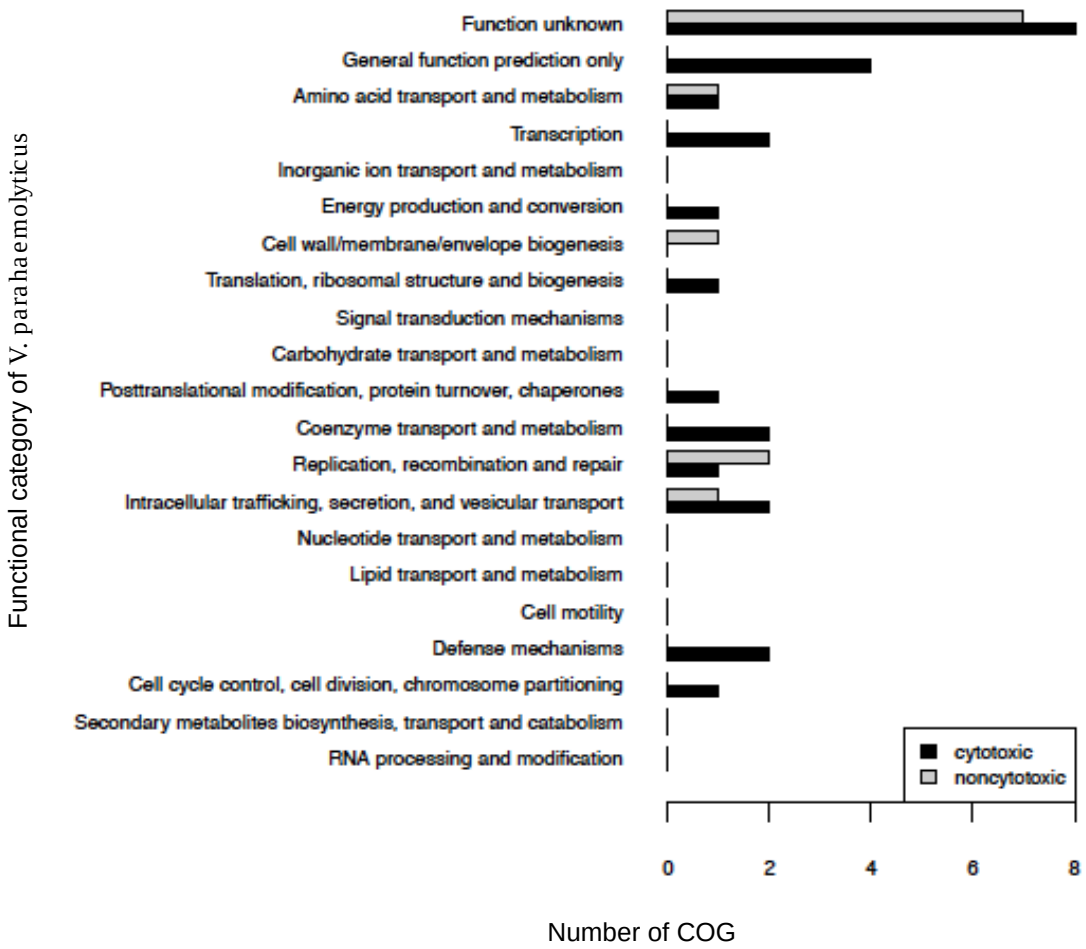


Figure 17: NOG counts without core and pan genome

The graph showed the functional groups of COGs without non-cytotoxic isolates present were “general function prediction only”, “transcription”, “energy production”, “posttranslational modification, protein turnover, chaperones”, “coenzyme transport and metabolism”, “defense mechanism”, and “cell cycle control, cell division, chromosome partitioning”. The most enriched functional group was “general function prediction only” with an average of six COGs. Further, both cytotoxic and non-cytotoxic sequences grouped in the category “function unknown” with an average of eight and seven genes, respectively. Unique non-cytotoxic sequences were categorized only in “cell wall/ membrane/envelope biogenesis” or increased compared to the cytotoxic isolates’ sequences in “replication, recombination and repair”.

The average number of SNPs across the *V. parahaemolyticus* genomes are displayed in (Figure 18). Generally, SNPs were more frequently found in the cytotoxic strains.

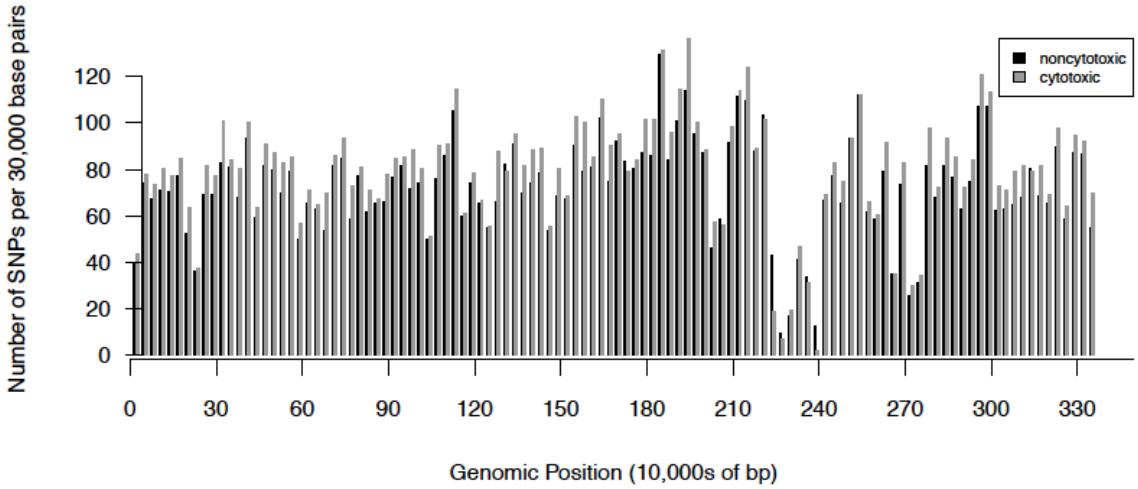


Figure 18: Comparison of SNPs in cytotoxic and non-cytotoxic isolates

5.7.3. Phylogenetic tree

The phylogenetic tree was created using the whole genome concatenated sequences and consisted of three main clusters with cluster C as the largest cluster (Figure 19). Cluster A and B only grouped non-cytotoxic isolates. Cluster C subdivided into three subclusters, C1, of which C2 and C3 emerged. While C2 groups mostly non-cytotoxic isolates, C3 was determined to hold the majority of cytotoxic isolates. Thus, a clustering based on the phenotype was observed.

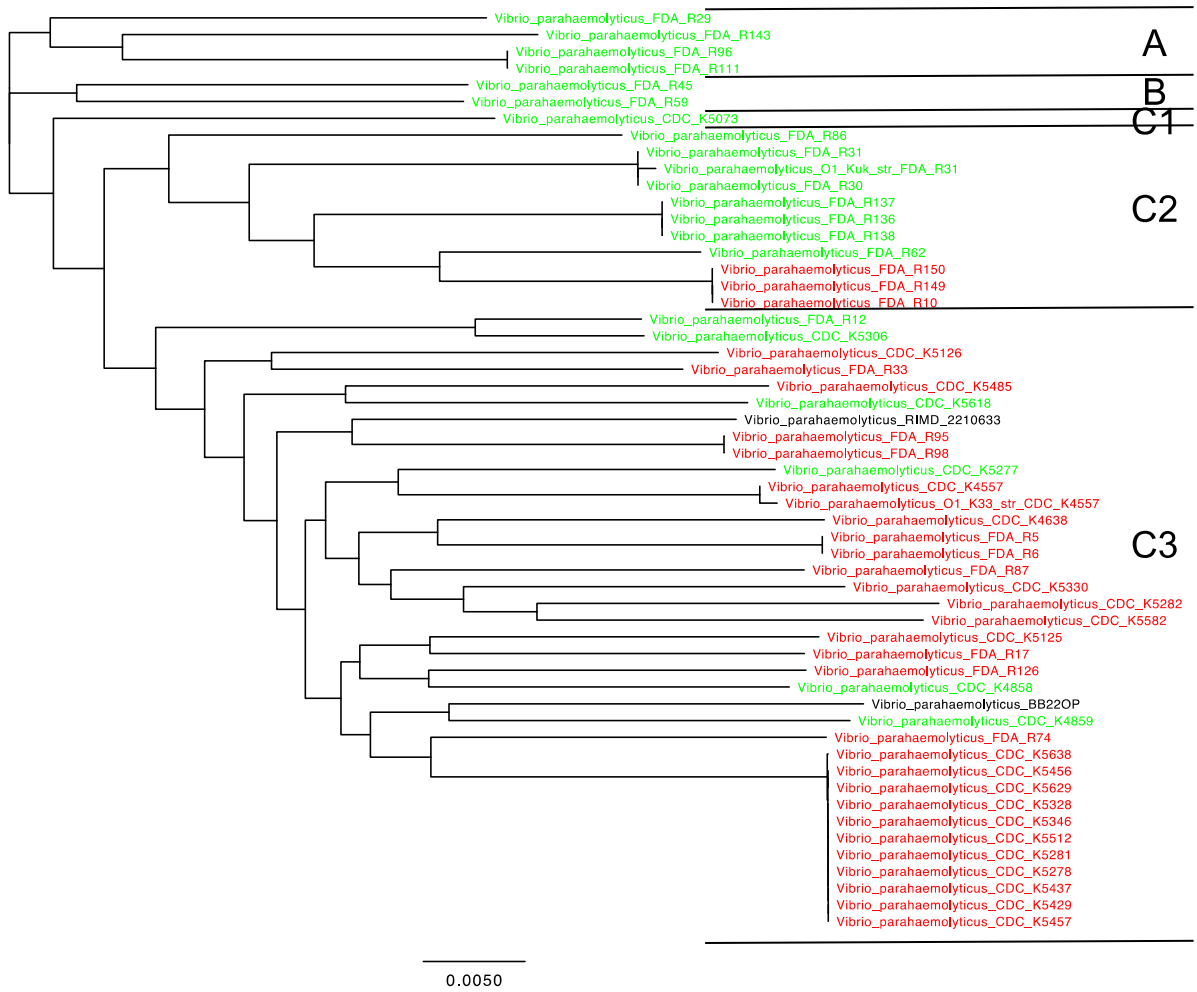


Figure 19: Rooted tree built of the alignment of concatenated sequences from the cytotoxicity data after RAxML bootstrapping; red: cytotoxic isolates, green: non-cytotoxic isolates, black: closed *V. parahaemolyticus* genomes from NCBI Genbank.

6. Discussion

6.1. Serological, biochemical and virulence characterization of 144 *V. parahaemolyticus* isolates

This section used a set of 144 *V. parahaemolyticus* isolates, including clinical isolates from across North America obtained from July 2006 to November 2007 and oyster isolates obtained from market oysters collected across the United States in 2007. Initially, these isolates were characterized for their biochemical activity, serotype, and virulence gene profile (*tdh*, *trh*, and T3SS genes) to examine their diversity. Concurrently, the perspective of this study was to identify differences between clinical and oyster isolates in order to provide insights into strain virulence on a basic molecular and/or biochemical level. Thereby, it is of great importance to recognize that there are most likely pathogenic strains among the group of oyster isolates with *tdh* and/or *trh*. As such, both oyster isolates with virulence genes and oyster isolates negative for both *tdh* and *trh*, were included in the analysis.

All strains were identified as *V. parahaemolyticus* by the presence of the species-specific gene, *tlh*. Identification of isolates based on the API 20E test was frequently erroneous as reported in previous studies (Martinez-Urtaza et al., 2006; Croci et al., 2007). The API 20E single test units were still used to biochemically characterize these isolates. Forty-six and 57% of the clinical and oyster isolates, respectively, had the ability to produce β-galactosidase (ONPG test). Historically, over 90% of *V. parahaemolyticus* strains were negative for ONPG (Kaysner and Depaola, 2004). Since the application of these test strips has decreased due to the unreliable results for species identification, the presence of OPNG may have been unreported. Previously, the presence of the *trh* gene and urease production were linked (Suthienkul et al., 1995; Iida et al., 1997). In this study, 95 of 96 (99%) isolates that produced urease also harbored the *trh* gene. Since the presence of the *ure* gene was not tested, it is possible that the three *trh*-positive strains negative for urease production

simply did not express the gene for urease production under the applied experimental conditions.

The most frequent serotype in this study was O1:Kut. This serotype is one of the serogroups associated with pandemic *V. parahaemolyticus* strains (Chowdhury et al., 2000; Chowdhury et al., 2004b). However, it has been reported that serotype O1:Kut alone is not a reliable indicator for a pandemic lineage (Chowdhury et al., 2004b). Other pandemic serotypes, such as O3:K6, O4:K68, and O1:K25, were also identified among the clinical isolates of this study, but were not found in any of the oyster isolates. In any case, this data set indicates a small number of O1:Kut strains in U.S. market oysters. The second most prevalent serotype was O4:K12, a common serotype causing human illness in Washington State (Depaola et al., 2003). In 2012, *V. parahaemolyticus* outbreaks occurred along the Northeast coast of the United States and in Spain with strains isolated from patients of the serotype O4:K12 or O4:Kut. Therefore, this strain was able to spread across the globe (Martinez-Urtaza et al., 2013a; Newton et al., 2014a). Further outbreak cases occurred in New York in 2013 associated with this specific strain (Newton et al., 2014a). All eleven of the O4:K12 isolates from 2006/2007 were from patient stool specimens. Only three were reported in the state of Washington and one by the state of New York, the majority of others were reported by states that may have received seafood products from Washington or New York even though the outbreaks had not occurred at the point of isolate collection. In this study thirteen additional serotypes were found only in clinical isolates, while nine serotypes were unique to oyster isolates. Consequently, serotype was the trait least shared by clinical and oyster isolates in this portion of the study. Throughout this thesis on virulence characterization, further experiments were planned to determine whether certain serotypes, or groups of serotypes, could indicate pathogenic potential.

The hemolysin genes, *tdh* and *trh*, are generally accepted as the main indicators of virulence for *V. parahaemolyticus* (Su and Liu, 2007; Broberg et al., 2011). Furthermore, it has been reported strains containing T3SS2 can be more virulent than those lacking the system (Caburlotto et al., 2010). Since more than 90% of clinical *V. parahaemolyticus* strains

carrying *tdh* and/or *trh* and T3SS2 were isolated from stool specimens, the hypothesis that these hemolysin genes and the presence of T3SS2 are predictive of food-borne illness risk is supported (Park et al., 2004b; Okada et al., 2009). Nevertheless, 27% of the clinical *V. parahaemolyticus* isolates of this study were negative for *tdh*, *trh*, and T3SS2. Therefore, these isolates would be considered avirulent. As such, these results question the reliability of the *tdh*, *trh*, and T3SS2 genes as indicators of virulence.

The T3SS1 is reported to be present in all strains of *V. parahaemolyticus* (Park et al., 2004b; Noriea et al., 2010). However, 10 of the *V. parahaemolyticus* isolates in this study were missing the genes VP1686 and/or VP1694 (*vscF*) of the four T3SS1 genes targeted by simplex or multiplex PCR. VP1686, which codes for the effector protein VopS, blocks the signaling cascade regulating the actin cytoskeleton of the host cell, which leads to cell rounding (Woolery et al., 2010). VP1694 encodes the needle subunit protein, which forms the needle complex of the T3SS1 (Wang et al., 2015). Since some weak amplification was observed by simplex PCR, it is possible that these strains do possess the VP1686 and VP1694 genes but have a divergent sequence that cannot be amplified efficiently with the current primer set. Based on the absence of VPA1362 (*vopB2*) in environmental isolates that contained other T3SS2α genes, a previous report suggested that *vopB2* may be a more reliable indicator of virulence than *tdh* (Noriea et al., 2010). The T3SS2α specific gene *vopB2* was amplified in all nine of the *tdh*$^+$/*trh*$^-$ clinical isolates, but not present in either of the two oyster *tdh*$^+$/*trh*$^-$ isolates, supporting Noriea et al.'s study (Noriea et al., 2010). As reported in chapter 5.1 all clinical and oyster *tdh*$^-$/*trh*$^+$ isolates amplified all four genes specific for T3SS2β (Noriea et al., 2010). Concurrently, all clinical *tdh*$^+$/*trh*$^+$ isolates and all but one oyster *tdh*$^+$/*trh*$^+$ isolate amplified all four T3SS2β genes, although a previous report suggested the that the T3SS2α and T3SS2β are absent in isolates positive for both *tdh* and *trh* (Noriea et al., 2010). This is the first study to report the presence of T3SS2β in *tdh*$^+$/*trh*$^+$ isolates. The diversity of the isolate set, as well as the use of simplex PCR to examine all isolates, might explain the difference to previous reports.

Conclusively, this study was able to characterize a diverse set of *V. parahaemolyticus* clinical and oyster isolates for biochemical differences and the distribution of pathogenicity factors. More than 25% of the clinical isolates were *tdh*$^-$/*trh*$^-$ and did not possess T3SS2. These results were among the first to indicate that the virulence of *V. parahaemolyticus* is more complex than historically believed, whereas *tdh* and/or *tdh* and T3SS2 genes are not necessarily indicative of pathogenic potential. As 17 of the serotypes were found only in clinical isolates, the serotype was a distinguishing feature of the clinical isolates. However, the variety of serotypes may be too wide for use as an indicator of virulence. This study reveals a higher level of complexity in the virulence mechanism of *V. parahaemolyticus*. Therefore, the reliability of the long-standing virulence markers as well as potential new insights into the pathogenicity mechanism will be further discussed in the following chapters.

6.2. Characterization of 144 *V. parahaemolyticus* isolates by fingerprinting and phylogenetic methods

As the findings of chapters 5.1 and 6.1 raised the question about the reliability of *tdh*, *trh*, and T3SS genes as pathogenicity markers, the need for more detailed pathogenicity investigations of *V. parahaemolyticus* arose. Therefore, higher discriminatory analyses were applied to these isolates with the goal of elucidating more reliable predictors of *V. parahaemolyticus* virulence. For general information about the relations on a genetic level, the 144 *V. parahaemolyticus* isolates were examined via the fingerprinting and phylogenetic methods PFGE, DGREA, ISR-1, MLST, and MLVA. Concurrently, further objectives of these studies were to identify which method(s) are most suitable for discrimination of *V. parahaemolyticus*.

PFGE is widely used for subtyping *V. parahaemolyticus* (Centers for Disease and Prevention, 2009; Staley and Harwood, 2010). In outbreak investigations PFGE is utilized in combination with serotyping (Newton et al., 2014a). For subtyping and source tracking of foodborne pathogens, in general, PFGE has been the “gold standard” method (Barrett et al., 2006). The primary enzyme of the PFGE protocol for *V. parahaemolyticus* is *Sfi*I (Kam et al.,

2008). However, based on previous studies a combination of patterns generated by *Sfi*I and *Not*I can be used for greater discrimination (Swaminathan et al., 2001; Cooper et al., 2006). In this thesis, the composite data set of both enzyme restriction patterns showed greater discrimination than the analyses of individual restriction patterns. Previous studies described that only 16-35% of *V. parahaemolyticus* isolates produced unique patterns (Wong et al., 2007; Dauros et al., 2011). However, only two oyster isolates from the same sample shared indistinguishable two-enzyme fingerprints in this study. Therefore, the current observations are more similar to a previous report of 94% diverse isolates producing unique pattern combinations as well as indistinguishable isolates sharing a common serotype when analyzing data from both restriction enzymes (Kam et al., 2008). Furthermore, isolates tended to cluster by their serotype. These observations added to the data presented in chapter 5.1, describe a high diversity of the isolate set and identify serotype as the most promising characteristic to distinguish between clinical and oyster isolates. The results also support the utility of PFGE in epidemiologic and phylogenetic investigations, as unrelated strains were clearly differentiated by this method.

ISR-1 has been applied to *V. parahaemolyticus* and *V. vulnificus* for subtyping (González-Escalona et al., 2005; Gonzalez-Escalona et al., 2007a; Hoffmann et al., 2010). However, compared to PFGE, it has not been used as frequently. Previously, ISR-1 patterns created four subgroups (I, II, II, and IV) when applying an analysis range of bands between 300 and 800 bp. In this thesis, more subgroups were identified and a different frequency of previously identified subgroups was observed. While pattern II was found to be as the most frequent group in a past study (Gonzalez-Escalona et al., 2006), pattern III was the most frequent pattern observed in this study of the patterns previously reported. This is likely a result of the diverse set of the isolates utilized, since Gonzalez-Escalona et al. mainly investigated pandemic isolates with the serotypes O3:K6 and O4:K68 (Gonzalez-Escalona et al., 2006). Using an expanded 50 to 2500 bp analysis for ISR-1, clusters separated based on the source of isolates (oyster and clinical) indicating a possible relatedness to pathogenic potential. A few oyster isolates are located in the clusters primarily composed of clinical

isolates, potentially indicating higher virulence of those isolates. Vice versa, some clinical isolates were grouped in the clusters primarily composed of oyster isolates, signaling a lower virulence potential. These isolates of clinical origin might be more opportunistic or have been isolated from immunocompromised individuals. Since the largest cluster held an equal mix of clinical and oyster isolates, these isolates are hypothesized to be of moderate virulence potential. Referring to these results, ISR-1 using the broader analysis scale might function as a potential screening method for differential virulence. Compared to PFGE, the ISR-1 method was generally less discriminatory. However, using a larger product size range provided a similar level of discrimination and was the only analysis to differentially cluster clinical and oyster isolates. Therefore, ISR-1 may be suggested as a rapid subtyping tool for indication of pathogenic potential. Throughout this thesis further investigations were undertaken and compared to the results of this method.

A previous study demonstrated that DGREA and PFGE clustered *V. parahaemolyticus* isolates in a similar manner (Fuenzalida et al., 2006). However, in this study DGREA did not cluster isolates similarly as the other methods mentioned above with the exception of a few serogroups. In general, the results of the DGREA analysis did not indicate a clear differentiation among serotypes, or any other characteristic. Fuenzalida et al. examined mostly pandemic O3:K6 isolates; conversely, the isolate panel of this thesis work presents higher diversity. These findings imply that DGREA has high discriminatory power that might be of use in phylogenetic studies. However, since almost 20% of the isolates were untypeable by *Nae*I, our results suggest DGREA is not a reliable method for subtyping *V. parahaemolyticus*. Therefore, DGREA was not incorporated into any further discussion sections of this thesis.

A subset of 58 of the 144 *V. parahaemolyticus* isolates was utilized for further analysis with MLST and HRM-MLVA. In the beginning of this thesis, the seven gene MLST protocol was utilized for this panel of isolates (Gonzalez-Escalona et al., 2008). Over time, *in silico* MLST was substituted for the basic MLST protocol due to the use of rapid and cost effective whole genome sequencins (WGS) (Haendiges et al., 2015). Therefore, an extended discussion on

in silico MLST after next generation sequencing of the larger set of *V. parahaemolyticus* isolates is presented in chapter 6.4 of this document.

Initially, the MLST protocol using the sequences of seven housekeeping genes (Gonzalez-Escalona et al., 2008) was selected due to the availability of a public database (http://pubmlst.org/vparahaemolyticus). Another MLST method utilizing ten housekeeping genes was available (Yan et al., 2011). Although both methods show the same level of discrimination, only the scheme of Gonzalez-Escalona et al. provides a widely applied public data depository to further compare results of this study to other described STs. At the point of publication, 22 novel STs added beneficially to the diversity of the MLST database. ST3, ST32, and ST36 were the most frequently occurring STs, as well as in the public database. The fourth most often ST observed was ST676, one of the novel STs reported in this study. ST3 and ST36 wer previously describes as part of clonal complexes CC3 and CC36, respectively (Gonzalez-Escalona et al., 2008). CC3 possessed ST3 as the ancestral ST along with ST27, ST42, and ST51 as SLVs. However, none of these STs from CC3 were identified in this subset of *V. parahaemolyticus* isolates. ST3 and ST36 were further correlated to outbreaks in the United States and Chile, including the cases of illness at the Northeastern coast of the US in 2012 (Fuenzalida et al., 2006; Martinez-Urtaza et al., 2013a; Gonzalez-Escalona et al., 2015). Turner et al. suggested ST3 to be of higher virulence potential, since a previous study using MLST on a set of clinical and environmental isolates from the Pacific Northwest region of the United States showed that some environmental isolates possessed this ST (Turner et al., 2013). In this study, no direct relationship to environmental or other clinical clades could be observed. However, chapter 6.4 will draw back on this discussion point.

For further subtyping and discrimination of similar PFGE patterns and STs, an HRM-MLVA method was developed. The HRM-MLVA method was based on an existing MLVA assay, which uses capillary electrophoresis (CE). The CE detection method counts the actual number of tandem repeats, whereas HRM cannot. However, the HRM analysis can still detect allelic variation and differentiate between otherwise indistinguishable strains, without

being able to determine the exact number of repeats. Isolates of ST3, ST32, and ST36 were in the same PFGE cluster. However, each isolate formed a unique HRM melting curve combination. Therefore, the resolving power of HRM-MLVA is similar to previous reports of CE-MLVA; Chilean isolates sharing a DGREA pattern and ST3 could be differentiated by CE-MLVA (Gonzalez-Escalona et al., 2008; Harth-Chu et al., 2009). The data of this section (chapter 5.3.2) demonstrates that the HRM-MLVA method established offers a similar discrimination compared to previous reports for the CE-MLVA method and is suitable for examination of *V. parahaemolyticus* isolates. Furthermore, the use of HRM-MLVA on the LightCycler®480 eliminates the step of electrophoretic detection, consequently reducing the potential for cross contamination by PCR amplicons. The loci amplified by MLVA are coding proteins. VPTR4 is a putative hemolysin, while VPTR3 is a putative collagenase (Kimura et al., 2008). Most of these genes could be amplified in the current isolate panel with the exception of VPTR7 (Multi A), VP1-10 (Multi B), and VPTR6 (Multi C). Nearly all clinical isolates failed to amplify the VP1-10 locus and approximately half failed to amplify the locus VPTR7. A failure to amplify VPTR7 from some shellfish isolates has been reported previously (Harth-Chu et al., 2009). Fewer than 20% of isolates amplified VPTR6 in this study. Additionally, a previous study found VPTR6 to be one of the few loci with high genetic diversity (Harth-Chu et al., 2009). Together, these data suggest that these two loci might not be suitable targets for future MLVA studies, especially for environmental isolate screening. Nonetheless, the HRM-MLVA method successfully discriminated between otherwise indistinguishable *V. parahaemolyticus* isolates.

CDC has utilized a combination of PFGE followed by MLVA to discriminate between closely related isolates in form of epidemiological investigations of STEC O157 (Hyytiä-Trees et al., 2006). Also, a similar approach combining MLST and MLVA has been used as an epidemiological tool for distinguishing between clones of *Listeria monocytogenes* (Chenal-Francisque et al., 2013). However, a combined method of PFGE, MLST, and HRM-MLVA has not been reported previously for *V. parahaemolyticus*.

By applying MLST to the isolates, identity and phylogenetic relatedness could be defined. Furthermore, the established HRM-MLVA method was able to differentiate between isolates within the same PFGE cluster or ST. This MLVA/MLST approach could be, in combination with PFGE, suitable for outbreak or evolutionary investigations.

In summary, the combination of PFGE and MLST with MLVA is a promising tool for outbreak investigations or subtyping *V. parahaemolyticus*. ISR-1 was able to distinguish between clinical and oyster isolates and might be a possible method for screening of virulence potential.

6.3. Discussion of the phenotypic examination of *V. parahaemolyticus* isolates using cytotoxicity assays

A requirement for the identification of a new pathogenicity marker is the determination of a virulent phenotype. Currently, mammalian cell cytotoxicity assays are the most frequently used methods to investigate virulent phenotypes of pathogenic bacteria. However, there are very few cytotoxicity assays to understand the virulence mechanism of *V. parahaemolyticus* (Le Roux et al., 2015). In one of the previous studies, an *in vitro* cell culture model demonstrated that *V. parahaemolyticus* strains could invade and adhere to non-phagocytic cells, especially intestinal cells. There, *V. parahaemolyticus* forms a reactive oxygen species (ROS) based host resistance for its own survival inside the intestinal tract and thereby supports tissue infection by disrupting the anti-oxidative stress response of the host (El-Malah et al., 2014). Another *in vitro* cell culture model was able to detect pathogenic *V. parahaemolyticus* strains while showing a differentiation of clinical and non-clinical isolates via a cytotoxicity assay in different cell lines (Yeung et al., 2007). However, this study only looked at a small set of isolates and did not include the virulence gene *trh*. In earlier studies, HeLa cells have been utilized to investigate cytotoxicity of *V. parahaemolyticus* (Zhou et al., 2010; Liu et al., 2013).

As part of this thesis another cytotoxicity assay was developed. The relative cytotoxicity was measured by detecting the LDH release of the host cell lines HeLa and Caco-2. The purpose was to evaluate which cell line of the two utilized is the most appropriate for a cytotoxicity assay and can a correlation between isolate cytotoxicity with the virulence gene profile of *tdh* and/or *trh* and the T3SS2 be observed. Since not only the presence, but also the expression, of virulence genes is of importance, this study examined the effect of expression of TDH, measured by the KAP-RPLA kit, towards cytotoxicity, which has not been previously reported for *V. parahaemolyticus*. Furthermore, this was the first study looking at correlations of isolates with naturally diverse virulence gene profiles, while earlier studies primarly focused heavily on knockout models.

Since Caco-2 cells are intestinal cells and build tight junctions while growing a confluent monolayer, this cell line was chosen as one of the target models for this cytotoxicity assay (Sambuy et al., 2005). As the second target cell line, HeLa cells were selected. HeLa cells originate from cervical adenocarcinoma and are the most widely used human cancer cell line in biomedical research fields and previous *V. parahaemolyticus* research (Masters, 2002; Grimm, 2004; Yeung et al., 2007). For test *V. parahaemolyticus* isolates, the previously characterized isolates from 2007 as well as an additional set of 20 *V. parahaemolyticus* strains isolated from shellfish (oyster and clams) and clinical cases in 2012 were examined.

As a significant difference between the cytotoxicity of shellfish and clinical isolates was observed in Caco-2 cells, but not HeLa cells, the Caco-2 cell response may more accurately reflect the pathogenic potential of *V. parahaemolyticus* isolates. *V. parahaemolyticus* affects the intestine (Broberg et al., 2011), which might explain the significant difference in cytotoxicity between oyster and clinical isolates in Caco-2 cells in this study. Based on the significant difference between oyster and clinical isolates and the fact that HeLa cells, even though they are epithelial cells as well, originated from cervical carcinoma and not from intestinal tissue, we hypothesize that HeLa cells are not the most fitting cell line to study *V. parahaemolyticus* effect on human pathogenicity. Seeing a difference in clinical and oyster isolates is of particular interest as described in chapter 6.1 and 6.2. Furthermore, recent

studies for *V. vulnificus* showed a distribution of environmental and clinical isolates correlated to their virulence gene profile (Rosche et al., 2005). Similar findings for *V. parahaemolyticus* could lead to identification of new virulence factors, based on the assumption that environmental isolates are generally less pathogenic than clinical isolates. In the Caco-2 cell cytotoxicity assay, as most oyster isolates were non-cytotoxic, one may assume these are less virulent strains while the cytotoxic environmental strains are the more pathogenic group. The cytotoxic clinical isolates, on the other side, consist of isolates with higher virulence potential. The non-cytotoxic clinical strains might have originated from immunocompromised patients and, therefore, show less pathogenic potential.

The statistical analysis suggests that associations of virulence factors with cytotoxic potential are moderated by the isolates origin. Among clinical isolates only the presence of T3SS2α was significantly associated with cytotoxicity in the Caco-2 cell assay, even though a significant association was observed overall between cytotoxic potential and both *tdh* and T2SS2α. Therefore, it was suggested that while cytotoxicity is, in some cases, associated with the presence of virulence genes, additional factors contribute to cytotoxic potential, and presumably virulence. The results of the KAP-RPLA kit, indicating TDH production, further strengthens this conclusion. A previous study using cytotoxicity assays targeting HeLa, Henle 407, L2, mouse-macrophage and Caco-2 cells presented similar findings where *tdh*$^{+}$ and *tdh*$^{-}$ strains showed no significant difference in cytotoxicity (Yeung et al., 2007); however, that study did not investigate the correlation towards *trh* and whether the isolates expressed TDH.

With a particular focus on outbreak-related strains and the influence of serotype on pathogenicity, the question was if the isolate's serotype is correlated to cytotoxicity. As mentioned in chapter 6.1 outbreaks occurred in the United States related to the serogroup O4:Kut and O4:K12 in 2012 and 2013 (Newton et al., 2012; Martinez-Urtaza et al., 2013a). Therefore, comparisons were focused on the effect of isolates with these serotypes as well as the serotypes of the pandemic group (O1:K25, O1:Kut, O3:K6, and O4:K68). Over 90% of clinical O4:K12 and O4:Kut isolates were cytotoxic for both cell lines; with a significant

positive association with cytotoxic potential in both cell assays. Of the pandemic serotypes ~80% of the O1:Kut clinical isolates (n=11) were cytotoxic or semi-cytotoxic to both cell lines, and were one of the most common serotypes in this study. However, of the shellfish isolates with serotypes O1:Kut (n=7) and O4:Kut (n=5), 50% and ~60% were cytotoxic, respectively. Further in the pandemic serogroup, the one O1:K25 isolate was semi-cytotoxic (in HeLa cells) or cytotoxic (in Caco cells), while the one O4:K68 isolate was semi-cytotoxic or non-cytotoxic in the HeLa or Caco cell assay, respectively. In this study, only 50% and 25% of O3:K6 isolates (n=4) were cytotoxic to HeLa and Caco-2 cells, respectively. These serotypes associated with the pandemic clade did not show a statistically significant association with cytotoxicity towards HeLa (p=0.205) or Caco-2 cells (p=0.100). In contrast to these results, a previous study comparing clinical O3:K6 and non-O3:K6 isolates showed higher cytotoxic effects for O3:K6 isolates (Yeung et al., 2007). As the current data set consists of a more diverse collection of serotypes, the correlation of cytotoxicity and serotype is more complex, and likely the reason for the discrepant findings. This data suggests that the serotypes O4:K12/O4:Kut, as known outbreak related serovars, are particularly influential to the determination of a significant association with serotype and cytotoxicity among this collection of isolates overall. The serotype alone may not be predictive of strain virulence; however, this analysis cannot definitely exclude the involvement of serotype in pathogenicity.

In conclusion, in this study cytotoxicity assays targeting HeLa and Caco-2 cells were developed and utilized to examine 164 *V. parahaemolyticus* isolates. Based on the data, Caco-2 cells seem to be the more suitable cell line for this type of cytotoxicity assay. Furthermore, the effect of the presence/absence of the different virulence genes and serotypes on the pathogenic potential of the different *V. parahaemolyticus* isolates was investigated. While cytotoxicity may be partially associated with *tdh* and T3SS2, the results suggest that there are other factors, which contribute to the cytotoxicty, and presumably pathogenicity, of this organism. These and additional *V. parahaemolyticus* isolates can be further examined to identify the reason for differential cytotoxicity reactions. Since a group of less virulent oyster isolates and a group with more virulent clinical isolates were found using

Caco-2 cells, these isolate sets can be compared to their whole genome sequence to identify genetic differences, which could play a role in pathogenicity (chapter 5.7).

Reasons why *V. parahaemolyticus* strains with identical genetic toxin profile produce variable amounts of cytotoxicity and whether additional markers can be identified for the differentiation between high and low cytotoxic strains are still unknown. Complex and dynamic regulatory processes, the condition of the host, as well as the composition of the food may contribute to high variability in the virulence mechanism.

6.4. Discussion of the next-generation sequencing analysis of *V. parahaemolyticus* isolates

The first two complete bacterial genome sequences were published in 1995. Nowadays by using third-generation DNA sequencing, it is possible to completely sequence a bacterial genome in a short time frame (Land et al., 2015). In the last few years' next generation sequencing (NGS) has been applied to better understand population dynamics and mechanisms contributing to increased virulence among foodborne bacterial pathogens. For instance, the outbreak strain of *E. coli* O104:H4 in Germany could be differentiated by NGS from other O104:H4 strains (Rasko et al., 2011). Furthermore, NGS demonstrated a close relationship of the Haitian *V. cholerae* outbreak strains to strains isolated in Bangladesh and no relatedness to isolates from South America (Chin et al., 2011). The availability of additional *V. parahaemolyticus* genomes in public databases, such as NCBI, can assist future analysis in identifying new genes related to pathogenicity and source tracking in outbreak scenarios (Haendiges et al., 2015). Previous sequencing and characterization of an environmental *V. parahaemolyticus* isolate not carrying the virulence-associated genes has demonstrated a significant genetic similarity to disease-associated clinical isolates. It was suggested additional genome sequencing of a diverse set of clinical and environmental *V. parahaemolyticus* isolates with different serotypes and virulence gene combinations might lend further insight into the ability to transition from an environmental niche and to emerge as

pathogens (Hazen et al., 2015). Compared to PFGE and MLST, NGS provides a much higher resolution (Haendiges et al., 2015). Since PFGE, ISR-1, DGREA and MLST have presented *V. parahaemolyticus* as a very diverse organism, WGS was used to explore this diversity and relationships in between this species. These investigations provide the framework for more detailed studies on elucidation of the pathogenicity mechanism of *V. parahaemolyticus*.

With this study, the sequences add beneficially to the sequences currently available in Genbank. This is the first large-scale study showing the diversity of an isolate set from all coastlines of the United States. Studies before were limited to local areas in the United States (Xu et al., 2015) and small sets of sequences (Li et al., 2014). In a broader spectrum *V. parahaemolyticus* has shown a distribution in gene pools across the globe (Cui et al., 2015). At the beginning of this study, the perspective was that utilizing a broad set of isolates, which has been beneficial for population studies, could be of use for a study of virulence relatedness, as a small number of isolates might not encompass the diversity of *V. parahaemolyticus* strains. The large data set improves the probability for identifying differences in the *V. parahaemolyticus* genome of several strains.

From the 144 isolates used for WGS, the contig counts ranged from 73 – 297 in the resultant sequences. Since *V. parahaemolyticus* is an organism with two chromosomes, with eleven copies of rRNA operons, and similar regions, assembly of raw sequences is complicated. These rRNA coding regions can cause assembly difficulties in second-generation sequencers, such as the Illumina HiSeq, due to the uncertainty of their exact position. Therefore, this number of sequences presents such a wide range of contigs (Miyamoto et al., 2014).

Second-generation sequencing produces draft genomes with a higher number of contigs and varying quality can present major challenges in genome assembly and closure (Land et al., 2014). However, third-generation sequencing technologies, such as the PacBio RSII platform, offers longer read lengths and can thereby provide closed, or finished, genomes (Miyamoto et al., 2014). In addition to the 132 shotgun sequences generated in these

studies, two closed genomes completed during this thesis add beneficially to the public database for *V. parahaemolyticus*.

The WGS data were used to identify STs, which were further differentiated by kSNP analysis and phylogeny. However, this data set showed a high diversity regarding the number of STs identified in the isolates. As described in chapter 6.2 a subset of the isolates used in this part of the thesis were investigated by MLST. When examining the larger set of 144 isolates by *in silico* MLST, the most frequent STs are still ST3 and ST36. Turner et al. reported that ST3 in oyster isolates is associated with pathogenic potential (Turner et al., 2013). While all ST3 isolates were of clinical origin in this research and no direct relation to oyster isolates was present, a confirmation of that report cannot be made. Furthermore, in the past, *recA* of *V. parahaemolyticus* has been shown to be untypeable by MLST (Gonzalez-Escalona et al., 2015; Ludeke et al., 2015). A recent study found a fragmentation of the original lineage-specific *recA* gene due to the insertion of a 30 kb long genomic island isolates from Peru (Gonzalez-Escalona et al., 2015). After reconstruction, that isolate was identified as a precursor to the ST189, a ST previously only reported in Asia. This finding supports the existence of recurrent transoceanic spreading of pathogenic *V. parahaemolyticus*. Genetic spreading complicates the study of virulence mechanisms. In this study's isolate set four strains were untypeable for *recA,* and showed an insertion in the genomic island. However, this genomic island differed from the genomic island previously reported by Gonzalez et al.. After further analysis of the new genomic insertion in the *recA* gene, the genomic islands were ascribed to two novel STs (3 isolates) and ST1184 (1 isolate), but not ST189. ST1184 occurred in *V. parahaemolyticus* isolates in the United States in 2012, but has not been recorded since February 2016 according to the PubMLST database. Therefore, isolates from the United States are showing the same complication of genetic spreading for pathogenicity studies, especially since the genetic variation in the *recA* gene is apparently becoming more diverse. The ML tree from the kSNP matrix revealed an intermixing between isolates mainly based on their isolation source (clinical or oyster isolates) (Figure 14).

However, some clusters were predominantly of isolates from one source. Furthermore, serotype and ST were the base of the assembly of the subgroups. Nevertheless, as observed in cluster C1 and C2, predominately clinical isolate clusters can contain a small number of oyster isolates, while clusters predominately of oyster isolates contain few clinical isolates (cluster A1). Similarities within the clusters are mainly based on their reporting/harvest state or serogroup. Since some of these isolates share the same serotype or reporting state, the hypothesis could be made these isolates could be related through illness and source of infection as seen in previous reports (McLaughlin et al., 2005). However, the necessary patient data and information on cause of infection for each strain are missing on these isolates to confirm this hypothesis. As mentioned before, isolates of the serotype O4:K12/Kut and of ST36 were associated with outbreaks in 2012 and 2013 across the East coast of the United States (Martinez-Urtaza et al., 2013a; Newton et al., 2014a). Isolates grouped in the same clade of the kSNP tree with the same ST and serotype as the outbreak strains. Since these isolates of clinical origin formed their own cluster, the hypothesis can be made that these isolates are of high virulence, assuming isolates from patients are the more pathogenic strains compared to isolates from an oyster. Although, it should be noted that some oyster isolates could possess a high virulence potential. Additionally, the clade with a mixture of clinical isolates and several related oyster isolates emerged from this group of O4:K12/Kut patient isolates. This supports the assumption that these isolates are genetically different from those in the oyster clade (A), providing an indication of pathogenic potential. On the other side, the phenomenon of clinical isolates clustering with the majority of oyster isolates can serve as an argument for the theory of presence of less virulent, or opportunistic, clinical isolates that may have originated from immunocompromised patients. However, the results of this large of data set concluded that the large sample size actually complicates the analysis, smaller sets may be needed as there could be spatial or temporal differences that lead to differential virulence. It is also possible that differential virulence is influenced by gene regulation, rather than just genomic content.

The determination of virulence potential and finding an environmental isolate as a pathogen is not easily done, since current models for disease and infection are limited (Xu et al., 2015). Therefore, this thesis also included cell culture assays and attempted to correlate cytotoxic phenotype with genomic data (chapter 5.7).
As described in the previous chapters, (chapter 5.2 and 6.2) ISR-1, PFGE, and MLST were applied to this isolate set as subtyping tools. Although fingerprinting methods have been used for a long time, thoroughness and resolution of these methods are limited. Since the analysis of ISR-1 with an examination range between 50 and 2500 bp could differentiate between clinical and oyster isolates the results were compared to the kSNP phylogeny (Figure 20).
ISR-1 is less discriminatory than PFGE (Ludeke et al., 2014); this is also observed in the association with the kSNP phylogeny. The differentiation of clinical and oyster isolates is more distinct in the kSNP ML tree than in the ISR-1 dendrogram.
Since PFGE grouped isolates based on their serotype the combined PFGE patterns were also compared to the kSNP analysis (Figure 21). PFGE was, as described in chapter 5.2 and 6.2, unable to distinguish between certain strains; kSNP analysis on the other hand provides stronger insights into diversity of strains as supported by (Haendiges et al., 2015). Also, indistinguishable PFGE patterns of *Salmonella enteridis* could be differentiated by kSNP analysis (Allard et al., 2013). However, PFGE cannot provide details of the genetic relationships among isolates (Foxman et al., 2005).
Multigene alignments for MLST analysis can be very long, but kSNP matrices are built of SNPs and are therefore shorter and easier to process. Furthermore, multigene alignments in a diverse set of sequences can cause loss of important data (Allard et al., 2013; Timme et al., 2013). As mentioned above MLST was also unable to distinguish between certain strains. However, the kSNP analysis can differentiate between highly related or identical ST (Figure 22).

KSNP analyses can overcome the challenges of ISR-1, MLST and PFGE. However, MLST and PFGE are still well applied pre-selection tools, especially in outbreak cases and/or to screen a large collection of isolates in a short time period.

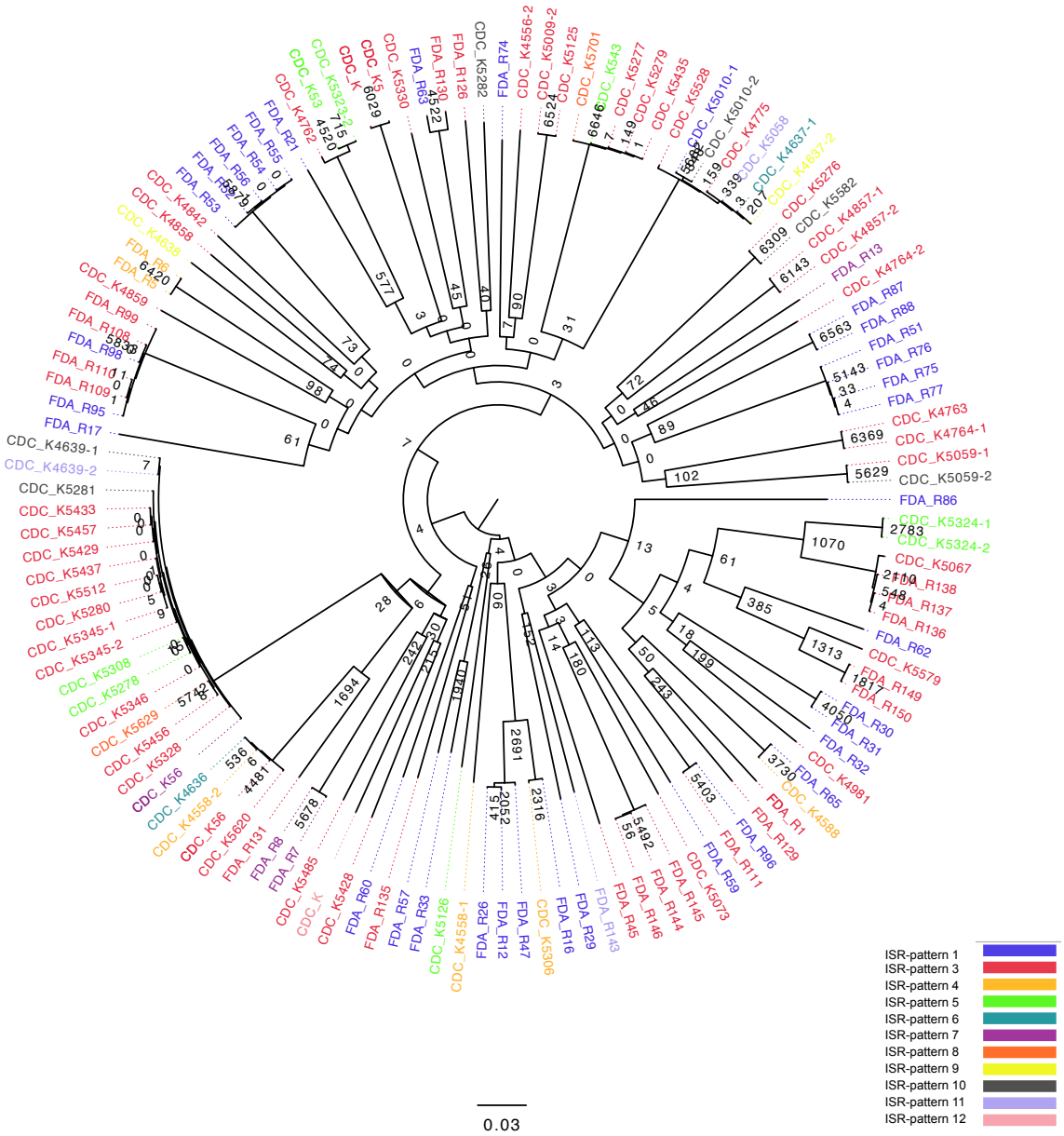


Figure 20: Comparison of ISR_1 patterns (50 – 2500 bp) towards kSNP phylogeny, visualized in FigTree 4.1.2., the Isolate_ID is colour-coded by ISR_1 pattern (see figure legend), Isolate_ID: FDA = oyster isolate, CDC = clinical isolate.

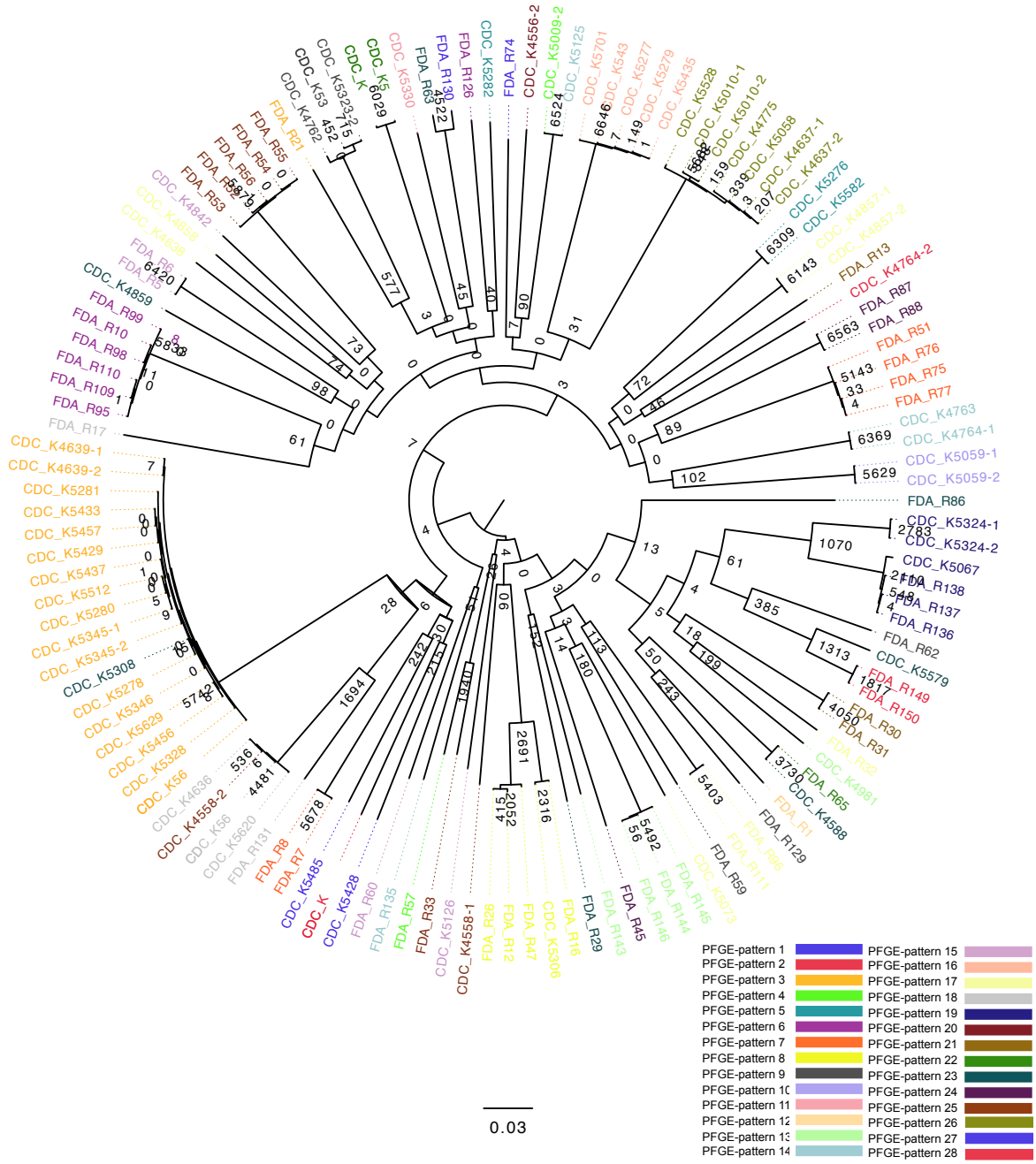


Figure 21: Comparison of general PFGE patterns of the combined analysis towards kSNP phylogeny, visualized in FigTree 4.1.2., the Isolate_ID is colour-coded by general PFGE pattern (see figure legend), Isolate_ID: FDA = oyster isolate, CDC = clinical isolate.

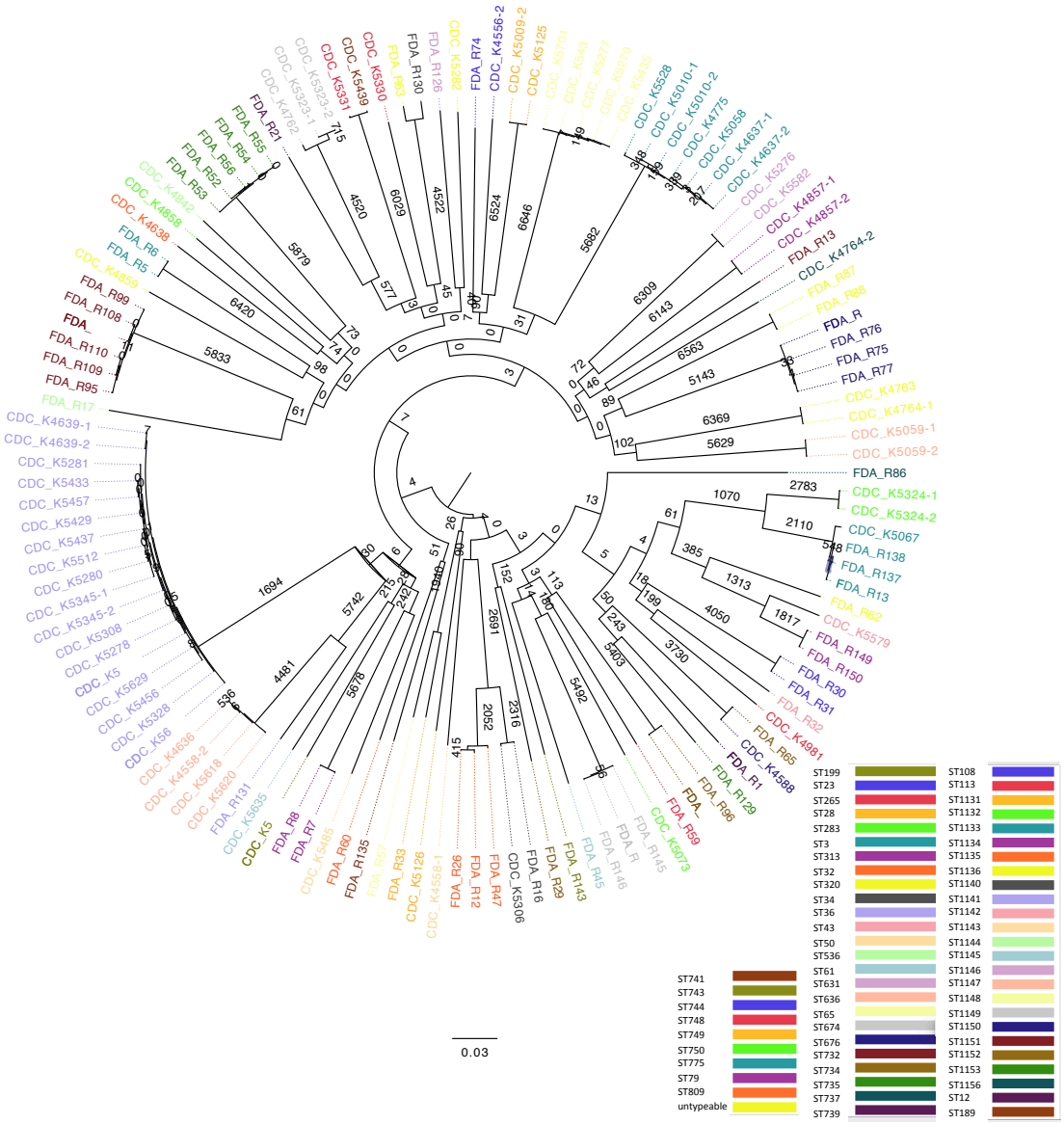


Figure 22: Comparison of general MLST ST towards kSNP phylogeny, visualized in FigTree 4.1.2., the Isolate_ID is colour-coded by general PFGE pattern (see figure legend), Isolate_ID: FDA = oyster isolate, CDC = clinical isolat

Further in this study, the presence of the existing virulence genes *tdh* and *trh* was examined in the WGS data by applying bioinformatics techniques, initially only for confirmation of previous real-time PCR results. However, inconsistencies emerged when real-time PCR results were compared to NPGAAP annotations and BLAST results against SwissProt sequences, so other approaches were investigated, including Interproscan and Blasting against sequences chosen based on previous detailed biochemical characterization of the protein (Yanagihara et al., 2010; Ohnishi et al., 2011). The NPGAAP annotations indicated

many of our strains contained a hemolysin activation protein, but none possessed the *tdh* or *trh* genes. Interproscan results generally agreed with the real-time PCR results except that Interpro groups the two proteins into a TDH family, consistent with previous studies indicating that TRH is similar to TDH (Nishibuchi et al., 1989). Since Interpro pools TDH and TRH into one group, it was expected that TDH would be present in all *tdh*$^+$/*trh*$^+$, *tdh*$^-$/*trh*$^+$, and *tdh*$^+$/*trh*$^-$ isolates, which was generally true. However, this tool does not differentiate between the two proteins. While Swiss-Prot's *trh* sequence only correlated with *tdh*$^+$/*trh*$^+$ isolates, *tdh*_2 mostly with *tdh*$^+$/*trh*$^+$, *tdh*$^-$/*trh*$^+$, and *tdh*$^+$/*trh*$^-$ isolates and *tdh*_1 was negative for all isolates. However, a one-way BLAST analysis using *tdh*_2 and *trh* sequences previously subjected to detailed previous work gave results with strong agreement to real-time PCR results (Yanagihara et al., 2010; Ohnishi et al., 2011). While tools such as BLAST have been used to do sequence similarity searches for decades this study serves as a reminder that it is not only important to use appropriately stringent e-values (Altschul et al., 1990; Altschul et al., 1997), but also to select comparison sequences from public databases carefully. A phylogenetic analysis of *tdh* and *trh* sequences has shown high variability in these genes among several *V. parahaemolyticus* strains (Bhowmik et al., 2014) making annotation of these genes difficult. In addition, previous studies suggest transcription levels of these genes may vary, which would make linking them as causal factors to pathogenicity based on mere presence-absence troublesome (Bhowmik et al., 2014).

Genome annotation is mostly done by automated pipelines. However, the accuracy of such automation has been questioned since the start of WGS; errors could occur on different levels such as during sequencing, as a result of gene-calling procedures, or in the process of assigning gene functions (Poptsova and Gogarten, 2010). A previous study comparing *E. coli* sequences found errors in the earlier annotated genomes, and the biggest concern is the need to correct these errors quickly in the large databases since these are used as references (Poptsova and Gogarten, 2010). Based on the examples of the *tdh* and *trh* genes, along with the historical perspective of annotation issues in the public databases, the presence/absence data for other genes, and putative proteins, in this study cannot be made

confidently at this point. In earlier stages of whole genome analysis, errors in microbial genomes have been demonstrated for *Vibrio fischeri*, mainly originating from point mutations and insertions; these small nucleotide changes can lead to large protein errors (Mandel et al., 2008). Distinguishing homologous relationships between proteins from sequence similarity is still a challenging research topic (Ochoa et al., 2015). However, the difficulties in the computational analysis of such a large dataset might indicate that a search only based on gene presence/absence might not be the best approach. On the amino acid level, single nucleotide changes can influence the protein outcome tremendously. Since sequencing of *V. parahaemolyticus* strains has been less frequent in the last couple of years, compared to other organisms like *Escherichia coli* and *Salmonella*, there is a lack of well-annotated closed genomes and proteogenomic research. Since some proteins of unknown function are classified based on similarities to other better characterized proteins from other species in the annotation software, the actual function of a *V. parahaemolyticus* protein might be unidentified or misidentified due to lack of characterization (Stein, 2001). A comparison of proteins across species might be helpful for functional annotation, since genes can undergo evolutionary changes in functional domains (Stein, 2001). Manual curation of protein sequences could improve existing annotations, especially in the case of misleading annotations. PCR can be used to distinguish different alleles of single genes, however, it is not totally reliable for calling sequences different genes without confirmed protein expression experiments.

6.5. Discussion of the comparison of next-generation sequencing and phenotypic analysis of *V. parahaemolyticus* isolates

To gain a better understanding of the virulence mechanism of *V. parahaemolyticus*, it is important to determine their phenotypic features. Phenotypic information can been inferred directly from the whole genome sequence (Amaral et al., 2014). Phenotype here mostly relates to diagnostic features such as indole production, Voges-Proskauer reaction, or

fermentation of sugars. However, the focus of this study's characterization was in the unknown part of *V. parahaemolyticus'* phenotype. Here, phenotype functioned as the ability to induce a cytotoxic effect in cell culture. The analysis of chapter 5.7 was conducted using a smaller set of sequences to simplify the computational analysis. A problem in identifying the appropriate number of genomes for analysis is that small isolate sets seem to only reflect a section of the full diversity of *V. parahaemolyticus*. Increasing the number of genomes increases the probability of finding more unique genes (Li et al., 2014). These unique genes can just by chance be in only the phenotype of interest, while a smaller and smaller number of genes would correspond exclusively to a given phenotype as the genome count increases. This is the first report on a comparative genome-phenotype analysis for *V. parahaemolyticus*. It has been shown that investigating the genetic diversity of environmental isolates not carrying the known virulence-associated genes can yield insight into the emergence of human disease-associated *V. parahaemolyticus* (Hazen et al., 2015). Studies investigating virulence factors of *V. cholerae* concluded that the study and comparison of the genomic sequences between pathogens and their non-virulent counterparts could help discover genes encoding both the classical virulence factors and those encoding novel virulence factors (Mukherjee et al., 2014). After phylogenetic analysis of the concatenated WGS sequences, a separation based on the cytotoxicity of isolates was observed (Figure 19). After sequencing *V. vulnificus* strains, possible virulence factors could be identified because of their presence in mouse-virulent strains only (Gulig et al., 2010). Similarly, the sequences from this study were further analyzed for specific genes eventually related to cytotoxicity by forming homologous groups (Figure 15 and Figure 16).

In this chapter, no specific genes could be identified in all of the cytotoxic isolates that were not present in the non-cytotoxic isolates and the other way around. A similar phenomenon has been seen before for *V. parahaemolyticus* due to significant genetic similarities between clinical and environmental isolates of the serotype O3:K6 using a smaller isolate set (Hazen et al., 2015). However, groups of genes could be assigned in correlation to phenotypic results (Figure 17). COG databases have been a popular tool for functional annotation; their

use has shown the possibility of identifying missed genes and improving genome annotation. The availability of functional assignments for some conserved proteins of the category “function unknown” have been reported (Galperin et al., 2015). Since numerous proteins of unknown functions occurred in the *V. parahaemolyticus* genomes of this study, the annotation software generally classified them into more manageable groups or families, in some cases based on similarities to better characterized proteins of other species (Stein, 2001). In a previous study, many of the genes identified exclusively to a particular genome were hypothetical or were similar to genes of mobile genetic elements including plasmids and phage, highlighting the contribution of mobile elements such as plasmids and phage to the diversification of *V. parahaemolyticus* (Hazen et al., 2015). Here, groups of genes exclusive to cytotoxic isolates were assigned to “general function prediction only”, “transcription”, “energy production and conversion”, “posttranslational modification, protein turnover, chaperones”, “coenzyme transport and metabolism”, “defense mechanism” and “cell cycle control, cell division, chromosome partitioning”; while the only group exclusively to non-cytotoxic sequences was “cell wall/membrane/envelope biogenesis”.

In a previous report, new prophage elements were found harboring genes responsible for fitness and survival in specific environments (Li et al., 2014). *V. cholerae*, *E. coli* and *S.* typhimurium, for example, undergo an adaptive stress response while they pass the gastric acid barrier. This allows the organisms to survive the acidic environment and enhance virulence later. The adaptive stress response functions as a fitness factor (Faruque et al., 2004). Since genes of cytotoxic isolates are present in the functional group for defense mechanism, similar fitness factors might be present in *V. parahaemolyticus*. However, most of the groups of gene function are related to protein chemistry, which emphasizes once more the need for additional research on the protein level for *V. parahaemolyticus*. As an example, a transcriptional activator ToxT regulates the main virulence factors of *V. cholerae*, cholera toxin and toxin-regulated pilus (Weber and Klose, 2011). Therefore, genes belonging to the functional groups of translation or transcription or post-translational modification could function as potential virulence factors.

The SNP comparison of cytotoxic and non-cytotoxic sequences revealed that cytotoxic strains possess more SNPs than non-cytotoxic isolates. As described in chapter 6.4, a differentiation based on SNPs was also observed between clinical and oyster isolates. In bacterial pathogens, SNPs have been discovered that belong to the category of pathogenicity-enhancing mutations in regulatory or structural genes that could provide a selective advantage during infection or long-term evolution of virulence (Sokurenko et al., 1999). Studies of variations of adhesins in *E. coli* and *S.* typhimurium have shown the impact of SNPs on host specificity up to pathogenic behavior (Weissman et al., 2003). Therefore, the SNPs found in the *V. parahaemolyticus* isolates may further influence the pathogenic potential of the organism.

Altogether, a differentiation of cytotoxic and non-cytotoxic isolates could be made based on the phylogenetic analysis of concatenated sequences of a subset of *V. parahaemolyticus* isolates. Even though no specific genes could be found in all of the cytotoxic isolates that were not present in the non-cytotoxic isolates and vice versa, groups of genes could be assigned in correlation to phenotypic results. This data adds beneficially to future research, since areas of interest have been narrowed down to certain gene groups. Genes of the functional groups might be able to function as additional virulence factors. Further, identified SNPs in cytotoxic strains can add to the understanding of the pathogenic potential of *V. parahaemolyticus*.

7. Prospect

Current findings have added beneficially to the search for new pathogenicity factor(s) for *V. parahaemolyticus*. As the levels of *V. parahaemolyticus* infections have been increasing with the years, it is of great importance to be able to calculate the risk of pathogenic potential as well as using more reliable virulence markers for outbreak investigations and source tracking. Therefore, it would be of great interest to further examine the virulence mechanism of this organism. Identified genetic differences minimized to actual multiple genes or a single gene can be characterized in further cell culture experiments. These genes can be confirmed as contributing to virulence through knock-out experiments.

Furthermore, this research work has revealed some difficulties annotating certain genes of the *V. parahaemolyticus* whole genome sequences. Closing further genomes would add beneficially to the public databases and would support further investigations to solve the annotation hurdles. Moreover, the use of protein analysis to confirm annotated genes as the expressed protein and its sequence should be applied. Comparative proteogenomics could be an approach to find genetic differences between virulent and non-virulent strains. It is of great importance to connect the genome content to protein function. The influence of the surrounding ecosystem of environmental *V. parahaemolyticus* strains upon protein expression can be investigated by applying comparative proteogenomics as well. Experimenting with protein expression in different growth conditions, such as in cell culture models, can also further characterize clinical strains.

Phenotypic investigations can be taken to an advanced level by developing animal models. Animal models have been reported to be the closest model to predict the phenotype of virulence. A model has been successfully developed for *V. vulnificus,* encouraging the design of one for *V. parahaemolyticus*.

8. Materials and Methods

8.1. Bacterial strains and cell lines

8.1.1. *Vibrio parahaemolyticus* isolates from 2006/2007

In 2006/2007 a retail study for oysters was conducted in Northern America. Sixty-seven *V. parahaemolyticus* isolates were collected from this study and are listed in Table 10.

Table 10: Oyster isolates used during this thesis work

Isolate ID	Month	State	Isolate ID	Month	State
FDA_R2	January	Texas	FDA_R74	August	Virginia
FDA_R5	March	Texas	FDA_R75	August	Virginia
FDA_R6	March	Texas	FDA_R76	August	Virginia
FDA_R7	March	Texas	FDA_R77	August	Virginia
FDA_R8	March	Texas	FDA_R86	August	Florida
FDA_R10	April	Florida	FDA_R87	August	Florida
FDA_R12	April	Louisiana	FDA_R88	August	Florida
FDA_R13	April	Louisiana	FDA_R94	August	Prince-Edward-Islands (Canada)
FDA_R16	May	Florida	FDA_R95	August	Prince-Edward-Islands (Canada)
FDA_R17	May	Florida	FDA_R96	August	Prince-Edward-Islands (Canada)
FDA_R21	May	Texas	FDA_R97	August	Prince-Edward-Islands (Canada)
FDA_R26	June	New Jersey	FDA_R98	August	Prince-Edward-Islands (Canada)
FDA_R29	June	Florida	FDA_R99	August	Prince-Edward-Islands (Canada)
FDA_R30	June	Florida	FDA_R100	August	Prince-Edward-Islands (Canada)
FDA_R31	June	Louisiana	FDA_R108	August	Prince-Edward-Islands (Canada)
FDA_R32	June	Louisiana	FDA_R109	August	Prince-Edward-Islands (Canada)
FDA_R33	June	Louisiana	FDA_R110	August	Prince-Edward-Islands (Canada)
FDA_R42	July	Washington	FDA_R111	August	Prince-Edward-Islands (Canada)
FDA_R45	July	Washington	FDA_R125	October	Florida
FDA_R47	July	Alabama	FDA_R126	October	Florida
FDA_R51	July	Alabama	FDA_R129	October	Florida
FDA_R52	July	Washington	FDA_R130	October	Florida
FDA_R53	July	Washington	FDA_R131	October	Florida
FDA_R54	July	Washington	FDA_R135	November	South Carolina
FDA_R55	July	Washington	FDA_R136	November	South Carolina
FDA_R56	July	Washington	FDA_R137	November	South Carolina
FDA_R57	July	Washington	FDA_R138	November	South Carolina
FDA_R59	July	Maine	FDA_R143	December	Florida

Table 10: *continued*

Isolate ID	Month	State	Isolate ID	Month	State
FDA_R60	July	Maine	FDA_R144	December	Florida
FDA_R61	July	Maine	FDA_R145	December	Florida
FDA_R62	July	Maine	FDA_R146	December	Florida
FDA_R63	July	Maine	FDA_R149	April	Florida
FDA_R64	July	Maine	FDA_R150	April	Florida
FDA_R65	July	Maine			

Additionally, 77 isolates were added from clinical sources from 2007. These were provided from the CDC (Table 11).

Table 11: Clinical isolates used in this thesis

Isolate ID	Reporting Date	Reporting State	Source of Isolate	Isolate ID	Reporting Date	Reporting State	Source of Isolate
CDC_ K4556-1	October (2006)	Louisiana	Wound	CDC_ K5280	July (2007)	Washington	Stool
CDC_ K4556-2	October (2006)	Louisiana	Wound	CDC_ K5281	July (2007)	Washington	Stool
CDC_ K4557	September (2006)	Louisiana	Stool	CDC_ K5282	May (2007)	Hawaii	Other
CDC_ K4558-1	August (2006)	Louisiana	Wound	CDC_ K5306	July (2007)	Georgia	Stool
CDC_ K4558-2	August (2006)	Louisiana	Wound	CDC_ K5308	May (2007)	Arkansas	Stool
CDC_ K4588	July (2006)	Maine	Stool	CDC_ K5323-1	no date reported	Virginia	Other
CDC_ K4636	September (2006)	New York	Stool	CDC_ K5323-2	no date reported	Virginia	Other
CDC_ K4637-1	October (2006)	New York	Stool	CDC_ K5324-1	June (2007)	Virginia	Stool
CDC_ K4637-2	October (2006)	New York	Stool	CDC_ K5324-2	June (2007)	Virginia	Stool
CDC_ K4638	September (2006)	New York	Stool	CDC_ K5328	no date reported	Indiana	Stool
CDC_ K4639-1	October (2006)	New York	Stool	CDC_ K5330	April (2007)	Texas	Other
CDC_ K4639-2	October (2006)	New York	Stool	CDC_ K5331	August (2007)	Georgia	Stool
CDC_ K4760	July (2006)	Virginia	Blood	CDC_ K5345-1	August (2007)	Iowa	Stool
CDC_ K4762	August (2006)	Virginia	Other	CDC_ K5345-2	August (2007)	Iowa	Stool
CDC_ K4763	August (2006)	Virginia	Stool	CDC_ K5346	August (2007)	Pennsylvania	Stool
CDC_ K4764-1	October (2006)	Virginia	Stool	CDC_ K5428	July (2007)	Nevada	Stool
CDC_ K4764-2	October (2006)	Virginia	Stool	CDC_ K5429	August (2007)	Nevada	Stool
CDC_ K4775	February (2007)	Georgia	Stool	CDC_ K5433	July (2007)	Washington	Stool
CDC_ K4842	October (2006)	Maryland	Stool	CDC_ K5435	August (2007)	Washington	Stool

Table 11: *continued*

Isolate ID	Reporting Date	Reporting State	Source of Isolate	Isolate ID	Reporting Date	Reporting State	Source of Isolate
CDC_ K4857-1	January (2007)	Hawaii	Stool	CDC_ K5436	no date reported	Washington	Stool
CDC_ K4857-2	January (2007)	Hawaii	Stool	CDC_ K5437	September (2007)	Washington	Stool
CDC_ K4858	September (2006)	Hawaii	Stool	CDC_ K5438	September (2007)	Washington	Stool
CDC_ K4859	February (2007)	Hawaii	Other	CDC_ K5439	September (2007)	Washington	Stool
CDC_ K4981	March (2007)	Oklahoma	Other	CDC_ K5456	no date reported	Washington	Stool
CDC_ K5009-1	August (2006)	Maryland	Stool	CDC_ K5457	August (2007)	Washington	Stool
CDC_ K5009-2	August (2006)	Maryland	Stool	CDC_ K5485	July (2007)	North Carolina	Other
CDC_ K5010-1	September (2006)	Maryland	Stool	CDC_ K5512	June (2007)	Oklahoma	Stool
CDC_ K5010-2	September (2006	Maryland	Stool	CDC_ K5528	October (2007)	Georgia	Stool
CDC_ K5058	May (2007)	Texas	Stool	CDC_ K5579	no date reported	Indiana	Stool
CDC_ K5059-1	May (2007)	Texas	Other	CDC_ K5582	October (2007)	Georgia	Stool
CDC_ K5059-2	May (2007)	Texas	Other	CDC_ K5615	August (2007)	New York	Stool
CDC_ K5067	April (2007)	South Dakota	Stool	CDC_ K5618	August (2007)	New York	Other
CDC_ K5073	March (2007)	Maryland	Stool	CDC_ K5620	August (2007)	New York	Stool
CDC_ K5125	June (2007)	Mississippi	Other	CDC_ K5621	September (2007)	New York	Stool
CDC_ K5126	May (2007)	Mississippi	Stool	CDC_ K5629	November (2007)	Georgia	Stool
CDC_ K5276	April (2007)	New York	Stool	CDC_ K5635	September (2007)	Maryland	Wound
CDC_ K5277	no date reported	Washington	Stool	CDC_ K5638	no date reported	Maryland	Stool
CDC_ K5278	June (2007)	Washington	Stool	CDC_ K5701	September (2007)	Oregon	Stool
CDC_ K5279	no date reported	Washington	Stool				

8.1.2. *Vibrio parahaemolyticus* isolates from 2012

Isolates were collected from oysters and clams in several harvest areas associated with a *V. parahaemolyticus* outbreak in New York and Connecticut in summer of 2012. Fifty-five of these isolates, 44 from oysters and 11 from clams, were selected as a representative sample set based on their results after hybridization probing for *tlh*, *tdh* and *trh* (Table 12).

Table 12: Outbreak associated isolates from New York and Connecticut and their origin

Isolate ID	Shellfish Type	State	Harvest Area
Vp-NY-12-1	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-2	Clam	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-3	Clam	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-4	Clam	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-5	Oyster	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-6	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-7	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-8	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-9	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-10	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-11	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-12	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-13	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-14	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-15	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-16	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-17	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-18	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-19	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-20	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-21	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-22	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-23	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-24	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-25	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-26	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-27	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-28	Oyster	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-29	Oyster	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-30	Oyster	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-31	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-32	Oyster	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-33	Oyster	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-34	Clam	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-35	Oyster	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-36	Clam	New York	Oyster Bay Harbor, Lot #5
Vp-NY-12-37	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-38	Oyster	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-39	Clam	New York	Oyster Bay Harbor, Lot #5
Vp-NY-12-40	Oyster	New York	Oyster Bay Harbor, Lot #2

Table 12: *continued*

Isolate ID	Shellfish Type	State	Harvest Area
Vp-NY-12-41	Oyster	New York	Oyster Bay Harbor, Lot #2
Vp-NY-12-42	Oyster	New York	Oyster Bay Harbor, Lot #4
Vp-NY-12-43	Clam	New York	Cold Spring Harbor
Vp-CT-12-1	Clam	Connecticut	301 E plant C
Vp-CT-12-2	Clam	Connecticut	Natural Bed
Vp-CT-12-3	Clam	Connecticut	Greenwich L-546H
Vp-CT-12-4	Oyster	Connecticut	Branford L-301E
Vp-CT-12-5	Oyster	Connecticut	Norwalk 213
Vp-CT-12-6	Oyster	Connecticut	367 Westport
Vp-CT-12-7	Clam	Connecticut	Westport -L-333
Vp-CT-12-8	Oyster	Connecticut	Westport -598
Vp-CT-12-9	Oyster	Connecticut	Norwalk L-100
Vp-CT-12-10	Oyster	Connecticut	158-L-5955
Vp-CT-12-11	Oyster	Connecticut	103-L-213
Vp-CT-12-12	Oyster	Connecticut	103-L-255

The FDA also received clinical isolates from different states in summer of 2012 (Table 13). These isolates were included in some of the analysis in this thesis (chapter 5.5 and 5.6).

Table 13: Clinical isolates from different states in 2012

Isolate ID	Reporting State	Source of isolate
MA___12EN2941	Maine	Stool
MA___12EN2945	Maine	Stool
CA___M12X02735	New York	Stool
M12017000	Pennsylvannia	unknown
M12014845	Pennsylvannia	unknown
M12014686	Pennsylvannia	unknown
MDA12108395	Maryland	unknown
MDA12104560	Maryland	unknown
MDA12135369	Maryland	unknown
MDA12131319	Maryland	unknown
12MP010932	Wisconsin	unknown
IDR1200025542	New York	unknown

8.1.3. Cell lines

8.1.3.1. HeLa cells

HeLa cells originated from cervical adenocarcinoma. The cancer sample was taken from the cervix of a young black lady named Henrietta Lacks (Schmitz, 2011). The name of the cell line was taken from the first two letters of her name (Masters, 2002). HeLa cells are used as a cell line in all biomedical research fields to study biochemical pathways in human cells and are the most widely utilized human cancer cell line. Originally HeLa was used in cancer research (Masters, 2002).

8.1.3.2. Caco cells

Caco-2-cells were obtained from colorectal adenocarcinoma of a 72 year old male (Fogh et al., 1977; Fogh et al., 1979). Like HeLa cells, Caco-2 is an epithelial cell line. Although they originated from colonic cancer, Caco-2 cells have specific functions of enterocytes of the small intestine, composed of duodenum, jejunum and ileum (Pinto et al., 1983). Once they build a monolayer, they behave similarly to the intestinal epithelial barrier and with that Caco-2 cells are a good choice for studying effects of enteric pathogens (Lievin-Le and Servin, 2013).

8.2. Methods

8.2.1. Storage of bacterial cultures

For bacterial isolates derived from sample analysis, the FDA Bacteriological Analytical Manual recommends storage at -80˚C in tryptic soy broth (TSB) containing 1% NaCl and 24% glycerol (Kaysner and Depaola, 2004). Glycerol stabilizes frozen bacteria and prevents damage to cell membranes (Srivastava, 2003) (chapter 11.2.3, protocol 11.2.3.1).

8.2.2. Identification and Isolation of *Vibrio parahaemolyticus*

The development of solid culture media by Robert Koch paved the way for isolating pure bacterial cultures. However, solid media had one disadvantage: Bacteria in small numbers

were difficult to isolate. Therefore, Martinus Beijerinck developed enrichment culturing utilizing liquid media. The liquid media was designed based on the organism's growth conditions to increase their proportion in the total bacterial population. After incubation the bacterial culture was plated on solid media for isolation (Wheelis, 2011). Since *V. parahaemolyticus* is a facultatively anaerobic and gram-negative bacterium, they prefer alkaline growth conditions and presence of high levels of bile salt. Therefore, alkaline peptone water (APW) is usually utilized for isolating *Vibrio* species. Additionally, media for *V. parahaemolyticus* testing should contain 2 – 3% sodium chloride (NaCl) (Kaysner and Depaola, 2004). For enrichment, seafood samples were homogenized; with oysters or clams twelve animals were pooled and blended. In a MPN format, bacterial cells were enumerated. Thiosulfate-citrate-bile salts-sucrose agar (TCBS) is commonly used as selective media for *V. parahaemolyticus* isolation. TCBS inhibits growth of most non-vibrios while *V. parahaemolyticus* grows in form of green colonies due to a lack of pH change indicated by bromthymol blue (Kobayashi et al., 1963). Through the presence of oxgall, a naturally occurring substance containing bile salts and sodium cholate, gram-positive bacteria are inhibited.

However, the classical identification procedure takes up to four days (McCarthy et al., 1999). Therefore, a DNA probe colony hybridization assay was developed in 1999 to enumerate *V. parahaemolyticus* in oysters at harvest (McCarthy et al., 1999). It is based on direct plating of homogenized oyster tissue onto a nutrient medium and performing colony lifts to transfer the colonies to a filter that can be tested by DNA gene probes to detect total (*tlh* gene) and pathogenic (*tdh* and *trh* genes) *V. parahaemolyticus* (McCarthy et al., 1999; McCarthy et al., 2000; Nordstrom et al., 2006). Furthermore, this method can be applied to isolated colonies for gene identification.

As DNA probes non-isotopic probes were used (sequences attached in chapter 11.2.2 in Table 60). These non-isotopic probes were recommended by the FDA Bacterial Analytical Manual (BAM), since they are easy to use and to store compared to radioactive probes (McCarthy et al., 1999; Nordstrom et al., 2006). The oligonucleotide

is labeled with the enzyme alkaline phosphatase, which uses nitroblue tetrazolium (NBT)/5-bromo-4-chloro-3-indolyl phosphate (BCIP) as a substrate. The colony hybridization process takes advantage of this natural mechanism (Tijssen, 1993). Filters containing potassium phosphate were used as a solid phase. Alkaline phosphatase dephosphorylates BCIP and forms through an intermediate ketone 5,5'-dibromo-4,4'-dichloro-indigo while releasing hydrogen. By reducing this hydrogen NBT results in an insoluble purple diformazan (Tijssen, 1993) (chapter 11.2.3, protocol 11.2.3.9). Additionally, species identification can be performed using PCR and real-time PCR (chapter 8.2.4).

8.2.3. DNA-Extraction

To utilize DNA based technologies extraction methods aiming for high quality DNA concentrations are of need. The quality is measured using levels of fragmentation or chemical purity (Mülhardt, 2009). In general, DNA isolation takes advantage of two characteristics of DNA: One, it reversibly precipitates using Cetyltrimethylammoniumbromide or isopropanol, and can therefore, be separated from other parts of the food matrix; two, DNA binds under high salt conditions to silica materials such as coluums or magnetic beads (Haase, 2012). The purity of DNA extracts can be determined photometrically through the absorption quotient A260 nm/A280 nm (Haase, 2012).

For *Vibrio* DNA extractions, solid-phase nucleic acid extraction is the base of most of the available commercial extraction kits (Tan and Yiap, 2009). The solid phase absorbs nucleic acid in the extraction process depending on the pH and salt content of the buffer. Solid-phase extraction usually consists of cell lysis, nucleic acids adsorption, washing, and elution (Mülhardt, 2009). It is often performed using a spin column (Tan and Yiap, 2009). The initial step in a solid phase extraction process is to condition the column for sample adsorption by using a buffer at a particular pH to convert the surface or functional groups on the solid phase into a particular chemical form (Tan and Yiap, 2009). The lysed sample can now be

applied to the column. The nucleic acids will absorb to the column under the conditions of high pH and salt concentration of the binding solution (Tan and Yiap, 2009). Proteins can interfere with the specific bond to the column surface. Therefore, a washing step removes the proteins by using washing buffer containing a competitive agent (Tan and Yiap, 2009). For the final elution step, TE buffer or water is usually used to release the desired nucleic acid from the column (Tan and Yiap, 2009) (chapter 11.2.3, protocols 11.2.3.4 and 11.2.3.5).

8.2.4. Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was developed by Kary Mullis in 1986 (Mullis et al., 1986). PCR is based on the amplification of part of a DNA strand, the target region. To utilize this technique the sequence of the target organism needs to be known (Mullis et al., 1986). Short single-stranded oligonucleotides are designed to target the specific area of the gene of interest. These oligonucleotides are called primers and are also the starting point of DNA replication (Mullis et al., 1986). Also needed are dNTPs, which are the four nucleotide triphosphates, a heat-stable polymerase, and magnesium ions in the buffer (Mullis et al., 1986). *Taq*-polymerase is the most commonly used DNA polymerase. It was isolated from the thermophilic eubacterium *Thermus aquaticus* (Terpe, 2013). The buffer maintains the pH and contains the Mg^{2+}-ions that bind to dNTPs and oligonucleotides (Mülhardt, 2009)

DNA amplification follows a temperature program to ensure the following steps: 1. denaturation of the DNA by heating the reaction mix to 94 - 98˚C for 20 – 30 seconds, 2. annealing of the specific primers to the single-stranded DNA, and 3. elongation of DNA by a DNA polymerase. These three steps are part of repetitive cycles. Therefore, amplified products accumulate exponentially (Mullis et al., 1986). After a certain number of cycles the amplification products reach a plateau effect. Accumulated end products slow down the reaction and unspecific products start to develop by primer annealing and reannealing of already ampflified products (Mülhardt, 2009). The melting process should fully separate the DNA strands of the template. Partially separated structures would rapidly reanneal when

temperature was dropped and will not be primed (Mülhardt, 2009). The required melting temperature and time of melting depends on the length and the template sequence. Similar conditions affect the annealing temperature. The temperature should be two degrees below the melting temperature of the primers used (Mülhardt, 2009).

Multiplex-PCR is a variant of PCR in which two or more target sequences can be amplified by including more than one pair of primers in the same reaction. Multiplex PCR has the potential to save time and effort in the laboratory. However, it is important that the primers do not amplify homolog sequences (Mülhardt, 2009) (chapter 11.2.3, protocol 11.2.3.10).

At the end of the PCR the PCR products are usually separated by agarose gel electrophoresis (Mülhardt, 2009). Agarose is a linear polymer that is extracted from seaweed (Serwer, 1983). After boiling in a buffer solution agarose forms a gel by hydrogen bonding. The liquid agarose solution can be poured into a gel-forming chamber. By inserting a comb into the liquid gel, wells are formed in the solidified gel in which the PCR product can be loaded. After the gel is solidified, it is transferred to an electrophoresis chamber loaded with with buffer. Usually, tris-acetate-EDTA buffer (TAE) or tris-borate-EDTA buffer (TBE) are used. To ensure that the sample sinks into the wells they are mixed with a dense solution, the loading buffer (Mülhardt, 2009).

Depending of the concentration of the agarose in the buffer the gel forms pores. These pores regulate the separation process based on the size of the PCR products: By applying an electric field to the gel large DNA-fragments move more slowly through the agarose gel than small fragments. The pores are less resistant to smaller DNA-fragments (Serwer, 1983). For electrophoresis of PCR products, an agarose concentration of 0.8-2% is usually recommended to assure an optimal separation and visualization of the bands (Mülhardt, 2009). To visualize the PCR products on the agarose gel, dyes such as ethidium bromide or SYBR Green are used. These dyes can be applied to the gel solution during preparation or the solidified gel can be stained by soaking in a dyed buffer after the electrophoresis (Mülhardt, 2009).

8.2.4.1. Real-time PCR

As conventional PCR only gives a qualitative answer, for quantification purposes real-time PCR was developed. Real-time PCR can be used for qualitative and quantitative assays (Mülhardt, 2009). Due to the plateau effect at the end of the PCR reaction, real-time PCR can detect and quantifiy even small numbers of target DNA only in the exponential phase during the PCR process. Here, the generated products are directly proportional to the amount of template prior to the start of the PCR process (Mülhardt, 2009).

Real-time PCR uses different detection methods: intercalation fluorophoric dyes or target specific fluorescent probes. One of the nonsequence-specific fluorogenic DNA-binding dyes is SYBR Green 1 which intercalates into all dsDNA (Morrison et al., 1998). SYBR Green 1 exhibits little fluorescence when in solution, but emits a strong fluorescent signal once it is bound to dsDNA (Morrison et al., 1998). The advantage of this technique is that it is relatively cheap and it can be used with any pair of primers for any target. However, as the presence of any dsDNA generates fluorescence, specificity of this assay is greatly decreased due to amplification of nonspecific PCR products and primer-dimers (Ririe et al., 1997).

The target specific probes take advantage of the 5‘ to 3‘ exonuclease activity of the *Taq*-polymerase (Holland et al., 1991). A probe is a small oligonucleotide sequence, which is designed to hybridize within the target sequence. These probes are usually labeled with certain fluorophores for detection purposes. As a general principle the probe has a reporter fluorescent dye at the 5´ end and a quencher dye attached to the 3´ end. While the probe is intact, the close proximity of the quencher significantly decreases the fluorescence emitted by the reporter dye. A fluorescence signal is only emitted on cleavage of the probe, based on the fluorescence resonance energy transfer (FRET) principle (Cardullo et al., 1988). A hydrolysis probe often used in real-time PCR and part of the PCR assays of this thesis are “TaqMan-style” probes (Heid et al., 1996). Cleavage of the probe by *Taq*-polymerase during PCR separates the reporter and quencher dyes, thereby increasing the fluorescence. Additionally, this removes the probe from the target strand, allowing primer extension to continue to the end of template strand, thereby not interfering with the exponential

accumulation of PCR product. Various fluorescent reporter dyes are in use including 6-carboxyfluorescein (FAM), tetrachloro-6-carboxy-fluorescein (TET), or hexacholoro-6-carboxyfluorescein (HEX). Quenchers include either 6-carboxytetramethylrhodamine (TAMRA) or 4-(dimethylaminoazo)benzene-4-carboxylic acid (DABCYL). The increase in fluorescence intensity is directly proportional to the amount of amplicon produced and can be detected in real time by a modified thermocycler (Giulietti et al., 2001; Ginzinger, 2002). The computer software constructs amplification plots using the fluorescence emission data that are collected during the PCR amplification. Certain points of this amplification plot influence the results: baseline, threshold and C_t-value. The baseline is defined as the PCR cycles in which a reporter fluorescent signal is accumulating in the absence of amplification. The threshold is chosen based on the variability of the baseline. It is calculated as ten times the standard deviation of the average signal of the baseline fluorescent signal between cycles three to 15. A fluorescent signal that is detected above the threshold is considered an amplification signal that can be used to define the threshold cycle (C_t) for a sample. The C_t-value is defined as the fractional PCR cycle number at which the reporter fluorescence is greater than the minimal detection level. Therefore, C_t measures the starting copy number instead of an endpoint measurement of the amount of accumulated PCR product. During the exponential phase none of the reaction components is limiting and therefore Ct values are very reproducible for replicate reactions with the same starting copy number (Mülhardt, 2009).

For the quantification of the amount of target samples two different methods are used: absolute quantification and relative quantification compared to a standard gene (Haase, 2012).

Absolute quantification uses standard curves by measuring C_t based on a serial dilution of a known target amount. Through linear regression the target concentration of the sample can be calculated (Mülhardt, 2009). Depending on the food matrices the real-time PCR assay needs individual design. The efficiency should be between 89.57 and 110.12% (slope

between -3.6 and -3.1) and the linearity expressed through the coefficient of determination R > 0.98 (Haase, 2012).

Relative quantification or relative threshold method uses a reference control quantifying differences on the expression level of the same specific target between different samples. The target concentration is being normalized to this reference control or calibrator and results are in a fold-change output.

For *V. parahaemolyticus* a real-time PCR assay was developed and evaluated to detect the *tdh* gene (Blackstone et al., 2003). To avoid the occurrence of false negatives resulting from PCR inhibition by the sample matrix, an internal amplification control (IAC) is part of each individual reaction mixture. Positive and negative controls run normally with every PCR master mix to ensure the integrity of the reagents (Hoorfar et al., 2004) (chapter 11.2.3, protocol 11.2.3.10).

8.2.5. Subtyping

8.2.5.1. Pulsed Field Gel Electrophoresis

Pulsed-field gel electrophoresis (PFGE) was developed in 1983 by Schwartz and Cantor (Schwartz et al., 1983; Schwartz and Cantor, 1984). For this method DNA is incorporated by whole cells transferred into agarose. After in situ cell lysis the whole bacterial DNA is digested with an infrequent cutting restriction endonuclease. Due to the use of a low frequency cutting enzyme large restriction fragments are formed, which cannot be separated by conventional electrophoresis. Therefore, a pulsed-field electrophoresis is applied to the fragments. The so called pulsed-field has a system setup with 24 electrodes arranged in a hexagon with an alternating electric field in 120˚ increments (Kaufmann, 1998; Herschleb et al., 2007). The resulting patterns can be analyzed with specific computer software.

PFGE has developed as a genome characterization tool. It has been widely applied to construct genomic clone libraries, detection of genomic polymorphisms and rearrangements

in organisms. Additionally, PFGE serves as a diagnostic strain-typing tool for pathogenic microbes (Herschleb et al., 2007).

PFGE discriminates among serotypes of *V. parahaemolyticus* and is widely used for subtyping of this organism (Centers for Disease and Prevention, 2009) (chapter 11.2.3, protocols 11.2.3.12 and 11.2.3.17).

8.2.5.2. Direct Genome Restriction Enzyme Analysis (DGREA)

Direct Genome Restriction Analysis (DGREA) was first used to subtype *V. parahaemolyticus* in 2006 (Fuenzalida et al., 2006). DGREA applies a high-cutting frequency restriction endonuclease, *Nae*I, and differentiates isolates according to their fragmentation pattern. First, DNA was extracted from overnight cultures and then digested with the restriction enzyme for two hours. The fragments are separated on a polyacrylamide gel and visualized using silver staining (Fuenzalida et al., 2006; Gonzalez-Escalona et al., 2007b). Silver staining utilizes the reaction of silver ions binding to the DNA bases, which were then selectively reduced by formaldehyde under alkaline conditions (Bassam et al., 1991; Bassam and Gresshoff, 2007).

Here, too, resulting gel patterns can be analyzed with specific computer software. DGREA has also been applied to *V. vulnificus* (Gonzalez-Escalona et al., 2007b) (chapter 11.2.3, protocols 11.2.3.13, 11.2.3.14 and 11.2.3.17).

8.2.5.3. Intergenic Spacer Region (ISR-1) Typing

As mentioned in chapter 3.3.1 *V. parahaemolyticus* has eleven rRNA operons. Each operon includes intergenic spacer regions between the 16S and 23S rRNA genes, which differ in in sizes and nucleotides sequences. The intergenic spacer regions encode tRNAs (Maeda et al., 2000; Makino et al., 2003). Through lateral gene transfer or recombination the same intergenic spacer regions are different between isolates of the same species (González-Escalona et al., 2005; Gonzalez-Escalona et al., 2006). The ISR-1 typing uses these differences between isolates. The resulting PCR product will be a mixture of fragments,

which is electrophoresed in a polyacrylamide gel and visualized by silver staining. The result is a complex banding pattern that provides a isolate-specific profile (González-Escalona et al., 2005; Gonzalez-Escalona et al., 2006) (chapter 11.2.3, protocols 11.2.3.15, 11.2.3.14 and 11.2.3.17).

8.2.6. Phylogenetics

8.2.6.1. Multiple Variable Tandem Repeat Analysis

Genomic DNA consists of a high number of repeats. They are different in their size and complexity and can be located in the same genomic area or spread out across the genome. These repeats are described as variable number of tandem repeats (VNTRs) (Lindstedt, 2005). Multiple Variable Tandem Repeat Analysis (MLVA) is utilizing the presence of VNTRs and can therefore provide insight into the diversity of bacterial species (Lindstedt, 2005). For the bacterial species presented here this is achieved by performing PCR of the VNTR loci followed by accurate sizing of the PCR products on an automated DNA sequencer. The assessed product size is used to calculate the number of repeat units in each locus (Lindstedt, 2005).

MLVA is used by CDC as complementary method to PFGE, since it is able to differentiate between indistinguishable PFGE patterns for bacteria (Hayat et al., 1993). Generally, the PCR products from MLVA are separated by sequencing or capillary electrophoresis (CE) (Lindstedt et al., 2013). However, differentiation of amplification products using high resolution melt (HRM) analysis has been described for MLVA assays in other organisms (Fortini et al., 2007).

Melting curve analysis in general is used to comfirm amplification products after real-time PCR with SYBR Green. The fluorescene is plotted against the temperature. A characteristic melting peak at the melting temperature (T_m) of the amplicon will distinguish it from amplification artifacts that melt at lower temperatures at broader peaks (Ririe et al., 1997). The developement of HRM realized the detection of minimal differences in melting

tempartures of two amplicons i. e. SNP's, mutations and methylations (Vossen et al., 2009). Usually, SYBR Green is used in singleplex reactions; however, when coupled with melting point analysis, it can be used for multiplex reactions (Lyon, 2001) (chapter 11.2.3, protocols 11.2.3.18 and 11.2.3.17).

8.2.6.2. Multi-Locus Sequence Typing (MLST)

Multi Locus Sequence Typing (MLST) is based on examination of the bacterial genome at usign sequence variation of multiple housekeeping genes (loci). The housekeeping genes are amplified via PCR and the products sequenced afterwards using Sanger methods. For each unique sequence at each locus a unique allele number is assigned. These designated numbers are incorporated into an allelic profile or a so called ST. By using alleles and STs as the unit of comparison and each single nucleotide change leads to an allelic change, this method provides a simple way to compare bacterial genomes (Maiden, 2006; Maiden et al., 2013). For the results interpretation three groups are defined: CC, SLV, and DLV. A CC is defined as at least six of seven alleles being identical (Feil et al., 2004). A SLV shares a single locus with another ST, while a DLV differs in two loci (Gonzalez-Escalona et al., 2008). Reference databases have been developed for data comparison and analysis. Therefore, MLST has become an important tool for investigations in vertical and horizontal gene transfers as well as epidemiology and evolution of bacterial pathogens (Maiden, 2006; Gonzalez-Escalona et al., 2008).

MLST was first used for the pathogen *Neisseria meningitides* (Maiden et al., 1998). Since then MLST has been applied to various bacteria including *V. parahaemolyticus* (Gonzalez-Escalona et al., 2008). The MLST scheme for *V. parahaemolyticus* utilizes seven housekeeping genes on chromosome I and II (Table 14). The nucleic acid sequences can be analyzed and stored in a public database (http://pubmlst.org/vparahaemolyticus) (Gonzalez-Escalona et al., 2008).

Table 14: Housekeeping genes and their fragment sizes of the *V. parahaemolyticus* MLST scheme

Chromosome	Locus	Fragment size [bp]
I	*dnaE*	557
	gyrB	592
	recA	729
II	*dtdS*	458
	pntA	430
	pyrC	493
	tnaA	423

Since the rapid increase of whole genome sequencing, MLST can now be applied to full genomes of bacteria. Additionally, more than seven loci of several bacterial strains can be compared to each other (Maiden et al., 2013). The consensus sequences of the *de novo* assembled genomes were analyzed through the MLST database for V. parahaemolyticus (http://pubmlst.org/vparahaemolyticus/) (Gonzalez-Escalona et al., 2008; Jolley and Maiden, 2010).

For the analysis of phylogenetic relations between each ST the eBURST software (http://eburst.mlst.net/) was used (Feil et al., 2004; Spratt et al., 2004). Additionally, a ML tree using the concatenated sequences was constructed using Mega 6 software with the Kimura 2-parameter model and 1000 bootstrap resamplings (Kimura, 1980; Tamura et al., 2013) (chapter 11.2.3, protocol 11.2.3.19).

8.2.7. Cell culture

8.2.7.1. Procedure of cell culture

Cell culture involves the removal of cells or tissue from an animal, human or plant, for growth in an artificial environment. If cells are removed prior to cultivation, the term cell culture is used. Small cell culture experiments started as early as 1907 when Ross Harrison showed how frog embryo nerve fibre grew *in vitro* (Davis, 2002). In the 1950s with the establishment of HeLa cells, the first human cell line was on the market (Masters, 2002). The development of specific cell culture media by Earle, Hanks, Dulbecco and Ham improved culturing of cell

lines immensely. These formulations are used as standard cell culture media still today (Schmitz, 2011). Today a variety of cell lines are commercially available. In this study, the HeLa and Caco-2 cell lines described above in section 8.1.3 were used.

Human and animal cells will only divide a certain number of times until cell division stops, known as the Hayflick limit (Davis, 2002). Therefore, culturing cell lines is divided into three different stages: Phase I, phase II and phase III. Phase I includes growing the primary tissue or cells directly from the organism. After cells build a confluent monolayer, they stop growing due to a contact inhibition. Hence, cells require subculturing while removing them, by physical scrapping or treatment with trypsin, and transferring them into fresh culture media. At this point, phase II starts and is distinguished by several subculturing steps until cells begin to divide less. Hereby phase III will be induced. At this point the cell dividing potential is exhausted and new subcultures need to be cultivated from cell line stocks of the primary culture (Davis, 2002; Schmitz, 2011) (chapter 11.2.3, protocols 11.2.3.20 and 11.2.3.21).

8.2.7.2. Cytotoxicity assays

Analyzing the effects on cell death in cell culture is an important component of bacteriological research. The most widely used experimental method for studying natural cellular cytotoxicity reactions is the ^{51}Cr release microcytotoxicity assay (Brunner et al., 1968). The radioisotope is released upon cell lysis (Brunner et al., 1968). However, this method has some drawbacks including: the time required for preincubation with the isotope, spontaneous release of ^{51}Cr during longer assay periods, and the expense, disposal and safety concerns associated with use of the isotope (Korzeniewski and Callewaert, 1983).

As an alternative, the amount of target cell killing over a given period of time can be determined by evaluation of remaining viable cells after their exposure to vital dyes (Parish and Mullbacher, 1983). Another type of assay is based on measurement of a cytoplasmic enzyme that is released following target cell death so that its activity in the cellular supernatant indicates the proportion of killed cells. This approach is both faster and more economical than procedures involving pre- or post-labeling of target cells (Korzeniewski and

Callewaert, 1983). Enzyme release assays have been described for alkaline phosphatase as well as for LDH. LDH is released upon cell lysis. The half-life of LDH released from cells into the surrounding medium is approximately 9 hours. Released LDH in culture supernatants is measured with a coupled enzymatic assay, which results in the conversion of a tetrazolium salt, iodonitrotetrazolium violet (INT), into a red formazan product. INT, lactate and NAD+ function as substrates in the presence of diaphorase (Decker and Lohmann-Matthes, 1988) (Figure 23). The generation of the red formazan product is proportional to the amount of LDH released and therefore the number of lysed cells. Visible wavelength absorbance data are collected at 490 nm using a standard 96-well plate reader (Decker and Lohmann-Matthes, 1988).

All variables in the cytotoxicity data set are categorical, and most non ordinal, since the percentage of relative toxicity is not considered a quantitative variable. Therefore, a contingency table analysis was used followed by tests of association between variables in contingency tables (Pearson, 1904; Plackett, 1983) (chapter 11.2.3, protocols 11.2.3.22 and 11.2.3.23).

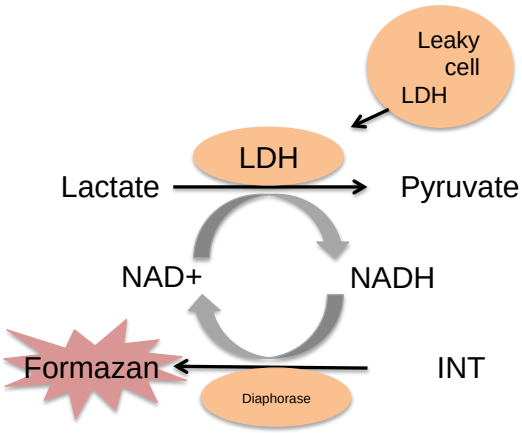


Figure 23: Release of LDH from damaged cells is measured by supplying lactate, NAD+ and INT as substrates in the presence of diaphorase. Generation of a red formazan product is proportional to the amount of LDH released and therefore the number of lysed cells (following (Promega, 2016).

8.2.8. Sequencing

To fully understand the DNA of an organism, it is of great interest to determine the exact order of the four bases in the genome (Stranneheim and Lundeberg, 2012). Today, sequencing already has a long history: Generally early sequencing techniques can be divided in radioactive and non-radioactive sequencing (Mülhardt, 2009). In 1977 Maxam and Gilbert started with sequencing history using radioactively-labeled nucleotides (Maxam and Gilbert, 1977). The introduction of the Sanger chain-termination sequencing method realized a more rapid and more accurate way of sequencing and has been used ever since (Sanger et al., 1977a; Sanger et al., 1977b). However, in the last few years massive changes appeared in the sequencing world; parallel sequencing allowed a faster and more detailed genome research. These novel sequencing methods using single molecules of DNA and real-time detection have become a frequently used tool in research (Stranneheim and Lundeberg, 2012).

8.2.8.1. Sanger Sequencing

Sanger Sequencing takes advantage of chain-terminating nucleotides, the dideoxy-nucleotides (Sanger et al., 1977a; Sanger et al., 1977b; Stranneheim and Lundeberg, 2012). Similar to the PCR principle the target DNA undergoes the process of denaturing, annealing with primers and extension by a *Taq*-polymerase. In addition to dNTPs the reaction mixture includes 2', 3'-didesoxynucleotides (ddNTPs). The ddNTPs lack a hydroxyl group at the third carbon atom of ribose that is necessary for binding the next nucleotide (Sanger et al., 1977a; Sanger et al., 1977b). When a ddNTP is incorporated into the complementary DNA strand instead of a dNTP, the extension process is terminated. Therefore, fragments of varying length are synthesized terminated by a chain-terminating nucleotide (Sanger et al., 1977a; Sanger et al., 1977b). Adding only a fraction of the terminating nucleotide ensures the statiscally random incorporation of dideoxynucleotides in a small subset of molecules (Stranneheim and Lundeberg, 2012). For detection purposes the ddNTPs are labeled with a

fluorophore, with different fluorophores used for each nucleotide. Therefore, each nucleotide at the end of each fragment can be detected through the fluorescent signal. The generated 3'-terminated DNA templates are then heat-denatured and fractionated by capillary gel-electrophoresis (Stranneheim and Lundeberg, 2012).

8.2.8.2. Next Generation Sequencing

The first bacterial genome was sequenced in 1995 of the species *Haemophilus influenza* (Dark, 2013). To date, 6855 closed bacterial genomes, nine of which are *V. parahaemolyticus*, exist in the Genomes Online Database (Pagani et al., 2012). The application of whole genome sequencing has tremendously increased since the commercialization of high-throughput sequencing. Those techniques make bacterial genome sequencing significantly cheaper and faster (Dark, 2013). There are several so called next generation sequencing techniques: Pyrosequencing (454), Sequencing by Oligo Ligation Detection (SOLiD), Solexa technology (MySeq and HiSeq), semiconductor technology (Ion Personal Genome Machine) and Single Molecule Real-time Sequencing (PacBio).

Pyrosequencing measures the converted amount of released pyrophosphate (PPi) after deoxynucleotide incorporation (Dark, 2013). The PPi is metabolized into ATP by ATP sulfurylase stimulating the conversion of luciferin to oxyluciferin. Oxyluciferin emits visible light at 560 nm detected by a photomultiplier (Ahmadian et al., 2006).

The **SOLiD** technique is based on ligation chemistry. DNA fragments are adapted to beads and located on these they are multiplied by emulsion PCR. Transferred onto a glass surface, primers and a set of dinucleotide combinations labeled with four different fluorophors, one dye for each combination, are added to the DNA fragments. Sequencing occurs by sequential rounds of hybridization and ligation. The sequence is determined by analyzing the fluorescence from two ligation processes of each nucleotide position (Morozova and Marra, 2008).

Solexa technology is based on synthesis of single-molecule arrays with reversible terminators. Using an adapter single-stranded DNA fragments are attached to a flow-cell or

single-molecule array. Therefore, the end of the single DNA molecule bends over and hybridizes with a complementary adapter on the surface creating a bridge. Functioning as templates these bridges are being amplified and form clusters of copies. These clusters are sequenced using reversible fluorophore-labeled nucleotides. The instrument measures the fluorescence at each nucleotide addition step (Morozova and Marra, 2008).

The **semiconductor technology** detects the hydrogen ions released during deoxynucleotide incorporation. This technique is used by Life Technologies for their IonTorrent Personal Genome Machine (PGM). Prior to sequencing, the genomic DNA is fragmented, ligated to adapters, and tagged with barcodes. The adaptor-ligated DNA fragments, so called „library“, is amplified via emulsion-PCR onto beads. After template-positive beads are enriched in a magnetic bead procedure, the sample is transferred onto a semiconductor chip. This chip consists of individual sensor wells connected to ion sensitive field-effect transistors, which are responsive to hydrogen ions. Sample beads are loaded into chip wells, with just one bead in each well. During the sequencing, all four nucleotides flow stepwise through the chip. By incorporating a complementary nucleotide into the template a single proton is released from the base and is detected by the sensor in form of a pH change. The pH, converted into voltage, calls a base and a sequence is built on repetition of these measurements (Rothberg et al., 2011; Dark, 2013).

Single Molecule Real-time Sequencing (SMRT) is used by Pacific Biosciences (PacBio, Menlo Park. CA, USA). The sequencing reaction takes place on so called SMRT chips consisting of thousands of zero-mode waveguides (ZMW), small cavities with DNA polymerase embedded in the bottom. DNA is synthesized from a single-stranded DNA template using fluorophore-labeled nucleotides. Each nucleotide has a different fluorophore attached to the phosphate group. During incorporation of a nucleotide into the complementary template the fluorophore emits light and identifies the base in real-time (Gupta, 2008; Schadt et al., 2010).

During this thesis, 144 draft shotgun sequences from clinical and oyster isolates from 2007 were completed. The sequences were generated using the Illumina HiSeq 2000 (Solexa

technology) with PE100 plus index read within the University of California at Davis 100K Pathogen Genome Project. Additionally, two closed genomes were sequenced within the 100K Pathogen Genome Project as well using the PacBio RSII sequencing platform (SMRT technology) (chapter 11.2.3, protocols 11.2.3.24 and 11.2.3.25)

8.2.8.3. Genome assembly and annotation

After generating sequence data from many small DNA fragments, these parts of the genome need to be positioned to form the full genomic sequence. This type of puzzle is called genome assembly. Generally, computer programs use fragments of single and paired reads. Single reads are short sequenced fragments. These fragments can be connected through overlapping regions into a continuous sequence known as a 'contig' (Baker, 2012). Paired reads have about the same length as single reads; however, they originated from either end of DNA fragments, which are too long to be sequenced at once (Baker, 2012).

Furthermore, genome assembly can be differentiated into reference assembly or *de novo* assembly (Pop, 2009). In a reference assembly, the fragments are mapped to a reference genome. However, even if a reference genome is available, significantly different regions of the sequenced genome can only be reconstructed using *de novo* assembly (Pop, 2009).

De novo assemblies aligns reads without a reference genome sequence, usually based on similarities of overlapping sequences. This technique relies on different algorithms to construct the genome sequence (Pop, 2009). Later, the assembled contigs can be linked into 'scaffolds', which are ordered assemblies of contigs with gaps in-between (Baker, 2012).

Two algorithmic strategies of *de novo* assemblies are overlap-layout-consensus (OLC) and de Bruijn graphs (DBG). OLC assembles by comparing all the reads to one another and computing which ones overlap with each other. Following the overlap step, the layout step positions the reads precisely based on the overlap, producing a most likely multiple alignment of all reads. This multi-alignment is then used to produce the consensus, the final DNA sequence (Pop, 2009). DBG use k-mers, which are sequences of a particular length k and shorter than the entire reads appearing as a consecutive substring of one of these reads

(Baker, 2012). The DBG cycle approach can be generalized to make use of these k-mers by building a graph. First, nodes are created from k-mers from a set of reads: Hence, a k-mer of a certain length was definded as a string and was split into smaller substrings, so called left- and right mers. By connecting one substring to another with a directed edge on a graph a DBG is built. An edge corresponds to an overlap between two k-mers. More precisely, it corresponds to a k-mer from the input (Compeau et al., 2011). The genome sequence is rebuilt by an eulerian walk through the DBG.

The raw reads of the sequences completed for this thesis, were deplexed after the sequencing reaction for further analysis. The genomic sequences were *de novo* assembled using SPAdes version 3.1.1 software (Nurk et al., 2013). SPAdes is based on identification of edges and resolving complex bulges in de Bruijn graphs (Nurk et al., 2013).

Once the genome is assembled, the locations of genes and coding regions need to be identified. Annotation softwares and algorithms search for known genetic markers and landmarks and map these to the sequenced genome (Stein, 2001). Finding landmarks includes searching short sequences, such as PCR-based genetic markers, and longer sequences, such as restriction-fragment length polymorphism markers, using a program such as BLASTN (Altschul et al., 1990). The draft genomes of this thesis were annotated in the publically available database NCBI Genbank using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP, http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html) (Klimke et al., 2009).

8.2.8.4. Phylogeny and SNPs

The goal of phylogenetics is to infer or estimate evolutionary relationships between and within species. This includes both information about order of branching and information about timing of events. Phylogenetics can also help understanding emerging disease (Baxevanis and Ouellette, 2004). For the display of the results phylogenetic trees or diagrams are helpful tools to present these relationships of different isolates (Baxevanis and Ouellette, 2004).

In general, phylogenetic analysis consists of four different steps: First, a multiple alignment of all sequences; second, the choice of the right substitution model; third, the tree; and lastly, the evaluation. Different phylogenetic methods exist and treat the data set differently (Baxevanis and Ouellette, 2004). The following paragraphs describe the most widely and in this thesis applied phylogenetic methods.

A multiple sequence alignment arranges three or more sequences based on the sum of similarities towards an ancestral sequence using computational algorithms such as ClustalW or MUSCLE (Bacon and Anderson, 1986; Thompson et al., 1994; Edgar, 2004; Edgar and Batzoglou, 2006). Models of nucleotide substitution specify the rates that the nucleotides mutate into other nucleotides compared to the ancestral sequence. Models of nucleotide substitution are Jukes Cantor- and Kimura-2-parameter model. Jukes Cantor treats all substitutions and base frequencies equally likely (Jukes and Cantor, 1969), the Kimura model on the other hand calls transitions and transversions at different rates and base frequencies equally (Kimura, 1980).

A phylogenetic tree can be displayed in different ways. It is composed of nodes, vertices, edges, arcs and branches (Baxevanis and Ouellette, 2004). The external nodes are usually associated with labels/isolate IDs. Trees can be rooted or unrooted. (Baxevanis and Ouellette, 2004). Tree building methods can be separated into distance based and character-based methods (Baxevanis and Ouellette, 2004). Distance methods use the dissimilarity between two sequences to create a tree (Swofford et al., 1996). Concerning distance data, methods that are used as a clustering approach are Neighbor-Joining (NJ), unweighted pair-group method using arithmetic averages (UPMGA), and minimum evolution (ME). The NJ method pairs neighboring taxonomic units with the lowest genetic similarity in a star like tree that minimizes the total branch length at each stage of clustering. The genetic distances of the sequences are calculated again and the next two neighboring taxa are connected until the star like tree is disolved (Saitou and Nei, 1987). The UPGMA method is a clustering mechanism. It joins tree branches on the criterion of greatest similarity among pairs and

averages of joined pairs (Sneath and Sokal, 1973). The ME method assigns each tree a score on the distance between internal and external nodes and minimizes branch lengths by minimizing this distance (Rzhetsky and Nei, 1992).

Character-based methods use the characters, the nucleotides, in all steps of the analysis (Baxevanis and Ouellette, 2004). Character data are maximum parsimony and ML, (Baxevanis and Ouellette, 2004). Maximum parsimony chooses the tree, which requires the fewest substitutions to explain the data. The numbers of substitutions are summed over sites giving a score for each tree topology. The topology having the smallest total number of substitutions represents the estimate of phylogeny (Edwards and Cavalli-Sforza, 1964; Cavalli-Sforza and Edwards, 1967). ML calculates the probability that a pattern of variation at a site is caused by a particular substitution process. The likelihood is the product of the sums of probabilities of each possible reconstruction (Felsenstein, 1981).

For tree evaluation the bootstrap value is used. The bootstrap estimate for phylogenetic analysis is derived from repeated resampling from the original sequence data with replacement nucleotides consituting a new sequence (Felsenstein, 1985). To estimate confidence levels of internal branches in phylogenetic analysis a tree is then reconstructed with these new sequences using the same tree building method as before. This new tree is then compared to the original tree by assigning scores (Felsenstein, 1985). A RAxML rapid bootstrapping run was also used for trees built based on WGS data (Stamatakis, 2014).

The kSNP script indexes each genome into a group of k-mers. Here, a k-mer is the flanking sequence including the SNP; using a k-value of 25 means that 12 bases on each side will flank the SNP. The software enumerates all these k-mers for each genome. Afterwards, the program filters non-unique k-mers, determines which k-mers have a variable position in the middle, and depending on which parameters are used (core, marjority, or all matrix) it includes or excludes those SNPs. A reference of finished genomes to identify putative regions of a genome that the k-mer/SNP might reside can be used (Gardner and Hall, 2013) (chapter 11.2.3, protocol 11.2.3.26).

8.2.8.5. Large scale BLAST score ratio (LS-BSR) *analysis*

To discriminate between differences in the genome of cytotoxic and non-cytotoxic isolates, homologs (groups of genes) were assembled, anticipating that some homologs would only be present in all of the cytotoxic sequences and not in the non-cytotoxic. These comparative groups of the *V. parahaemolyticus* genome content were built using large scale BLAST score ratio (LS-BSR) (Sahl et al., 2014; Hazen et al., 2015). The LS-BSR pipeline is used to compare the gene content of a relatively large number of bacterial genomes and based on the generated matrix of relatedness of all CDSs the results can be visualized as a cluster (Sahl et al., 2014). Thereby, the clusters could contain either cytotoxic (≥ 75%) or non-cytotoxic (≤ 50%) isolates, but not both at the same time. The semi-cytotoxic isolates could be among either cluster. The clusters were interpreted as bar charts using the software R version 3.1.2. (Team, 2015).

8.2.8.6. Identification of functional categories of COGs and average number of SNPs

To investigate which clusters of orthologous groups COGs could cause a difference in clustering in the phylogenetic tree, general gene functions of the core-, pan-, cytotoxic and non-cytotoxic genomes were determined and compared. The pan-genome describes the full complement of all gene families and specific genes in a set of strains. In addition it includes the core-genome containing the indispensable genes shared by all, and the dispensable genome containing the genes not shared by all strains (Medini et al., 2005; Tettelin et al., 2005). Therefore, cytotoxic and non-cytotoxic genomes are part of the pan genomes. In detail, clusters of homologs were verified and those results are described below. Predicted open reading frames (ORFs) were found in assemblies of our sequence data from isolates of *V. parahaemolyticus* using Prodigal (Hyatt et al., 2010) and saved as nucleotide sequences. ORFs were then blasted (Altschul et al., 1997) in both directions against predicted protein sequences from Vibrio_parahaemolyticus_RIMD2210633, Vibrio_parahaemolyticus_CDC_K4557, Vibrio_parahaemolyticus_ FDA_R31, and

Vibrio_parahaemolyticus_BB22OP, which were downloaded from NCBI. If the highest scoring pair was the same for a reciprocal blast and the e-value was below $10e^{-50}$ in both directions the sequences were kept as pairwise reciprocal best hit (PRBH). PRBHs were then clustered based on shared members. Each cluster was predicted to be a group of homologous sequences. The numbers of cytotoxic and non-cytotoxic strains from each of these clusters were quantified. Clusters including only cytotoxic or non-cytotoxic isolates were categorized and tabulated. Clusters were placed in COG (Clusters of Orthologous Groups) (Tatusov et al., 1997) functional categories using NCBI Gene IDs from the annotated predicted proteins. Evolutionary genealogy of genes: Non-supervised Orthologous Groups (EggNOG) is a database of orthologous groups of genes (Powell et al., 2014). EggNOG was used to map from each cluster functional categories to COG ID's (http://eggnog.embl.de/version_4.0.beta.).

Additionally, the average number of SNPs across the *V. parahaemolyticus* genomes was determined. The sequences were mapped to the publically available closed genome of the strain Vibrio_parahaemolyticus_O1_Kuk_str_FDA_R31 (chapter 5.4), which is non-cytotoxic. Each value was divided by the number of cytotoxic or non-cytotoxic isolates as appropriate to correct for difference in the number of strains.

9. List of chemicals

Table 15: List of chemicals

Chemical	Hazard symbol (German)	H-Sentence	P-Sentence	Manufacturer
1X MgCl2	-	-	-	Invitrogen, Carlsbad, CA, USA
2-Mercaptoethanol	Gefahr - giftig oder sehr giftig, ätzend, gesundheitsgefährdend, umweltgefährlich	301+331-310-315-317-318-373-410	273-280-302+352-304+340-305+351+338-308+310	Sigma-Aldrich, St. Louis, MO, USA
2-Propanol	Gefahr - Leichtentzündlich, Achtung	225-319-336	210-233-305+351+338	Fisher Scientific, Pittsburgh, PA, USA
Acetic acid	Gefahr - leicht-/hochentzündlich, ätzend	226-290-314	210-280-301+330+331-305+351+338-308+310	Fisher Scientific, Pittsburgh, PA, USA
Acetone	Gefahr - leicht-/hochentzündlich, Achtung	225-319-336	210-233-305+351+338	Sigma-Aldrich, St. Louis, MO, USA
Agar	-	-	-	Fisher Scientific, Pittsburgh, PA, USA
Agarose, Molecular Biology Grade	-	-	-	Fisher Scientific, Pittsburgh, PA, USA
Ampicillin	Gefahr - Achtung, gesundheitsgefährdend	315-317-319-334-335	261-280-305+351+338-342+311	Fisher Scientific, Pittsburgh, PA, USA
Bovine serum albumin (BSA)	-	-	-	Sigma-Aldrich, St. Louis, MO, USA
Bromthymolblue, 0.04% aqueous solution	-	-	-	ACROS, New Jersey, USA
Chloroform	Gefahr - Giftig oder sehr giftig, gesundheitsgefährdend	351-361d-331-302-372-319-315	302+352-314	Fisher Scientific, Pittsburgh, PA, USA
D-glucose solution, 200 g/L	-	-	-	Gibco, Carlsbad, CA, USA
D-Mannit	-	-	-	ACROS, New Jersey, NJ, USA
Dimethyl sulfoxide	-	-	-	Sigma-Aldrich, St. Louis, MO, USA
Dimethylpolysiloxane	-	413	-	Sigma-Aldrich, St. Louis, MO, USA

Table 15: *continued*

Chemical	Hazard symbol (German)	H-Sentence	P-Sentence	Manufacturer
Disodium hydrogen orthophosphate	-	-	-	Sigma-Aldrich, St. Louis, MO, USA
Dnase I	-	-	-	Invitrogen, Carlsbad, CA, USA
Dulbecco's Modified Eagle's medium (DMEM)	-	-	-	Gibco, Carlsbad, CA, USA
Ethanol, 200 proof anhydrous	Gefahr - Leichtentzündlich, Achtung	225-319	210-240-305+351+338-403+233	Warner graham company, cockeysville, MD, USA
Ethidium bromide 10mg/mL	Gefahr - Sehr giftig, gesundheitsgefährdend	330-341-302-315-319-335	281-302+352-305+351+338-304+340-309-310	Invitrogen, Carlsbad, CA, USA
Ethylenediaminetetra-acetic acid (EDTA)	Achtung	319	305+351+338	Fisher Scientific, Pittsburgh, PA, USA
Fetal bovine serum (FBS)	-	-	-	Gibco, Carlsbad, CA, USA
Formaldehyde	Gefahr – Giftig, ätzend, gesundheitsgefährdend	301+311+331, 314-317-335-341-350-370	201-260-280-301+310-330-303+361+353-304+340-310-305+351+338-308+310-403+233	Fisher Scientific, Pittsburgh, PA, USA
GlutaMAX, 200 mM L-alanyl-L-glutamine dipeptide in 0.85% NaCl, 100X	-	-	-	Gibco, Carlsbad, CA, USA
Glycerol	-	-	-	Fisher Scientific, Pittsburgh, PA, USA
Hydrochloric acid	Gefahr - ätzend, Achtung	290-314-335	234-260-304+340-303+361+353 305+351+338-309+311-501	Fisher Scientific, Pittsburgh, PA, USA
LB broth, Miller, granulated	-	-	-	Fisher Scientific, Pittsburgh, PA, USA
Methanol	Gefahr - leicht-/hochentzündlich, giftig oder sehr giftig, gesundheitsgefährdend	225-331-311-301-370	210-233-280-302+352	J. T. Baker Chemical Co., Phillipsburg, NJ, USA
N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid (HEPES), 1 M	-	-	-	Gibco, Carlsbad, CA, USA
***Nae*I restriction enzyme**	-	-	-	Promega, Madison, WI, USA
Non-essential amino acids (NEAA), 100X	-	-	-	Gibco, Carlsbad, CA, USA

Table 15: *continued*

Chemical	Hazard symbol (German)	H-Sentence	P-Sentence	Manufacturer
***Not*I restriction enzyme**	-	-	-	Promega, Madison, WI, USA
PCR Nucleotide Mix	-	-	-	Roche Diagnostics, Indianapolis, IN, USA
Polypeptone	-	-	-	Becton, Dickinson and Company, Sparks, MD, USA
Potassium chloride, ≥99%	-	-	-	MP Biomedicals, Solon, OH, USA
Potassium dihydrogen orthophosphate, for molecular biology, ≥98.0%	-	-	-	Sigma-Aldrich, St. Louis, MO, USA
Proteinase K	Gefahr - Achtung, gesundheitsgefährdend	315-319-334-335	261-305+351+338-342+311	Sigma-Aldrich, St. Louis, MO, USA
***Sfi*I restriction enzyme**	-	-	-	Promega, Madison, WI, USA
Sodium bicarbonate, 7.5% in solution	-	-	-	Gibco, Carlsbad, CA, USA
Sodium chloride	-	-	-	Fisher Scientific, Pittsburgh, PA, USA
Sodium hydroxide	Gefahr- ätzend	290-314	280-301+330+331-305+351+338-308+310	Fisher Scientific, Pittsburgh, PA, USA
Sodium pyruvate, 100 mM	-	-	-	Gibco, Carlsbad, CA, USA
Sodium acetate	-	-	-	Sigma-Aldrich, St. Louis, MO, USA
***Taq* DNA polymerase**	-	-	-	Roche Diagnostics, Indianapolis, IN, USA
Thiosulfate-citrate-bile salts-sucrose agar (TCBS)	-	-	-	Becton, Dickinson and Company, Sparks, MD, USA
Trypsin-EDTA, 0.5%	-	-	-	Gibco, Carlsbad, CA, USA
Tryptic soy agar (TSA)	-	-	-	Fisher Scientific, Pittsburgh, PA, USA

Table 15: *continued*

Chemical	Hazard symbol (German)	H-Sentence	P-Sentence	Manufacturer
Tween 20	-	-	-	Fisher Scientific, Pittsburgh, PA, USA
Versene, 0.2 g EDTA(Na4) /L of Phosphate Buffered Saline (PBS), 0.48 mM	-	-	-	Gibco, Carlsbad, CA, USA
***Xba*I restriction enzyme**	-	-	-	Promega, Madison, WI, USA

10. References

Ahmadian, A., Ehn, M., and Hober, S. (2006). Pyrosequencing: history, biochemistry and future. *Clin Chim Acta* 363**,** 83-94.

Allard, M.W., Luo, Y., Strain, E., Pettengill, J., Timme, R., Wang, C., Li, C., Keys, C.E., Zheng, J., Stones, R., Wilson, M.R., Musser, S.M., and Brown, E.W. (2013). On the evolutionary history, population genetics and diversity among isolates of *Salmonella enteritidis* PFGE pattern JEGX01.0004. *PLoS One* 8**,** e55254.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215**,** 403-410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25**,** 3389-3402.

Amaral, G.R., Dias, G.M., Wellington-Oguri, M., Chimetto, L., Campeao, M.E., Thompson, F.L., and Thompson, C.C. (2014). Genotype to phenotype: identification of diagnostic vibrio phenotypes using whole genome sequences. *Int J Syst Evol Microbiol* 64**,** 357-365.

Bacon, D.J., and Anderson, W.F. (1986). Multiple sequence alignment. *Journal of molecular biology* 191**,** 153-161.

Baker, M. (2012). De novo genome assembly: what every biologist should know. *Nat Meth* 9**,** 333-337.

Banakar, V., Constantin De Magny, G., Jacobs, J., Murtugudde, R., Huq, A., Wood, R.J., and Colwell, R.R. (2011). Temporal and spatial variability in the distribution of *Vibrio vulnificus* in the Chesapeake Bay: a hindcast study. *Ecohealth* 8**,** 456-467.

Barrett, T.J., Gerner-Smidt, P., and Swaminathan, B. (2006). Interpretation of pulsed-field gel electrophoresis patterns in foodborne disease investigations and surveillance. *Foodborne.Pathog.Dis.* 3**,** 20-31.

Barua, D. (1992). "History of Cholera," in *Cholera,* eds. D. Barua & W. Greenough, Iii. Springer US), 1-36.

Bassam, B.J., Caetano-Anolles, G., and Gresshoff, P.M. (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* 196**,** 80-83.

Bassam, B.J., and Gresshoff, P.M. (2007). Silver staining DNA in polyacrylamide gels. *Nat Protoc* 2**,** 2649-2654.

Baxevanis, A.D., and Ouellette, B.F.F. (2004). *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins.* Wiley.

Bhowmik, S.K., Pazhani, G.P., and Ramamurthy, T. (2014). Phylogenetic and *in silico* functional analyses of thermostable-direct hemolysin and *tdh*-related encoding genes in *Vibrio parahaemolyticus* and other Gram-negative bacteria. *Biomed Res Int* 2014**,** 576528.

Bhunia, A. (2007). *Foodborne Microbial Pathogens: Mechanisms and Pathogenesis.* Springer Sciene & Business Media.

Bikandi, J., and San Millan, R. (2012). "Discriminatory Power Calculator".).

Blackstone, G.M., Nordstrom, J.L., Vickery, M.C., Bowen, M.D., Meyer, R.F., and Depaola, A. (2003). Detection of pathogenic *Vibrio parahaemolyticus* in oyster enrichments by real time PCR. *J.Microbiol.Methods* 53**,** 149-155.

Breidenbach, J., and Frank, C. (2012). "Informationsbroschüre zu Nicht-Cholera-Vibrionen in Deutschland". Robert-Koch-Institut).

Broberg, C.A., Calder, T.J., and Orth, K. (2011). *Vibrio parahaemolyticus* cell biology and pathogenicity determinants. *Microbes.Infect.* 13**,** 992-1001.

Brock, T.D. (1999). *Robert Koch.* American Society of Microbiology.

Brunner, K.T., Mauel, J., Cerottini, J.C., and Chapuis, B. (1968). Quantitative assay of the lytic action of immune lymphoid cells on 51-Cr-labelled allogeneic target cells in vitro; inhibition by isoantibody and by drugs. *Immunology* 14**,** 181-196.

Bryant, T.N., Lee, J.V., West, P.A., and Colwell, R.R. (1986). A probability matrix for the identification of species of *Vibrio* and related genera. *J Appl Bacteriol* 61**,** 469-480.

Burdette, D.L., Seemann, J., and Orth, K. (2009). Vibrio VopQ induces PI3-kinase-independent autophagy and antagonizes phagocytosis. *Mol.Microbiol.* 73**,** 639-649.

Burdette, D.L., Yarbrough, M.L., Orvedahl, A., Gilpin, C.J., and Orth, K. (2008). *Vibrio parahaemolyticus* orchestrates a multifaceted host cell infection by induction of autophagy, cell rounding, and then cell lysis. *Proc.Natl.Acad.Sci.U.S.A* 105**,** 12497-12502.

Caburlotto, G., Lleo, M.M., Hilton, T., Huq, A., Colwell, R.R., and Kaper, J.B. (2010). Effect on human cells of environmental *Vibrio parahaemolyticus* strains carrying type III secretion system 2. *Infection and Immunity* 78**,** 3280-3287.

Cardullo, R.A., Agrawal, S., Flores, C., Zamecnik, P.C., and Wolf, D.E. (1988). Detection of nucleic acid hybridization by nonradiative fluorescence resonance energy transfer. *Proc Natl Acad Sci U S A* 85**,** 8790-8794.

Cavalli-Sforza, L.L., and Edwards, A.W. (1967). Phylogenetic analysis. Models and estimation procedures. *American journal of human genetics* 19**,** 233.

Centers for Disease, C., and Prevention (2009). Standard Opreating Procedure for PFGE of *Vibrio parahaemolyticus CDC PulseNet protocols*.

Chenal-Francisque, V., Diancourt, L., Cantinelli, T., Passet, V., Tran-Hykes, C., Bracq-Dieye, H., Leclercq, A., Pourcel, C., Lecuit, M., and Brisse, S. (2013). Optimized Multilocus variable-number tandem-repeat analysis assay and its complementarity with pulsed-field gel electrophoresis and multilocus sequence typing for *Listeria monocytogenes* clone identification and surveillance. *J Clin Microbiol* 51**,** 1868-1880.

Chin, C.S., Sorenson, J., Harris, J.B., Robins, W.P., Charles, R.C., Jean-Charles, R.R., Bullard, J., Webster, D.R., Kasarskis, A., Peluso, P., Paxinos, E.E., Yamaichi, Y., Calderwood, S.B., Mekalanos, J.J., Schadt, E.E., and Waldor, M.K. (2011). The origin of the Haitian cholera outbreak strain. *N Engl J Med* 364**,** 33-42.

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., and 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**,** 319-327.

Chowdhury, N.R., Chakraborty, S., Ramamurthy, T., Nishibuchi, M., Yamasaki, S., Takeda, Y., and Nair, G.B. (2000). Molecular evidence of clonal *Vibrio parahaemolyticus* pandemic strains. *Emerg.Infect.Dis.* 6**,** 631-636.

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

Compeau, P.E.C., Pevzner, P.A., and Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nat Biotech* 29**,** 987-991.

Cooper, K.L., Luey, C.K., Bird, M., Terajima, J., Nair, G.B., Kam, K.M., Arakawa, E., Safa, A., Cheung, D.T., Law, C.P., Watanabe, H., Kubota, K., Swaminathan, B., and Ribot, E.M. (2006). Development and validation of a PulseNet standardized pulsed-field gel electrophoresis protocol for subtyping of *Vibrio cholerae. Foodborne.Pathog.Dis.* 3**,** 51-58.

Crim, S.M., Griffin, P.M., Tauxe, R., Marder, E.P., Gilliss, D., Cronquist, A.B., Cartter, M., Tobin-D'angelo, M., Blythe, D., Smith, K., Lathrop, S., Zansky, S., Cieslak, P.R., Dunn, J., Holt, K.G., Wolpert, B., Henao, O.L., Centers for Disease, C., and Prevention (2015). Preliminary incidence and trends of infection with pathogens transmitted commonly through food - Foodborne Diseases Active Surveillance Network, 10 U.S. sites, 2006-2014. *MMWR Morb Mortal Wkly Rep* 64**,** 495-499.

Croci, L., Suffredini, E., Cozzi, L., Toti, L., Ottaviani, D., Pruzzo, C., Serratore, P., Fischetti, R., Goffredo, E., Loffredo, G., Mioni, R., and Vibrio Parahaemolyticus Working, G. (2007). Comparison of different biochemical and molecular methods for the identification of *Vibrio parahaemolyticus*. *J Appl Microbiol* 102**,** 229-237.

Cui, Y., Yang, X., Didelot, X., Guo, C., Li, D., Yan, Y., Zhang, Y., Yuan, Y., Yang, H., Wang, J., Wang, J., Song, Y., Zhou, D., Falush, D., and Yang, R. (2015). Epidemic Clones,

Oceanic Gene Pools, and Eco-LD in the Free Living Marine Pathogen *Vibrio parahaemolyticus*. *Mol Biol Evol* 32**,** 1396-1410.

Dark, M.J. (2013). Whole-genome sequencing in bacteriology: state of the art. *Infect Drug Resist* 6**,** 115-123.

Dauros, P., Bello, H., Dominguez, M., Hormazabal, J.C., and Gonzalez, G. (2011). Characterization of *Vibrio parahaemolyticus* strains isolated in Chile in 2005 and in 2007. *J.Infect.Dev.Ctries.* 5**,** 502-510.

Davis, J.M. (2002). "Basic Cell Culture". (New York: Oxford University Press).

Dechet, A.M., Yu, P.A., Koram, N., and Painter, J. (2008). Nonfoodborne Vibrio infections: an important cause of morbidity and mortality in the United States, 1997-2006. *Clin Infect Dis* 46**,** 970-976.

Decker, T., and Lohmann-Matthes, M.L. (1988). A quick and simple method for the quantitation of lactate dehydrogenase release in measurements of cellular cytotoxicity and tumor necrosis factor (TNF) activity. *J Immunol Methods* 115**,** 61-69.

Depaola, A., Kaysner, C.A., Bowers, J., and Cook, D.W. (2000). Environmental investigations of *Vibrio parahaemolyticus* in oysters after outbreaks in Washington, Texas, and New York (1997 and 1998). *Appl.Environ.Microbiol.* 66**,** 4649-4654.

Depaola, A., Ulaszek, J., Kaysner, C.A., Tenge, B.J., Nordstrom, J.L., Wells, J., Puhr, N., and Gendel, S.M. (2003). Molecular, serological, and virulence characteristics of *Vibrio parahaemolyticus* isolated from environmental, food, and clinical sources in North America and Asia. *Appl.Environ.Microbiol.* 69**,** 3999-4005.

Dickinson, G., Lim, K.Y., and Jiang, S.C. (2013). Quantitative microbial risk assessment of pathogenic vibrios in marine recreational waters of southern california. *Appl Environ Microbiol* 79**,** 294-302.

Ebel, E.D., and Williams, M.S. (2015). When Are Qualitative Testing Results Sufficient To Predict a Reduction in Illnesses in a Microbiological Food Safety Risk Assessment? *J Food Prot* 78**,** 1451-1460.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32**,** 1792-1797.

Edgar, R.C., and Batzoglou, S. (2006). Multiple sequence alignment. *Current opinion in structural biology* 16**,** 368-373.

Edwards, A., and Cavalli-Sforza, L. (1964). Reconstruction of evolutionary trees in Phenetic and phylogenetic classification. *Systematics Association Publication*.

El-Malah, S.S., Yang, Z., Hu, M., Li, Q., Pan, Z., and Jiao, X. (2014). *Vibrio parahaemolyticus* strengthens their virulence through modulation of cellular reactive oxygen species in vitro. *Front Cell Infect Microbiol* 4**,** 168.

European Food Safety Authority, E.C.O.D., Prevention, and Control "The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2012".).

Eversole, A.G. (1987). "Species profiles: Life Histories and Environmental Requirements of Coastal Fishes and Invertebrates (South Atlantic) - Hard Clam". US Fish and Wildlife Service Biological Report).

Faruque, S.M., Nair, G.B., and Mekalanos, J.J. (2004). Genetics of stress adaptation and virulence in toxigenic *Vibrio cholerae*. *DNA Cell Biol* 23**,** 723-741.

Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P., and Spratt, B.G. (2004). eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* 186**,** 1518-1530.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17**,** 368-376.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution***,** 783-791.

Fogh, J., Fogh, J.M., and Orfeo, T. (1977). One hundred and twenty-seven cultured human tumor cell lines producing tumors in nude mice. *J.Natl.Cancer Inst.* 59**,** 221-226.

Fogh, J., Orfeo, T., Tiso, J., and Sharkey, F.E. (1979). Establishment of human colon carcinoma lines in nude mice. *Exp.Cell Biol.* 47**,** 136-144.

Fortini, D., Ciammaruconi, A., De, S.R., Fasanella, A., Battisti, A., D'amelio, R., Lista, F., Cassone, A., and Carattoli, A. (2007). Optimization of high-resolution melting analysis for low-cost and rapid screening of allelic variants of *Bacillus anthracis* by multiple-locus variable-number tandem repeat analysis. *Clin.Chem.* 53**,** 1377-1380.

Foxman, B., Zhang, L., Koopman, J.S., Manning, S.D., and Marrs, C.F. (2005). Choosing an appropriate bacterial typing technique for epidemiologic studies. *Epidemiol Perspect Innov* 2**,** 10.

Fuenzalida, L., Hernandez, C., Toro, J., Rioseco, M.L., Romero, J., and Espejo, R.T. (2006). *Vibrio parahaemolyticus* in shellfish and clinical samples during two large epidemics of diarrhoea in southern Chile. *Environ.Microbiol.* 8**,** 675-683.

Fujino, T., Miwatani, T., Takeda, Y., and Tomaru, A. (1969). A thermolabile direct hemolysin of *Vibrio parahaemolyticus. Biken.J.* 12**,** 145-148.

Fukui, T., Shiraki, K., Hamada, D., Hara, K., Miyata, T., Fujiwara, S., Mayanagi, K., Yanagihara, K., Iida, T., Fukusaki, E., Imanaka, T., Honda, T., and Yanagihara, I. (2005). Thermostable direct hemolysin of *Vibrio parahaemolyticus* is a bacterial reversible amyloid toxin. *Biochemistry* 44**,** 9825-9832.

Galan, J.E., and Wolf-Watz, H. (2006). Protein delivery into eukaryotic cells by type III secretion machines. *Nature* 444**,** 567-573.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* 43**,** D261-269.

Galtstoff, P.S. (1964). *The American oyster Crassostrea virginica Gmelin.*

Gardner, S.N., and Hall, B.G. (2013). When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. *PLoS.One.* 8**,** e81760.

Garrity, G., Brenner, D.J., Krieg, N.R., and Staley, J.R. (2007). *Bergey's Manual of Systematic Bacteriology.* Springer Science & Business Media.

Ginzinger, D.G. (2002). Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Hematol* 30**,** 503-512.

Giulietti, A., Overbergh, L., Valckx, D., Decallonne, B., Bouillon, R., and Mathieu, C. (2001). An overview of real-time quantitative PCR: applications to quantify cytokine gene expression. *Methods* 25**,** 386-401.

Gonzalez-Escalona, N., Gavilan, R.G., Brown, E.W., and Martinez-Urtaza, J. (2015). Transoceanic Spreading of Pathogenic Strains of *Vibrio parahaemolyticus* with Distinctive Genetic Signatures in the recA Gene. *PLoS.One.* 10**,** e0117485.

Gonzalez-Escalona, N., Jaykus, L.A., and Depaola, A. (2007a). Typing of *Vibrio vulnificus* strains by variability in their 16S-23S rRNA intergenic spacer regions. *Foodborne.Pathog.Dis.* 4**,** 327-337.

Gonzalez-Escalona, N., Martinez-Urtaza, J., Romero, J., Espejo, R.T., Jaykus, L.A., and Depaola, A. (2008). Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. *J.Bacteriol.* 190**,** 2831-2840.

González-Escalona, N., Romero, J., and Espejo, R.T. (2005). Polymorphism and gene conversion of the 16S rRNA genes in the multiple rRNA operons of *Vibrio parahaemolyticus*. *FEMS microbiology letters* 246**,** 213-219.

Gonzalez-Escalona, N., Romero, J., Guzman, C.A., and Espejo, R.T. (2006). Variation in the 16S-23S rRNA intergenic spacer regions in *Vibrio parahaemolyticus* strains are due to indels nearby their tRNAGlu. *FEMS Microbiol.Lett.* 256**,** 38-43.

Gonzalez-Escalona, N., Whitney, B., Jaykus, L.A., and Depaola, A. (2007b). Comparison of direct genome restriction enzyme analysis and pulsed-field gel electrophoresis for typing of *Vibrio vulnificus* and their correspondence with multilocus sequence typing data. *Appl.Environ.Microbiol.* 73**,** 7494-7500.

Gosling, E. (2008). *Bivalve Molluscs Biology, Ecology and Culture.* John Wiley & Sons.

Grimm, S. (2004). The art and design of genetic screens: mammalian culture cells. *Nat.Rev.Genet.* 5**,** 179-189.

Gulig, P.A., De Crecy-Lagard, V., Wright, A.C., Walts, B., Telonis-Scott, M., and Mcintyre, L.M. (2010). SOLiD sequencing of four *Vibrio vulnificus* genomes enables comparative genomic analysis and identification of candidate clade-specific virulence genes. *BMC.Genomics* 11**,** 512.

Gupta, P.K. (2008). Single-molecule DNA sequencing technologies for future genomics research. *Trends Biotechnol* 26**,** 602-611.

Haase, I.V., F; Fischer, M (2012). Trendbericht Lebensmittelchemie 2011: Molekularbiologische Methoden in der Lebensmittelanalytik. *Nachrichten aus der Chemie* 60.

Haendiges, J., Timme, R., Allard, M.W., Myers, R.A., Brown, E.W., and Gonzalez-Escalona, N. (2015). Characterization of *Vibrio parahaemolyticus* clinical strains from Maryland (2012-2013) and comparisons to a locally and globally diverse *V. parahaemolyticus* strains by whole-genome sequence analysis. *Front Microbiol* 6**,** 125.

Ham, H., and Orth, K. (2012). The role of type III secretion System 2 in *Vibrio parahaemolyticus* pathogenicity. *J.Microbiol.* 50**,** 719-725.

Han, T.J., and Chai, T.J. (1992). Electrophoretic and chemical characterization of lipopolysaccharides of *Vibrio parahaemolyticus*. *J Bacteriol* 174**,** 3140-3146.

Harth-Chu, E., Espejo, R.T., Christen, R., Guzman, C.A., and Hofle, M.G. (2009). Multiple-locus variable-number tandem-repeat analysis for clonal identification of *Vibrio parahaemolyticus* isolates by using capillary electrophoresis. *Appl.Environ.Microbiol.* 75**,** 4079-4088.

Hayat, U., Reddy, G.P., Bush, C.A., Johnson, J.A., Wright, A.C., and Morris, J.G., Jr. (1993). Capsular types of *Vibrio vulnificus*: an analysis of strains from clinical and environmental sources. *J Infect Dis* 168**,** 758-762.

Hazen, T.H., Lafon, P.C., Garrett, N.M., Lowe, T.M., Silberger, D.J., Rowe, L.A., Frace, M., Parsons, M.B., Bopp, C.A., Rasko, D.A., and Sobecky, P.A. (2015). Insights into the environmental reservoir of pathogenic *Vibrio parahaemolyticus* using comparative genomics. *Front Microbiol* 6**,** 204.

Heid, C.A., Stevens, J., Livak, K.J., and Williams, P.M. (1996). Real time quantitative PCR. *Genome Res* 6**,** 986-994.

Herschleb, J., Ananiev, G., and Schwartz, D.C. (2007). Pulsed-field gel electrophoresis. *Nat.Protoc.* 2**,** 677-684.

Higa, N., Toma, C., Koizumi, Y., Nakasone, N., Nohara, T., Masumoto, J., Kodama, T., Iida, T., and Suzuki, T. (2013). *Vibrio parahaemolyticus* effector proteins suppress inflammasome activation by interfering with host autophagy signaling. *PLoS.Pathog.* 9**,** e1003142.

Hiyoshi, H., Kodama, T., Iida, T., and Honda, T. (2010). Contribution of *Vibrio parahaemolyticus* virulence factors to cytotoxicity, enterotoxicity, and lethality in mice. *Infection and Immunity* 78**,** 1772-1780.

Hiyoshi, H., Kodama, T., Saito, K., Gotoh, K., Matsuda, S., Akeda, Y., Honda, T., and Iida, T. (2011). VopV, an F-actin-binding type III secretion effector, is required for *Vibrio parahaemolyticus* -induced enterotoxicity. *Cell Host.Microbe* 10**,** 401-409.

Hoffmann, M., Brown, E.W., Feng, P.C., Keys, C.E., Fischer, M., and Monday, S.R. (2010). PCR-based method for targeting 16S-23S rRNA intergenic spacer regions among Vibrio species. *BMC.Microbiol.* 10**,** 90.

Holland, P.M., Abramson, R.D., Watson, R., and Gelfand, D.H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci U S A* 88**,** 7276-7280.

Honda, T., Abad-Lapuebla, M.A., Ni, Y.X., Yamamoto, K., and Miwatani, T. (1991). Characterization of a new thermostable direct haemolysin produced by a Kanagawa-phenomenon-negative clinical isolate of *Vibrio parahaemolyticus. J.Gen.Microbiol.* 137**,** 253-259.

Honda, T., Ni, Y.X., and Miwatani, T. (1988). Purification and characterization of a hemolysin produced by a clinical isolate of Kanagawa phenomenon-negative *Vibrio parahaemolyticus* and related to the thermostable direct hemolysin. *Infection and Immunity* 56**,** 961-965.

Hoorfar, J., Malorny, B., Abdulmawjood, A., Cook, N., Wagner, M., and Fach, P. (2004). Practical considerations in design of internal amplification controls for diagnostic PCR assays. *J Clin Microbiol* 42**,** 1863-1868.

Hsieh, Y.C., Liang, S.M., Tsai, W.L., Chen, Y.H., Liu, T.Y., and Liang, C.M. (2003). Study of capsular polysaccharide from *Vibrio parahaemolyticus. Infection and Immunity* 71**,** 3329-3336.

Huehn, S., Eichhorn, C., Urmersbach, S., Breidenbach, J., Bechlars, S., Bier, N., Alter, T., Bartelt, E., Frank, C., Oberheitmann, B., Gunzer, F., Brennholt, N., Boer, S., Appel, B., Dieckmann, R., and Strauch, E. (2014). Pathogenic vibrios in environmental, seafood and clinical sources in Germany. *Int.J.Med.Microbiol.*

Hunter, P.R., and Gaston, M.A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J.Clin.Microbiol.* 26**,** 2465-2466.

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11**,** 119.

Hyytiä-Trees, E., Smole, S.C., Fields, P.A., Swaminathan, B., and Ribot, E.M. (2006). Second generation subtyping: a proposed PulseNet protocol for multiple-locus variable-number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 (STEC O157). *Foodborne.Pathog.Dis.* 3**,** 118-131.

Iguchi, T., Kondo, S., and 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**,** 287-292.

Iida, T., Suthienkul, O., Park, K.S., Tang, G.Q., Yamamoto, R.K., Ishibashi, M., Yamamoto, K., and Honda, T. (1997). Evidence for genetic linkage between the *ure* and *trh* genes in *Vibrio parahaemolyticus*. *J Med Microbiol* 46**,** 639-645.

Iida, T., and Yamamoto, K. (1990). Cloning and expression of two genes encoding highly homologous hemolysins from a Kanagawa phenomenon-positive *Vibrio parahaemolyticus* T4750 strain. *Gene* 93**,** 9-15.

Janda, J.M., Newton, A.E., and Bopp, C.A. (2015). Vibriosis. *Clin Lab Med* 35**,** 273-288.

Jeannotte, R.E., Kong, N., Ng, W., and Weimer, B.C. (2014). "High-Throughput Analysis of Foodborne Bacterial Genomic DNA Using Agilent 2200 TapeStation and Genomic DNA ScreenTape System". Agilent Technologies.

Johnson, N.B., Hayes, L.D., Brown, K., Hoo, E.C., and Ethier, K.A. (2014). CDC National Health Report: Leading Causes of Morbidity and Mortality and Associated Behavioral Risk and Protective Factors-United States, 2005-2013. *MMWR Surveill Summ.* 63**,** 3-27.

Jolley, K.A., and Maiden, M.C. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11**,** 595.

Jones, J.L., Ludeke, C.H.M., Garett, N., Fischer, M., Bopp, C.A., and Depaola, A. (2012). Biochemical, Serological, and Virulence Characterization of Clinical and Oyster *Vibrio parahaemolyticus* Isolates. *J.Clin.Microbiol.* 50**,** 2343-2352.

Jukes, T.H., and Cantor, C.R. (1969). Evolution of protein molecules. *Mammalian protein metabolism* 3**,** 132.

Kam, K.M., Luey, C.K., Parsons, M.B., Cooper, K.L., Nair, G.B., Alam, M., Islam, M.A., Cheung, D.T., Chu, Y.W., Ramamurthy, T., Pazhani, G.P., Bhattacharya, S.K., Watanabe, H., Terajima, J., Arakawa, E., Ratchtrachenchai, O.A., Huttayananont, S., Ribot, E.M., Gerner-Smidt, P., and Swaminathan, B. (2008). Evaluation and validation of a PulseNet standardized pulsed-field gel electrophoresis protocol for subtyping *Vibrio parahaemolyticus*: an international multicenter collaborative study. *J.Clin.Microbiol.* 46**,** 2766-2773.

Kaufmann, M.E. (1998). Pulsed-Field Gel Electrophoresis. *Molecular Bacteriology* 15**,** 33-50.

Kaysner, C.A., and Depaola, A. (2004). Vibrio. *FDA Bacterial Analytical Manual*.

Kimura, B., Sekine, Y., Takahashi, H., Tanaka, Y., Obata, H., Kai, A., Morozumi, S., and Fujii, T. (2008). Multiple-locus variable-number of tandem-repeats analysis distinguishes *Vibrio parahaemolyticus* pandemic O3:K6 strains. *J.Microbiol.Methods* 72**,** 313-320.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J.Mol.Evol.* 16**,** 111-120.

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

Klimke, W., Agarwala, R., Badretdin, A., Chetvernin, S., Ciufo, S., Fedorov, B., Kiryutin, B., O'neill, K., Resch, W., Resenchuk, S., Schafer, S., Tolstoy, I., and Tatusova, T. (2009). The National Center for Biotechnology Information's Protein Clusters Database. *Nucleic Acids Res.* 37**,** D216-D223.

Kobayashi, T., Enomoto, S., Sakazaki, R., and Kuwahara, S. (1963). A new selective Isolation medium for the Vibrio group; on a modified Nakanishi's medium (TCBS agar medium. *Nihon Saikingaku Zasshi* 18**,** 387-392.

Kong, N., Thao, K., Ng, W., Kim, K.S., Korlach, J., Hickey, L., Kelly, L., Lappin, S., and Weimer, B.C. (2014). "Automation of PacBio SMRTbellT 10 kb Library Preparation on Agilent Bravo NGS Workstation". Agilent Technologies.

Korzeniewski, C., and Callewaert, D.M. (1983). An enzyme-release assay for natural cytotoxicity. *J Immunol Methods* 64**,** 313-320.

Krachler, A.M., Ham, H., and Orth, K. (2011). Outer membrane adhesion factor multivalent adhesion molecule 7 initiates host cell binding during infection by gram-negative pathogens. *Proc.Natl.Acad.Sci.U.S.A* 108**,** 11614-11619.

Krachler, A.M., Ham, H., and Orth, K. (2012). Turnabout is fair play: use of the bacterial Multivalent Adhesion Molecule 7 as an antimicrobial agent. *Virulence.* 3**,** 68-71.

Krachler, A.M., and Orth, K. (2011). Functional characterization of the interaction between bacterial adhesin multivalent adhesion molecule 7 (MAM7) protein and its host cell ligands. *J.Biol.Chem.* 286**,** 38939-38947.

Kraeuter, J.N., and Castagna, M. (2001). *Biology of the Hard Clam.*

Krämer, J. (2007). *Lebensmittel-Mikrobiologie.* Ulmer.

Kueh, C.S., and Chan, K.Y. (1985). Bacteria in bivalve shellfish with special reference to the oyster. *J.Appl.Bacteriol.* 59**,** 41-47.

Land, M., Hauser, L., Jun, S.R., Nookaew, I., Leuze, M.R., Ahn, T.H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., and Ussery, D.W. (2015). Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* 15**,** 141-161.

Land, M.L., Hyatt, D., Jun, S.R., Kora, G.H., Hauser, L.J., Lukjancenko, O., and Ussery, D.W. (2014). Quality scores for 32,000 genomes. *Stand Genomic Sci* 9**,** 20.

Laohaprertthisan, V., Chowdhury, A., Kongmuang, U., Kalnauwakul, S., Ishibashi, M., Matsumoto, C., and Nishibuchi, M. (2003). Prevalence and serodiversity of the pandemic clone among the clinical strains of *Vibrio parahaemolyticus* isolated in southern Thailand. *Epidemiol Infect* 130**,** 395-406.

Le Roux, F., Wegner, K.M., Baker-Austin, C., Vezzulli, L., Osorio, C.R., Amaro, C., Ritchie, J.M., Defoirdt, T., Destoumieux-Garzon, D., Blokesch, M., Mazel, D., Jacq, A., Cava, F., Gram, L., Wendling, C.C., Strauch, E., Kirschner, A., and Huehn, S. (2015). The emergence of Vibrio pathogens in Europe: ecology, evolution, and pathogenesis (Paris, 11-12th March 2015). *Front Microbiol* 6**,** 830.

Li, L., Wong, H.C., Nong, W., Cheung, M.K., Law, P.T., Kam, K.M., and Kwan, H.S. (2014). Comparative genomic analysis of clinical and environmental strains provides insight into the pathogenicity and evolution of *Vibrio parahaemolyticus. BMC.Genomics* 15**,** 1135.

Lievin-Le, M.V., and Servin, A.L. (2013). Pathogenesis of human enterovirulent bacteria: lessons from cultured, fully differentiated human colon cancer cell lines. *Microbiol.Mol.Biol.Rev.* 77**,** 380-439.

Lindstedt, B.A. (2005). Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis* 26**,** 2567-2582.

Lindstedt, B.A., Torpdahl, M., Vergnaud, G., Le, H.S., Weill, F.X., Tietze, E., Malorny, B., Prendergast, D.M., Ni, G.E., Lista, R.F., Schouls, L.M., Soderlund, R., Borjesson, S., and Akerstrom, S. (2013). Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012. *Euro.Surveill* 18**,** 20385.

Liu, M., Yan, M., Liu, L., and Chen, S. (2013). Characterization of a novel zinc transporter ZnuA acquired by *Vibrio parahaemolyticus* through horizontal gene transfer. *Front Cell Infect.Microbiol.* 3**,** 61.

Liverman, A.D., Cheng, H.C., Trosky, J.E., Leung, D.W., Yarbrough, M.L., Burdette, D.L., Rosen, M.K., and Orth, K. (2007). Arp2/3-independent assembly of actin by Vibrio type III effector VopL. *Proc.Natl.Acad.Sci.U.S.A* 104**,** 17117-17122.

Ludeke, C.H., Fischer, M., Lafon, P., Cooper, K., and Jones, J.L. (2014). Suitability of the molecular subtyping methods intergenic spacer region, direct genome restriction analysis, and pulsed-field gel electrophoresis for clinical and environmental *Vibrio parahaemolyticus* isolates. *Foodborne.Pathog.Dis.* 11**,** 520-528.

Ludeke, C.H., Gonzalez-Escalona, N., Fischer, M., and Jones, J.L. (2015). Examination of clinical and environmental *Vibrio parahaemolyticus* isolates by multi-locus sequence typing (MLST) and multiple-locus variable-number tandem-repeat analysis (MLVA). *Front Microbiol* 6**,** 564.

Lynch, T., Livingstone, S., Buenaventura, E., Lutter, E., Fedwick, J., Buret, A.G., Graham, D., and Devinney, R. (2005). *Vibrio parahaemolyticus* disruption of epithelial cell tight junctions occurs independently of toxin production. *Infection and Immunity* 73**,** 1275-1283.

Lyon, E. (2001). Mutation detection using fluorescent hybridization probes and melting curve analysis. *Expert Rev Mol Diagn* 1**,** 92-101.

Maeda, T., Takada, N., Furushita, M., and Shiba, T. (2000). Structural variation in the 16S-23S rRNA intergenic spacers of *Vibrio parahaemolyticus. FEMS Microbiol.Lett.* 192**,** 73-77.

Maiden, M.C. (2006). Multilocus sequence typing of bacteria. *Annu.Rev.Microbiol.* 60**,** 561-588.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., and Spratt, B.G. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc.Natl.Acad.Sci.U.S.A* 95**,** 3140-3145.

Maiden, M.C., Van Rensburg, M.J., Bray, J.E., Earle, S.G., Ford, S.A., Jolley, K.A., and Mccarthy, N.D. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat.Rev.Microbiol.* 11**,** 728-736.

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., and Iida, T. (2003). Genome sequence of *Vibrio parahaemolyticus* : a pathogenic mechanism distinct from that of *V. cholerae. Lancet* 361**,** 743-749.

Malham, S.K., Rajko-Nenow, P., Howlett, E., Tuson, K.E., Perkins, T.L., Pallett, D.W., Wang, H., Jago, C.F., Jones, D.L., and Mcdonald, J.E. (2014). The interaction of human microbial pathogens, particulate material and nutrients in estuarine environments and their impacts on recreational and shellfish waters. *Environ Sci Process Impacts* 16**,** 2145-2155.

Mandel, M.J., Stabb, E.V., and Ruby, E.G. (2008). Comparative genomics-based investigation of resequencing targets in *Vibrio fischeri*: focus on point miscalls and artefactual expansions. *BMC Genomics* 9**,** 138.

Martinez-Urtaza, J., Baker-Austin, C., Jones, J.L., Newton, A.E., and Depaola, A. (2013a). Spread of Pacific Northwest *Vibrio parahaemolyticus* Strain. *N.Engl.J.Med.* 369.

Martinez-Urtaza, J., Baker-Austin, C., Jones, J.L., Newton, A.E., Gonzalez-Aviles, G.D., and Depaola, A. (2013b). Spread of Pacific Northwest *Vibrio parahaemolyticus* strain. *N Engl J Med* 369**,** 1573-1574.

Martinez-Urtaza, J., Lozano-Leon, A., Vina-Feas, A., De Novoa, J., and Garcia-Martin, O. (2006). Differences in the API 20E biochemical patterns of clinical and environmental *Vibrio parahaemolyticus* isolates. *FEMS Microbiol Lett* 255**,** 75-81.

Martinez-Urtaza, J., Simental, L., Velasco, D., Depaola, A., Ishibashi, M., Nakaguchi, Y., Nishibuchi, M., Carrera-Flores, D., Rey-Alvarez, C., and Pousa, A. (2005). Pandemic *Vibrio parahaemolyticus* O3:K6, Europe. *Emerg.Infect.Dis.* 11**,** 1319-1320.

Masters, J.R. (2002). HeLa cells 50 years on: the good, the bad and the ugly. *Nat.Rev.Cancer* 2**,** 315-319.

Matsuda, S., Kodama, T., Okada, N., Okayama, K., Honda, T., and Iida, T. (2010). Association of *Vibrio parahaemolyticus* thermostable direct hemolysin with lipid rafts is essential for cytotoxicity but not hemolytic activity. *Infection and Immunity* 78**,** 603-610.

Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 74**,** 560-564.

McCarthy, S.A., Depaola, A., Cook, D.W., Kaysner, C.A., and Hill, W.E. (1999). Evaluation of alkaline phosphatase- and digoxigenin-labelled probes for detection of the thermolabile hemolysin (*tlh*) gene of *Vibrio parahaemolyticus. Lett.Appl.Microbiol.* 28**,** 66-70.

McCarthy, S.A., Depaola, A., Kaysner, C.A., Hill, W.E., and Cook, D.W. (2000). Evaluation of nonisotopic DNA hybridization methods for detection of the *tdh* gene of *Vibrio parahaemolyticus. J.Food Prot.* 63**,** 1660-1664.

McLaughlin, J.B., Depaola, A., Bopp, C.A., Martinek, K.A., Napolilli, N.P., Allison, C.G., Murray, S.L., Thompson, E.C., Bird, M.M., and Middaugh, J.P. (2005). Outbreak of *Vibrio parahaemolyticus* gastroenteritis associated with Alaskan oysters. *N.Engl.J.Med.* 353**,** 1463-1470.

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr Opin Genet Dev* 15**,** 589-594.

Menzel, W. (1991). *Estuarine and Marine Bivalve Mollusk Culture.* Taylor & Francis.

Merino, S., Shaw, J.G., and Tomas, J.M. (2006). Bacterial lateral flagella: an inducible flagella system. *FEMS Microbiol.Lett.* 263**,** 127-135.

Miyamoto, M., Motooka, D., Gotoh, K., Imai, T., Yoshitake, K., Goto, N., Iida, T., Yasunaga, T., Horii, T., Arakawa, K., Kasahara, M., and Nakamura, S. (2014). Performance comparison of second- and third-generation sequencers using a bacterial genome with two chromosomes. *BMC Genomics* 15**,** 699.

Morozova, O., and Marra, M.A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92**,** 255-264.

Morris, J.G. (2014). Control strategies for vibrios: borrowing from the Japanese experience. *Epidemiol Infect* 142**,** 2248-2250.

Morrison, T.B., Weis, J.J., and Wittwer, C.T. (1998). Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques* 24**,** 954-958, 960, 962.

Mukherjee, M., Kakarla, P., Kumar, S., Gonzalez, E., Floyd, J.T., Inupakutika, M., Devireddy, A.R., Tirrell, S.R., Bruns, M., He, G., Lindquist, I.E., Sundararajan, A., Schilkey, F.D., Mudge, J., and Varela, M.F. (2014). Comparative genome analysis of non-toxigenic non-O1 versus toxigenic O1. *Genom Discov* 2**,** 1-15.

Mülhardt, C. (2009). *Der Experimentator: Molekularbiologie / Genomics.* Spektrum Akademischer Verlag.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 51 Pt 1**,** 263-273.

Murray, P.R., Rosenthal, K.S., and Pfaller, M.A. (2002). *Medical Microbiology.* Elsevier Health Sciences.

Nair, G.B., Ramamurthy, T., Bhattacharya, S.K., Dutta, B., Takeda, Y., and Sack, D.A. (2007). Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. *Clin Microbiol Rev* 20**,** 39-48.

Newton, A., Kendall, M., Vugia, D.J., Henao, O.L., and Mahon, B.E. (2012). Increasing rates of vibriosis in the United States, 1996-2010: review of surveillance data from 2 systems. *Clin.Infect.Dis.* 54 Suppl 5**,** S391-S395.

Newton, A.E., Garrett, N., Stroika, S.G., Halpin, J.L., Turnsek, M., and Mody, R.K. (2014a). Increase in *Vibrio parahaemolyticus* infections associated with consumption of Atlantic Coast shellfish--2013. *MMWR Morb.Mortal.Wkly.Rep.* 63**,** 335-336.

Newton, A.E., Garrett, N., Stroika, S.G., Halpin, J.L., Turnsek, M., Mody, R.K., Centers for Disease, C., and Prevention (2014b). Increase in *Vibrio parahaemolyticus* infections associated with consumption of Atlantic Coast shellfish--2013. *MMWR Morb Mortal Wkly Rep* 63**,** 335-336.

Neyen, C., and Lemaitre, B. (2016). Sensing Gram-negative bacteria: a phylogenetic perspective. *Curr Opin Immunol* 38**,** 8-17.

Nishibuchi, M., and Kaper, J.B. (1995). Thermostable direct hemolysin gene of *Vibrio parahaemolyticus*: a virulence gene acquired by a marine bacterium. *Infection and Immunity* 63**,** 2093-2099.

Nishibuchi, M., Taniguchi, T., Misawa, T., Khaeomanee-Iam, V., Honda, T., and Miwatani, T. (1989). Cloning and nucleotide sequence of the gene (*trh*) encoding the hemolysin related to the thermostable direct hemolysin of *Vibrio parahaemolyticus. Infection and Immunity* 57**,** 2691-2697.

Nordstrom, J.L., Rangdale, R., Vickery, M.C., Phillips, A.M., Murray, S.L., Wagley, S., and Depaola, A. (2006). Evaluation of an alkaline phosphatase-labeled oligonucleotide probe for the detection and enumeration of the thermostable-related hemolysin (*trh*) gene of *Vibrio parahaemolyticus. J.Food Prot.* 69**,** 2770-2772.

Noriea, N.F., Iii, Johnson, C.N., Griffitt, K.J., and Grimes, D.J. (2010). Distribution of type III secretion systems in *Vibrio parahaemolyticus* from the northern Gulf of Mexico. *J.Appl.Microbiol.* 109**,** 953-962.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., Prjibelski, A.D., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., Clingenpeel, S.R., Woyke, T., Mclean, J.S., Lasken, R., Tesler, G., Alekseyev, M.A., and Pevzner, P.A. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J.Comput.Biol.* 20**,** 714-737.

Ochoa, A., Storey, J.D., Llinas, M., and Singh, M. (2015). Beyond the E-Value: Stratified Statistics for Protein Domain Prediction. *PLoS Comput Biol* 11**,** e1004509.

Ohnishi, K., Nakahira, K., Unzai, S., Mayanagi, K., Hashimoto, H., Shiraki, K., Honda, T., and Yanagihara, I. (2011). Relationship between heat-induced fibrillogenicity and hemolytic activity of thermostable direct hemolysin and a related hemolysin of *Vibrio parahaemolyticus FEMS Microbiol.Lett.* 318**,** 10-17.

Okada, N., Iida, T., Park, K.S., Goto, N., Yasunaga, T., Hiyoshi, H., Matsuda, S., Kodama, T., and Honda, T. (2009). Identification and characterization of a novel type III secretion system in trh-positive Vibrio parahaemolyticus strain TH3996 reveal genetic lineage and diversity of pathogenic machinery beyond the species level. *Infection and Immunity* 77**,** 904-913.

Okada, N., Matsuda, S., Matsuyama, J., Park, K.S., De Los Reyes, C., Kogure, K., Honda, T., and Iida, T. (2010). Presence of genes for type III secretion system 2 in *Vibrio mimicus* strains. *BMC Microbiol* 10**,** 302.

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

Pagani, I., Liolios, K., Jansson, J., Chen, I.M., Smirnova, T., Nosrat, B., Markowitz, V.M., and Kyrpides, N.C. (2012). The Genomes OnLine Database (GOLD) v.4: status of

genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 40**,** D571-579.

Parish, C.R., and Mullbacher, A. (1983). Automated colorimetric assay for T cell cytotoxicity. *J Immunol Methods* 58**,** 225-237.

Park, K.S., Arita, M., Iida, T., and Honda, T. (2005). vpaH, a gene encoding a novel histone-like nucleoid structure-like protein that was possibly horizontally acquired, regulates the biogenesis of lateral flagella in trh-positive Vibrio parahaemolyticus TH3996. *Infection and Immunity* 73**,** 5754-5761.

Park, K.S., Ono, T., Rokuda, M., Jang, M.H., Iida, T., and Honda, T. (2004a). Cytotoxicity and enterotoxicity of the thermostable direct hemolysin-deletion mutants of Vibrio parahaemolyticus *Microbiol.Immunol.* 48**,** 313-318.

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

Pearson, K. (1904). *Mathematical contributions to the theory of evolution.* Dulau and co.

Pinto, M., Robine-Leon, S., Appay, M., Kedinger, M., Triadou, N., Dussaulx, E., Lacroix, B., Simon-Assmann, P., Haffen, K., Fogh, J., and Zweibaum, A. (1983). Enterocyte-like Differentiation and Polarization of the Human Colon Carcinoma Cell Line Caco-2 in Culture. *Biol.Cell* 47**,** 323-330.

Plackett, R.L. (1983). Karl Pearson and the chi-squared test. *International Statistical Review/Revue Internationale de Statistique***,** 59-72.

Pop, M. (2009). Genome assembly reborn: recent computational challenges. *Brief Bioinform* 10**,** 354-366.

Poptsova, M.S., and Gogarten, J.P. (2010). Using comparative genome analysis to identify problems in annotated microbial genomes. *Microbiology* 156**,** 1909-1917.

Potasman, I., Paz, A., and Odeh, M. (2002). Infectious outbreaks associated with bivalve shellfish consumption: a worldwide perspective. *Clin Infect Dis* 35**,** 921-928.

Powell, J.L. (1999). Vibrio species. *Clin.Lab Med.* 19**,** 537-552, vi.

Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., Gabaldon, T., Rattei, T., Creevey, C., Kuhn, M., Jensen, L.J., Von Mering, C., and Bork, P. (2014). eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res* 42**,** D231-239.

Promega (2016). "CytoTox 96 Non-Radioactive Cytotoxicity Assay Protocol", (ed.) P. Corporation. (USA).

Rasko, D.A., Webster, D.R., Sahl, J.W., Bashir, A., Boisen, N., Scheutz, F., Paxinos, E.E., Sebra, R., Chin, C.S., Iliopoulos, D., Klammer, A., Peluso, P., Lee, L., Kislyuk, A.O., Bullard, J., Kasarskis, A., Wang, S., Eid, J., Rank, D., Redman, J.C., Steyert, S.R., Frimodt-Moller, J., Struve, C., Petersen, A.M., Krogfelt, K.A., Nataro, J.P., Schadt, E.E., and Waldor, M.K. (2011). Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med* 365**,** 709-717.

Ririe, K.M., Rasmussen, R.P., and Wittwer, C.T. (1997). Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem* 245**,** 154-160.

Rosche, T.M., Yano, Y., and Oliver, J.D. (2005). A rapid and simple PCR analysis indicates there are two subgroups of *Vibrio vulnificus* which correlate with clinical or environmental isolation. *Microbiol.Immunol.* 49**,** 381-389.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., Hoon, J., Simons, J.F., Marran, D., Myers, J.W., Davidson, J.F., Branting, A., Nobile, J.R., Puc, B.P., Light, D., Clark, T.A., Huber, M., Branciforte, J.T., Stoner, I.B., Cawley, S.E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J.A., Namsaraev, E., Mckernan, K.J., Williams, A., Roth, G.T., and Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475**,** 348-352.

Rzhetsky, A., and Nei, M. (1992). A simple method for estimating and testing minimum-evolution trees. *Mol Biol Evol* 9**,** 945-967.

Sahl, J.W., Caporaso, J.G., Rasko, D.A., and Keim, P. (2014). The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ* 2**,** e332.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4**,** 406-425.

Sakazaki, R., Tamura, K., Kato, T., Obara, Y., and Yamai, S. (1968). Studies on the enteropathogenic, facultatively halophilic bacterium, Vibrio parahaemolyticus . 3. Enteropathogenicity. *Jpn.J.Med.Sci.Biol.* 21**,** 325-331.

Sakurai, J., Matsuzaki, A., and Miwatani, T. (1973). Purification and characterization of thermostable direct hemolysin of Vibrio parahaemolyticus *Infection and Immunity* 8**,** 775-780.

Sambuy, Y., De, A.I., Ranaldi, G., Scarino, M.L., Stammati, A., and Zucco, F. (2005). The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics. *Cell Biol.Toxicol.* 21**,** 1-26.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocombe, P.M., and Smith, M. (1977a). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265**,** 687-695.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977b). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74**,** 5463-5467.

Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.A., Roy, S.L., Jones, J.L., and Griffin, P.M. (2011). Foodborne illness acquired in the United States--major pathogens. *Emerg.Infect.Dis.* 17**,** 7-15.

Schadt, E.E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. *Hum Mol Genet* 19**,** R227-240.

Schmitz, S. (2011). *Der Experimentator Zellkultur.* Spektrum Akademischer Verlag.

Schwartz, D.C., and Cantor, C.R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 37**,** 67-75.

Schwartz, D.C., Saffran, W., Welsh, J., Haas, R., Goldenberg, M., and Cantor, C.R. (1983). New techniques for purifying large DNAs and studying their properties and packaging. *Cold Spring Harb.Symp.Quant.Biol.* 47 Pt 1**,** 189-195.

Serwer, P. (1983). Agarose gels: Properties and use for electrophoresis. *Electrophoresis* 4**,** 375-382.

Shinoda, S. (2011). Sixty years from the discovery of *Vibrio parahaemolyticus* and some recollections. *Biocontrol.Sci.* 16**,** 129-137.

Sneath, P.H., and Sokal, R.R. (1973). *Numerical taxonomy. The principles and practice of numerical classification.*

Sokurenko, E.V., Hasty, D.L., and Dykhuizen, D.E. (1999). Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol* 7**,** 191-195.

Spratt, B.G., Hanage, W.P., Li, B., Aanensen, D.M., and Feil, E.J. (2004). Displaying the relatedness among isolates of bacterial species -- the eBURST approach. *FEMS Microbiol Lett* 241**,** 129-134.

Srivastava, S. (2003). *Understanding Bacteria.* Springer Science & Business Media.

Staley, C., and Harwood, V.J. (2010). The use of genetic typing methods to discriminate among strains of *Vibrio cholerae*, *V. parahaemolyticus*, and *V. vulnificus. J.AOAC Int.* 93**,** 1553-1569.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30**,** 1312-1313.

Stein, L. (2001). Genome annotation: from sequence to biology. *Nat Rev Genet* 2**,** 493-503.

Stranneheim, H., and Lundeberg, J. (2012). Stepping stones in DNA sequencing. *Biotechnol J* 7**,** 1063-1073.

Su, Y.C., and Liu, C. (2007). *Vibrio parahaemolyticus*: a concern of seafood safety. *Food Microbiol* 24**,** 549-558.

Suthienkul, O., Ishibashi, M., Iida, T., Nettip, N., Supavej, S., Eampokalap, B., Makino, M., and Honda, T. (1995). Urease production correlates with possession of the *trh* gene in *Vibrio parahaemolyticus* strains isolated in Thailand. *J Infect Dis* 172**,** 1405-1408.

Swaminathan, B., Barrett, T.J., Hunter, S.B., and Tauxe, R.V. (2001). PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg.Infect.Dis.* 7**,** 382-389.

Swofford, D.L., Olsen, G.J., Waddell, P.J., and Hillis, D.M. (1996). Phylogenetic inference.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol.Biol.Evol.* 30**,** 2725-2729.

Tan, S.C., and Yiap, B.C. (2009). DNA, RNA, and protein extraction: the past and the present. *BioMed Research International* 2009.

Taniguchi, H., Hirano, H., Kubomura, S., Higashi, K., and Mizuguchi, Y. (1986). Comparison of the nucleotide sequences of the genes for the thermostable direct hemolysin and the thermolabile hemolysin from *Vibrio parahaemolyticus. Microb.Pathog.* 1**,** 425-432.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997). A genomic perspective on protein families. *Science* 278**,** 631-637.

Team, R.C. (2015). "R: A Language and Environment for Statistical Computing". 3.1.2. ed.: R Foundation for Statistical Computing.

Terpe, K. (2013). Overview of thermostable DNA polymerases for classical PCR applications: from molecular and biochemical fundamentals to commercial systems. *Appl.Microbiol.Biotechnol.* 97**,** 10243-10254.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., Deboy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit Y Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'connor, K.J., Smith, S.,

Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R., and Fraser, C.M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* 102**,** 13950-13955.

Thompson, F.L., Iida, T., and Swings, J. (2004). Biodiversity of vibrios. *Microbiol Mol Biol Rev* 68**,** 403-431, table of contents.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 22**,** 4673-4680.

Tijssen, P. (1993). *Hybridization with Nucleic Acid Probes, Part II: Part II. Probe Labeling and Hybridization Techniques.* Elsevier Science.

Timme, R.E., Pettengill, J.B., Allard, M.W., Strain, E., Barrangou, R., Wehnes, C., Van Kessel, J.S., Karns, J.S., Musser, S.M., and Brown, E.W. (2013). Phylogenetic diversity of the enteric pathogen *Salmonella* enterica subsp. enterica inferred from genome-wide reference-free SNP characters. *Genome Biol.Evol.* 5**,** 2109-2123.

Tiruvayipati, S., Bhassu, S., Kumar, N., Baddam, R., Shaik, S., Gurindapalli, A.K., Thong, K.L., and Ahmed, N. (2013). Genome anatomy of the gastrointestinal pathogen, *Vibrio parahaemolyticus* of crustacean origin. *Gut Pathog.* 5**,** 37.

Trosky, J.E., Li, Y., Mukherjee, S., Keitany, G., Ball, H., and Orth, K. (2007). VopA inhibits ATP binding by acetylating the catalytic loop of MAPK kinases. *J.Biol.Chem.* 282**,** 34299-34305.

Trosky, J.E., Mukherjee, S., Burdette, D.L., Roberts, M., Mccarter, L., Siegel, R.M., and Orth, K. (2004). Inhibition of MAPK signaling pathways by VopA from *Vibrio parahaemolyticus J.Biol.Chem.* 279**,** 51953-51957.

Turner, J.W., Paranjpye, R.N., Landis, E.D., Biryukov, S.V., Gonzalez-Escalona, N., Nilsson, W.B., and Strom, M.S. (2013). Population structure of clinical and environmental

Vibrio parahaemolyticus from the Pacific Northwest coast of the United States. *PLoS.One.* 8**,** e55726.

Vossen, R.H., Aten, E., Roos, A., and Den Dunnen, J.T. (2009). High-resolution melting analysis (HRMA): more than just sequence variant screening. *Hum Mutat* 30**,** 860-866.

Wang, R., Zhong, Y., Gu, X., Yuan, J., Saeed, A.F., and Wang, S. (2015). The pathogenesis, detection, and prevention of *Vibrio parahaemolyticus*. *Front Microbiol* 6**,** 144.

Ward, D.R., and Hackney, C.A. (1991). *Microbiology of Marine Food Products.* Springer US.

Weber, G.G., and Klose, K.E. (2011). The complexity of ToxT-dependent transcription in *Vibrio cholerae*. *Indian J Med Res* 133**,** 201-206.

Weissman, S.J., Moseley, S.L., Dykhuizen, D.E., and Sokurenko, E.V. (2003). Enterobacterial adhesins and the case for studying SNPs in bacteria. *Trends Microbiol* 11**,** 115-117.

West, P.A. (1989). The human pathogenic vibrios--a public health update with environmental perspectives. *Epidemiol.Infect.* 103**,** 1-34.

Wheelis, M. (2011). *Principles of Modern Microbiology.* Jones & Bartlett Publishers.

Who, W.H.O. (2006). "Five keys to Safer Food Manual". (Geneva, Switzerland).

Who, W.H.O. (2015). "WHO estimates of the global burden of foodborne diseases: foodborne disease burden epidemiology reference group 2007-2015". (Geneva, Switzerland).

Wong, H.C., Liu, S.H., Chiou, C.S., Nishibuchi, M., Lee, B.K., Suthienkul, O., Nair, G.B., Kaysner, C.A., and Taniguchi, H. (2007). A pulsed-field gel electrophoresis typing scheme for Vibrio parahaemolyticus isolates from fifteen countries. *Int.J.Food Microbiol.* 114**,** 280-287.

Woolery, A.R., Luong, P., Broberg, C.A., and Orth, K. (2010). AMPylation: Something Old is New Again. *Front Microbiol* 1**,** 113.

World Health Organisation, F.a.a.O.O.T.U.N. (2009). "Risk Characterization of Microbiological Hazards in Food Guidelines", in: *Microbiological Risk Assessment Series.*).

Xu, F., Ilyas, S., Hall, J.A., Jones, S.H., Cooper, V.S., and Whistler, C.A. (2015). Genetic characterization of clinical and environmental Vibrio parahaemolyticus from the Northeast USA reveals emerging resident and non-indigenous pathogen lineages. *Front Microbiol* 6**,** 272.

Xu, M., Iida, T., Yamamoto, K., Takarada, Y., Miwatani, T., and Honda, T. (1994). Demonstration and characterization of simultaneous production of a thermostable direct hemolysin (TDH/I) and a TDH-related hemolysin (TRHx) by a clinically isolated Vibrio parahaemolyticus strain, TH3766. *Infection and Immunity* 62**,** 166-171.

Yan, Y., Cui, Y., Han, H., Xiao, X., Wong, H.C., Tan, Y., Guo, Z., Liu, X., Yang, R., and Zhou, D. (2011). Extended MLST-based population genetics and phylogeny of Vibrio parahaemolyticus with high levels of recombination. *Int.J.Food Microbiol.* 145**,** 106-112.

Yanagihara, I., Nakahira, K., Yamane, T., Kaieda, S., Mayanagi, K., Hamada, D., Fukui, T., Ohnishi, K., Kajiyama, S., Shimizu, T., Sato, M., Ikegami, T., Ikeguchi, M., Honda, T., and Hashimoto, H. (2010). Structure and functional characterization of *Vibrio parahaemolyticus* thermostable direct hemolysin. *J.Biol.Chem.* 285**,** 16267-16274.

Yarbrough, M.L., Li, Y., Kinch, L.N., Grishin, N.V., Ball, H.L., and Orth, K. (2009). AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling. *Science* 323**,** 269-272.

Yeung, P.S., Wiedmann, M., and Boor, K.J. (2007). Evaluation of a tissue culture-based approach for differentiating between virulent and avirulent Vibrio parahaemolyticus strains based on cytotoxicity. *J Food Prot* 70**,** 348-354.

Zhang, L., Krachler, A.M., Broberg, C.A., Li, Y., Mirzaei, H., Gilpin, C.J., and Orth, K. (2012). Type III effector VopC mediates invasion for Vibrio species. *Cell Rep.* 1**,** 453-460.

Zhang, L., and Orth, K. (2013). Virulence determinants for *Vibrio parahaemolyticus* infection. *Curr.Opin.Microbiol.*

Zhou, X., Konkel, M.E., and Call, D.R. (2010). Vp1659 is a Vibrio parahaemolyticus type III secretion system 1 protein that contributes to translocation of effector proteins needed to induce cytolysis, autophagy, and disruption of actin structure in HeLa cells. *J.Bacteriol.* 192**,** 3491-3502.

11. Appendix

11.1. Attachment to chapter 5

11.1.1. Attachment to chapter 5.1

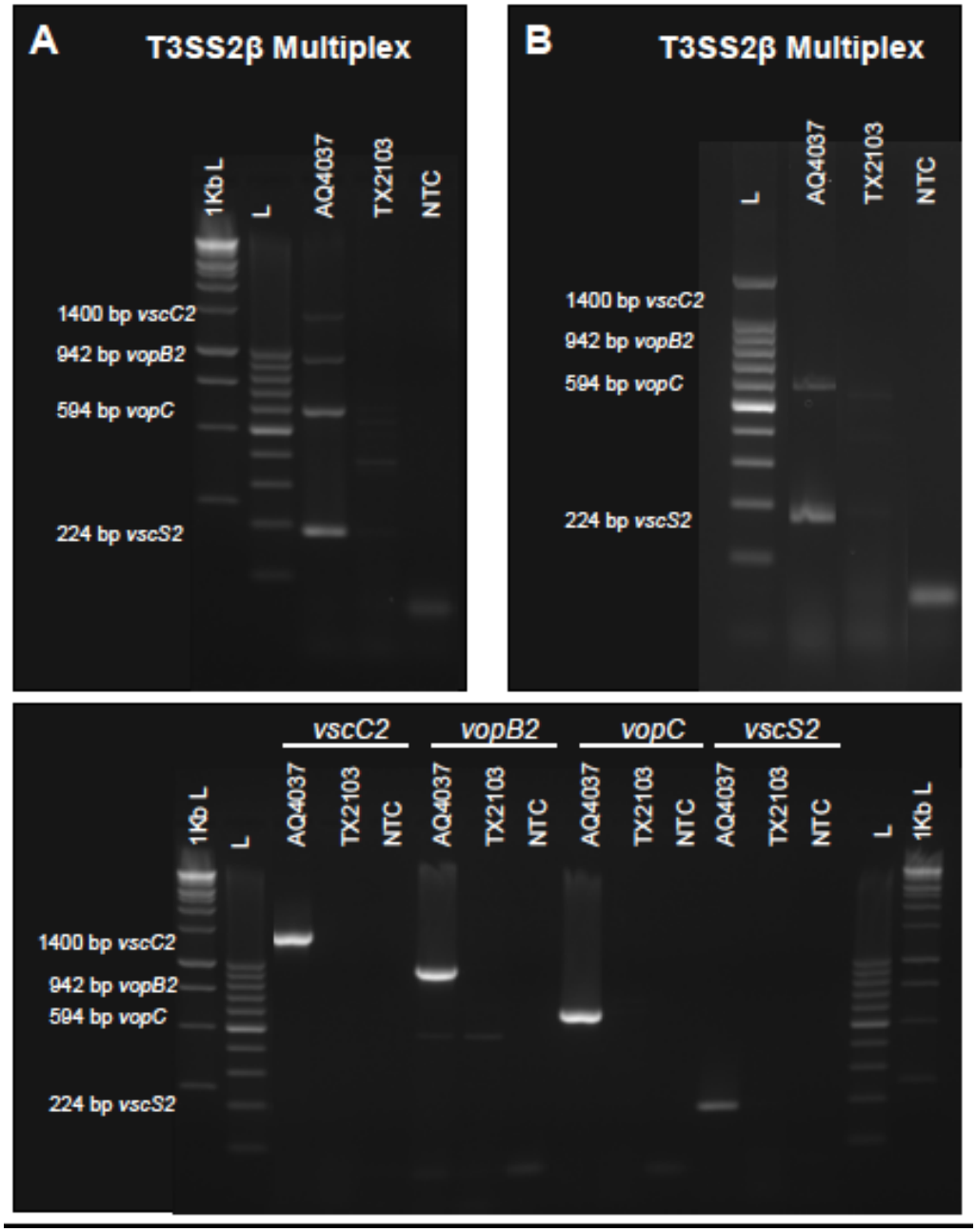


Figure 24: Agarose gel of the T3SS2β Multiplex PCR (Jones et al., 2012)

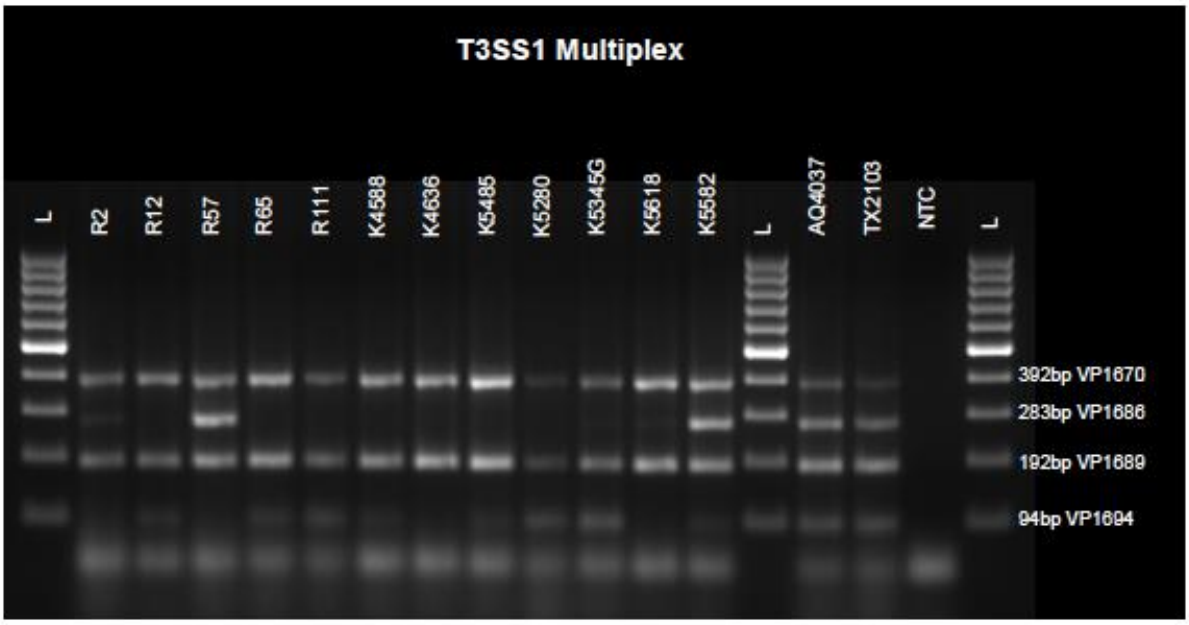


Figure 25: Representative strains for the T3SS1 Multiplex PCR (Jones et al., 2012)

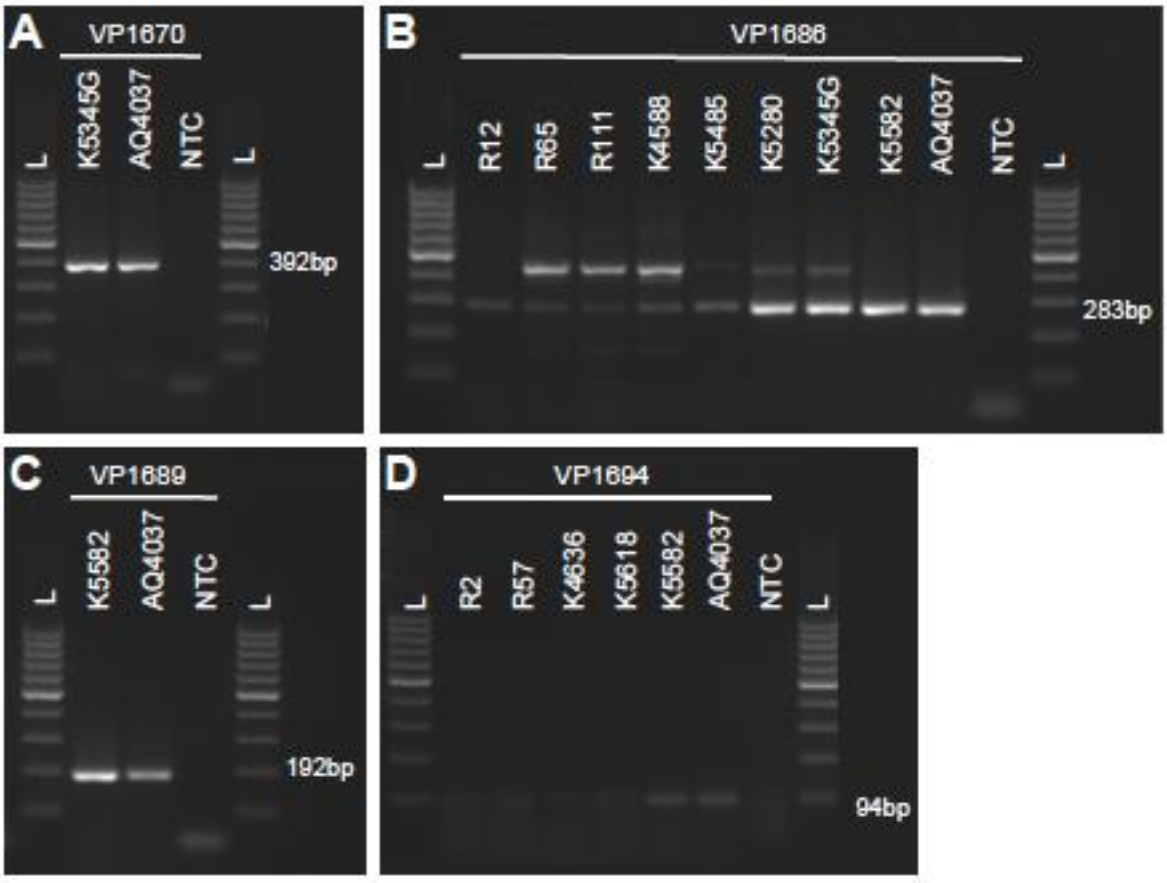


Figure 26: Representative strains for the simplex T3SS1 PCR (Jones et al., 2012)

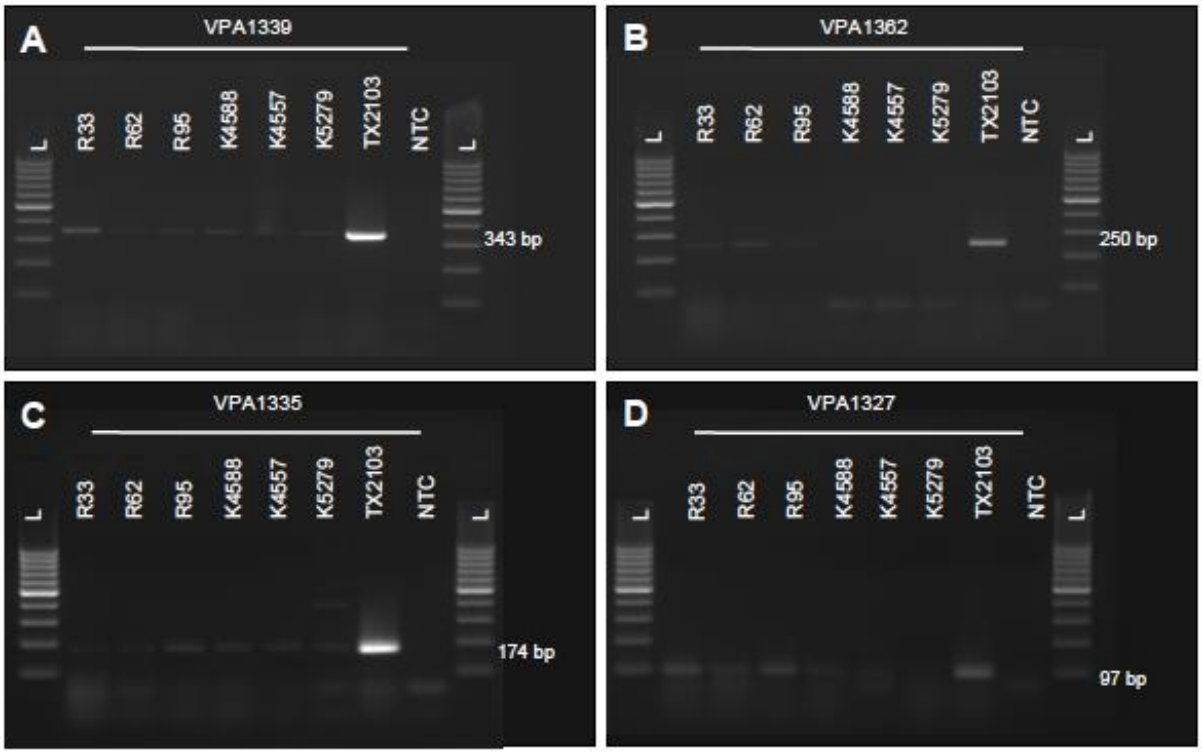


Figure 27: Representative strains for the T3SS2α Singleplex PCR (Jones et al., 2012)

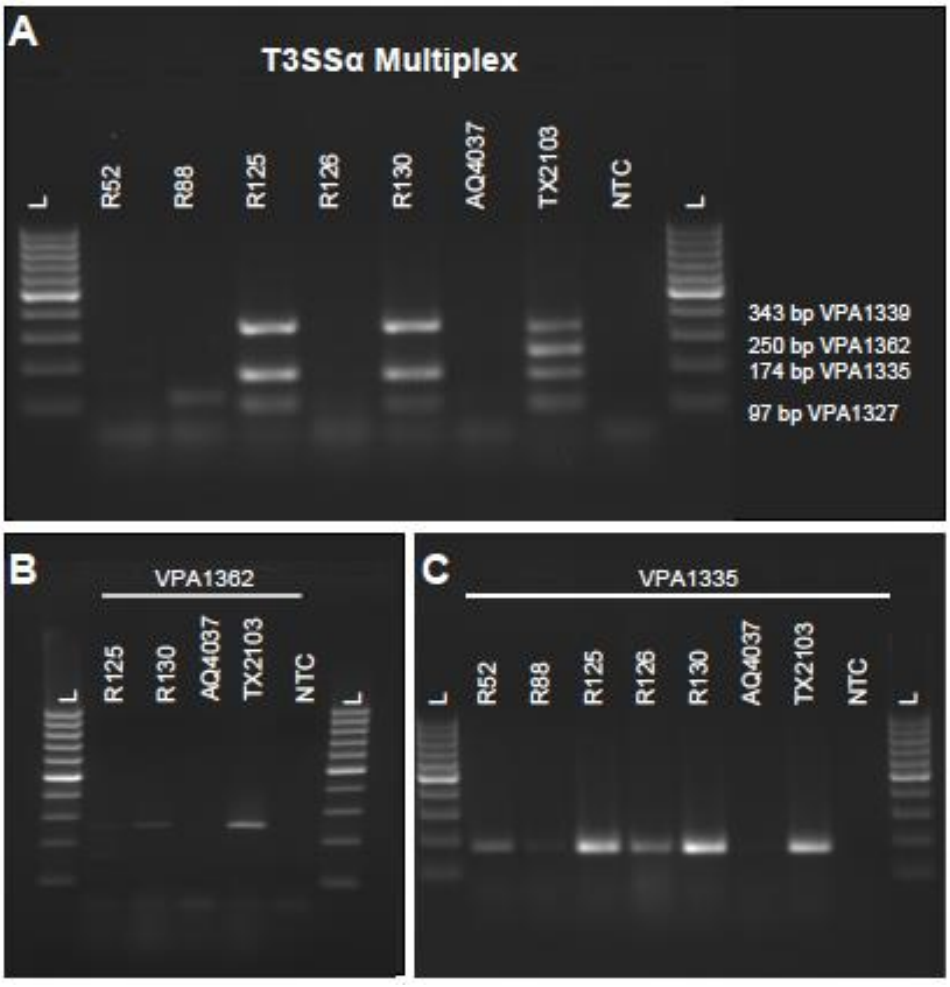


Figure 28: Representative strains for the T3SS2α Multiplex PCR (Jones et al., 2012)

11.1.2. Attachment to chapter 5.2

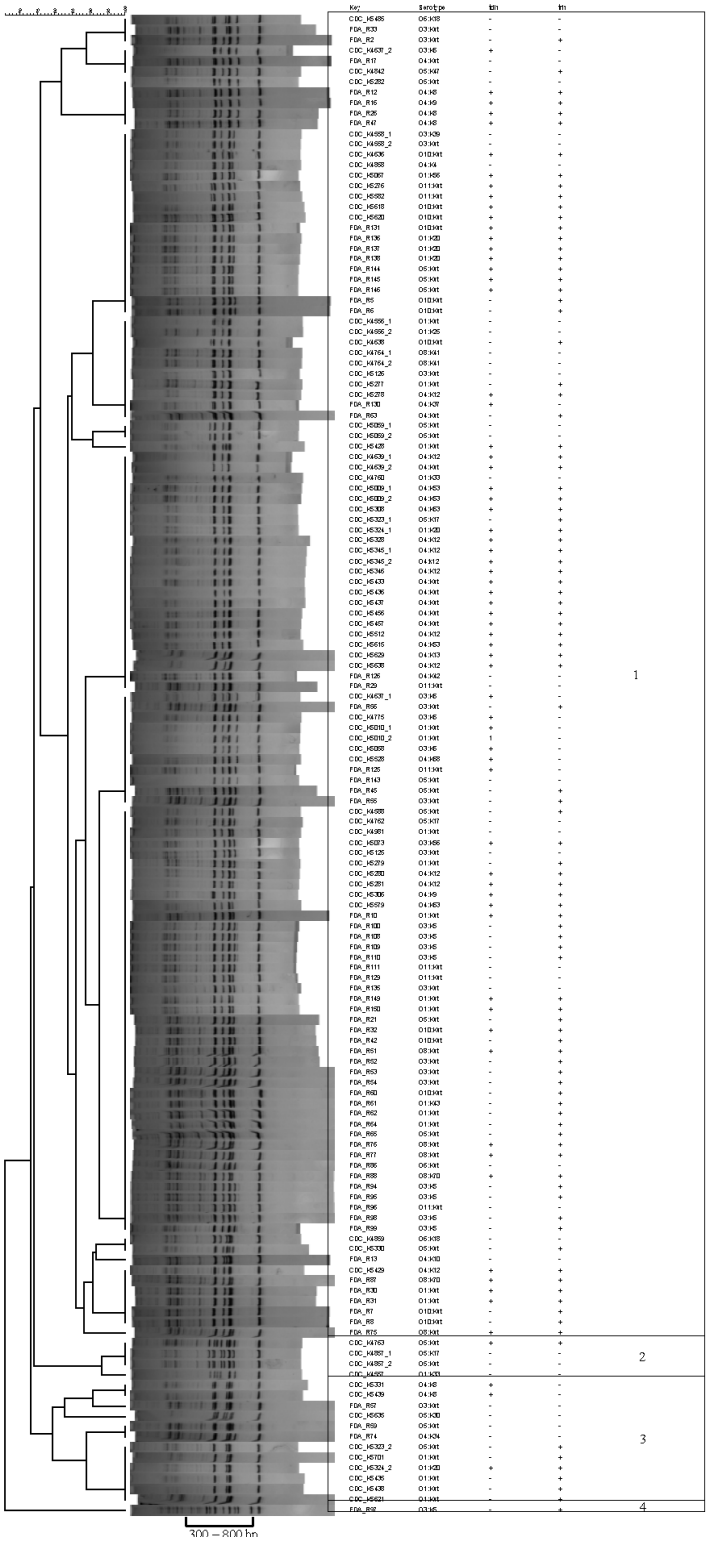


Figure 29: ISR dendrogram of the 144 *V. parahaemolyticus* isolates within an analysis range of 300 – 800 bp. The dendrogram was built with BioNumerics software 6.6 using Dice correlation with 0.5% optimization and 1% tolerance and the unweighted pair group method using arithmetic averages (UPGMA). Clusters were defined at 76% similarity (Ludeke et al., 2015).

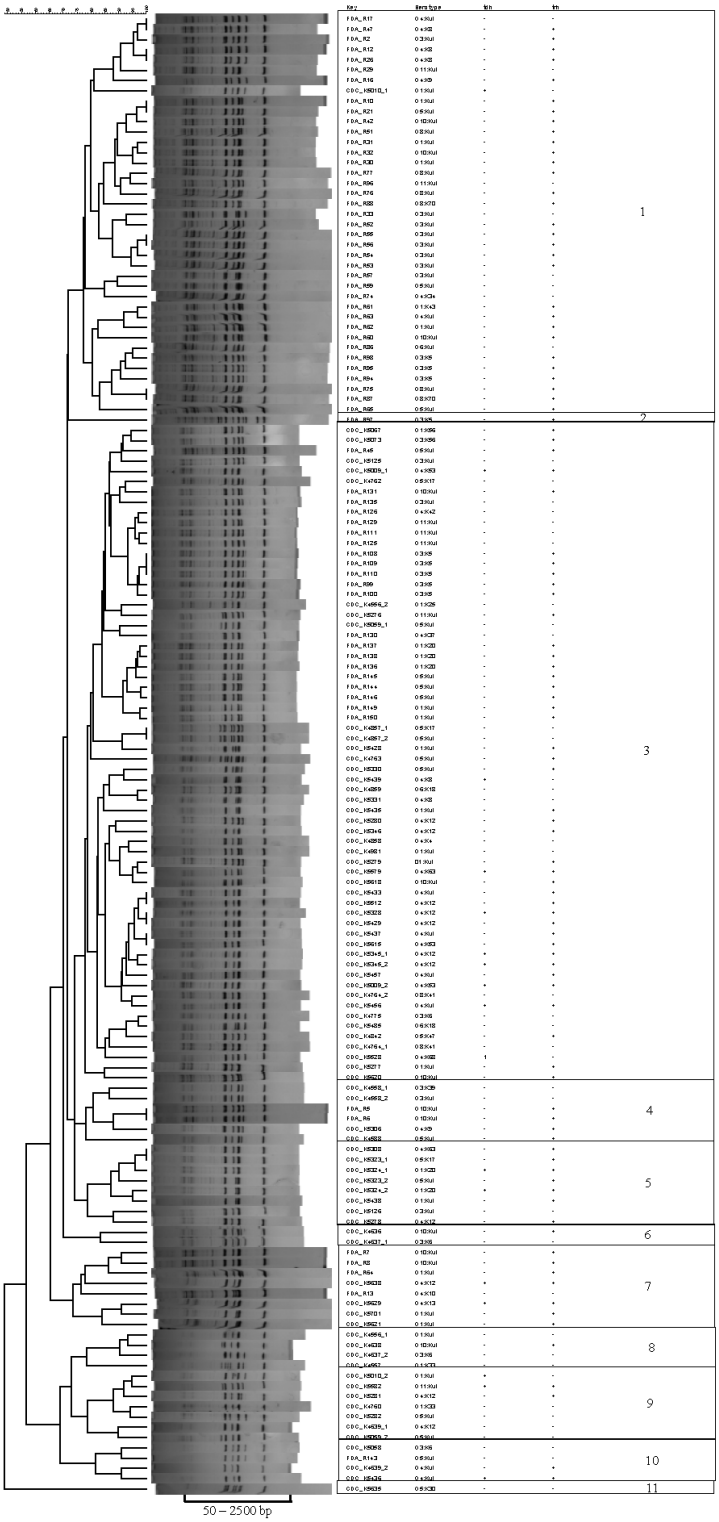


Figure 30: ISR dendrogram of the 144 *V. parahaemolyticus* isolates within an analysis range of 50 – 2500 bp. The dendrogram was built with BioNumerics software 6.6 using Dice correlation with 0.5% optimization and 1% tolerance and the unweighted pair group method using arithmetic averages (UPGMA). Clusters were defined at 76% similarity (Ludeke et al., 2015).

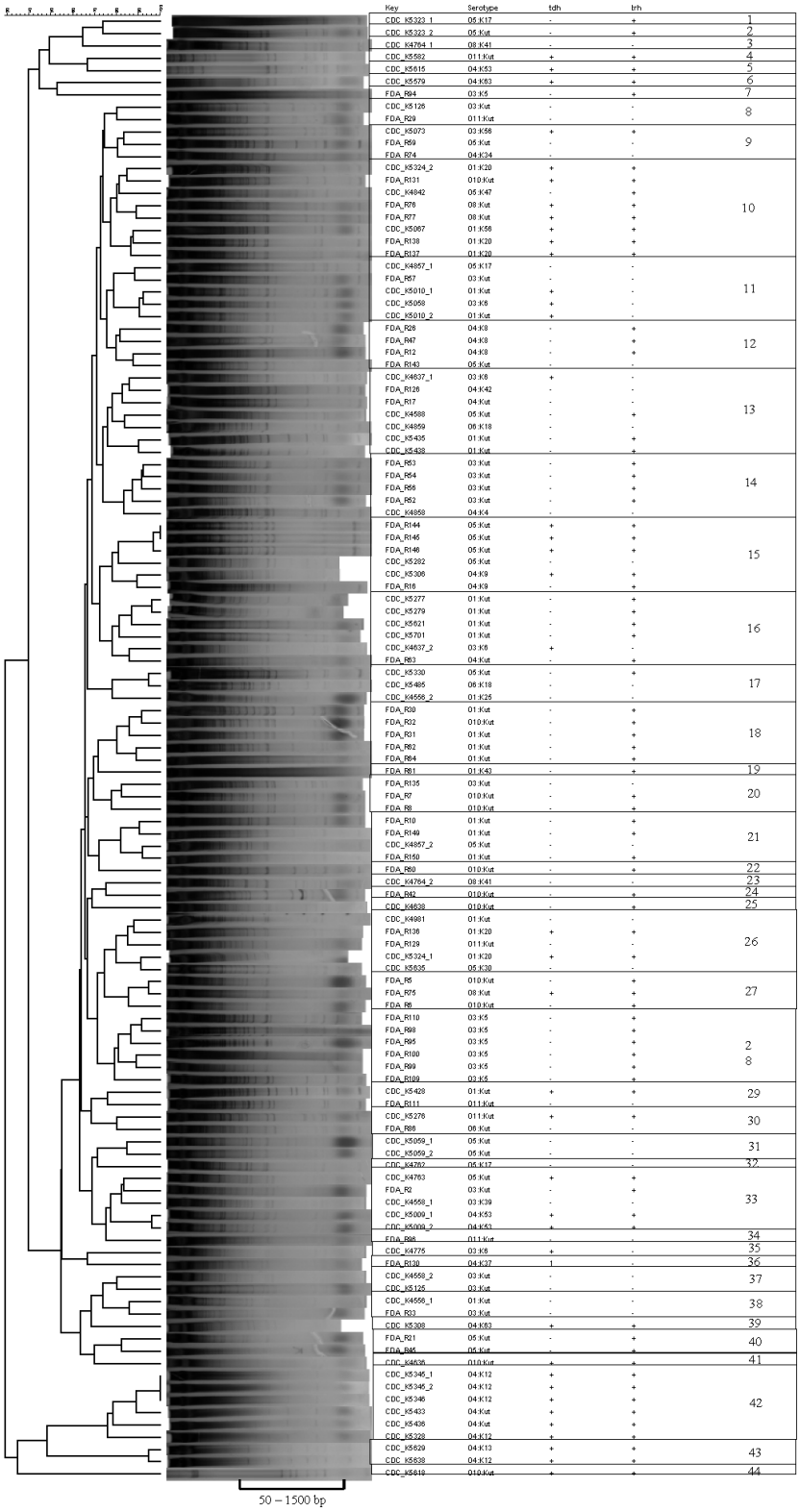


Figure 31: DGREA dendrogram of the 144 *V. parahaemolyticus* isolates within an analysis range of 50 – 1500 bp. The dendrogram was built with BioNumerics software 6.6 using Dice correlation with 0.5% optimization and 1% tolerance and the unweighted pair group method using arithmetic averages (UPGMA). Clusters were defined at 76% similarity (Ludeke et al., 2015).

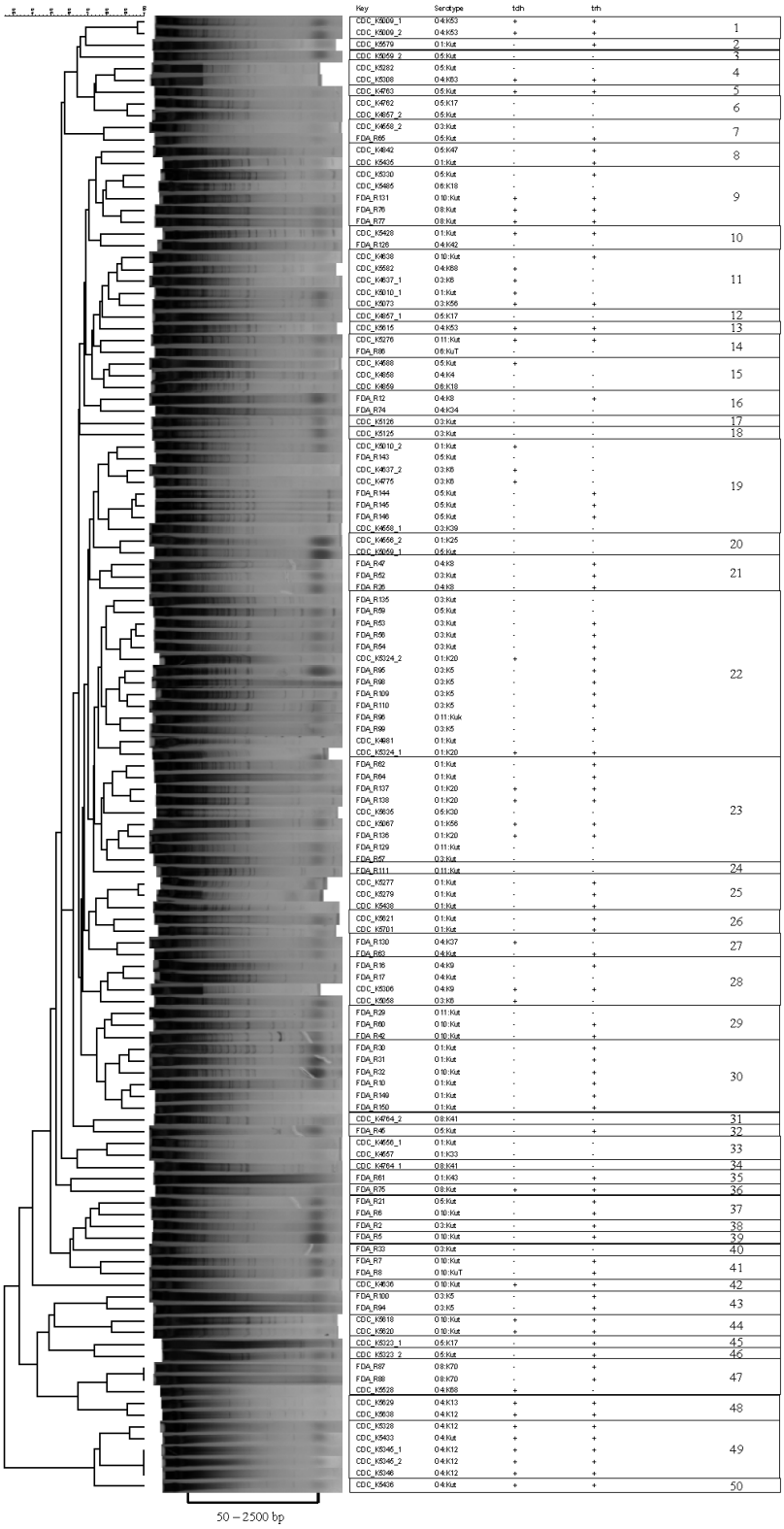


Figure 32: DGREA dendrogram of the 144 *V. parahaemolyticus* isolates within an analysis range of 50 – 2500 bp. The dendrogram was built with BioNumerics software 6.6 using Dice correlation with 0.5% optimization and 1% tolerance and the unweighted pair group method using arithmetic averages (UPGMA). Clusters were defined at 76% similarity (Ludeke et al., 2015).

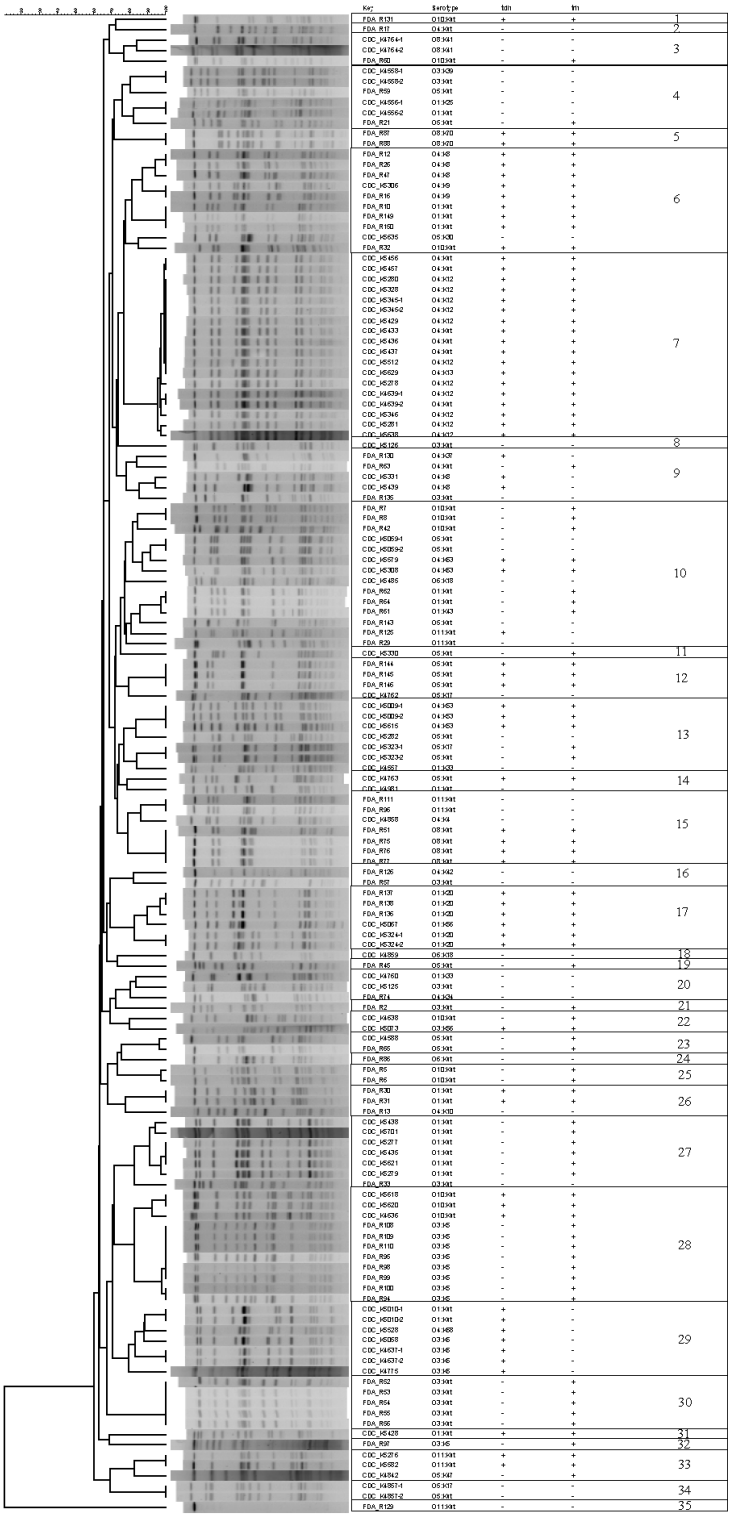


Figure 33: PFGE dendrogram of the 144 *V. parahaemolyticus* isolates digested with the enzyme *Sfi*I. The dendrogram was built with BioNumerics software 6.6 using Dice correlation with 1% optimization and 1.5% tolerance and the unweighted pair group method using arithmetic averages (UPGMA). Clusters were defined at 76% similarity (Ludeke et al., 2015).

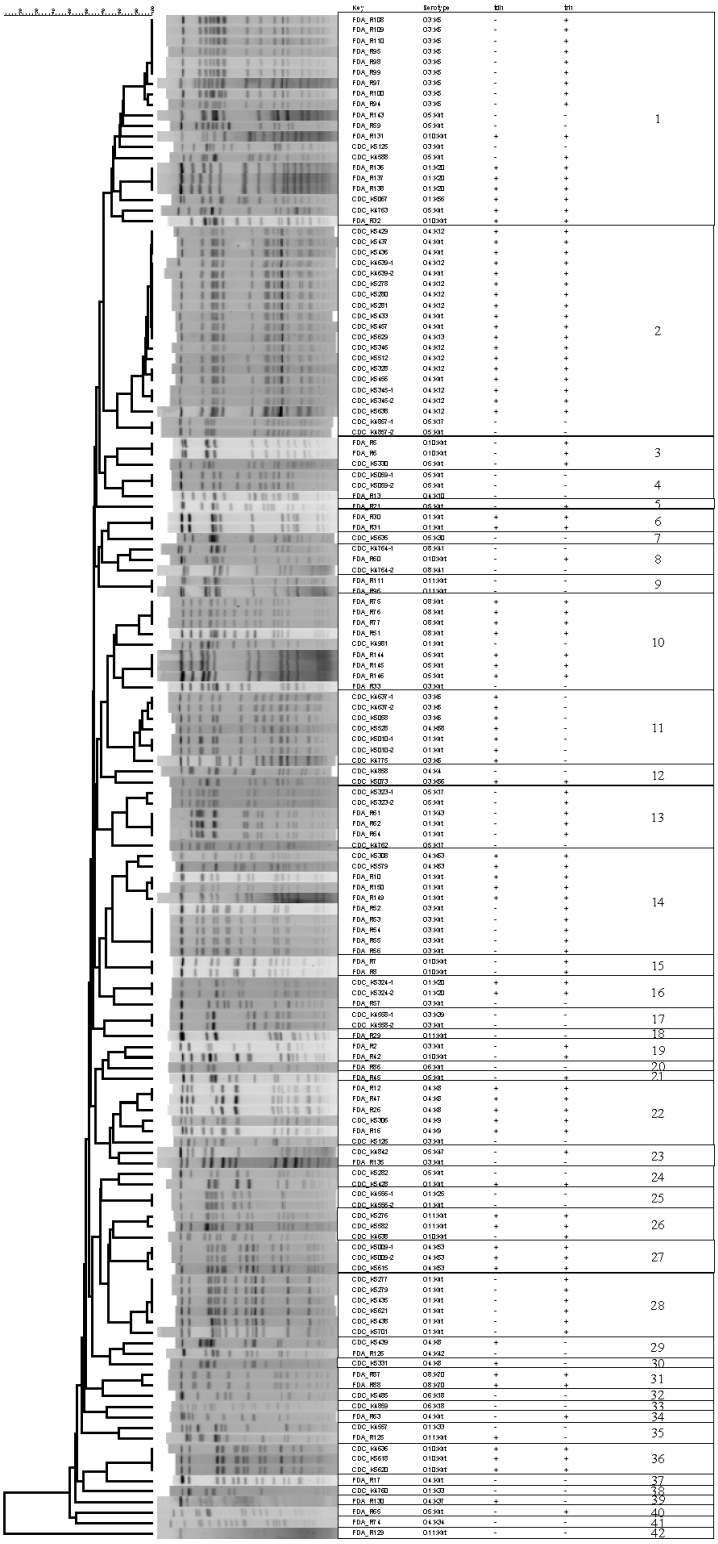


Figure 34: PFGE dendrogram of the 144 *V. parahaemolyticus* isolates digested with the enzyme *Not*I. The dendrogram was built with BioNumerics software 6.6 using Dice correlation with 1% optimization and 1.5% tolerance and the unweighted pair group method using arithmetic averages (UPGMA). Clusters were defined at 76% similarity (Ludeke et al., 2015).

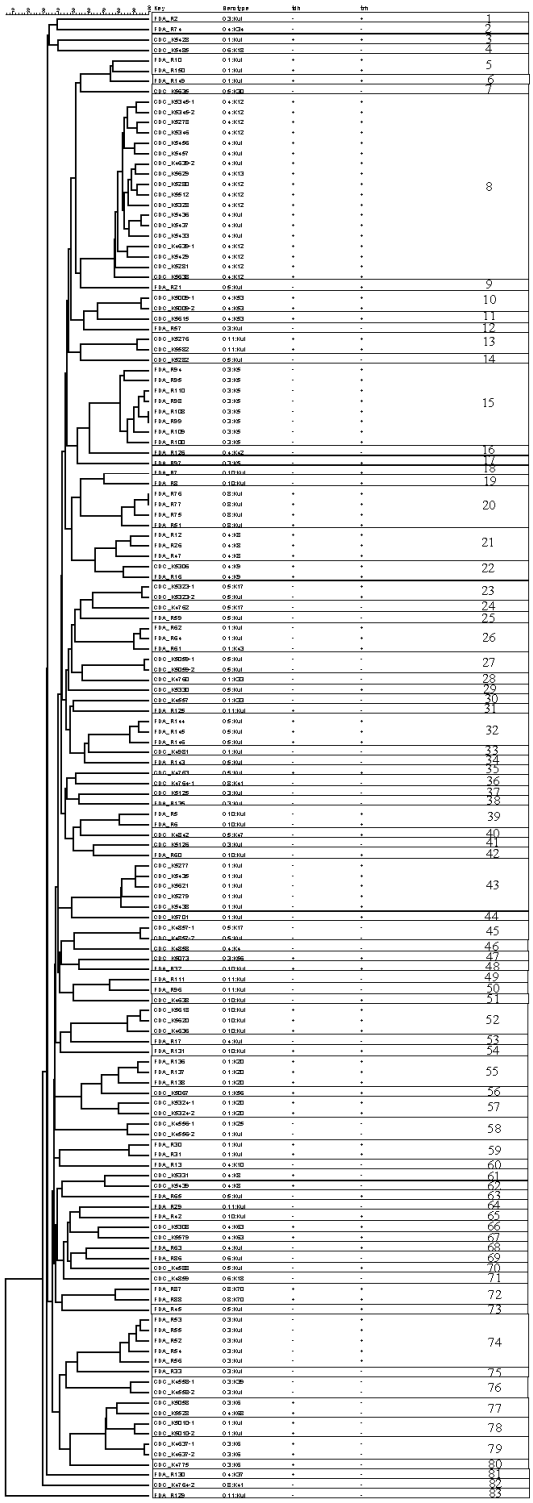


Figure S7: Combined PFGE dendogram using both enzymes of the 144 *V. parahaemolyticus* isolates. Dendograms containing *SfiI* and *NotI* combined comparisons were constructed using the average from experiments (SfiI and NotI) and the unweighted pair group method using arithmetic averages (UPGMA) . Clusters were defined at 76% similarity.

Figure 35: Combined PFGE dendrogram of the 144 *V. parahaemolyticus* isolates. Dendrograms containing *Sfi*I and *Not*I combined comparisons were constructed using the average from experiments (*Sfi*I and *Not*I) and the unweighted pair group method using arithmetic averages (UPGMA). Clusters were defined at 76% similarity (Ludeke et al., 2015).

11.1.3. Attachment to chapter 5.3

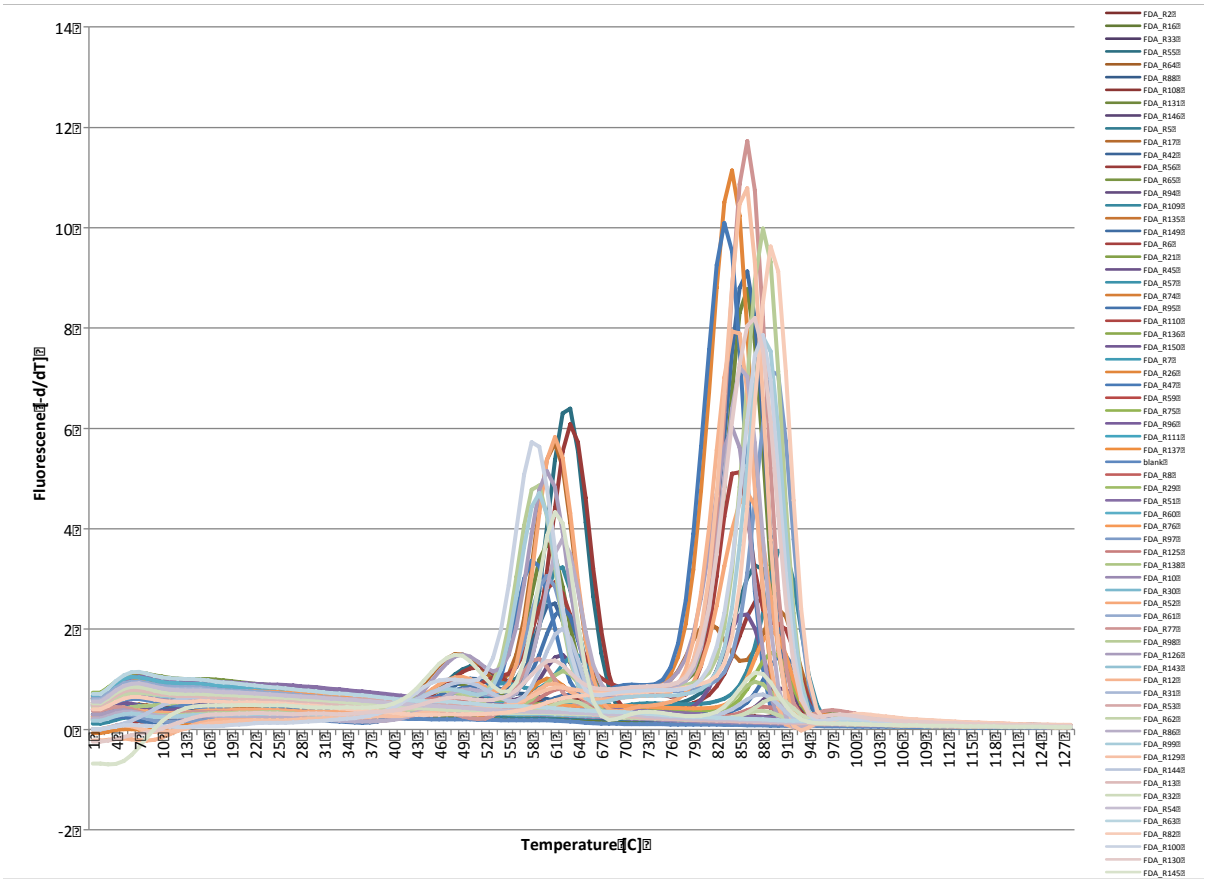


Figure 36: HRM-MLVA melting curves of oyster isolates (Multi A)

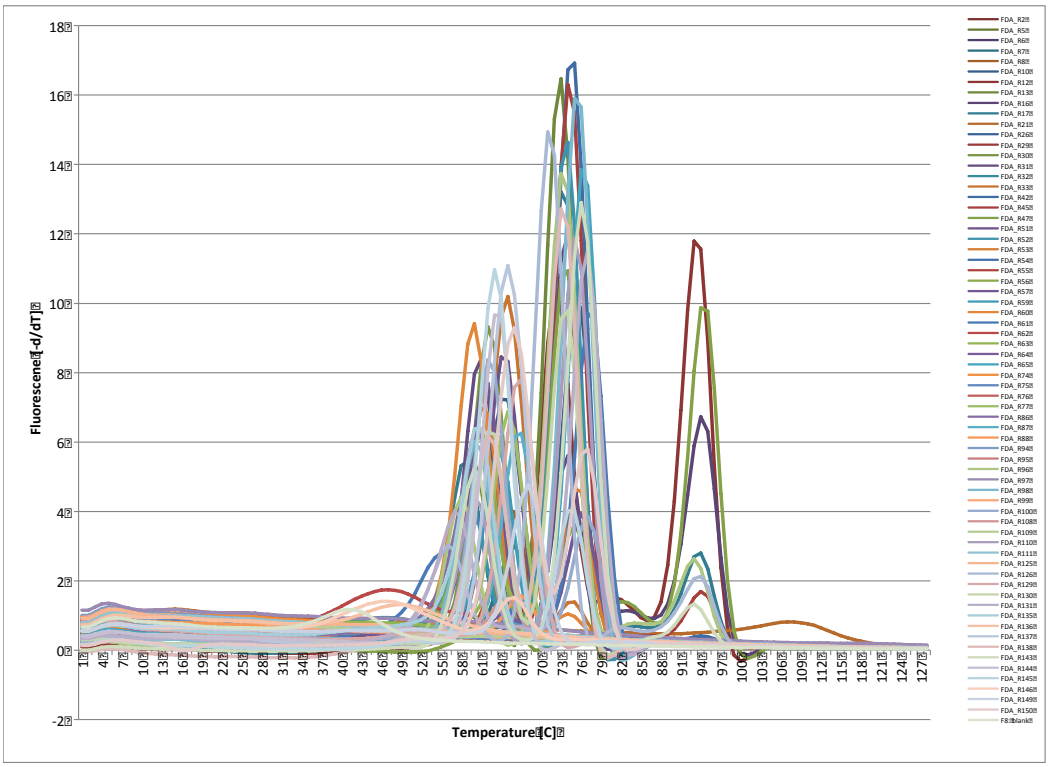


Figure 37: HRM-MLVA melting curves of oyster isolates (Multi B)

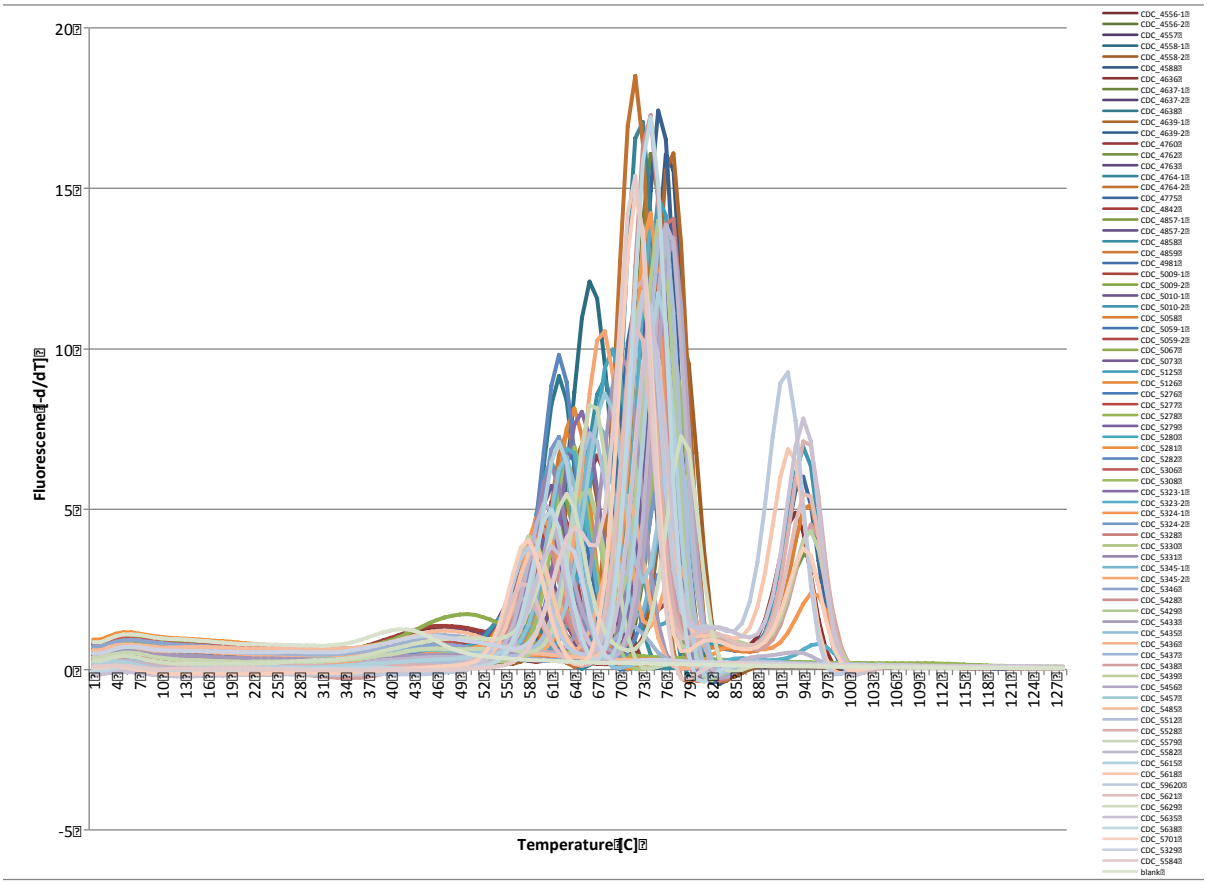


Figure 38: HRM-MLVA melting curves of clinical isolates (Multi B)

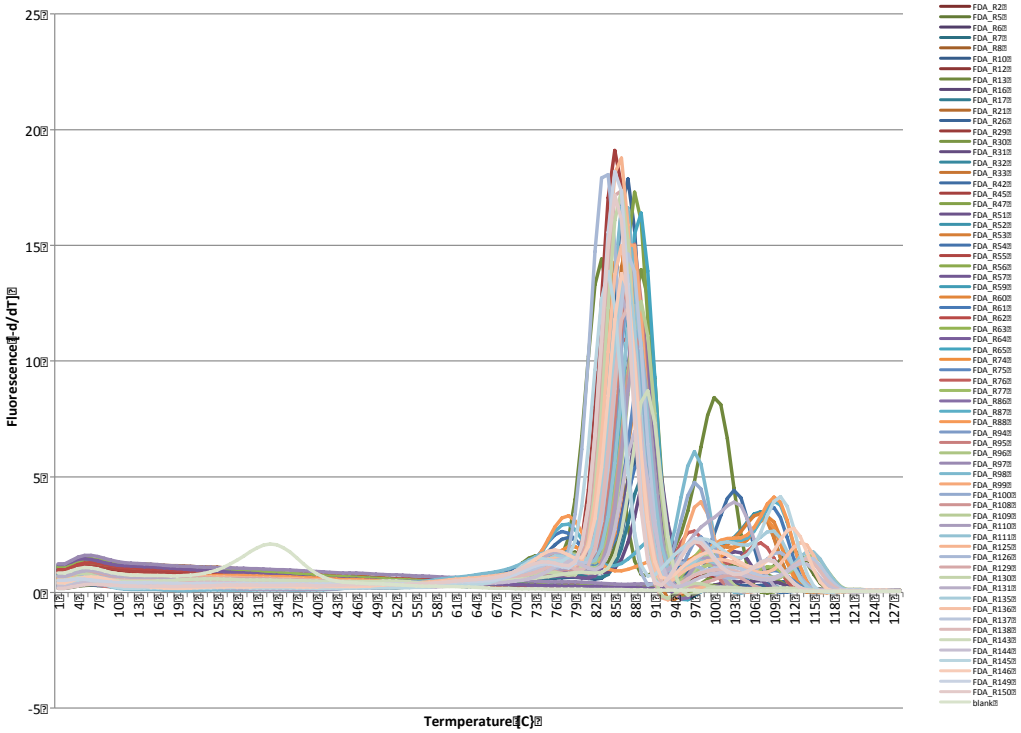


Figure 39: HRM-MLVA melting curves of oyster isolates (Multi C)

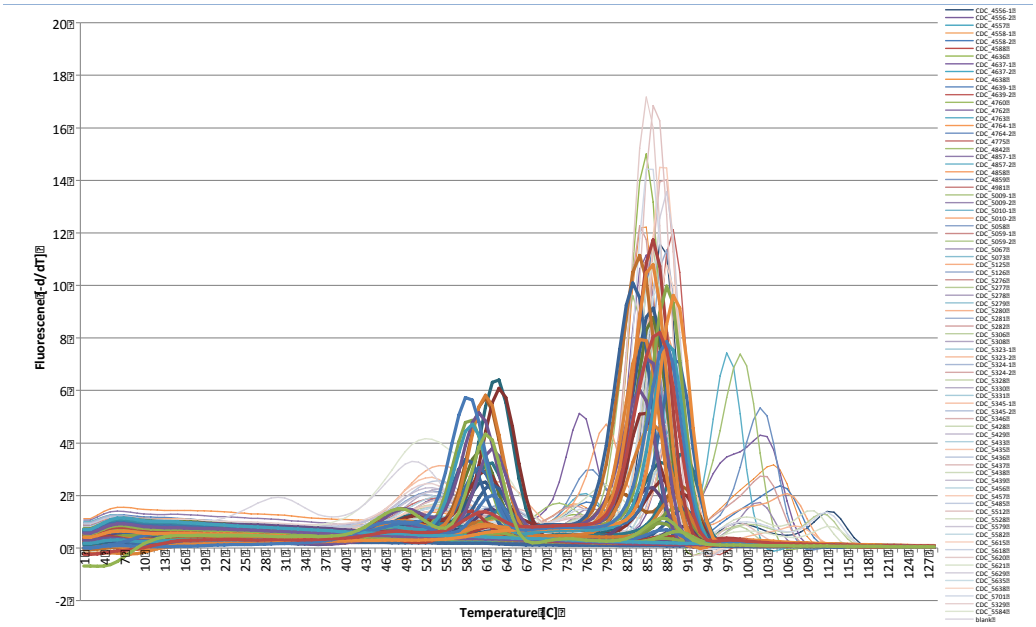


Figure 40: HRM-MLVA melting curves of clinical isolates (Multi C)

11.1.4. Attachment to chapter 5.4

Table 16: Accession numbers of the closed *V. parahaemolyticus* genomes

Closed sequence of *V. parahaemolyticus*	Accession number
Chromosome I FDA_R31	CP006004
Chromosome II FDA_R31	CP006005
Chromosome I CDC_K4557	CP006008
Chromosome II CDC_K4557	CP006007

11.1.5. Attachment to chapter 5.5

Table 17: Contingency table of clinical/shellfish and cytotoxic potential in the HeLa cell assay

Cytotoxic potential	clinical	shellfish
Cytotoxic	48	34
Semi-cytotoxic	19	18
Non-cytotoxic	22	23

p=0.534

Table 18: Contingency table of tdh and cytotoxic potential in the HeLa cell assay

Cytotoxic potential	*tdh*$^+$	*tdh*$^-$
Cytotoxic	45	36
Semi-cytotoxic	17	21
Non-cytotoxic	20	25

p=0.372

Table 19: Contingency table of trh and cytotoxic potential in the HeLa cell assay

Cytotoxic potential	*trh*$^+$	*trh*$^-$
Cytotoxic	58	23
Semi-cytotoxic	25	13
Non-cytotoxic	31	14

p=0.809

Table 20: Contingency table of TDH and cytotoxic potential in the HeLa cell assay

Cytotoxic potential	TDH$^+$	TDH$^-$
Cytotoxic	28	37
Semi-cytotoxic	16	14
Non-cytotoxic	13	17

p=0.621

Table 21: Contingency table of TDH and cytotoxic potential in the HeLa cell assay

Cytotoxic potential	TDH^+	TDH^-
Cytotoxic	28	37
Semi-cytotoxic	16	14
Non-cytotoxic	13	17

p=0.621

Table 22: Contingency table of T3SS2α+ and cytotoxic potential in the HeLa cell assay

Cytotoxic potential	$T3SS2\alpha^+$	$T3SS2\alpha^-$
Cytotoxic	6	76
Semi-cytotoxic	4	33
Non-cytotoxic	1	44

p=0.288

Table 23: Contingency table of T3SS2β+ and cytotoxic potential in the HeLa cell assay

Cytotoxic potential	$T3SS2\beta^+$	$T3SS2\beta^-$
Cytotoxic	58	24
Semi-cytotoxic	24	13
Non-cytotoxic	30	15

p=0.786

Table 24: Contingency table of clinical/shellfish and cytotoxic potential in the Caco cell assay

Cytotoxic potential	clinical	shellfish
Cytotoxic	46	19
Semi-cytotoxic	18	24
Non-cytotoxic	25	32

p=0.003

Table 25: Contingency table of tdh and cytotoxic potential in the Caco cell assay

Cytotoxic potential	tdh^+	tdh^-
Cytotoxic	36	29
Semi-cytotoxic	14	28
Non-cytotoxic	32	25

p=0.043

Table 26: Contingency table of trh and cytotoxic potential in the Caco cell assay

Cytotoxic potential	trh^+	trh^-
Cytotoxic	46	19
Semi-cytotoxic	32	10
Non-cytotoxic	36	21

p=0.365

Table 27: Contingency table of TDH and cytotoxic potential in the Caco cell assay

Cytotoxic potential	TDH$^+$	TDH$^-$
Cytotoxic	21	27
Semi-cytotoxic	10	23
Non-cytotoxic	26	18

p=0.041

Table 28: Contingency table of T3SS2α+ and cytotoxic potential in the Caco cell assay

Cytotoxic potential	T3SS2α$^+$	T3SS2α$^-$
Cytotoxic	2	64
Semi-cytotoxic	1	40
Non-cytotoxic	8	49

p=0.023

Table 29: Contingency table of T3SS2β+ and cytotoxic potential in the Caco cell assay

Cytotoxic potential	T3SS2β$^+$	T3SS2β$^-$
Cytotoxic	46	20
Semi-cytotoxic	31	10
Non-cytotoxic	37	20

p=0.525

Table 30: Contingency table of *tdh* and cytotoxic potential of shellfish isolates in the HeLa cell assay

Cytotoxic potential	*tdh*$^+$	*tdh*$^-$
Cytotoxic	12	22
Semi-cytotoxic	7	11
Non-cytotoxic	9	14

p=0.946

Table 31: Contingency table of *trh* and cytotoxic potential of shellfish isolates in the HeLa cell assay

Cytotoxic potential	*trh*$^+$	*trh*$^-$
Cytotoxic	26	8
Semi-cytotoxic	12	6
Non-cytotoxic	17	6

p=0.747

Table 32: Contingency table of TDH and cytotoxic potential of shellfish isolates in the HeLa cell assay

Cytotoxic potential	TDH$^+$	TDH$^-$
Cytotoxic	12	14
Semi-cytotoxic	6	6
Non-cytotoxic	6	9

p=0.867

Table 33: Contingency table of T3SS2α+ and cytotoxic potential of shellfish isolates in the HeLa cell assay

Cytotoxic potential	T3SS2α$^+$	T3SS2α$^-$
Cytotoxic	1	33
Semi-cytotoxic	1	17
Non-cytotoxic	0	23

p=0.544

Table 34: Contingency table of T3SS2β+ and cytotoxic potential of shellfish isolates in the HeLa cell assay

Cytotoxic potential	T3SS2β$^+$	T3SS2β$^-$
Cytotoxic	26	8
Semi-cytotoxic	12	6
Non-cytotoxic	16	7

p=0.719

Table 35: Contingency table of *tdh* and cytotoxic potential of shellfish isolates in the Caco cell assay

Cytotoxic potential	*tdh*$^+$	*tdh*$^-$
Cytotoxic	7	12
Semi-cytotoxic	5	19
Non-cytotoxic	16	16

p=0.083

Table 36: Contingency table of *trh* and cytotoxic potential of shellfish isolates in the Caco cell assay

Cytotoxic potential	*trh*$^+$	*trh*$^-$
Cytotoxic	13	6
Semi-cytotoxic	21	3
Non-cytotoxic	21	11

p=0.160

Table 37: Contingency table of TDH and cytotoxic potential of shellfish isolates in the Caco cell assay

Cytotoxic potential	TDH$^+$	TDH$^-$
Cytotoxic	7	6
Semi-cytotoxic	4	16
Non-cytotoxic	13	8

p=0.019

Table 38: Contingency table of T3SS2α+ and cytotoxic potential of shellfish isolates in the Caco cell assay

Cytotoxic potential	T3SS2α$^+$	T3SS2α$^-$
Cytotoxic	0	20
Semi-cytotoxic	0	23
Non-cytotoxic	2	30

p=0.251

Table 39: Contingency table of T3SS2β+ and cytotoxic potential of shellfish isolates in the Caco cell assay

Cytotoxic potential	T3SS2β$^+$	T3SS2β$^-$
Cytotoxic	14	6
Semi-cytotoxic	20	3
Non-cytotoxic	22	10

p=0.265

Table 40: Contingency table of *tdh* and cytotoxic potential of clinical isolates in the HeLa cell assay

Cytotoxic potential	*tdh*$^+$	*tdh*$^-$
Cytotoxic	34	14
Semi-cytotoxic	9	10
Non-cytotoxic	11	11

p=0.103

Table 41: Contingency table of *trh* and cytotoxic potential of clinical isolates in the HeLa cell assay

Cytotoxic potential	*trh*$^+$	*trh*$^-$
Cytotoxic	33	15
Semi-cytotoxic	12	7
Non-cytotoxic	14	8

p=0.868

Table 42: Contingency table of TDH and cytotoxic potential of clinical isolates in the HeLa cell assay

Cytotoxic potential	TDH$^+$	TDH$^-$
Cytotoxic	16	24
Semi-cytotoxic	10	6
Non-cytotoxic	7	10

p=0.289

Table 43: Contingency table of T3SS2α+ and cytotoxic potential of clinical isolates in the HeLa cell assay

Cytotoxic potential	T3SS2α$^+$	T3SS2α$^-$
Cytotoxic	5	43
Semi-cytotoxic	3	16
Non-cytotoxic	1	21

p=0.490

Table 44: Contingency table of T3SS2β+ and cytotoxic potential of clinical isolates in the HeLa cell assay

Cytotoxic potential	T3SS2β$^+$	T3SS2β$^-$
Cytotoxic	32	16
Semi-cytotoxic	12	7
Non-cytotoxic	14	8

p=0.949

Table 45: Contingency table of *tdh* and cytotoxic potential of clinical isolates in the Caco cell assay

Cytotoxic potential	*tdh*$^+$	*tdh*$^-$
Cytotoxic	29	13
Semi-cytotoxic	9	7
Non-cytotoxic	16	10

p=0.622

Table 46: Contingency table of *trh* and cytotoxic potential of clinical isolates in the Caco cell assay

Cytotoxic potential	*trh*$^+$	*trh*$^-$
Cytotoxic	33	13
Semi-cytotoxic	11	7
Non-cytotoxic	15	10

p=0.530

Table 47: Contingency table of TDH and cytotoxic potential of clinical isolates in the Caco cell assay

Cytotoxic potential	TDH$^+$	TDH$^-$
Cytotoxic	14	22
Semi-cytotoxic	6	6
Non-cytotoxic	13	10

p=0.401

Table 48: Contingency table of T3SS2α+ and cytotoxic potential of clinical isolates in the Caco cell assay

Cytotoxic potential	T3SS2α$^+$	T3SS2α$^-$
Cytotoxic	2	44
Semi-cytotoxic	1	17
Non-cytotoxic	6	19

0.025

Table 49: Contingency table of T3SS2β+ and cytotoxic potential of clinical isolates in the Caco cell assay

Cytotoxic potential	T3SS2β$^+$	T3SS2β$^-$
Cytotoxic	32	14
Semi-cytotoxic	11	7
Non-cytotoxic	15	10

p=0.665

11.1.6. Attachment to chapter 5.6

Table 50: Coverage, genome length, contig count and gene count of 132 *V. parahaemolyticus* shotgun sequences; FDA indicates oyster origin, CDC indicates clinical origin.

Isolate_ID	Coverage depth	Total length	Contig count	Gene count
FDA_R5	43	5185921	95	4.792
FDA_R6	80	5087599	107	4.659
FDA_R7	44	5052796	138	4.640
FDA_R8	52	5200929	119	4.807

Table 50: *continued*

Isolate_ID	Coverage depth	Total length	Contig count	Gene count
FDA_R10	38	5072012	158	4.665
FDA_R12	43	5051895	116	4.618
FDA_R13	135	5122134	161	4.694
FDA_R16	56	5051817	114	4.638
FDA_R17	60	5010966	100	4.576
FDA_R21	56	5124206	80	4.685
FDA_R26	119	5067308	134	4.629
FDA_R29	40	4946371	104	4.525
FDA_R30	45	5132857	131	4.724
FDA_R31	102	5126812	132	4.710
FDA_R32	27	5059594	105	4.607
FDA_R33	109	5043795	94	4.586
FDA_R42	39	4981892	105	4.522
FDA_R45	130	4973945	101	4.533
FDA_R47	66	5064518	151	4.653
FDA_R51	34	5049334	105	4.632
FDA_R52	84	5189100	86	4.693
FDA_R53	68	5191609	101	4.715
FDA_R54	48	5189492	96	4.715
FDA_R55	62	5191687	100	4.718
FDA_R56	111	5191187	101	4.715
FDA_R57	72	5100768	101	4.652
FDA_R59	38	4953641	128	4.512
FDA_R60	77	5264840	126	4.802
FDA_R62	34	5176254	126	4.777
FDA_R63	44	5125149	102	4.684
FDA_R65	46	5065957	111	4.652
FDA_R74	90	5041370	100	4.605
FDA_R75	84	5014250	87	4.578
FDA_R76	108	5017829	82	4.588
FDA_R77	74	5014463	76	4.581
FDA_R86	70	4894547	116	4.454
FDA_R87	57	5180717	111	4.794
FDA_R88	34	5184181	90	4.781
FDA_R95	44	5123173	103	4.657
FDA_R96	41	4931576	111	4.466
FDA_R98	41	5125449	102	4.676
FDA_R99	74	5120561	96	4.647
FDA_R108	68	5156159	93	4.700
FDA_R109	34	5124297	93	4.665
FDA_R110	54	5123801	96	4.673
FDA_R111	102	4925276	95	4.459
FDA_R125	138	5078459	150	4.628
FDA_R126	103	5138446	115	4.714

Table 50: *continued*

Isolate_ID	Coverage depth	Total length	Contig count	Gene count
FDA_R129	71	4897642	115	4.452
FDA_R130	76	5057199	108	4.594
FDA_R131	41	5140538	132	4.688
FDA_R135	34	5066023	141	4.634
FDA_R136	78	5098402	139	4.651
FDA_R137	71	5107378	127	4.638
FDA_R138	65	5108840	148	4.657
FDA_R143	280	4947784	148	4.510
FDA_R144	48	5085869	181	4.641
FDA_R145	64	5074938	156	4.624
FDA_R146	50	5084933	172	4.649
FDA_R149	61	5067880	152	4.641
FDA_R150	25	5066369	149	4.663
CDC_K4556-2	74	5149094	135	4.735
CDC_K4557	35	5074019	94	4.619
CDC_K4558-1	34	5032007	120	4.611
CDC_K4558-2	88	5335587	155	4.897
CDC_K4588	57	5014336	118	4.610
CDC_K4636	35	5340949	138	4.889
CDC_K4637-1	59	5163379	150	4.755
CDC_K4637-2	59	5156414	133	4.740
CDC_K4638	47	5077332	96	4.619
CDC_K4639-1	61	5092291	122	4.654
CDC_K4639-2	57	5078734	114	4.658
CDC_K4762	78	5103196	108	4.676
CDC_K4763	43	5287071	129	4.843
CDC_K4764-1	43	5276586	114	4.813
CDC_K4764-2	52	5033430	111	4.616
CDC_K4775	71	5158365	121	4.699
CDC_K4842	47	5194684	115	4.791
CDC_K4857-1	37	5038683	103	4.569
CDC_K4857-2	47	5039775	108	4.592
CDC_K4858	56	4961237	89	4.534
CDC_K4859	192	5180859	107	4.732
CDC_K4981	118	4928549	135	4.500
CDC_K5009-2	94	5125537	130	4.681
CDC_K5010-1	59	5122961	146	4.694
CDC_K5010-2	94	5118895	134	4.676
CDC_K5058	60	5114568	84	4.620
CDC_K5059-1	83	4959751	80	4.543
CDC_K5059-2	458	4953954	123	4.513
CDC_K5067	70	5028824	144	4.585
CDC_K5073	85	5082152	117	4.604
CDC_K5125	71	5134866	134	4.699

Table 50: *continued*

Isolate_ID	Coverage depth	Total length	Contig count	Gene count
CDC_K5126	112	5143760	152	4.680
CDC_K5276	86	5168475	136	4.774
CDC_K5277	71	5165700	109	4.699
CDC_K5278	69	5067491	112	4.639
CDC_K5279	76	5172883	93	4.683
CDC_K5280	87	5083288	110	4.662
CDC_K5281	102	5077351	115	4.661
CDC_K5282	50	4963427	79	4.506
CDC_K5306	122	5045861	117	4.636
CDC_K5308	138	5095192	142	4.669
CDC_K5323-1	30	5220531	98	4.795
CDC_K5323-2	57	5207450	84	4.779
CDC_K5324-1	56	5272305	294	4.941
CDC_K5324-2	70	5060178	174	4.654
CDC_K5328	85	5090169	118	4.661
CDC_K5330	43	5168109	142	4.757
CDC_K5331	90	5067258	95	4.631
CDC_K5345-1	49	5148490	112	4.719
CDC_K5345-2	70	5125535	105	4.687
CDC_K5346	53	5087978	126	4.661
CDC_K5428	20	5092714	139	4.639
CDC_K5429	111	5079150	118	4.653
CDC_K5433	47	5073057	118	4.635
CDC_K5435	73	5168466	105	4.718
CDC_K5437	122	5068632	126	4.636
CDC_K5438	114	5154409	123	4.705
CDC_K5439	36	5031538	106	4.616
CDC_K5456	119	5080809	114	4.656
CDC_K5457	72	5075290	98	4.625
CDC_K5485	23	5151277	128	4.719
CDC_K5512	92	5082991	122	4.670
CDC_K5528	128	5101848	108	4.645
CDC_K5579	107	5072634	169	4.659
CDC_K5582	54	5174332	104	4.798
CDC_K5618	25	5107768	147	4.668
CDC_K5620	24	5099609	131	4.658
CDC_K5629	106	5077945	119	4.659
CDC_K5635	118	5070117	110	4.636
CDC_K5638	29	5144324	125	4.733
CDC_K5701	24	5157591	93	4.707

Table 51: Accession numbers of the *V. parahaemolyticus* shotgun sequences

Isolate ID	WGS Accession	Isolate ID	WGS Accession
FDA_R5	MIQI00000000	CDC_K4636	MISV00000000
FDA_R6	MIQJ00000000	CDC_K4637-1	MISW00000000
FDA_R7	MIQK00000000	CDC_K4637-2	MISX00000000
FDA_R8	MIQL00000000	CDC_K4638	MISY00000000
FDA_R10	MIQM00000000	CDC_K4639-1	MISZ00000000
FDA_R12	MIQN00000000	CDC_K4639-2	MITA00000000
FDA_R13	MIQO00000000	CDC_K4762	MITB00000000
FDA_R16	MIQP00000000	CDC_K4763	MITC00000000
FDA_R17	MIQQ00000000	CDC_K4764-1	MITD00000000
FDA_R21	MIQR00000000	CDC_K4764-2	MITE00000000
FDA_R26	MIQS00000000	CDC_K4775	MITF00000000
FDA_R29	MIQT00000000	CDC_K4842	MITG00000000
FDA_R30	MIQU00000000	CDC_K4857-1	MITH00000000
FDA_R31	MIQV00000000	CDC_K4857-2	MITI00000000
FDA_R32	MIQW00000000	CDC_K4858	MITJ00000000
FDA_R33	MIQX00000000	CDC_K4859	MITK00000000
FDA_R42	MIQY00000000	CDC_K4981	MITL00000000
FDA_R45	MIQZ00000000	CDC_K5009-2	MITM00000000
FDA_R47	MIRA00000000	CDC_K5010-1	MITN00000000
FDA_R51	MIRB00000000	CDC_K5010-2	MITO00000000
FDA_R52	MIRC00000000	CDC_K5058	MITP00000000
FDA_R53	MIRD00000000	CDC_K5059-1	MITQ00000000
FDA_R54	MIRE00000000	CDC_K5059-2	MITR00000000
FDA_R55	MIRF00000000	CDC_K5067	MITS00000000
FDA_R56	MIRG00000000	CDC_K5073	MITT00000000
FDA_R57	MIRH00000000	CDC_K5125	MITU00000000
FDA_R59	MIRI00000000	CDC_K5126	MITV00000000
FDA_R60	MIRJ00000000	CDC_K5276	MITW00000000
FDA_R62	MIRK00000000	CDC_K5277	MITX00000000
FDA_R63	MIRL00000000	CDC_K5278	MITY00000000
FDA_R65	MIRM00000000	CDC_K5279	MITZ00000000
FDA_R74	MIRN00000000	CDC_K5280	MIUA00000000
FDA_R75	MIRO00000000	CDC_K5281	MIUB00000000
FDA_R76	MIRP00000000	CDC_K5282	MIUC00000000
FDA_R77	MIRQ00000000	CDC_K5306	MIUD00000000
FDA_R86	MIRR00000000	CDC_K5308	MIUE00000000
FDA_R87	MIRS00000000	CDC_K5323-1	MIUF00000000
FDA_R88	MIRT00000000	CDC_K5323-2	MIUG00000000
FDA_R95	MIRU00000000	CDC_K5324-1	MIUH00000000
FDA_R96	MIRV00000000	CDC_K5324-2	MIUI00000000
FDA_R98	MIRW00000000	CDC_K5328	MIUJ00000000
FDA_R99	MIRX00000000	CDC_K5330	MIUK00000000
FDA_R108	MIRY00000000	CDC_K5331	MIUL00000000

Table 51: *continued*

Isolate ID	WGS Accession	Isolate ID	WGS Accession
FDA_R109	MIRZ00000000	CDC_K5345-1	MIUM00000000
FDA_R110	MISA00000000	CDC_K5345-2	MIUN00000000
FDA_R111	MISB00000000	CDC_K5346	MIUO00000000
FDA_R125	MISC00000000	CDC_K5428	MIUP00000000
FDA_R126	MISD00000000	CDC_K5429	MIUQ00000000
FDA_R129	MISE00000000	CDC_K5433	MIUR00000000
FDA_R130	to be assigned	CDC_K5435	MIUS00000000
FDA_R131	MISF00000000	CDC_K5437	MIUT00000000
FDA_R135	MISG00000000	CDC_K5438	MIUU00000000
FDA_R136	MISH00000000	CDC_K5439	MIUV00000000
FDA_R137	MISI00000000	CDC_K5456	MIUW00000000
FDA_R138	MISJ00000000	CDC_K5457	MIUX00000000
FDA_R143	MISK00000000	CDC_K5485	MIUY00000000
FDA_R144	MISL00000000	CDC_K5512	MIUZ00000000
FDA_R145	MISM00000000	CDC_K5528	MIVA00000000
FDA_R146	MISN00000000	CDC_K5579	MIVB00000000
FDA_R149	MISO00000000	CDC_K5582	MIVC00000000
FDA_R150	MISP00000000	CDC_K5618	MIVD00000000
CDC_K4556-2	MISQ00000000	CDC_K5620	MIVE00000000
CDC_K4557	MISR00000000	CDC_K5629	MIVF00000000
CDC_K4558-1	MISS00000000	CDC_K5635	MIVG00000000
CDC_K4558-2	MIST00000000	CDC_K5638	MIVH00000000
CDC_K4588	MISU00000000	CDC_K5701	MIVI00000000

Table 52: Isolates used in this study and their MLST allele profile and according sequence types

Isolate ID	Serotype	*tlh*	*tdh*	*trh*	*dnaE*	*gyrB*	*recA*	*dtdS*	*pntA*	*pyrC*	*tnaA*	ST
CDC_K4637-1	O3:K6	+	+	-	3	4	19	4	29	4	22	3
CDC_K4637-2	O3:K6	+	+	-	3	4	19	4	29	4	22	3
CDC_K4775	O3:K6	+	+	-	3	4	19	4	29	4	22	3
CDC_K5010-1	O1:Kuk	+	+	-	3	4	19	4	29	4	22	3
CDC_K5010-2	O1:Kuk	+	+	-	3	4	19	4	29	4	22	3
CDC_K5058	O3:K6	+	+	-	3	4	19	4	29	4	22	3
CDC_K5528	O4:K68	+	+	-	3	4	19	4	29	4	22	3
FDA_R21	O5:Kuk	+	-	+	9	21	15	13	4	10	26	12
FDA_R30	O1:Kuk	+	+	+	17	16	13	36	15	31	26	23
FDA_R31	O1:Kuk	+	+	+	17	16	13	36	15	31	26	23
FDA_R33	O3:Kuk	+	-	-	17	16	13	7	24	16	20	28
FDA_R12	O4:K8	+	+	+	20	25	15	6	7	11	4	32
FDA_R26	O4:K8	+	+	+	20	25	15	6	7	11	4	32
FDA_R47	O4:K8	+	+	+	20	25	15	6	7	11	4	32
CDC_K5306	O4:K9	+	+	+	20	25	15	13	7	11	5	34

Table 52: *continued*

Isolate ID	Sero-type	*tlh*	*tdh*	*trh*	*dnaE*	*gyrB*	*recA*	*dtdS*	*pntA*	*pyrC*	*tnaA*	ST
FDA_R16	O4:K9	+	+	+	20	25	15	13	7	11	5	34
CDC_K4639-1	O4:K12	+	+	+	21	15	1	23	23	21	16	36
CDC_K4639-2	O4:Kuk	+	+	+	21	15	1	23	23	21	16	36
CDC_K5278	O4:K12	+	+	+	21	15	1	23	23	21	16	36
CDC_K5280	O4:K12	+	+	+	21	15	1	23	23	21	16	36
CDC_K5281	O4:K12	+	+	+	21	15	1	23	23	21	16	36
CDC_K5308	O4:K63	+	+	+	21	15	1	23	23	21	16	36
CDC_K5328	O4:K12	+	+	+	21	15	1	23	23	21	16	36
CDC_K5345-1	O4:K12	+	+	+	21	15	1	23	23	21	16	36
CDC_K5345-2	O4:K12	+	+	+	21	15	1	23	23	21	16	36
CDC_K5346	O4:K12	+	+	+	21	15	1	23	23	21	16	36
CDC_K5429	O4:K12	+	+	+	21	15	1	23	23	21	16	36
CDC_K5433	O4:Kuk	+	+	+	21	15	1	23	23	21	16	36
CDC_K5437	O4:Kuk	+	+	+	21	15	1	23	23	21	16	36
CDC_K5456	O4:Kuk	+	+	+	21	15	1	23	23	21	16	36
CDC_K5457	O4:Kuk	+	+	+	21	15	1	23	23	21	16	36
CDC_K5512	O4:K12	+	+	+	21	15	1	23	23	21	16	36
CDC_K5629	O4:K13	+	+	+	21	15	1	23	23	21	16	36
CDC_K5638	O4:K12	+	+	+	21	15	1	23	23	21	16	36
CDC_K5579	O4:K63	+	+	+	23	29	10	7	14	24	2	43
CDC_K5485	O6:K18	+	-	-	29	5	22	12	20	22	25	50
FDA_R45	O5:Kuk	+	-	+	37	14	14	9	14	34	26	61
CDC_K5277	O1:Kuk	+	-	+	39	9	27	39	3	37	30	65
CDC_K5279	O1:Kuk	+	-	+	39	9	27	39	3	37	30	65
CDC_K5435	O1:Kuk	+	-	+	39	9	27	39	3	37	30	65
CDC_K5438	O1:Kuk	+	-	+	39	9	27	39	3	37	30	65
CDC_K5701	O1:Kuk	+	-	+	39	9	27	39	3	37	30	65
CDC_K4857-1	O5:K17	+	-	-	35	43	38	21	31	35	37	79
CDC_K4857-2	O5:Kuk	+	-	-	35	43	38	21	31	35	37	79
FDA_R74	O4:K34	+	-	-	26	58	53	19	28	9	26	108
FDA_R59	O5:Kuk	+	-	-	63	62	25	54	4	56	45	113
CDC_K5439	O4:K8	+	+	-	11	48	3	48	26	48	26	189
CDC_K5428	O1:Kuk	+	+	+	22	28	17	13	8	19	14	199
CDC_K5330	O5:Kuk	+	-	+	11	48	107	48	26	48	26	265
CDC_K5331	O4:K8	+	+	-	11	48	107	48	26	48	26	265
CDC_K4858	O4:K4	+	-	-	27	84	127	139	54	124	37	283
FDA_R10	O1:Kuk	+	+	+	142	29	10	7	4	24	20	313
FDA_R149	O1:Kuk	+	+	+	142	29	10	7	4	24	20	313
FDA_R150	O1:Kuk	+	+	+	142	29	10	7	4	24	20	313
FDA_R87	O8:K70	+	+	+	145	177	140	158	4	132	104	320
FDA_R88	O8:K70	+	+	+	145	177	140	158	4	132	104	320
FDA_R17	O4:Kuk	+	-	-	14	30	49	11	49	11	13	536
CDC_K5276	O11:Kuk	+	+	+	222	128	21	69	46	236	12	631
CDC_K5582	O11:Kuk	+	+	+	222	128	21	69	46	236	12	631

Table 52: *continued*

Isolate ID	Serotype	*tlh*	*tdh*	*trh*	*dnaE*	*gyrB*	*recA*	*dtdS*	*pntA*	*pyrC*	*tnaA*	ST
CDC_K4558-2	O3:Kuk	+	-	-	223	106	31	221	45	171	165	636
CDC_K4636	O10:Kuk	+	+	+	223	106	31	221	45	171	165	636
CDC_K5618	O10:Kuk	+	+	+	223	106	31	221	45	171	165	636
CDC_K5620	O10:Kuk	+	+	+	223	106	31	221	45	171	165	636
CDC_K4762	O5:K17	+	-	-	83	82	73	83	4	77	58	674
CDC_K5323-1	O5:K17	+	-	+	83	82	73	83	4	77	58	674
CDC_K5323-2	O5:Kuk	+	-	+	83	82	73	83	4	77	58	674
FDA_R51	O8:Kuk	+	+	+	60	106	31	72	66	62	65	676
FDA_R75	O8:Kuk	+	+	+	60	106	31	72	66	62	65	676
FDA_R76	O8:Kuk	+	+	+	60	106	31	72	66	62	65	676
FDA_R77	O8:Kuk	+	+	+	60	106	31	72	66	62	65	676
FDA_R13	O4:K10	+	-	-	241	330	205	253	28	22	188	732
FDA_R29	O11:Kuk	+	-	-	235	22	25	273	164	254	20	734
FDA_R52	O3:Kuk	+	-	+	4	13	11	38	18	46	23	735
FDA_R53	O3:Kuk	+	-	+	4	13	11	38	18	46	23	735
FDA_R54	O3:Kuk	+	-	+	4	13	11	38	18	46	23	735
FDA_R55	O3:Kuk	+	-	+	4	13	11	38	18	46	23	735
FDA_R56	O3:Kuk	+	-	+	4	13	11	38	18	46	23	735
FDA_R86	O6:Kuk	+	-	-	45	336	143	7	171	255	36	737
FDA_R125	O11:Kuk	+	+	-	17	331	235	23	33	137	94	739
FDA_R135	O3:Kuk	+	-	-	26	16	41	224	31	32	23	741
FDA_R143	O5:Kuk	+	-	-	17	64	137	60	94	11	51	743
CDC_K4556-2	O1:Kuk	+	-	-	31	82	236	35	23	26	51	744
CDC_K4981	O1:Kuk	+	-	-	17	327	13	8	172	32	181	748
CDC_K5009-2	O4:K53	+	+	+	5	71	238	162	26	11	107	749
CDC_K5125	O3:Kuk	+	-	-	5	71	238	162	26	11	107	749
CDC_K5073	O3:K56	+	+	+	17	57	52	285	44	28	36	750
CDC_K5067	O1:K56	+	+	+	31	16	13	36	33	11	19	775
FDA_R136	O1:K20	+	+	+	31	16	13	36	33	11	19	775
FDA_R137	O1:K20	+	+	+	31	16	13	36	33	11	19	775
FDA_R138	O1:K20	+	+	+	31	16	13	36	33	11	19	775
CDC_K4557	O1:K33	+	-	-	28	4	82	88	63	187	1	799
CDC_K4638	O10:Kuk	+	-	+	215	344	144	76	48	232	26	809
CDC_K5126	O3:Kuk	+	-	-	17	323	3	19	21	32	20	1131
CDC_K5324-1	O1:K20	+	+	+	56	16	13	286	14	11	19	1132
CDC_K5324-2	O1:K20	+	+	+	56	16	13	286	14	11	19	1132
FDA_R5	O10:Kuk	+	-	+	214	329	30	19	26	69	26	1133
FDA_R6	O10:Kuk	+	-	+	214	329	30	19	26	69	26	1133
FDA_R7	O10:Kuk	+	-	+	137	212	19	139	129	46	14	1134
FDA_R8	O10:Kuk	+	-	+	137	212	19	139	129	46	14	1134
FDA_R60	O10:Kuk	+	-	+	63	326	231	191	48	120	24	1135
FDA_R62	O1:Kuk	+	-	+	31	327	13	157	14	3	20	1136
FDA_R130	O4:K37	+	+	-	19	295	295	223	136	11	13	1140
FDA_R131	O10:Kuk	+	+	+	51	57	75	353	45	78	231	1141

Table 52: *continued*

Isolate ID	Serotype	*tlh*	*tdh*	*trh*	*dnaE*	*gyrB*	*recA*	*dtdS*	*pntA*	*pyrC*	*tnaA*	ST
FDA_R32	O10:Kuk	+	+	+	26	118	13	321	33	32	155	1142
CDC_K4558-1	O3:K39	+	-	-	10	323	296	15	14	160	2	1143
CDC_K4842	O5:K47	+	-	+	3	53	297	13	50	334	187	1144
CDC_K5635	O5:K30	+	-	-	158	131	298	287	128	43	189	1145
FDA_R126	O4:K42	+	-	-	36	285	292	250	26	227	26	1146
CDC_K5059-1	O5:Kuk	+	-	-	292	154	25	50	73	35	23	1147
CDC_K5059-2	O5:Kuk	+	-	-	292	154	25	50	73	35	23	1147
FDA_R57	O3:Kuk	+	-	-	293	16	13	57	24	16	20	1148
FDA_R144	O5:Kuk	+	+	+	59	417	95	361	14	32	232	1149
FDA_R145	O5:Kuk	+	+	+	59	417	95	361	14	32	232	1149
FDA_R146	O5:Kuk	+	+	+	59	417	95	361	14	32	232	1149
CDC_K4588	O5:Kuk	+	-	+	56	16	237	362	33	59	20	1150
FDA_R108	O3:K5	+	-	+	47	328	299	13	2	256	23	1151
FDA_R109	O3:K5	+	-	+	47	328	299	13	2	256	23	1151
FDA_R110	O3:K5	+	-	+	47	328	299	13	2	256	23	1151
FDA_R95	O3:K5	+	-	+	47	328	299	13	2	256	23	1151
FDA_R98	O3:K5	+	-	+	47	328	299	13	2	256	23	1151
FDA_R99	O3:K5	+	-	+	47	328	299	13	2	256	23	1151
FDA_R111	O11:Kuk	+	-	-	3	16	192	363	206	335	2	1152
FDA_R65	O5:Kuk	+	-	+	3	16	192	363	206	335	2	1152
FDA_R96	O11:Kuk	+	-	-	3	16	192	363	206	335	2	1152
FDA_R129	O11:Kuk	+	-	-	49	16	13	364	33	31	2	1153
FDA_R42	O10:Kuk	+	-	+	17	323	6	23	171	29	155	1155
CDC_K4764-2	O8:K41	+	-	-	294	39	300	167	106	11	26	1156
CDC_K4763	O5:Kuk	+	+	+	291	177	*untyp	125	18	175	26	untyp.
CDC_K4764-1	O8:K41	+	-	-	291	177	untyp.	125	18	175	26	untyp.
CDC_K4859	O6:K18	+	-	-	19	4	untyp.	19	74	214	24	untyp.
CDC_K5282	O5:Kuk	+	-	-	19	217	89	175	untyp.	62	51	untyp.
FDA_R63	O4:Kuk	+	-	+	19	295	untyp.	223	136	11	13	untyp.

*untyp. = untypeable

Table 53: Results of the presence/absence analysis of *tlh*, *tdh*, and *trh*

	PCR results			BLAST (selected sequences)			NPGAAP	Interpro	BLAST (Swiss-Prot)		
Isolate ID	*tlh*	*tdh*	*trh*	*tlh*	*tdh*2	*trh*	hemolysin activation protein	TDH	*trh*	*tdh*_1	*tdh*_2
FDA_R5	+	-	+	+	-	+	+	+	-	-	+
FDA_R6	+	-	+	+	-	+	+	+	-	-	+
FDA_R7	+	-	+	+	-	+	-	-	-	-	-
FDA_R8	+	-	+	+	-	+	-	-	-	-	-
FDA_R10	+	+	+	+	+	+	+	+	+	-	+
FDA_R12	+	+	+	+	+	+	+	+	+	-	+
FDA_R13	+	-	-	+	-	-	-	-	-	-	-
FDA_R16	+	+	+	+	+	+	+	+	+	-	+
FDA_R17	+	-	-	+	-	-	-	-	-	-	-
FDA_R21	+	-	+	+	-	+	+	+	-	-	+
FDA_R26	+	+	+	+	+	+	+	+	+	-	+
FDA_R29	+	-	-	+	-	-	-	-	-	-	-
FDA_R30	+	+	+	+	+	+	+	+	+	-	+
FDA_R31	+	+	+	+	+	+	+	+	+	-	+
FDA_R32	+	+	+	+	+	+	+	+	+	-	+
FDA_R33	+	-	-	+	-	-	-	-	-	-	-
FDA_R42	+	-	+	+	-	+	-	+	-	-	+
FDA_R45	+	-	+	+	-	+	-	+	-	-	+
FDA_R47	+	+	+	+	+	+	+	+	+	-	+
FDA_R51	+	+	+	+	+	+	+	+	+	-	+
FDA_R52	+	-	+	+	-	+	+	+	-	-	+
FDA_R53	+	-	+	+	-	+	+	+	-	-	+
FDA_R54	+	-	+	+	-	+	+	+	-	-	+
FDA_R55	+	-	+	+	-	+	+	+	-	-	+
FDA_R56	+	-	+	+	-	+	+	+	-	-	+
FDA_R57	+	-	-	+	-	-	-	-	-	-	-
FDA_R59	+	-	-	+	-	-	-	-	-	-	-
FDA_R60	+	-	+	+	-	+	+	+	-	-	+
FDA_R62	+	-	+	+	-	+	+	+	-	-	+
FDA_R63	+	-	+	+	-	+	+	+	-	-	+
FDA_R65	+	-	+	+	-	+	+	+	-	-	+
FDA_R74	+	-	-	+	-	-	-	-	-	-	-
FDA_R75	+	+	+	+	+	+	+	+	+	-	+
FDA_R76	+	+	+	+	+	+	+	+	+	-	+
FDA_R77	+	+	+	+	+	+	+	+	+	-	+
FDA_R86	+	-	-	+	-	-	-	-	-	-	-
FDA_R87	+	+	+	+	+	+	+	+	+	-	+
FDA_R88	+	+	+	+	+	+	+	+	+	-	+

Table 53: *continued*

	PCR results			BLAST (selected sequences)			NPGAAP	Interpro	BLAST (Swiss-Prot)		
Isolate ID	*tlh*	*tdh*	*trh*	*tlh*	*tdh*2	*trh*	hemolysin activation protein	TDH	*trh*	*tdh*_1	*tdh*_2
FDA_R95	+	-	+	+	-	+	+	+	-	-	+
FDA_R96	+	-	-	+	-	-	-	-	-	-	-
FDA_R98	+	-	+	+	-	+	+	+	-	-	+
FDA_R99	+	-	+	+	-	+	+	+	-	-	+
FDA_R108	+	-	+	+	-	+	+	+	-	-	+
FDA_R109	+	-	+	+	-	+	+	+	-	-	+
FDA_R110	+	-	+	+	-	+	+	+	-	-	+
FDA_R111	+	-	-	+	-	-	-	-	-	-	-
FDA_R125	+	+	-	+	+	-	+	+	-	-	+
FDA_R126	+	-	-	+	-	-	-	-	-	-	-
FDA_R129	+	-	-	+	-	-	-	-	-	-	-
FDA_R130	+	+	-	+	+	-	+	+	-	+	-
FDA_R131	+	+	+	+	+	+	+	+	+	-	+
FDA_R135	+	-	-	+	-	-	-	-	-	-	-
FDA_R136	+	+	+	+	+	+	+	+	+	-	+
FDA_R137	+	+	+	+	+	+	+	+	+	-	+
FDA_R138	+	+	+	+	+	+	+	+	+	-	+
FDA_R143	+	-	-	+	-	-	-	-	-	-	-
FDA_R144	+	+	+	+	+	+	+	+	+	-	+
FDA_R145	+	+	+	+	+	+	+	+	+	-	+
FDA_R146	+	+	+	+	+	+	+	+	+	-	+
FDA_R149	+	+	+	+	+	+	+	+	+	-	+
FDA_R150	+	+	+	+	+	+	+	+	+	-	+
CDC_K4556-2	+	-	-	+	-	-	-	-	-	-	-
CDC_K4557	+	-	-	+	-	-	-	-	-	-	-
CDC_K4558-1	+	-	-	+	-	-	-	-	-	-	-
CDC_K4558-2	+	-	-	+	+	+	+	+	+	-	+
CDC_K4588	+	-	+	+	-	+	+	+	-	-	+
CDC_K4636	+	+	+	+	+	+	+	+	+	-	+
CDC_K4637-1	+	+	-	+	+	-	+	+	-	-	+
CDC_K4637-2	+	+	-	+	+	-	+	+	-	+	+
CDC_K4638	+	-	+	+	-	+	+	+	-	-	+
CDC_K4639-1	+	+	+	+	+	+	+	+	+	-	+
CDC_K4639-2	+	+	+	+	+	+	+	+	+	-	+
CDC_K4762	+	-	-	+	-	-	-	-	-	-	-
CDC_K4763	+	+	+	+	+	+	+	+	+	-	+
CDC_K4764-1	+	-	-	+	+	+	+	+	+	-	+
CDC_K4764-2	+	-	-	+	-	-	-	-	-	-	-
CDC_K4775	+	+	-	+	+	-	+	+	-	+	-

Table 53: *continued*

	PCR results			BLAST (selected sequences)			NPGAAP	Interpro	BLAST (Swiss-Prot)		
Isolate ID	*tlh*	*tdh*	*trh*	*tlh*	*tdh*2	*trh*	hemolysin activation protein	TDH	*trh*	*tdh*_1	*tdh*_2
CDC_K4842	+	-	+	+	-	+	+	+	-	-	+
CDC_K4857-1	+	-	-	+	-	-	-	-	-	-	-
CDC_K4857-2	+	-	-	+	-	-	-	-	-	-	-
CDC_K4858	+	-	-	+	-	-	-	-	-	-	-
CDC_K4859	+	-	-	+	-	-	-	-	-	-	-
CDC_K4981	+	-	-	+	-	-	-	-	-	-	-
CDC_K5009-2	+	+	+	+	+	-	+	+	-	-	+
CDC_K5010-1	+	+	-	+	+	-	+	+	-	+	+
CDC_K5010-2	+	+	-	+	+	-	+	+	-	+	+
CDC_K5058	+	+	-	+	+	-	+	+	-	+	-
CDC_K5059-1	+	-	-	+	-	-	-	-	-	-	-
CDC_K5059-2	+	-	-	+	-	-	-	-	-	-	-
CDC_K5067	+	+	+	+	+	+	+	+	+	-	+
CDC_K5073	+	+	+	+	+	+	+	+	+	-	+
CDC_K5125	+	-	-	+	+	-	+	+	-	-	+
CDC_K5126	+	-	-	+	-	-	-	-	-	-	-
CDC_K5276	+	+	+	+	+	+	+	+	+	-	+
CDC_K5277	+	-	+	+	-	+	+	+	-	-	+
CDC_K5278	+	+	+	+	+	+	+	+	+	-	+
CDC_K5279	+	-	+	+	-	+	+	+	-	-	+
CDC_K5280	+	+	+	+	+	+	+	+	+	-	+
CDC_K5281	+	+	+	+	+	+	+	+	+	-	+
CDC_K5282	+	-	-	+	-	-	-	-	-	-	-
CDC_K5306	+	+	+	+	+	+	+	+	+	-	+
CDC_K5308	+	+	+	+	+	+	+	+	+	-	+
CDC_K5323-1	+	-	+	+	-	+	+	+	-	-	+
CDC_K5323-2	+	-	+	+	-	+	+	+	-	-	+
CDC_K5324-1	+	+	+	+	+	+	+	+	+	-	+
CDC_K5324-2	+	+	+	+	+	+	+	+	+	-	+
CDC_K5328	+	+	+	+	+	+	+	+	+	-	+
CDC_K5330	+	-	+	+	-	+	+	+	-	-	+
CDC_K5331	+	+	-	+	+	-	+	+	-	+	-
CDC_K5345-1	+	+	+	+	+	+	+	+	+	-	+
CDC_K5345-2	+	+	+	+	+	+	+	+	+	-	+
CDC_K5346	+	+	+	+	+	+	+	+	+	-	+
CDC_K5428	+	+	+	+	+	+	+	+	+	-	+
CDC_K5429	+	+	+	+	+	+	+	+	+	-	+
CDC_K5433	+	+	+	+	+	+	+	+	+	-	+
CDC_K5435	+	-	+	+	-	+	+	+	-	-	+

Table 53: *continued*

	PCR results			BLAST (selected sequences)			NPGAAP	Interpro	BLAST (Swiss-Prot)		
Isolate ID	*tlh*	*tdh*	*trh*	*tlh*	*tdh*2	*trh*	hemolysin activation protein	TDH	*trh*	*tdh*_1	*tdh*_2
CDC_K5437	+	+	+	+	+	+	+	+	+	-	+
CDC_K5438	+	-	+	+	-	+	+	+	-	-	+
CDC_K5439	+	+	-	+	+	-	+	+	-	+	-
CDC_K5456	+	+	+	+	+	+	+	+	+	-	+
CDC_K5457	+	+	+	+	+	+	+	+	+	-	+
CDC_K5485	+	-	-	+	-	-	-	-	-	-	-
CDC_K5512	+	+	+	+	+	+	+	+	+	-	+
CDC_K5528	+	+	-	+	+	-	+	+	-	+	+
CDC_K5579	+	+	+	+	+	+	+	+	+	-	+
CDC_K5582	+	+	+	+	+	+	+	+	+	-	+
CDC_K5618	+	+	+	+	+	+	+	+	+	-	+
CDC_K5620	+	+	+	+	+	+	+	+	+	-	+
CDC_K5629	+	+	+	+	+	+	+	+	+	-	+
CDC_K5635	+	-	-	+	-	-	-	-	-	-	-
CDC_K5638	+	+	+	+	+	+	+	+	+	-	+
CDC_K5701	+	-	+	+	-	+	+	+	-	-	+

11.1.7. Attachments to chapter 5.7

Table 54: Isolates used for LS-BSR analysis

Isolate ID	Cytotoxicity Caco	Isolate ID	Cytotoxicity Caco	Isolate ID	Cytotoxicity Caco
CDC_K4556-2	Semi-cytotoxic	FDA_R75	Semi-cytotoxic	CDC_K5330	cytotoxic
CDC_K4588	Semi-cytotoxic	FDA_R8	Semi-cytotoxic	CDC_K5345-1	cytotoxic
CDC_K4762	Semi-cytotoxic	FDA_R88	Semi-cytotoxic	CDC_K5345-2	cytotoxic
CDC_K4764-2	Semi-cytotoxic	FDA_R99	Semi-cytotoxic	CDC_K5346	cytotoxic
CDC_K4981	Semi-cytotoxic	CDC_K4557	cytotoxic	CDC_K5429	cytotoxic
CDC_K5010-1	Semi-cytotoxic	CDC_K4637-1	cytotoxic	CDC_K5456	cytotoxic
CDC_K5067	Semi-cytotoxic	CDC_K4638	cytotoxic	CDC_K5457	cytotoxic
CDC_K5276	Semi-cytotoxic	CDC_K4639-1	cytotoxic	CDC_K5485	cytotoxic
CDC_K5280	Semi-cytotoxic	CDC_K4639-2	cytotoxic	CDC_K5512	cytotoxic
CDC_K5579	Semi-cytotoxic	CDC_K4763	cytotoxic	CDC_K5582	cytotoxic
CDC_K5635	Semi-cytotoxic	CDC_K4764-1	cytotoxic	CDC_K5629	cytotoxic
FDA_R108	Semi-cytotoxic	CDC_K4842	cytotoxic	FDA_R10	cytotoxic
FDA_R109	Semi-cytotoxic	CDC_K4857-1	cytotoxic	FDA_R126	cytotoxic
FDA_R110	Semi-cytotoxic	CDC_K4857-2	cytotoxic	FDA_R145	cytotoxic
FDA_R135	Semi-cytotoxic	CDC_K5009-2	cytotoxic	FDA_R146	cytotoxic
FDA_R144	Semi-cytotoxic	CDC_K5010-2	cytotoxic	FDA_R149	cytotoxic

Table 54: *continued*

Isolate ID	Cytotoxicity Caco	Isolate ID	Cytotoxicity Caco	Isolate ID	Cytotoxicity Caco
FDA_R42	Semi-cytotoxic	CDC_K5059-1	cytotoxic	FDA_R150	cytotoxic
FDA_R47	Semi-cytotoxic	CDC_K5059-2	cytotoxic	FDA_R16	cytotoxic
FDA_R52	Semi-cytotoxic	CDC_K5125	cytotoxic	FDA_R17	cytotoxic
FDA_R53	Semi-cytotoxic	CDC_K5126	cytotoxic	FDA_R21	cytotoxic
FDA_R54	Semi-cytotoxic	CDC_K5278	cytotoxic	FDA_R33	cytotoxic
FDA_R55	Semi-cytotoxic	CDC_K5281	cytotoxic	FDA_R5	cytotoxic
FDA_R56	Semi-cytotoxic	CDC_K5282	cytotoxic	FDA_R6	cytotoxic
FDA_R57	Semi-cytotoxic	CDC_K5308	cytotoxic	FDA_R74	cytotoxic
FDA_R60	Semi-cytotoxic	CDC_K5323-1	cytotoxic	FDA_R87	cytotoxic
FDA_R65	Semi-cytotoxic	CDC_K5323-2	cytotoxic	FDA_R95	cytotoxic
FDA_R7	Semi-cytotoxic	CDC_K5328	cytotoxic	FDA_R98	cytotoxic

Table 55: Isolates used for RAxML tree, COG functions and SNP analysis and their cytotoxic values, red: cytotoxic(75,01-100), green: non-cytotoxic (0-50)

Isolate ID	Cytotoxicity Caco	Isolate ID	Cytotoxicity Caco
FDA_R5	103,33	CDC_K4557	101,26
FDA_R6	95,49	CDC_K4638	94,08
FDA_R10	98,72	CDC_K4858	-0,027
FDA_R12	11,00	CDC_K4859	18,23
FDA_R17	93,40	CDC_K5073	40,57
FDA_R29	-0,06	CDC_K5125	102,21
FDA_R30	5,14	CDC_K5126	96,76
FDA_R31	5,76	CDC_K5277	17,73
FDA_R33	76,14	CDC_K5278	103,82
FDA_R45	-4,89	CDC_K5281	101,28
FDA_R59	-6,95	CDC_K5282	98,81
FDA_R62	-3,69	CDC_K5306	19,84
FDA_R74	84,89	CDC_K5328	103,14
FDA_R86	-0,21	CDC_K5330	99,60
FDA_R87	81,09	CDC_K5346	100,45
FDA_R95	93,69	CDC_K5429	83,41
FDA_R96	-1,44	CDC_K5437	102,59
FDA_R98	75,16	CDC_K5456	98,57
FDA_R111	-0,36	CDC_K5457	86,69
FDA_R126	81,47	CDC_K5485	82,85
FDA_R136	5,91	CDC_K5512	82,68
FDA_R137	7,22	CDC_K5582	77,26
FDA_R138	7,23	CDC_K5618	49,47

Table 55: *continued*

Isolate ID	Cytotoxicity Caco	Isolate ID	Cytotoxicity Caco
FDA_R143	-2,83	CDC_K5629	98,82
FDA_R149	87,43	CDC_K5638	85,20
FDA_R150	97,92		

11.2. Materials and Protocols

11.2.1. Equipment

Table 56: List of equipment used

Equipment	Model/ Name	Manufacturer
Autoclave	Amsco Eagle SG3021	STERIS Corporation, Mentor, OH, USA
	Amsco Lab 250 LV250	
Micro scales	PB503-S/Fact	Mettler Toldeo, Sanford, NC, USA
Scale	PM300	Mettler Toldeo, Sanford, NC, USA
pH-electrode	XL500	Thermo Scientific Forma, Rockville, MD, USA
Water purification system	Milli-Q Biocel system	EMD Millipore, Billerica, MA, USA
Stirring hotplate	Isotemp Basic Stirring Hotplate	Thermo Scientific Forma, Rockville, MD, USA
Shaking waterbath	Precision reciprocal shaking bath	Thermo Scientific Forma, Rockville, MD, USA
Heatblock	Isotemp heatblock	Thermo Scientific Forma, Rockville, MD, USA
Microscope	Axiovert 35	Carl Zeiss Microscopy, Jena, Germany
Thermocycler	peqSTAR thermocycler	Peqlab, Erlangen, Germany
Real-time-PCR-Cycler	AB7500 Fast	LifeTechnologies, Carlsbad, CA, USA
	Smart Cycler	Cepheid, Sunnyvale, CA, USA
	Light Cycler 480	Roche Applied Science, USA
Whole Genome Sequencer	IonTorrent Personal Genome Machine Illumina HiSeq 2000 sequencer PacBio RSII sequencer	LifeTechnologies, Carlsbad, CA, USA Illumina Inc., San Diego, CA, USA Pacific Biosciences, Menlo Park, CA, USA

Table 56: *continued*

Equipment	Model/ Name	Manufacturer
PFGE equipment	CHEF-DR® III Variable Angle System	Bio-Rad, Hercules, CA, USA
Microwave	GE Countertop Microwave	GE Appliances, Rapid City, SD, USA
Centrifuge	Scientific Sorvall RC 12BP Plus floor centrifuge	Thermo Scientific Forma, Rockville, MD, USA
Microfuge	Centrifuge 5415D	Eppendorf, New York, NY, USA
Tabletop microfuge	VWR Galaxy MiniStar Microcentrifuge	Thermo Scientific Forma, Rockville, MD, USA
Tissue culture incubator	Series II Water Jacketed CO2 Incubator	Thermo Scientific Forma, Rockville, MD, USA
Plate reader	Tecan Infinite F200 Pro	Tecan, San Jose, CA, USA
Agarose gel chamber	Subcell GT Electrophoresis system	Bio-Rad, Hercules, CA, USA
Polyacrylamide gel chamber	Criterion Vertical electrophoresis cell	Bio-Rad, Hercules, CA, USA
	2200 Tapestation	Agilent Technologies, Santa Clara, CA, USA
Sonicator	Covaris g-TUBE	Covaris, Woburn, MA, USA

Table 57: List of commercial kits used in this thesis

Description	Name	Manufacturer
DNA extraction kits	DNeasy Blood&Tissue Kit Wizard Genomic DNA Purification Kit QIAamp DNA minikit	QIAgen, Valencia, CA, USA Promega, Madison, WI, USA QIAgen, Valencia, CA, USA
Gel extraction kit	QIAquick Gel Extraction Kit	QIAgen, Valencia, CA, USA
PCR purification kit	QIAquick PCR purification Kit	QIAgen, Valencia, CA, USA
Cytotoxicity kit	CytoTox96® NonRadioactive Cytotoxicity kit	Promega, Madison, WI, USA
Library construction	PacBio SMRTbell 10kb Library preparation kit	Pacific Biosciences, Menlo Park, CA, USA
Biochemical diagnostic strips	API 20E diagnostic strips	bioMérieux, Durham, NC, USA
Melting curve analysis kit	LightCycler 480 High Resolution Melting Master	Roche, Indianapolis, IN, USA

Table 58: List of media and buffers used in this thesis

Name	Abbreviation	Ingredients	Supplier
Tris-Borate EDTA Buffer (5X)	TBE buffer	0.45 M Tris, 0.45 M boric acid, 0.01 M EDTA (pH = 8)	Thermo Scientific Forma, Rockville, MD, USA
Tris-EDTA Buffer (10X)	TE buffer	10 mM Tris,1 mM EDTA (pH 8.0)	Thermo Scientific Forma, Rockville, MD, USA
Cell suspension buffer	CSB	100 mM Tris,100 mM EDTA (pH 8.0)	Prepared before use
Cell Lysis Buffer	CLB	50 mM Tris, 50 mM EDTA (pH 8.0) + 1% Sarcosyl	Prepared before use
Tissue-culture phosphate buffered saline	TC-PBS	140.0 mmol NaCl, 2.6 mmol KCl, 0.9 mmol KH_2PO_4 and 6.4 mmol Na_2HPO_4 (pH 7.5)	Prepared before use
Lysis solution	LS	20 g NaOH, 87 g NaCl in 1 L of water	Prepared before use
Ammonium acetate	$C_2H_7NO_2$	154 g $C_2H_7NO_2$ in 1 L of water	Prepared before use
Alkaline peptone water	APW	10 g Bacto Peptone, 10 g NaCl in 1 L of water (pH 8.5)	Becton Dickinson, Franklin Lakes, NJ, USA
20X Standard Saline Citrate solution	20X SSC	175.4 g NaCl, 88.2 g $Na_3C_6H_5O_7$ •$2H_2O$ in 1 L of water	Prepared before use
1X Standard Saline Citrate/sodium dodecyl sulfate solution	1X SSC/SDS	50 mL 20X SSC, 10 g SDS in 1 L of water	Prepared before use

Table 57: *continued*

Name	Abbreviation	Ingredients	Supplier
3X Standard Saline Citrate/sodium dodecyl sulfate solution	3X SSC/SDS	150 mL 20X SSC, 10 g SDS in 1 L of water	Prepared before use
Ethanol-acetic acid (10x)	Eth-Ac	95 mL Ethanol mixed with 5 ml acetic acid	Prepared before use
Hybridazation buffer	HB	0.5 g BSA, 1.0 g $NaC_{12}H_{25}SO_4$, 0.5 g PVP, 25 mL 20X SSC in 1 L of water	Prepared before use
Triptic soy agar	TSA	15.0 g Bacto tryptone, 5.0 g Bacto soytone, 5.0 g NaCl, 15 g agar in 1 L of water (pH 7.3)	Becton Dickinson, Franklin Lakes, NJ, USA
Triptic soy broth	TSB	17.0 g Bacto tryptone, 3.0 g Bacto soytone, 2.5 g glucose, 5.0 g NaCl, 2.5 g K_2HPO_4 in 1 L of water (pH 7.3)	Becton Dickinson, Franklin Lakes, NJ, USA
Luria bertani	LB	10 g tryptone, 5 g yeast, 10 g sodium chloride in 1 L of water	Becton Dickinson, Franklin Lakes, NJ, USA
Thiosulfate citrate bile salts sucrose agar	TCBS	89 g of TCBS powder in 1 L of water (pH 8.6)	Becton Dickinson, Franklin Lakes, NJ, USA
Triptic soy agar with 10% sheep blood	TSA sheep	40 g TSA, 100 mL sheep blood to 1 L of water	Becton Dickinson, Franklin Lakes, NJ, USA

Table 57: *continued*

Name	Abbreviation	Ingredients	Supplier
Triptic soy broth with 30% glycerol	TSB glycerol	30 g TSB, 300 mL Glycerol to 1 L of water	Prepared before use
phosphate buffered saline	PBS	7.65 g NaCl, 0.724 g Na_2HPO_4 (anhydrous), 0.210 g KH_2PO_4 in 1 L of water (pH 7.4)	Prepared before use
Silver nitrate solution (100X)	$AgNO_3$	0.9 g of silver nitrate in 5 mL of water	Prepared before use
Tris-Acetat-EDTA buffer (50X)	TAE buffer	2M tris-acetate and 0.050M EDTA	Thermo Scientific Forma, Rockville, MD, USA

Table 59: List of software used in this thesis

Name	Description	Manufacturer
BioNumerics 6.6	Analysis of electrophoresis patterns and sequences	Applied Maths, Austin, TX, USA
CLC Genomics Workbench	Analysis of NGS data	QIAgen, Valencia, CA, USA
Discriminatory power calculator	Software to calculate the discriminatory power	http://insilico.ehu.es/mini_tools/discriminatory_power/index.php
Mega 6	Tool for sequence alignment and phylogenetic tree	Suhdir Kumar, Arizona State University, Tempe, AZ, USA
kSNP v2	Tool for SNP discovery and phylogenetics	Shea Gardner, Computations/Global Security, Lawrence Livermore National Laboratory, Livermore, CA, USA
FigTree	Tool for phylogenetic trees	http://tree.bio.ed.ac.uk/software/figtree/
Sigma Plot version 12.5.	Software for statistical analysis	Systat Software, Inc., San Jose, CA, USA

11.2.2. List of primers used

Table 60: Sequences of the *V. parahaemolyticus* hybridization probes

Probe	Sequence 5' – 3'
trh	ACTTTGCTTTCAGTTTGCTATTGGCT
tdh	GGTTCTATTCCAAGTAAAATGTATTTG
tlh	AAAGCGGATTATGCAGAAGCACTG

Table 61: T3SS primers and sequences used in this thesis

Name	Sequence 5'-3'	Reference
VP1669F	TACCGAGTTGCCAACGTG	Noriea et al. 2010
VP1669R	GATTGTTCCGCGATTTCTTG	Noriea et al. 2010
VP1670F	ACCGATTACTCAAGGCGATG	Noriea et al. 2010
VP1670R	TACGTTGTTGGCGTGATTGT	Noriea et al. 2010
VP1686F	CAAAAGCGATCACAAAAGCA	Noriea et al. 2010
VP1686R	AGCGACTTAACGGCATCATC	Noriea et al. 2010
VP1689F	AAGGTTGGCAAAAAGCGTTA	Noriea et al. 2010
VP1689R	GCTCTTCAACGAGCCAAGAG	Noriea et al. 2010
VP1694F	ACGATGCGACCAACAGTGTA	Noriea et al. 2010
VP1694R	TTTTAATTGCATCGGTGACG	Noriea et al. 2010
VPA1327F	TGGCGAAAGAGCCATTAGAT	Noriea et al. 2010
VPA1327R	TCAACTCCAAATTCGCCTTC	Noriea et al. 2010
VPA1335F	ATGTAACGGCGGCTAGCTTA	Noriea et al. 2010
VPA1335R	CAAACTGTGTCAGTAGCACCA	Noriea et al. 2010
VPA1339F	GATTCGCGGAACTCAAGAAG	Noriea et al. 2010
VPA1339R	CTTGTCCGAGATCAACGTCA	Noriea et al. 2010
VPA1346F	GGCTCTGATCTTCGTGAA	Noriea et al. 2010
VPA1346R	GATGTTTCAGGCAACTCTC	Noriea et al. 2010
VPA1354F	AATTGGCCGAGCCAACTTT	Noriea et al. 2010
VPA1354R	GTCATCCGGTTCTTGTGTAA	Noriea et al. 2010
VPA1362F	CTGCAGGTATCGCATCTTCA	Noriea et al. 2010
VPA1362R	TTAGAACCAACCGACGAAGC	Noriea et al. 2010
β_vscC2_F	GTACTTTGCTGTCTAACC	Okada et al. 2009
β_vscC2_R	CTTACTCTTAACTTCCGACG	Okada et al. 2009
β_vscS2_F	TTGATGTTGTTTCGGCTAGC	Okada et al. 2009
β_vscS2_R	CCACCGCCGAACTCGGCTAACAAG	Okada et al. 2009
β_vopB2_F	GAGCCTGTTGCTCTATGGAGCCAGG	Okada et al. 2009
β_vopB2_R	CGACACAGAACGCAATGCTTGCTCG	Okada et al. 2009
β_vopC_F	AACCAACTTGCGACTAAATC	Okada et al. 2009
β_vopC_R	TCCCGACAGTTTTTCTGCAC	Okada et al. 2009

Table 62: Real-time PCR primer and probe sequences utilized in the multiplex assay (MGBNFQ: minor groove binding nonfluorescent quencher; TxRED: Texas Red; BHQ2: black hole quencher 2).

Primer or probe	Sequence (5'to 3')	Modifications
tdh forward	TCCCTTTTCCTGCCCC	-
tdh reverse	CGCTGCCATTGTATAGTCTTTATC	-
tdh probe	TGACATCCTACATGACTGTG	5' FAM to 3' MGBNFQ
trh forward	TTGCTTTCAGTTTGCTATTGGCT	-
trh reverse	TGTTTACCGTCATATAGGCGCTT	-
trh probe	AGAAATACAACAATCAAAACTGA	5' TET to 3' MGBNFQ
tlh forward	ACTCAACACAAGAAGAGATCGACAA	-
tlh reverse	GATGAGCGGTTGATGTCCAA	-
tlh probe	CGCTCGCGTTCACGAAACCGT	5' TxRED to 3' BHQ2
IAC forward	GACATCGATATGGGTGCCG	-
IAC reverse	CGAGACGATGCAGCCATTC	-
IAC probe	TCTCATGCGTCTCCCTGGTGAATGTG	5' Cy5 to 3' BHQ2

Table 63: MLVA primer sequences and their product sizes used in this thesis

Multiplex-PCR	Locus	Primer	5′-3′ Sequence	Pro-duct length [bp]	Melting temperature Tm [◦C]	Refer-ences
Multi A	VP1-11	VP1-11 F	CTGCCTGGAGAATTGGCTTA	854	95	(Harth-Chu et al., 2009)
		VP1-11 R	TGAGCCTGAAGCTGAAAACA			
	VP2-03	VP2-03 F	CATAAACGAGCGACACGAGA	168	57	
		VP2-03 R	GCGCAAAAATTCATTGTGATT			
	VPTR5	VPTR5 F	GCTGGATTGCTGCGAGTAAGA	204	82	(Kimura et al., 2008)
		VPTR5 R	AACTCAAGGGCTGCTTCGG			
	VPTR7	VPTR7-1F	TATCTACAAAGGTGGCGGAGAT	200	80	(Harth-Chu et al., 2009)
		VPTR7-1R	AAGGTGTTACTTGTTCCAGACG			
Multi B	VP1-17	VP1-17 F	TCAACACGAGCTTGATCACC	206	69	
		VP1-17 R	GAAATCCGGAGTACCTGCAA			
	VP1-10	VP1-10 F	CGTCTTGCTCGTGAACGTAA	955	94	
		VP1-10 R	TCATTAAGTCAGGCGTGCTG			
	VPTR1	VPTR1 F	TAACAACGCAAGCTTGCAACG	253	54	(Kimura et al., 2008)
		VPTR1 R	TCATTCTCGCCACATAACTCAGC			
	VPTR8	VPTR8 F	ACATCGGCAATGAGCAGTTG	301	89	
		VPTR8 R	AAGAGGTTGCTGAGCAAGCG			
Multi C	VP2-07	VP2-07 F	TGATTTTGAAGCAGCGAAGA	296	98, smaller peak at 74	(Harth-Chu et al., 2009)
		VP2-07 R	TTTGTGACTGCTGTCCTTGC			
	VPTR3	VPTR3 F	CGCCAGTAATTCGACTCATGC	331	77	(Kimura et al., 2008)
		VPTR3 R	AAGACTGTTCCCGTCGCTGA			
	VPTR4	VPTR4 F	AAACGTCTCGACATCTGGATCA	227	85	
		VPTR4 R	TGTTTGGCTATGTAACCGCTCA			
	VPTR6	VPTR6 F	TGTCGATGGTGTTCTGTTCCA	316	107, smaller peak 97, 72	
		VPTR6 R	CTTGACTTGCTCGCTCAGGAG			

11.2.3. Protocols

11.2.3.1. Storage of bacterial cultures

Materials

Reagents:
TSA
TSB broth containing 30% glycerol
Equipment:
UV-Vis Spectrometer (wavelength 600 nm)
Cuvettes

Protocol:

1. A single bacterial colony was streaked on to TSA media.
2. The plate culture was incubated overnight at 35°C.
3. Bacterial colonies were transferred with a cotton swab into TSB broth containing 30% glycerol.
4. The cultures were stored at -80°C.

11.2.3.2. Determining bacterial growth

Materials

Reagents:
TSB broth
Equipment:
UV-Vis Spectrometer (wavelength 600 nm)
Cuvettes

Protocol:

1. For growth determination 1 mL of culture was transferred to a cuvette.
2. The OD_{600} was measured in the UV/Vis spectrometer.

11.2.3.3. Biochemical characterization

Materials

Reagents:
Sterile destilled water
Equipment:
Sterile pasteur pipettes
Biomerieux API 20 NE kit contents:
Sterile oil
ONPG: test for β-galactosidase enzyme by hydrolysis of the substrate o-nitrophenyl-b-D-galactopyranoside

ADH: decarboxylation of the amino acid arginine by arginine dihydrolase
LDC: decarboxylations of the amino acid lysine by lysine decarboxylase
ODC: decarboxylations of the amino acid ornithine by ornithine decarboxylase
CIT: utilization of citrate as only carbon source
H2S: production of hydrogen sulfide
URE: test for the enzyme urease
TDA (Tryptophan deaminase): detection of the enzyme tryptophan deaminase: Reagent to put- Ferric Chloride.
IND: Indole Test-production of indole from tryptophan by the enzyme tryptophanase . Reagent- *Indole is detected by addition of Kovac's reagent*
VP: the Voges-Proskauer test for the detection of acetoin (acetyl methylcarbinol) produced by fermentation of glucose by bacteria utilizing the butylene glycol pathway
GEL: test for the production of the enzyme gelatinase which liquefies gelatin
GLU: fermentation of glucose (hexose sugar)
MAN: fermentation of mannose (hexose sugar)
INO: fermentation of inositol (cyclic polyalcohol)
SOR: fermentation of sorbitol (alcohol sugar)
RHA: fermentation of rhamnose (methyl pentose sugar)
SAC: fermentation of sucrose (disaccharide)
MEL: fermentation of melibiose (disaccharide)
AMY: fermentation of amygdalin (glycoside)
ARA: fermentation of arabinose (pentose sugar)
Ferric chloride
Kovacs reagent
40 % KOH (VP reagent 1)
α- Naphthol (VP Reagent 2)

Protocol:

1. Several isolated colonies were used to make a suspension in 0.85% saline.
2. API20E Biochemical Test Strip is commercially available, with 20 separate compartments.. (Bacteria will react with reagents in each compartment and

will give different colors which will help to identify bacteria to the species level).

3. The compartments of the API20E Biochemical Test Strip which contains dehydrated bacterial media/bio-chemical reagents were filled up with a pasteur pipette holding the bacterial suspension.
4. Sterile oil was added into the ADH, LDC, ODC, H2S and URE compartments.
5. Some drops of water were added to the tray of the API Test strip and was closed for the incubation at 37oC for 18 to 24 hours.
6. For some of the compartments, the change in color ist he result after 24 hours, but for some reagents have to added before reading.
7. The following reagents were added to these specific compartments:
8. TDA: One drop of Ferric Chloride was added.
    i. IND: One drop of Kovacs reagent was added.
    ii. VP: One drop of VP reagent 1 and one drop of VP Reagent 2 were added.
9. The results were compared to the API color chart. Positive results add up to a score code. The number combination led to the organism identification at apiweb.

11.2.3.4. DNA extraction QIAgen

Materials

Reagents:
Ethanol (100%)
RNase A (100 mL)
Primer TE
QIAgen DNeasy Blood & Tissue kit contents:
Buffer ATL
Buffer AL
Buffer AW1 (concentrated)
Buffer AW2 (concentrated)

Buffer AE
Proteinase K
DNeasy Mini Spin Columns in 2 mL Collection Tubes
Collection Tubes (2 mL)

Protocol:

1. Cells were harvested in a microfuge tube by centrifuging for 10 min at 5000 xg (7500 rpm), the supernatant was discarded.
2. The pellet was resuspended in 180 µL Buffer ATL.
3. The cell suspension was mixed with 20 µL of Proteinase K by vortexing and incubated at 56˚C until complete lysis.
4. (*for PGM Ion Torrent use*) For RNA-free genomic DNA 4 µL of RNase A were added, mixed by vortexing and incubated for 2 minutes.
5. The solution was vortexed for 15 seconds and 200 µL of buffer AL were added. Again the sample was mixed by vortexing and 200 µL ethanol were pipetted to the solution.
6. The mixture was transferred to a DNeasy mini spin column in a 2 mL collection tube and centrifuged for 1 minute at 6000 xg. The flow-through was discarded.
7. To the spin column 500 µL of buffer AW1 were added and the tube was centrifuged for 1 minute at 6000 xg. Again the flow-through was discarded.
8. The same procedure with AW2, 500 µL were added to the column and the tube was centrifuged for 3 minutes at 20000 xg. Again the flow-through was discarded.
9. The spin column was placed in a new 1.5 mL microcentrifuge tube and 200 µL of AE buffer were added (For sequencing purposes the DNA was eluted with 50 µL low TE). The mixture was incubated for 1 minute and and centrifuged for 1 minute at 6000 xg

11.2.3.5. DNA extraction Promega

Materials

Reagents:
Ethanol (70%)
Isopropanol
Primer TE
Equipment:
Microcentrifuge tubes (1.5 mL)
Water bath (80°C)
Water bath (37°C)
Promega Wizard Genomic DNA purification kit contents:
Cell lysis solution
Nuclei lysis solution
Protein precipitation solution
DNA rehydration solution
RNase A solution

Protocol:

1. One ml of an overnight culture was added to a 1.5ml microcentrifuge tube and centrifuged at 13,000–16,000 ×*g* for 2 minutes to pellet the cells. Afterwards the supernatant was removed.
2. The sample was incubated at 80°C for 5 minutes to lyse the cells; then it was cooled to room temperature.
3. The cell lysate was mixed with 3µl of RNase Solution. The tube was inverted 2–5 times to mix.
4. The sample was incubated at 37°C for 15–60 minutes. Afterwards the sample was cooled to room temperature.
5. To the RNase-treated cell lysate 200µl of Protein Precipitation Solution were added and vortexed vigorously at high speed for 20 seconds to mix the Protein Precipitation Solution with the cell lysate.
6. The sample was incubated on ice for 5 minutes.
7. The sample tube was centrifuged at 13,000–16,000 ×*g* for 3 minutes.

8. The supernatant containing the DNA was transferred to a clean 1.5 ml microcentrifuge tube containing 600µl of room temperature isopropanol.
9. The sample solution was gently mixed by inversion until the thread-like strands of DNA formed a visible mass.
10. The sample was centrifuge at 13,000–16,000 × *g* for 2 minutes.
11. The supernatant was carefully poured off. The tube was drained on clean absorbent paper and 600µl of room temperature 70% ethanol were added. The tube was gently inverted several times to wash the DNA pellet.
12. The sample was centrifuged at 13,000–16,000 × *g* for 2 minutes and the ethanol carefully aspirated.
13. The tube was drained on clean absorbent paper to allow the pellet to air-dry for 10–15 minutes.
14. Onehundred µl of DNA Rehydration Solution were added to the tube and incubated at 65°C for 1 hour. The solution was periodically mixed by gently tapping the tube.
15. The DNA was stored at 2–8°C.

11.2.3.6. Gel extraction

Materials

Reagents:
Ethanol (100%)
Isopropanol (100%)
QIAquick gel extraction kit contents:
Buffer QG
Buffer PE (concentrate)
Buffer EB
Loading dye
Mini Spin Columns in 2 mL Collection Tubes
Collection Tubes (2 mL)

Protocol:

1. The DNA fragment was excised from the agarose gel with a clean, sharp scalpel.
2. The gel slice was weighed in a colorless tube and 3 volumes Buffer QG were added to 1 volume gel.
3. The gel slice was incubated at 50 °C for 10 min (or until the gel slice had completely dissolved). The tube was vortexed every 2–3 min to help dissolve the gel.
4. After the gel slice has dissolved completely,1 gel volume of isopropanol was added to the sample and mixed.
5. A QIAquick spin column was placed in a provided 2 ml collection tube. To bind DNA, the sample was applied to the QIAquick column and centrifuged for 1 min until all the samples have passed through the column.
6. The flow-through was discarded and the QIAquick column was placed back into the same tube.
7. To wash, 750 µl Buffer PE were added to the QIAquick column and centrifuged for 1 min. The flow-through was discarded and the QIAquick column was placed back into the same tube.
8. The QIAquick column was centrifuged in the provided 2 ml collection tube for 1 min to remove residual wash buffer.
9. The QIAquick column was placed into a clean 1.5 ml microcentrifuge tube.
10. To elute DNA, 50 µl Buffer EB were added to the center of the QIAquick membrane and centrifuged the column for 1 min.

11.2.3.7. PCR product purification

Materials

Reagents:

Ethanol (100%)
3 M sodium acetate solution (pH 5)
QIAquick PCR purification kit contents:
Buffer PB
Buffer PE
Buffer EB
pH indicator I
Loading dye
Mini Spin Columns in 2 mL Collection Tubes
Collection Tubes (2 mL)

Protocol:

1. Five volumes of Buffer PB were added to 1 volume of the amplified reaction mixture and mixed.
2. To bind DNA, the sample was applied to a QIAquick spin column and centrifuged at 17,900 *xg* for 30–60 s.
3. The flow-through was discarded. The QIAquick column was placed back into the same collection tube.
4. To wash, 0.75 ml Buffer PE were added to the QIAquick column and centrifuged at 17,900 *xg* for 30–60 s.
5. The flow-through was discarded and the QIAquick column was placed back in the same collection tube. The column was centrifuged at 17,900 *xg* for an additional 1 min.
6. A QIAquick column was placed in a clean 1.5 ml microcentrifuge tube.
7. To elute DNA 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or water (pH 7.0–8.5) were added to the center of the QIAquick membrane and the column was centrifuged at 17,900 *xg* for 1 min.

11.2.3.8. Serology

Materials

Reagents:
3% NaCl solution
O and K antisera kit contents:
K antigens
O antigens

Protocol:

1. O-antigen typing:
    a. Prepare heavy colony suspension in 3% NaCl solution and autoclave at 121°C for 1 h.
    b. Add 1 µL loopful of the autoclaved suspension to one drop of each O antiserum and test for agglutination.
2. K antigen typing:
    a. Mix 1 µL loopful of plate growth into one drop of each pool of K antisera and test for agglutination.
    b. If an isolate agglutinated in any K pool, the individual antisera of that pool were tested.

11.2.3.9. Protocol DNA Probe Colony Hybridization for Confirmation of *V. parahaemolyticus* Isolates

Materials

Reagents:
T_1N_3
Lysis solution
Ammonium acetate solution
1X SSC buffer
Proteinase K solution
Hybridization buffer
1X SSC/1% SDS

3X SSC/1% SDS
NBT/BCIP solution

Equipment:

#541 Whatman filters
Petri dishes
Microwave
Wash containers
Shaker
Shaking water bath
Whirl-Pak bag

Protocol:

1. Isolates were transferred from the 96-well plate to a T_1N_3 plate using a multi prong replicator. The plate was incubated overnight at 35°C.
2. The control strips were prepared using *V. vulnificus*, a *tdh*- strain of *V. parahaemolyticus*, and a *tdh*+ strain of *V. parahaemolyticus*. These strains were spot inoculated in multiple lines on a T_1N_3 plate and incubated overnight. These filters were treated equally as sample filters.
3. Whatman filters were labeled with sample number, date, analyst initials, and probe to be hybridized with (*tlh, tdh,* or *trh*).
4. Each filter was placed label-side down on appropriate T_1N_3 plate. Afterwards each filter was transferred colony-side up to a petri dish lid containing 1 ml of lysis solution.
5. The filters were microwaved in petri dishes for 15-20 seconds/filter.
6. The filters were transferred to a plastic wash container (up to 30 filters were combined in one container) and neutralized with ammonium acetate (4 ml/filter) for 5 minutes on shaker at room temperature.
7. The ammonium acetate was poured off and filters rinsed two times with 1X SSC buffer (10 mL/filter), for two minutes each time.

8. One control strip for each set of five filters was added to the reaction to be hybridized with one probe.
9. The filters (up to 30) were placed in a plastic wash container of Proteinase K/SSC solution. The mixture was incubated for 30 min in a 42°C water bath.
10. The Proteinase K solution was decanted. Filters were rinsed three times in 1X SSC (10 ml/filter) for 10 minutes with shaking at 50 rpm.
11. Up to five proK-treated filters were placed in a Whirl-Pak bag, combined with 10 ml of pre-warmed hybridization buffer the bag was closed to exclude air. Filters were incubated for 30 minutes at 54°C in a shaking (50 rpm) water bath.
12. The buffer was poured from the bag and 10 ml fresh pre-warmed buffer/bag were added. Afterwards the probe (final conc. is 0.5 pmol/ml) was added to the bag with filters. Resealed the bags were incubated for one hour in a 54°C water bath with shaking.
13. Filters were placed in plastic wash container, 10 ml/filter 1X SSC/1% SDS (for *tlh* and *trh*) or 3X SSC/1% SDS (for *tdh*) were added and incubated in a 54°C water bath with shaking for 10 minutes. This step was repeated for second time.
14. Filters were rinsed five times for 5 minutes each in 1X SSC at room temperature on an orbital shaker (100 rpm).
15. In petri dish, 20 ml of NBT/BCIP solution were added. Up to five filters were transferred to the dish and incubated with gentle shaking at 35°C away from light.
16. After incubating filters were transferred to a plastic wash container and rinsed with water (10 ml/filter) with shaking for 10 minutes. The rinse was repeated twice to stop color development. Purple or brown spots were considered positive.

11.2.3.10. MasterMix set-up for *tdh*, *trh*, and *tlh* Multiplex-PCR using a SmartCycler

Materials

Equipment:

Smart Cycler real-time PCR

Eppendorf pipettes

Protocol:

1. The following reagents were added for the master mix:

Table 64: Master mix for *tdh*, *trh*, and *tlh* Multiplex-PCR using a SmartCycler

Component	Units	Stock concentration	Final volume	Volume [µl] /Reaction
PCR H_2O				10.68
Buffer	X	10	1.000	2.50
$MgCl_2$	mM	50	5.000	2.50
dNTPs (equal conc. of each)	mM	10	0.400	1.00
Forward Primer *tlh* 884F	µM	10	0.075	0.188
Reverse Primer *tlh* 1091R	µM	10	0.075	0.188
Forward Primer *trh* 20F	µM	10	0.200	0.500
Reverse Primer *trh* 292R	µM	10	0.200	0.500
Forward Primer *tdh* 89F	µM	10	0.200	0.500
Reverse Primer *tdh* 321R	µM	10	0,200	0.500
Forward Primer IC 46F	µM	10	0.075	0.188
Reverse Primer IC 186R	µM	10	0.075	0.188
Probe *tlh* TX Red	µM	10	0.150	0.375
Probe *trh* TET	µM	10	0.075	0.188
Probe *tdh* FAM	µM	10	0.075	0.188
Probe IC Cy5	µM	10	0.150	0.375
Internal Control DNA	µl			2.000
Platinum *Taq*	Units/ µl	5	2.25	0.45
Target template	µl			2.00
Volume of master mix (µl) in each tube		23.00		

2. 23.0 µl of the master mix were combined with 2.0 µl of the sample.

3. The follwing temperate program was used:

Table 65: Temperature program for *tdh*, *trh*, and *tlh* Multiplex-PCR using a SmartCycler

	Initial denaturation	45 cycles	Denaturation	Annealing/ Extension	
Temperature [°C]	95		95	59	-
Time (s)	60		5	45	-

11.2.3.11. Master mix set-up for the T3SS Multiplex-PCR

Materials

Equipment:
peqStar thermocycler
Eppendorf pipettes

Protocol:

1. The following reagents were added for the master mix:

Table 66: Master mix set-up for the T3SS1 Multiplex-PCR

Component	Units	Stock concentration	Final volume [µl]	Volume [µl] /Reaction
PCR-graded H_2O				9.90
Buffer (includes $MgCl_2$)	X	10	1.000	3.00
$MgCl_2$	mM	50	2.000	1.20
dNTPs	mM	10	0.200	0.60
Forward primer 1670F	µM	10	0.500	1.50
Reverse primer 1670R	µM	10	0.500	1.50
Forward primer 1686F	µM	10	0.500	1.50
Reverse primer 1686R	µM	10	0.500	1.50
Forward primer 1689F	µM	10	0.500	1.50
Reverse primer 1689R	µM	10	0.500	1.50
Forward primer 1694F	µM	10	0.500	1.50
Reverse primer 1694R	µM	10	0.500	1.50
Roche *Taq* polymerase	Units/ µl	5	1.500	0.30
Volume of master mix (µl) in each tube		27		
Target template	µl			3.00

Table 67: Master mix set-up for the T3SS2α PCR

Component	Units	Stock concentration	Final volume [µl]	Volume [µl] /Reaction
PCR-graded H_2O				9.90
Buffer (includes $MgCl_2$)	X	10	1.00	3.00
$MgCl_2$	mM	50	2.00	1.20
dNTPs	mM	10	0.200	0.60
Forward primer 1327F	µM	10	0.500	1.50
Reverse primer 1327R	µM	10	0.500	1.50
Forward primer 1335F	µM	10	0.500	1.50
Reverse primer 1335R	µM	10	0.500	1.50
Forward primer 1339F	µM	10	0.500	1.50
Reverse primer 1339R	µM	10	0.500	1.50
Forward primer 1362F	µM	10	0.500	1.50
Reverse primer 1362R	µM	10	0.500	1.50
Roche *Taq* polymerase	Units/ µl	5	1.500	0.30
Volume of master mix (µl) in each tube		27		
Target template	µl			3.00

Table 68: Master mix set-up for the T3SS2β PCR

Component	Units	Stock concentration	Final volume [µl]	Volume [µl] /Reaction
PCR-graded H_2O				9.60
Buffer (includes $MgCl_2$)	X	10	1	3.00
$MgCl_2$	mM	50	2	1.20
dNTPs	mM	10	0.300	0.90
Forward primer vscC2 F	µM	10	0.500	1.50
Reverse primer vsc C2 R	µM	10	0.500	1.50
Forward primer vscS2 F	µM	10	0.500	1.50
Reverse primer vscS2 R	µM	10	0.500	1.50
Forward primer vopB2 F	µM	10	0.500	1.50
Reverse primer vopB2 R	µM	10	0.500	1.50
Forward primer vopC F	µM	10	0.500	1.50
Reverse primer vopC R	µM	10	0.500	1.50
Roche *Taq* polymerase	Units/ µl	5	1.5	0.30
Volume of master mix (µl) in each tube		27		
Target template	µl			3.00

2. 27.0 µl of the master mix were combined with 3.0 µl of the sample.
3. The follwing temperature program was used:

Table 69: Temperature program for the T3SS1 and T3SS2α PCR

	Initial denaturation		Denaturation	Annealing	Extension	Final extension
Temperature [°C]	94	33 cycles	94	60	72	72
Time (s)	120		40	45	40	420

Table 70: Temperature program for the T3SS2β PCR

	Initial denaturation		Denaturation	Annealing	Extension	Final extension
Temperature [°C]	94	33 cycles	94	55	72	72
Time (s)	120		30	30	120	420

4. The PCR products were loaded on to a 2% agarose gel containing 0.2% ethidium bromide and were electrophorized for 1.5 hours at 120V.

11.2.3.12. PFGE protocol

Materials

Reagents:
TSA with sheep blood
Restriction enzyme *Not*I
Restriction enzyme *Sfi*I
Restriction enzyme *Xba*I
10X restriction buffer
Proteinase K solution (20 mg/mL)
BSA (10 mg/mL)
SeaKem Gold agarose
TE buffer
Cell suspension buffer (CSB)
Cell lysis buffer
TBE buffer
Ultrapure reagent water
Equipment:
PFGE equipment
Waterbath

50 mL Falcon tubes
Spectrophotometer
PFGE plug molds
Shaking incubator
Microfuge
Cotton swabs

Protocol:

1. An isolated colony was streaked from test cultures onto TSA with 5% defibrinated sheep blood (TSA-sheep) plates. Cultures were incubated at 35°C for 14-18 hours.
2. For plug preparation TE buffer and 0.5% agarose gel solution in TE buffer were assembled.
3. Two ml of CSB were transferred to small labeled tubes. Using a sterile cotton swab some of the growth from agar plate was removed and suspended in CSB.
4. The optical density of the cell suspensions was measured with a spectrophotometer at wavelength 610 nm. The concentrations were adjusted to a value of 0.45 by diluting with sterile CSB or by adding additional cells.
5. 400 µl of the adjusted cell suspensions were transferred to labeled 1.5-ml microcentrifuge tubes and 20 µl of Proteinase K solution were added to each tube and mix gently with pipet tip.
6. 400 µl melted 1% SeaKem Gold agarose were added to 400 µl cell suspension and mixed by gently pipetting the mixture up and down.
7. Immediately, part of mixture was pipetted into the equivalent well of the plug mold. Two plugs of each sample can be made from these amounts of cell suspension and agarose. Plugs needed to solidify at room temperature for 10-15 minutes.

8. In a 50 mL Falcon tube 5 mL of cell lysis buffe and 25 µL of Proteinase K were combined. The plugs were transferred from the plug mold with a spatula to appropriately labeled tube. The tubes were incubated in a 54-55°C shaking incubator for 1.5 - 2 hours with constant and vigorous agitation (150-175 rpm).
9. Sterile ultrapure water was pre-heated to 54-55°C for plug washing. The lysis buffer was poured off and the plug was washed two times with 10 – 15 mL of ultrapure water. At the same time TE buffer was pre-heated to 54-55°C for plug washing. Then the plugs were washed four times with TE buffer. Afterwards the plugs were stored in TE buffer at 4°C.
10. For the restriction enzyme digest 1.5 mL microfuge tubes were used. Plugs made from *Salmonella ser. Braenderup H9812* were utilized as control plugs.
11. To pre-incubate the plugs the appropriate 10X restriction buffer was diluted 1:10 with sterile ultrapure water and 200µl diluted restriction buffer (1X) were added to a 1.5-mL microcentrifuge tube. A 2.0 to 2.5 mm wide slice was cut from each test sample plug and the reference strain plug. The plugs slice was transferred to tube containing diluted restriction buffer. The sample and control plug slices were incubated in a water bath for 5-10 minutes (digested with SfiI at 50°C, digested with NotI and XbaI at 37°C). After incubation, the buffer was removed from the plug slice.
12. The restriction enzyme master mix was prepared according to the following table:

Table 71: Enzyme master mix for PFGE digest

Reagent	µL/plug slice	µL/10 plug slices	µL/15 plug slices
Ultrapure water	177	1770	2655
10X restriction buffer	20	200	300
BSA (10 mg/mL)	2	20	30
*Sfi*I or *Not*I	1	10	15
Total Volume	200	2000	3000

13. The plug slice was incubated in 200 µl restriction enzyme master mix for 4 hours in the appropriate waterbath (*Sfi*I at 50°C, *Not*I and *Xba*I at 37°C).
14. A 0.5X TBE dilution was prepared for both the gel and electrophoresis running buffer. A 1% SeaKem Gold agarose gel in 0.5X TBE was used for electrophoresis.
15. The plug slices were loaded on the bottom of the comb teeth and residual buffer was removed with a kimwipe.
16. The comb was positioned in the gel form; plug slices needed to align with the black platform of the gel cast. The still liquid agarose was poured carefully into the gel form. The comb was removed after the gel solidified.
17. The gel frame was transferred to the electrophoresis chamber containing 2 - 2.2 L freshly prepared 0.5X TBE.
18. Following electrophoresis conditions were used on CHEF DR-III:
    - Initial switch time: 10 s
    - Final switch time: 35 s
    - Voltage: 6V
    - Included Angle: 120°
    - Run time: 18-19 hours
19. After the electrophoresis the gel was stained with ethidium bromide for 20-30 minutes in covered container. Afterwards, the gel was destained in approximately 500 ml reagent grade water for 60-90 minutes. The gel image was then documented using an imaging system.
20. The gel image was analyzed using the BioNumerics software.

11.2.3.13. DGREA protocol

Materials

Reagents:
TSB plus 1% NaCl

Promega Wizard Genomic DNA Purification Kit
Restriction enzyme *Nae*I
Proteinase K solution
BSA
PCR-graded water
Pre-made 8% nondenaturating polyacrylamide gels
Equipment:
NanoDrop 1000
Heat block
PAGE electrophoresis chamber

Protocol:

1. Each isolate was inoculated in TSB plus 1% NaCl and incubated with shaking overnight at 35°C.
2. Two mL of the culture were used for DNA extraction with the Promega Wizard Genomic DNA Purification Kit (see protocol 11.2.3.5); DNA was suspended in 50 µL elution buffer.
3. DNA concentration of extracts was determined using a NanoDrop 1000.
4. 10 µg of DNA were digested with 10U of *Nae*I in a 20 µL reaction with PCR-graded water for 2 h at 37°C according to the following enzyme mix:

Table 72: Enzyme master mix for DGREA digest

Component	Stock concentration	Final volume [µl]	Volume [µl] /Reaction
Buffer	10	1	2.00
BSA			0.20
NaeI			1.00

5. Each reaction mixture was then treated with 0.0020 µg/µL Proteinase K for 1h at 37°C.
6. Resultant digest products (10 µL) were separated using an 8% nondenaturating polyacrylamide gel and electrophoresed for 3 h at 100 V. Bands were visualized by silver staining.

11.2.3.14. Silver staining protocol

Materials

Reagents:
Ethanol-acetic acid (1x dilution in water)
Silver nitrate (100x)
Formaldehyde (37%)
Sodium hydroxide (7.5%)
PCR-graded water
Equipment:
Microwave

Protocol:

1. The gel was covered with 50 ml of Eth-Ac (1x), microwaved for 20 seconds, and afterwards Eth-Ac was discarded.
2. Gels were covered with 50 ml $AgNO_3$ solution (1x), microwaved for 20 seconds, and afterwards $AgNO_3$ solution was discarded..
3. The gels were washed twice (10 seconds each) demineralized water.
4. The gels were developed in a mix of 0.5 ml formaldehyde 37%, 20 ml of NaOH 7.5% and 30 ml of demineralized water. The mixture was microwaved for 20 seconds and sat until the bands started to appear.
5. Afterwards the solution was discarded and Eth-Ac added for the fixing step.

11.2.3.15. ISR-1 typing protocol

Materials

Reagents:
PCR-graded water
Primer 16S.6
Primer 23S.1
LB plus 1% NaCl
PCR reagents

Equipment:

peqSTAR thermocycler
PAGE electrophoresis chamber

Protocol:

1. Each isolate was inoculated into LB plus 1% NaCl and incubated with shaking overnight at 35°C.
2. One milliliter of the overnight culture was removed, heated to 100°C for 10 minutes, and plunged into ice for 5 minutes.
3. The master mix was prepared as follows:

Table 73: Master mix for ISR-1 typing PCR

Component	Units	Stock concentration	Final volume [µl]	Volume [µl] /Reaction
PCR-graded H_2O				21.22
Buffer (containing $MgCl_2$)	X	10	1.0	3.00
$MgCl_2$	mM	25	1.5	
dNTPs	mM	10	0.125	0.38
Primer 16S.6	µM	10	0.500	0.60
Primer 23S.1	µM	10	0.500	0.60
Taq DNA polymerase	Units/µl	5	1	0.20
Master mix (µl) to each tube				26.00
Target template	µl			4.00

4. The following temperature program was used:

Table 74: Temperature program for ISR-1 typing PCR

	Initial denaturation		Denaturation	Annealing	Extension	Final extension
Temperature [°C]	95	30 cycles	95	58	72	72
Time (s)	60		60	60	60	600

5. The PCR products (1:5 dilutions) were subjected to a heteroduplex resolution reaction, which consisted of:

Table 75: Master mix for heteroduplex reaction

Component	Units	Stock concentration	Final volume [µl]	Volume [µl] /Reaction
PCR-graded H_2O				7.80
Buffer (containing $MgCl_2$)	X	10	1.0	3.00
$MgCl_2$	mM	50	1.5	0.30
dNTPs	mM	10	0.200	1.20
Primer 16S.6	µM	10	0.600	1.20
Primer 23S.1	µM	10	0.600	1.20
Taq DNA polymerase	Units/µl	5	0.05	0.30
Master Mix (µl) to each tube				15.00
Target template	µl			15.00

6. The PCR program for heteroduplex resolution was the same as for the initial amplification, except only one cycle was conducted.
7. Resultant digest products (10 µL) were separated using an 8% nondenaturating polyacrylamide gel and electrophoresed for 3 h at 100 V. Bands were visualized by silver staining.

11.2.3.16. Calculation of discriminatory power

Materials

Software:

Online discriminatory power calculator (Bikandi and San Millan, 2012)

Protocol:

1. The discriminatory power was calculated as previously described by Hunter and Gaston using the following formula (Hunter and Gaston, 1988):

$$D = 1 - \frac{1}{N\,(N-1)} \sum_{j=1}^{s} x_{j(x_j-1)}$$

D = discriminatory power

N = number of unrelated strains tested

s = number of different types

x_j = number of strains belonging to the j-type

11.2.3.17. Fingerprint analysis

Materials

Software:

BioNumerics software 5.1. and 6.6.

Protocol:

1. ISR-1 and DGREA fingerprints were analyzed with BioNumerics software 6.6. The dendrograms were formed using the Dice correlation with 0.5% optimization and 1% tolerance
2. PFGE fingerprints were analyzed with BioNumerics software 5.1. The dendrograms were constructed using the Dice correlation with 1.5% optimization and 1.5% tolerance. Dendrograms containing *Sfi*I and *Not*I combined comparisons were constructed using the average from experiments and the UPGMA.
3. MLVA melt curves were analyzed using BioNumerics software 6.6 with a customized script. This script compared the melting curves of each multiplex PCR, as well as a combination of all curves from the three multiplex PCRs. The combined dendrogram of all three multiplex PCR was built based on the Pearson correlation of average trend curves.

11.2.3.18. MLVA protocol

Materials

Reagents:

PCR-graded water
High resolution melting master kit
Primers
PCR reagents (table below)

High resolution master kit contents:

Master mix, 2x concentrated, including high resolution melting dye

$MgCl_2$ stock solution, 25 mM
PCR-graded H_2O
Equipment:
Roche Lightcycler
Eppendorf pipettes

Protocol:

1. Each isolate was inoculated into LB plus 1% NaCl and incubated with shaking overnight at 35°C.
2. One milliliter of the overnight culture was removed, heated to 100°C for 10 minutes, and plunged into ice for 5 minutes.
3. The master mixes were prepared as follows:

Table 76: Master mix set-up for MultiA Multiplex-PCR

Component	Units	Stock concentration	Final volume [µl]	Volume [µl] /Reaction
PCR-graded H_2O				0.20
Master Mix	X	2	1	10.00
$MgCl_2$	mM	25	2	1.60
Forward Primer VPTR-7	µM	10	0.200	0.400
Reverse Primer VPTR-7	µM	10	0.200	0.400
Forward Primer VPTR-5	µM	10	0.200	0.400
Reverse Primer VPTR-5	µM	10	0.200	0.400
Forward Primer VP1-11	µM	10	0.200	0.400
Reverse Primer VP1-11	µM	10	0.200	0.400
Forward Primer VPTR2	µM	10	0.200	0.400
Reverse Primer VPTR2	µM	10	0.200	0.400
Target template	µl			5.00
Master Mix (µl) to each tube		15.00		20.00

Table 77: Master mix set-up for MultiB Multiplex-PCR

Component	Units	Stock concentration	Final volume [µl]	Volume [µl] /Reaction
PCR-graded H_2O				0.20
Master Mix	X	2	1	10.00
$MgCl_2$	mM	25	2	1.60
Forward Primer VP1-17	µM	10	0.200	0.400
Reverse Primer VP1-17	µM	10	0.200	0.400
Forward Primer VPTR-1	µM	10	0.200	0.400
Reverse Primer VPTR-1	µM	10	0.200	0.400
Forward Primer VPTR8	µM	10	0.200	0.400
Reverse Primer VPTR8	µM	10	0.200	0.400
Forward Primer VP1-10	µM	10	0.200	0.400
Reverse Primer VP1-10	µM	10	0.200	0.400
Target template	µl			5.00
Master Mix (µl) to each tube		15.00		20.00

Table 78: Master mix set-up for MultiC Multiplex-PCR

Component	Units	Stock concentration	Final volume [µl]	Volume [µl] /Reaction
PCR-graded H_2O				0.20
Master Mix	X	2	1	10.00
$MgCl_2$	mM	25	2	1.60
Forward Primer V2-07	µM	10	0.400	1.000
Reverse Primer V2-07	µM	10	0.400	1.000
Forward Primer VPTR-4	µM	10	0.200	0.400
Reverse Primer VPTR-4	µM	10	0.200	0.400
Forward Primer VPTR-3	µM	10	0.200	0.400
Reverse Primer VPTR-3	µM	10	0.200	0.400
Forward Primer VPTR6	µM	10	0.200	0.400
Reverse Primer VPTR6	µM	10	0.200	0.400
Target template	µl			5.00
Master Mix (µl) to each tube		15.00		20.00

4. The following temperature program was used for the LightCycler:

Table 79: Temperature program for Multi A and B

	Initial denat.	20 cycles:	Denat.	Ann.	Ext.	10 cycles:	Denat.	Ann.	Ext.	Final ext.
Temperature [°C]	95		94	63 decrease 0.2°C/cycle	72		94	59	72	60
Time (s)	900		30	90	60		30	90	60	1800

Table 80: Temperature program of Multi C

	Initial denat.	30 cycles:	Denaturation	Annealing	Extension	Final ext.
Temperature [°C]	95		94	61	72	60
Time (s)	900		30	90	60	1800

5. The temperature program for the HRM analysis started at 95°C for 1 minute with a ramp of 4.4°C per second, followed by 40°C for 1 minute with a ramp of 2.2°C per second, 60°C for 1 second with a ramp of 4.4°C per second and a continuous step at 95°C.

11.2.3.19. MLST protocol

Materials

Reagents:
PCR-graded water
PCR purification kit
Primers
PCR reagents (see table)
Equipment:
peqSTAR thermocycler
Agarose electrophoresis chamber
Eppendorf pipettes

Protocol:

1. Each isolate was inoculated into TSB plus 1% NaCl and incubated with shaking overnight at 35°C.
2. One milliliter of the overnight culture was removed, heated to 100°C for 10 minutes, and plunged into ice for 5 minutes.
3. The master mix was set-u as follows:

Table 81: Master mix for MLST

Component	Units	Stock concentration	Final volume [µl]	Volume [µl] /Reaction
PCR-graded H_2O				16.74
Buffer	X	10	1	2.50
$MgCl_2$	mM	50	1.5	0.75
dNTPs	mM	10	0.125	0.31
Forward primer	µM	10	0.500	1.25
Reverse primer	µM	10	0.500	1.25
Taq DNA polymerase	Units/µl	5	1	0.20
Master Mix (ul) to Each Tube				23.00
Target Template	µl			2.00

4. The following temperature program was used:

Table 82: Temperature program for MLST

	Initial denaturation	30 cycles	Denaturation	Annealing	Extension
Temperature [°C]	96		96	58	72
Time (s)	60		60	60	60

5. The PCR products were separated by agarose electrophoresis and purified using a PCR purification kit with a total elution volume of 25µL.
6. The purified samples were sequenced on an ABI 3730 xl sequencer at McLab (South San Francisco, CA).
7. Sequences were analyzed with BioEdit software 7.1.9 (Abbott, Carlsbad, CA).

11.2.3.20. Growing HeLa and Caco cells

Materials

Reagents:

Dulbecco's Modified Eagle Medium (DMEM)
Fetal bovine serum (FBS)
glucose
sodium bicarbonate
N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid (HEPES)
Non-essential amino acids
Glutamax

Equipment:

Cell culture flasks
10 mL Falcon tubes
Filtered pipettes

Protocol:

1. Cells coming out of the nitrogen tank were thawed quickly in a 37˚C warm waterbath.
2. Under a fume hood the cell suspension was transferred to a conical tube and mixed with 5 mL of media
3. The cell suspension was centrifuged for 5 minutes to separate cells in media from the DMSO of the freezing media
4. Cells were rinsed with 5 mL of fresh media and an aliquot transferred to a 10 mL tissue culture flask. HeLa cells were grown in DMEM with 10% FBS, 4.5 g/mL glucose, 1X sodium pyruvate, 1X NEAA, 1X glutamax, 1X sodium bicarbonate and phenol red as a pH indicator. Caco-2 cells were grown in DMEM with 15% FBS, 4.5 g/mL glucose, 1X HEPES, 1X NEAA, 1X Glutamax, 1X sodium bicarbonate and phenol red as a pH indicator. The cytotoxicity assay medium was identical to the DMEM cultural medium except for a 5% FBS concentration and the lack of phenol red.
5. Cells suspensions were incubated in a humidified chamber at 37˚C and 5% CO_2 until confluence was reached.

11.2.3.21. Passaging cell cultures

Materials

Reagents:

DMEM with nutrions
Trypsin-EDTA

Versene
Equipment:
Cell culture flasks
Filtered pipettes

Protocol:

1. Cells coming out of the incubator are transferred to a laminar flow hood. The media was poured off.
2. Cells were rinsed with Versene, an EDTA solution, and then treated with 1X Trypsin-EDTA prior to passaging. The cell culture flask was incubated for one minute at 37˚C and 5% CO_2.
3. Fresh media was added and the cell suspension was transferred to a culture flask containing DMEM.
4. Cells suspensions were incubated in a humidified chamber at 37˚C and 5% CO_2 until confluence was reached.

11.2.3.22. Cytotoxicity assay protocol

Materials

Reagents:
PBS Tissue Culture
PBS Tissue Culture + 1%BSA
TSB plus 1% NaCl
DMEM without phenol red
Cytotoxicity kit contents:
Substrate Mix
Assay Buffer
LDH Positive Control
Lysis Solution (10X)
Stop Solution
Equipment:
Tecan plate reader

UV-Vis Spectrometer (wavelength 600 nm)
Multi pipettors

Protocol:

1. Target cells were added to experimental wells and incubated for 24 hours in a humidified chamber with 5% CO_2 at 37˚C.
2. Afterwards, the suspended cells were counted and infected with an appropriate dilution of a *V. parahaemolyticus* culture. For infection of HeLa and Caco-2 cells, 100 µL of the overnight culture were transferred into 2 mL fresh TSB + 1% NaCl and incubated for three hours at 35˚C to acquire an early log-phase culture. Following incubation, the optical density was measured at 600nm to estimate bacterial cell density using a UV-Vis Spectrometer as in protocol 11.2.3.2. The log-phase culture was diluted in PBS tissue culture to a calculated OD_{600} of 0.008 or 0.0008 for the HeLa or Caco-2 assays, respectively.
3. Cells for Target Cell Spontaneous LDH Release Control were added.
4. The culture medium for Volume Correction Control and for Culture Medium Background Control was added to the appropriate wells.
5. The effector cells for Effector Cell Spontaneous LDH Release Control and experiment wells were transferred.
6. The plate was centrifuged 250xg for 4 minutes.
7. The plate was incubated for 4 hours at 37˚C.
8. Lysis Solution (10X) was added to Target Cell Maximum LDH Release Control 45 minutes prior to centrifugation.
9. The plate was centrifuged at 250xg for 4 minutes.
10. Fifty µL of the supernatant were transferred to an enzymatic assay plate.

11. Fifty µL of a 1:5,000 dilution of LDH positive Control were added to separate well.
12. Fifty µL reconstituted substrate mix were added to each well of enzymatic assay plate.
13. The plate was incubated for 30 minutes at room temperature, protected from light.
14. Fifty µL of stop solution were transferred to each well.
15. For defoaming/reducing errors in optical reading 5 µL of silicon oil with a viscosity of 5 centistokes was pipetted to each well.
16. The absorbance was measured at 490 nm using a plate reader.

11.2.3.23. Statistical analysis cytotoxicity assay

Materials

Software:

Sigma Plot version 12.5.

Protocol:

1. Relative cytotoxicity measurements were scored as ordinal data, with three categories determined from evaluation of the current data: cytotoxic (>75%), semi-cytotoxic (25-75%), and non-cytotoxic (<25%).
2. Based on this scoring, differences in cytotoxicity between isolates of different origins (clinical vs. shellfish) and due to the presence (or expression) of different virulence genes or serotype were evaluated by contingency table analysis, for each cell line separately.
3. The statistical significance of observed differences, and associations with virulence genes and serotype, was determined by chi-square test of independence.

4. A significance level of 0.05 was used for determination of statistical significance in all comparisons.

11.2.3.24. Sequencing protocol using the Illumina HiSeq

Materials

Reagents:

PCR-graded water
TSA

Protocol:

1. The isolates were checked for purity, and stored at -80˚C until needed for DNA isolation as part of the 100K Pathogen Genome Project. The strains were regrown on TSA overnight at 37 ˚C. Single colonies were used for DNA extraction.
2. High molecular weight DNA was extracted from overnight cultures.
3. The library preparation and sequencing was performed at UC Davis, Davis, CA, USA.
4. The genomic sequences were *de novo* assembled using SPAdes version 3.1.1 software.
5. The draft genomes were annotated using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP, http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html).

11.2.3.25. Sequencing protocol using the PacBio

Materials

Reagents:

PCR-graded water
TSA

Equipment:
2200 TapeStation system

Protocol:

1. The isolates were checked for purity, and stored at -80˚C until needed for DNA isolation as part of the 100K Pathogen Genome Project. The strains were regrown on TSA overnight at 37˚C. Single colonies were used for DNA extraction.
2. The library preparation and sequencing was performed at UC Davis, Davis, CA, USA.
3. High molecular weight DNA was extracted from overnight cultures.
4. The DNA extracts were analyzed on a 2200 TapeStation system with the Genomic DNA ScreenTape (Agilent Technologies, Santa Clara, CA, USA). assay for integrity of high molecular weight gDNA (Jeannotte et al., 2014).
5. After evaluation of gDNA size and quantity, 10 µg was used for fragmentation using the Covaris g-TUBE device (Covaris, Woburn, MA, USA) following the manufacturer's instructions (Kong et al., 2014).
6. The fragmented gDNA was used for library construction with the PacBio SMRTbell 10kb Library preparation kit, which was normalized to 1 to 5 µg input.
7. Libraries were sequenced utilizing PacBio RSII and C2 chemistry with 100x coverage per the manufacturer's instructions.
8. For each isolate, the genomic sequence single-pass reads were de novo assembled using the Hierarchical Genome Assembly Process (HGAP) version 1.4 software (Pacific Biosciences) and were then annotated using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok) (Klimke et al., 2009).

11.2.3.26. kSNP protocol

Materials

Software:
kSNP v3
Software RAxML
FigTree
Equipment:
Computer system using Linux

Protocol:

1. The following parameters were used for the kSNP analysis: k-mer size was 25 and SNP locations were determined reference-free.
2. Post kSNP analysis bootstrap values (N = 1000) were added with the software for Randomized Axelerated Maximum Likelihood (RAxML) (http://sco.h-its.org/exelixis/web/software/raxml/) (Stamatakis, 2014).
3. From the kSNP matrix a maximum likelihood tree was built, visualizing the number of SNPs shared by each lineage in FigTree v1.4.2 (http://tree.bio.ed.ac.uk).

12. List of publications

- **Lüdeke CHM, Fischer M, Jones JL.** Cytotoxicity of *Vibrio parahaemolyticus* is independent of *tdh*, *trh*, or serotype. Under review
- **Lüdeke CHM, Gonzalez-Escalona N, Fischer M, Jones JL.** Examination of clinical and environmental *Vibrio parahaemolyticus* isolates by multi-locus sequence typing (MLST) and multiple-locus variable-number tandem-repeat analysis (MLVA). Froniters Microb.. 2015
- **Lüdeke CHM, Kong N, Weimer BC, Fischer M, Jones JL.** Complete Genome Sequences of a Clinica Isolate and an Environmental Isolate of *Vibrio parahaemolyticus*. Genome Announc.. 2015
- **Vinarao RT, Jones JL, Ragaza RJ, Lüdeke CHM, Burkhardt W, Cabigao JS, Bassig RA.** Incidence of *Vibrio parahaemolyticus* in Oysters along Nothern Manila Bay, Philippines. Proc. 9th Int. Conf. Molluscan Shellfish Safety. 2014 (Conference paper)
- **Jones JL, Lüdeke CHM, Bowers JC, DeRosia-Banick K, Carey DH, Hastback W**. *Vibrio cholerae*, *V. vulnificus*, and *V. parahaemolyticus* Abundance in Oysters (*Crassostrea virginica*) and Clams (*Mercenaria mercenaria*) from Long Island Sound. Appl. Environ. Microbiol. 2014
- **Lüdeke CHM, Fischer M, LaFon P, Cooper K, Jones JL**. Suitability of the Molecular Subtyping Methods Intergenic Spacer Region, Direct Genome Restriction Analysis, and Pulsed-Field Gel Electrophoresis for Clinical and Environmental *Vibrio parahaemolyticus* Isolates. Foodborne Pathog Dis 2014
- **Jones JL, Lüdeke CHM, Bowers J, Depaola A**. Comparison of Plating Media for Recovery of Total and Virulent Genotypes of *Vibrio vulnificus* in U.S. Market Oysters. Intl J Food Microbiol 2013

- **Jones JL, Lüdeke CHM, Garett N, Fischer M, Bopp CA, Depaola A**. Biochemical, Serological, and Virulence Characterization of Clinical and Oyster *Vibrio parahaemolyticus* Isolates. J Clin Microbiol 2012 Apr 25;**50**:2343-52

13. Acknowledgement

This dissertation work couldn't have been done without the support and encouragement of a lot of people for a variety of reasons. I owe my gratitude to all of them.

First, I would like to thank my mentor and friend, Dr. Jessica Jones, for her effort, encouragement, endless support and her ability to always find a sympathetic ear. With your knowledge and enthusiasm you made this dissertation work to a challenging but wonderful adventure.

I would also like to thank Dr. Capt. William Burkhardt for supervising this work at DSST and for giving me the chance to gain my doctorate in his branch and division. I'm very grateful for this opportunity and for his support. At this point I also like to mention Dr. Robert Dickey for creating the Ph.D. student position during his time in the lab.

My gratitude also belongs to my doctorate supervisor Prof. Dr. Markus Fischer from the University of Hamburg. Thank you for supervising my dissertation work from the far and the useful discussions for this dissertation.
I would also like to thank my committee, Prof. Dr. José Broekaert, Prof. Dr. Bernward Bisping, and Dr. Angelika Paschke-Kratzin.

For financial support, I thank the Center for Food Safety and Applied Nutrition (CFSAN) for providing the Oak Ridge Associated Universities (ORAU) scholarship in the Research Fellowship Program.

13 Acknowledgement

I would like to acknowledge the support of Dr. Lydia V. Rump and Dr. Narjol Gonzalez-Escalona, who fought to get me into the FDA intern program, which paved the way for my dissertation work. I'm forever thankful for your care, motivation and friendship.

I also want to thank all members of the Division of Seafood Science and Technology for their support, the great work environment and friendship, you made my time as a Ph.D. student very special. I would like to particularly thank Dr. Angelo DePaola, Capt. Kevin Calci, George Doup, Clinton Collier, Tony Previto and Thomas Kinsey for their help with experimental setup and media preparations, general advice and supportive discussions.

My friends in the United States, thank you for making this time in the United States so unforgettable and supporting me during those four years. Also to my friends in Germany, thank you for your long-distance cheerleading and cheer-up when it was needed the most. A special thank you goes to Olga Kraus and Katja Schilling, since the first day of University we have supported each other until today, I'm very thankful for your friendship.

Last but not least, I would like to thank my family, especially my parents and my sister, for letting me fulfill my dreams in the United States, but keeping me in your thoughts every day at home. Without your endless support, encouragement, patience and every visit throughout those four years this thesis would not exist and no words in the world can describe how thankful I am.

14. Eidesstattliche Versicherung

“Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben. Ich versichere, dass diese Dissertation nicht in einem früheren Promotionsverfahren eingereicht wurde.

Hamburg, den 27.06.2017

Catharina H. M. Lüdeke