# Molecular evolutionary analyses and epidemiology of *Vibrio parahaemolyticus* in Thailand
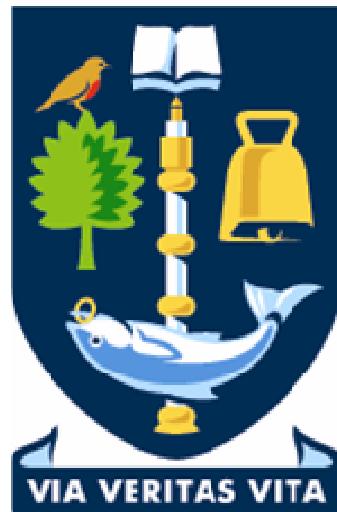


## CHONCHANOK THEETHAKAEW

**(BSc., MSc.)**

**Institute of Infection, Immunity and Inflammation**
**College of Medical, Veterinary and Life Sciences**
**University of Glasgow**

**September 2012**
**A thesis submitted for the degree of**
**Doctor of Philosophy**

# ABSTRACT

*Vibrio parahaemolyticus* is a seafood-borne pathogenic bacterium which is a major cause of gastroenteritis worldwide. In the present study, the genetic relationships and population structure of isolates originating from clinical and seafood production sources in Thailand were investigated by multilocus sequence typing (MLST). Nucleotide sequence variation of virulence-related genes including haemolysin and TTSS1 genes among Thai and worldwide isolates was also analyzed. The outer membrane proteome of *V. parahaemolyticus* isolate RIMD2210633 was predicted using bioinformatic approaches, and the outer membrane proteomes of eight isolates from different sources and representing different MLST sequence type characterized using proteomics.

The 101 Thai *V. parahaemolyticus* isolates examined were recovered from clinical samples (n=15), healthy human carriers (n=18), various fresh seafood (n=18), frozen shrimps (n=16), fresh-farmed shrimp tissue (n=18) and shrimp-farm water (n=16). Phylogenetic analysis revealed a high degree of genetic diversity within the *V. parahaemolyticus* population, although isolates recovered from clinical samples, farmed shrimp and water samples represented five distinct clusters. The majority of clinical isolates were resolved into two genetic clusters and none of these isolates were found to share sequence types (STs) with strains isolated from human carriers, seafood, or water. Similarly, STs representing human carrier isolates differed from those of clinical, seafood and water isolates. The limited genetic diversity of the clinical isolates suggested non-random selection for pathogenic strains, but the absence of such strains in local seafood raises questions about the likely source of infection. Extensive serotypic diversity occurred among isolates representing the same STs and

recovered from the same source at the same time point. Furthermore, evidence of interspecies horizontal gene transfer and intragenic recombination was observed at the *recA* locus in a large proportion of isolates; this has a substantial effect on the apparent phylogenetic relationships of the isolates. Notably, the majority of these recombinational exchanges occurred among clinical and carrier isolates, suggesting that the human intestinal tract is serving as a reservoir that is driving evolutionary change and leading to the emergence of new, potentially pathogenic strains. MLST was also applied to study genetic relationships between *V. parahaemolyticus* isolates from Thailand (n=101) and those from European countries (n=9). With the exception of the pandemic ST3 which was resolved from two isolates from Thai human carriers, two clinical isolates from England and a clinical isolate from Norway, none of the other European isolates examined in this study shared the same ST with the Thai isolates.

This study demonstrated that Thai human carrier isolates are capable of harbouring virulence-related genes including the haemolysin-encoding genes *tdhA, tdhS, trh1* and *trh2,* and the TTSS1-related genes *vcrD1, vscC2* and *VP1680,* that are present in clinical isolates. In particular, the Thai human carrier isolate VP132 shared identical TTSS1-related gene fragments with the pandemic *V. parahaemolyticus* serotype O3:K6 (RIMD2210633) and related strains (AQ3810, AQ4037, Peru466, AN5034 and K5030) of worldwide distribution.

A total of 117 outer membrane proteins (OMPs) were predicted from the genome of *V. parahaemolyticus* isolate RIMD2210633. A total 73 OMPs proteins were identified from eight *V. parahaemolyticus* isolates recovered from clinical samples, human carriers, oyster, shrimp tissue and water in Thailand. Of the 117 predicted OMPS, 32 were identified in eight strains by proteomic analysis.

OmpU, a non-specific porin protein, represents the most abundantly expressed protein in all eight isolates. OMPs involved in TTSSs (YscW, YscJ, YscC, PopN and VscC2) and iron uptake (IrgA, putative 83 Da decaheme outer membrane cytochrome C, PvuA1, PvuA2, LutA, FhuE, HutA and putative-regulated protein B) were predicted from the genome of *V. parahaemolyticus* isolate RIMD2210633, but were not recovered from any of the eight Thai isolates. The absence of TTSS and iron uptake related OMPs in the eight representative strains that were grown under *in vitro* conditions may suggest an important requirement for *in vivo* growth conditions to induce expression of important virulence factor-related OMPs in *V. parahaemolyticus*. There was no clear association between OMP profile and the source of isolation, ST or serotype. However, a high degree of variation of OMP profiles was observed in isolates from different sources as well as in the isolates representing the same ST.

This study demonstrated the usefulness of a multidisciplinary approach that includes MLST, virulence-related gene DNA sequence analysis, bioinformatic prediction and gel-based proteomic analyses for the study of molecular evolutionary relationships and the epidemiology of *V. parahaemolyticus* isolates from clinical and seafood production sources. The outcomes of this study highlight the role of human carriers as a reservoir of potentially pathogenic *V. parahaemolyticus* and this should be considered as one of the possible contamination sources in the surveillance of seafood safety.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# DEDICATION

This four year PhD study is dedicated to my parents, Dr. Pismai and Mr. Prasit Theethakaew, my sister, Ms Sitang Khanti, and my beloved country, THAILAND.

# ACKNOWLEDGEMENTS

# PRESENTATIONS / PUBLICATIONS

1. Theethakaew, C., Feil, E., Castillo-Ramírez, S., Aanensen, D., Suthienkul, O., Neil, D. M. and Davies, R. L. (2012). Population structure of *Vibrio parahaemolyticus* from clinical, human carriage and environmental sources in Thailand. *Applied Environmental Microbiology* (submitted paper).

2. Theethakaew, C., Feil, E., Castillo-Ramírez, S., Aanensen, D., Suthienkul, O., Neil, D. M. and Davies, R. L. Molecular evolutionary relationships of *Vibrio parahaemolyticus* isolates from multiple sources in Thailand. Oral presentation at the 1st Health Challenge Thailand conference, 2nd-3rd July 2011, London, United Kingdom.

3. Theethakaew, C., Neil, D. M., Suthienkul, O., Kapoor, K. and Davies, R. L. Identification of genetic variation of *Vibrio parahaemolyticus* isolates from multiple sources in Thailand by Multilocus Sequence Typing (MLST). Poster presentation (awarded $200 prize for outstanding graduate student poster) at the *Vibrios* in the environment 2010, 7th-12th November 2010, Mississippi, USA.

4. Theethakaew, C., Neil, D. M., Suthienkul, O., Kapoor, K. and Davies, R. L. Genetic relationships of *Vibrio parahaemolyticus* isolates recovered from clinical and environmental sources in Thailand. Poster presentation at the 9th International Meeting on Microbial Epidemiological Markers (IMMEM 9), 1st-4th September 2010, Wernigerode, Germany.

5. Theethakaew, C., Neil, D. M., Suthienkul, O. and Davies, R. L. Molecular evolution and epidemiology of *Vibrio parahaemolyticus* in Thailand. Poster presentation at the 3rd Trans-Atlantic Fisheries Technology Conference (TAFT), 15th -18th September 2009, Copenhagen, Denmark.

# DECLARATION

I hereby certify that this thesis has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

The research for this thesis was performed in the University of Glasgow between 2008 and 2012.

Chonchanok Theethakaew

# ABBREVIATIONS

| | |
|---|---|
| 1D-PAGE | One-dimentional gel electrophorysis |
| 2D-PAGE | Two-dimentional gel electrophorysis |
| aaST | Amino acid sequence type |
| ABC | ATP-binding cassette |
| AFLP | Amplified fragment length polymorphism |
| AP | Allelic profile |
| APS | Ammonium persulphate |
| ARDRA | Amplified ribosomal DNA restriction analysis |
| Asp | Aspartic acid |
| BAM | ß-barrel assembly machinery |
| BAPS | Bayesian analysis of population structure |
| BOMP | ß-barrel outer membrane protein predictor |
| BoNT | botulinum neurotoxin encoding genes |
| bp | Base pair |
| CC | Clonal complexes |
| Cefas | Centre of Environment, Fisheries, and Aquaculture Science |
| CFU | Colony forming unit |
| ChiRP | Chitin-regulated pilus |
| CoC | Code of Conduct |
| Codex | Codex Alimentarius Commission |
| CP | Chareon Pokphan |
| CPF | Charoen Pokphand Foods |
| CPS | Capsular polysaccharide |
| D | Doublet |
| Da | Dalton |
| DISC-PAGE | Discontinuous polyacrylamide gel electrophorysis |
| DNA | Deoxyribonucleic acid |
| EPEC | Enteropathogenic *Escherichia coli* |
| ERIC | Enterobacterial repetitive intergenic consensus sequence |
| ESI | Electrospray ionization |
| EU | European Union |
| FDA | Food and drug administration |

| | |
|---|---|
| Fig | Figure |
| FISH | Fluorescence in situ hybridization |
| FT-MS | Fourier transform ion cyclotron |
| g | gram |
| g-1 | per gram |
| GAP | Good Aquaculture Practice |
| GMP | Good Manufacturing Practice |
| GS-PCR | Gene specific PCR |
| $H_2O$ | Dihydrogen Monoxide (water) |
| HACCP | Hazard Analysis and Critical Control Points |
| HGT | Horizontal gene transfer |
| HMM | Hidden Markov model |
| HPA | Health Protection Agency |
| HPV | Hepatopancreatic virus |
| IDSC | Infectious Disease Surveillance Centre |
| IEF | Isoelectric focusing |
| IHHNV | Infectious hypodermal and hematopoietic necrosis virus |
| IL | Interleukin |
| $I^S_A$ | Index of Association |
| kb | Kilobase |
| kDa | Kilodalton |
| KP-negative | Kanagawa phenomenon negative |
| KP-positive | Kanagawa phenomenon positive |
| LAB | Lactic acid bacteria |
| LC-MS | Liquid chromatography ESI-MS systems |
| LC-MS/MS | Liquid chromatography with tandem mass spectrophotometry |
| LDH | Lecithin-dependent haemolysin |
| Lol | Localized lipoprotein |
| LPS | Lipopolysaccharide |
| M | Molar |
| m/z | Mass-to-charge ratio |
| MALDI | Matrix-assisted laser desorption/ionization |
| MAPK | Mitogen-activated protein kinases |
| MAPK | Mitogen-activated protein kinases |
| MBV | Monodon baculovirus |

| | |
|---|---|
| MCMBB | Markov Chain Model for Beta Barrels |
| MFP | Membrane fusion protein |
| MgSO$_4$ | Magnesium sulphate |
| MLEE | Multilocus enzyme electrophoresis |
| MLST | Multilocus sequence typing |
| mM | Milimolar |
| MPN | Most probable number |
| MRSA | Methicillin-resistant S. aureus |
| MSGS | Monodon slow growth syndrome |
| MSHA | Mannose-sensitive hemagglutinin |
| MSSA | methicillin-susceptable *S. aureus* |
| NaCl | Sodium chloride |
| NGS | Next-generation genome sequencing |
| NH$_4$HCO$_3$ | Ammonium bicarbonate |
| NCTC | National Collection of Type Cultures |
| $^o$C | Degrees Celsius |
| OMP | Outer membrane protein |
| OP | Opaque |
| ORF | Open reading frame |
| PA | Proteome Analyst |
| PCR | Polymerase chain reaction |
| PFGE | Pulse field gel electrophorysis |
| PGE | Pore gradient electrophorysis |
| PI | Pathogenicity island |
| pmol | Picomole |
| PSI-BLAST | Position-specific interative basic local alignment search tool |
| PSSM | Position specific scoring matrix |
| RAPD | Random amplified polymorphic DNA |
| RBF | Radial basis function |
| REP | Repetitive extragenic palindromes |
| RFLP | Restriction fragment length polymorphism |
| RIL | Rabbit ileal loop |
| rpm | Revolutions per minute |
| RS | Ribosomal gene spacer sequence |
| SDS | Sodium dodecyl sulphate |

| | |
|---|---|
| Ser | Serine |
| SLV | Single locus variant |
| SNP | Single nucleotide polymorphism |
| SPaseI | Signal peptidase I |
| SPaseII | Signal peptidase II |
| SPF | Specific pathogen-free |
| ST | Sequence types |
| TDH | Thermostable direct haemolysin |
| TEMED | N,N,N'N'- tetramethylethylene diamine |
| TOF | Time-of-flight |
| TR | Translucent |
| TRH | Thermostable direct haemolysin-related haemolysin |
| TSA | Tryptic Soy Agar |
| TSB | Tryptic Soy Broth |
| TTSS | Type three secretion system |
| UK | United Kingdom |
| USA | United States of America |
| v/v | Volume per volume |
| VBNC | Viable but non culturable |
| VF | Vibrioferrin |
| VPaI | *V. parahaemolyticus* pathogenicity island |
| w/v | Weight per volume |
| WHO | World Health Organization |
| WSSV | White spot syndrome virus |
| YHV | Yellow head virus |
| $\sigma/\theta$ | Recombination per mutation rate |
| ∂-VPH | Heat-stable haemolysin |
| µl | Microlitre |

# 1. INTRODUCTION

This thesis describes the application of a multidisciplinary approach, including nucleotide sequence and proteomic analyses, to the investigation of the molecular evolutionary relationships and epidemiology of a seafood-borne pathogenic bacterium, namely *Vibrio parahaemolyticus*. The isolates examined in this study were obtained from Thailand, where seafood is widely consumed. This chapter will introduce a number of topics that are relevant to this study and underlie the scientific approach that has been used.

The scale of the economic impact and production of white shrimp, a commercially important seafood species in Thailand, means that any aspect of the shrimp aquaculture industry that may also impact on the health of consumers is important to understand. One such potential threat arises from food poisoning due to infection by the bacterium *V. parahaemolyticus*, and it is therefore necessary to understand how this organism causes a problem in public health, and the microbial risk assessment procedures that are currently in place to detect it. Of particular relevance is the epidemiology of the disease, both worldwide and particularly in Thailand.

An understanding of the causation of food poisoning by *V. parahaemolyticus* requires knowledge of the potential virulence factors that may be involved, their modes of action and the triggers for their expression. In order to study these it is first necessary to deploy appropriate molecular methods to understand the relationships between various strains of the bacterium isolated from environmental, industrial, seafood and human sources. The molecular typing method chosen for use in the present study was multilocus sequence typing

(MLST). Such knowledge facilitates a more informed approach to the study of the expression of virulence factors in Gram-negative bacteria such as *V. parahaemolyticus*. Several of these virulence factors involve the outer membrane proteins (OMPs), and for this reason OMP analysis techniques were used in the present study (i.e. bioinformatics and gel-based proteomic approaches).

The relationships between the topics outlined above and reviewed in this chapter are summarized by the flow diagram in Fig. 1.1. These have informed the development of the aims and objectives of the scientific work that is described in the subsequent chapters of the thesis.
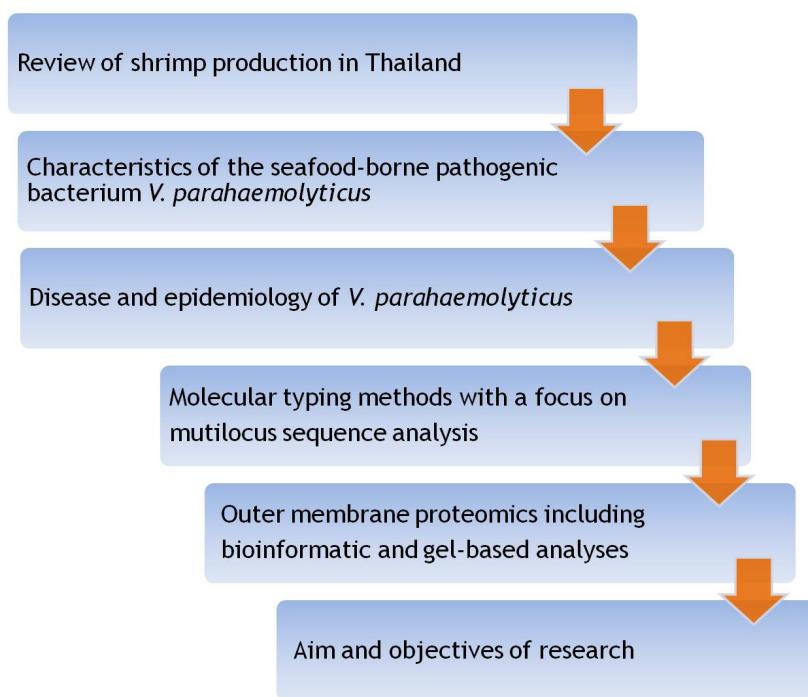


**Figure 1.1.** Diagram showing the structure of the introduction section of the present study

# 1.1 Overview of shrimp aquaculture in Thailand with respect to *V. parahaemolyticus* infection.

This section provides background information on seafood production in Thailand by using a commercially important species, white shrimp (*Penaeus vannamei*), as an example. A survey of the shrimp production process, including hatcheries, shrimp farms and a processing factory, was conducted through a field trip in Thailand. A review of the risk of contamination and microbial risk assessment of *V. parahaemolyticus* throughout the seafood production chain is also given.

## 1.1.1 Commercial value and economic impact of shrimp production

Thailand is one of the world's leading shrimp aquaculture countries. The geographical position of the country, being surrounded by the Andaman Sea and Gulf of Thailand, provides a large estuary area for shrimp farming, particularly along the East and South coasts (Fig. 1.9). Its moderate tropical climate with an average temperature of 32°C is suitable for warm-water aquaculture. Moreover, numerous innovative technologies that have been introduced to the country have contributed to the fast growth of several agricultural activities, including shrimp farming.

Over the last two decades, shrimp aquaculture has expanded to involve a total culture area of 71,887 hectares in 1993 (Tookwinas, 1993). The total area dedicated to shrimp aquaculture in the country increased to 75,736 hectares in 2003 (Tookwinas *et al.*, 2005). Due to the application of improved technologies, shrimp production per hectare has continuously increased to the present.

Thailand became the world leader in cultured shrimp production in 1991 with 152,000 tonnes annual production, but in 2001 this figure was overtaken by that of China. In 2009 Thailand was the second ranked country for shrimp aquaculture production, followed by Vietnam, Indonesia and India respectively (Fig. 1.2). However, in terms of exports, Thailand is ranked first of the major shrimp exporting countries, with a yearly export value of £1,524.4 million from 500,000 tonnes production in 2007 (Fig. 1.3) (FAO, 2010; Josupeit, 2008). Shrimp from Thailand are widely exported across the world, particularly to the USA, Japan, and Canada (Fig. 1.4). As a result, Thailand currently has the highest value of exported product in the world shrimp market.

Regarding the total of exported agricultural goods from the country, shrimp products including frozen shrimp and fresh shrimp rank number five by value (at £429.5 million per annum in 2010) among the other agricultural products such as rubber, rice, cassava, processed chicken, fish, etc. (Fig. 1.5). These data indicate the high value and economical importance of shrimp production among the agricultural products from Thailand. Moreover, data from the Thai Ministry of Commerce show that annual exportation of frozen shrimp from the country has continuously increased (Fig. 1.6).
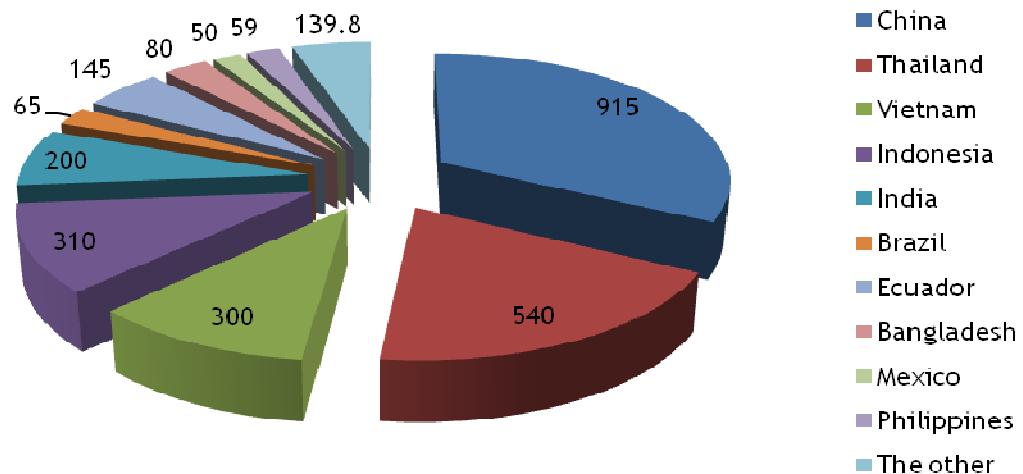
**Figure 1.2.** Estimated amount of world shrimp aquaculture production in 2009 by country (in 1000 tonnes) (FAO, 2010)
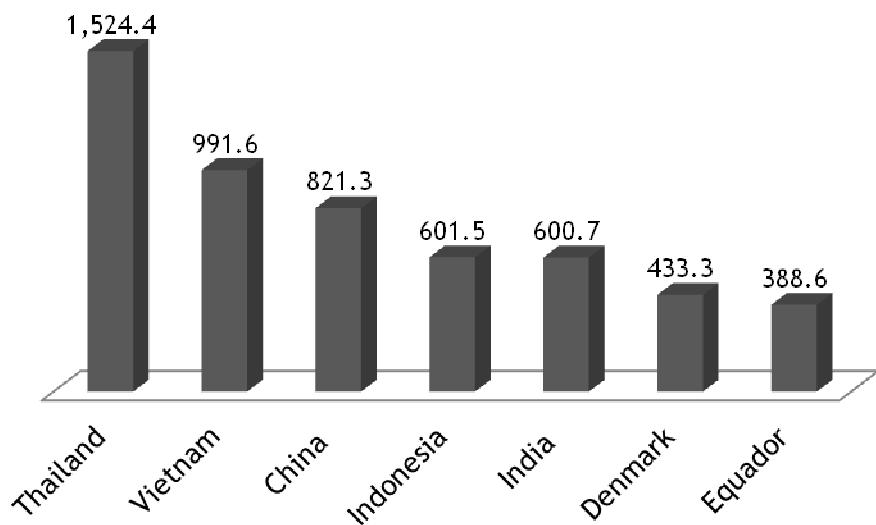


**Figure 1.3.** Yearly shrimp export value by major shrimp exporting countries in 2007 (in £ million) (FAO, 2010)

5

United State of America
Japan
Canada
United Kingdom
Spain
Korea
Belgium
France
Germany
Australia
the other
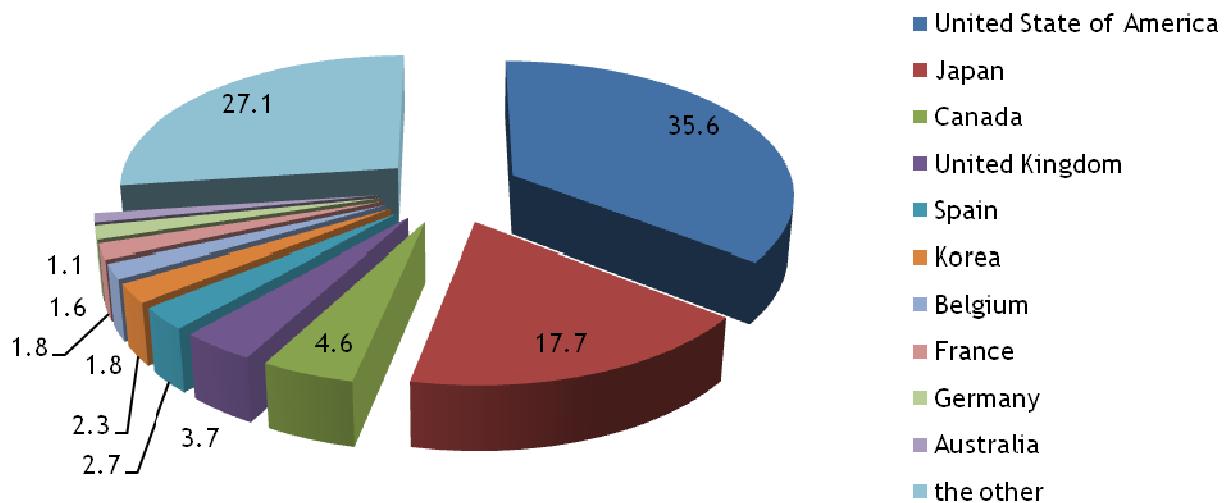
**Figure 1.4.** Proportion by value of the exported shrimp from Thailand to different purchasing countries (FAO, 2010)



| | |
|---|---|
| Rubber | 3887.9 |
| Rice | 2105.7 |
| Cassava | 794.2 |
| Processed chicken | 532.2 |
| Shrimp (fresh/frozen) | 429.5 |
| Fish (fresh/frozen) | 325.5 |
| Fruit (fresh/frozen/dried) | 213.8 |
| Squid (fresh/frozen) | 126.4 |
| Fish (fresh/frozen) | 115.1 |
| Vegetable (fresh/frozen) | 82.7 |
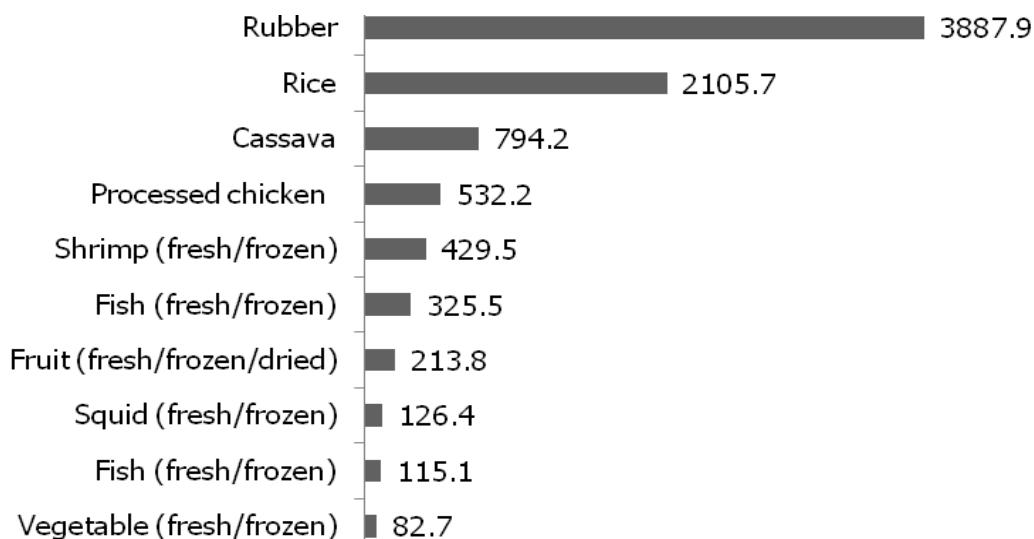
**Figure 1.5.** Annual export value of Thai agricultural products in 2010 (in £ million) (Ministry of commerce, Thailand, online at www2.ops3.moc.go.th)
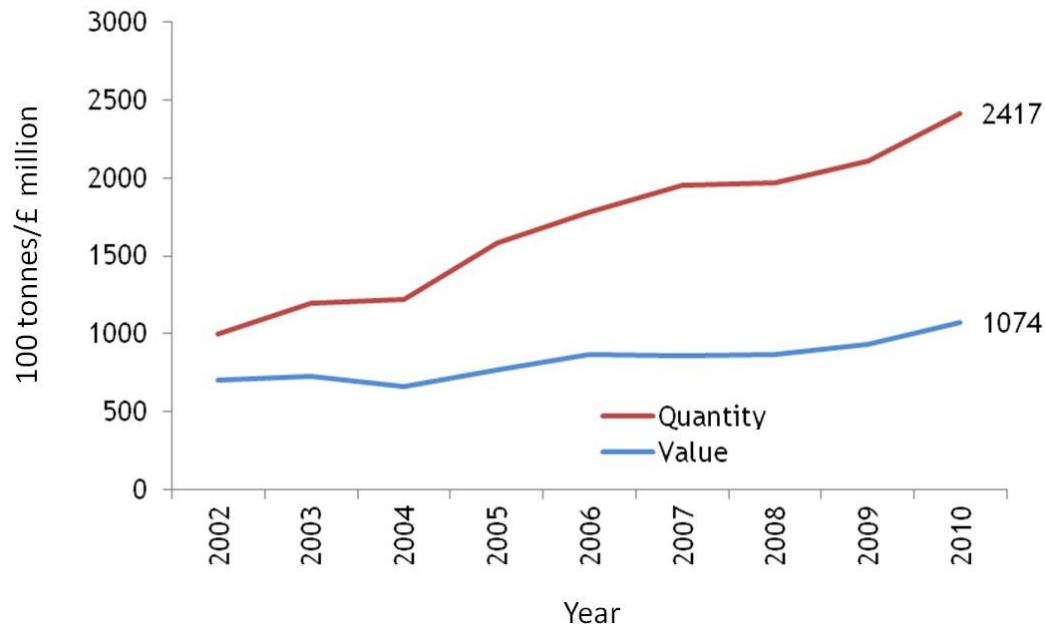
**Figure 1.6.** Growth of exported frozen shrimp in Thailand by quantity (100 tonnes) and value (£ million) (Ministry of Commerce, Thailand, online at www2.ops3.moc.go.th)

Traditionally, the Thai shrimp industry farmed wild stocks of the black tiger shrimp, *Penaeus monodon,* which has a natural distribution in the Indo-West-Pacific. However, in the early 1990s it was found that using these wild stocks posed a number of disease risks, such as monodon slow growth syndrome (MSGS) (Chayaburakul *et al.*, 2004), and various other viral diseases, e.g. white spot syndrome virus (WSSV) and infectious hypodermal and hematopoietic necrosis virus (IHHNV) (Flegel, 2006). The susceptibility of *P. monodon* to these diseases led to the introduction of an alternative species, *Penaeus vannamei* (now *Litopenaeus vannamei*), the so-called Pacific white shrimp (also known as the whiteleg shrimp), which is native to the eastern Pacific Ocean, from Mexico to northern Peru. In particular, in the late 1990s domesticated lines of specific pathogen-free (SPF) *P. vannamei* were developed in the USA for exportation to Asian countries, and use of these is the main reason that the shrimp aquaculture

industry recovered and expanded following the viral pandemics of the early 1990's (FAO, 2006; Lightner & Redman, 2010).

Moreover, because of its higher survival rate, faster growth with uniform size and tolerance to higher densities than *P. monodon, P. vannamei* was considered to be a more profitable species. Thus, *P. vannamei* has been extensively farmed, and since 2001 has become the most commercially-important seafood species in Thailand (Wyban, 2007). This increased concentration on *P. vannamei* production has occurred not only in Thailand, but also globally, so that production of *P. vannamei* has grown rapidly for the last two decades (Fig. 1.7). Of the main *P. vannamei* producing countries from Asia, Latin America and elsewhere, China and Thailand have remained the top two producers since 2005 (Table 1.1).
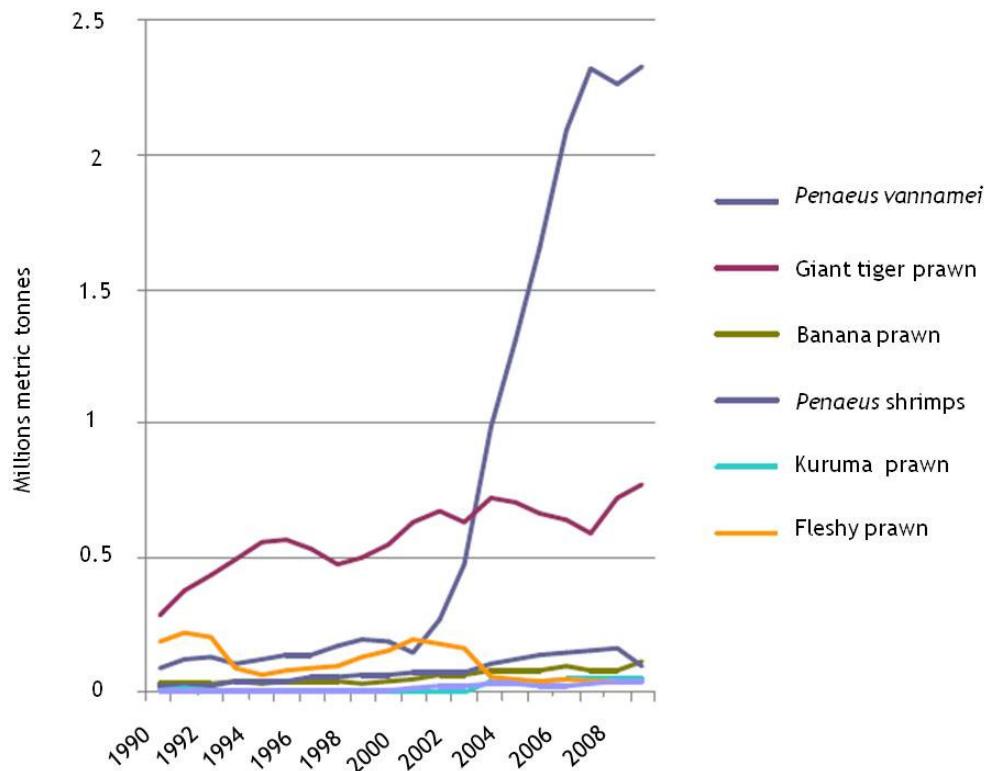
**Figure 1.7.** Trends in global cultured shrimp production from 1990-2008. Figure was adapted from Bondad-Reantaso *et al.* (2012)

**Table 1.1.** Aquaculture production of *P. vannamei* from the top ten global producers during 2005-2009 (Bondad-Reantaso *et al.*, 2012)

| Producers | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|
| China | 702484 | 887838 | 1065644 | 1062765 | 1118142 |
| Thailand | 374487 | 480061 | 508446 | 501394 | 535000 |
| Ecuador | 118500 | 149200 | 150000 | 150000 | 179100 |
| Indonesia | 103874 | 141649 | 164466 | 208648 | 170969 |
| Mexico | 90008 | 112495 | 111787 | 130201 | 125778 |
| Brazil | 63134 | 65000 | 65000 | 70251 | 65188 |
| Vietnam | 100000 | 150000 | 153000 | 38600 | 36000 |
| Columbia | 19000 | 21600 | 20300 | 18400 | 18100 |
| Nicaragua | 9633 | 10860 | 11097 | 14690 | 17362 |
| Peru | 8324 | 9257 | 11657 | 13314 | 13425 |
| Others | 60811 | 62155 | 55737 | 57083 | 48470 |
| World | 1650255 | 2090115 | 2317134 | 2265346 | 2327534 |

## 1.1.2 Shrimp aquaculture and the processing industry

The shrimp industry is composed of two main sectors, namely shrimp production and the supply of shrimp feed and equipment. Shrimp production starts at a hatchery, then shrimp larvae are grown on at a farm. The shrimp will then be distributed and some of them processed in a seafood plant. Feed, chemical supply and farming equipment manufacture are related activities that also comprise the shrimp production industry. The shrimp production chain in Thailand is shown schematically in Fig 1.8.
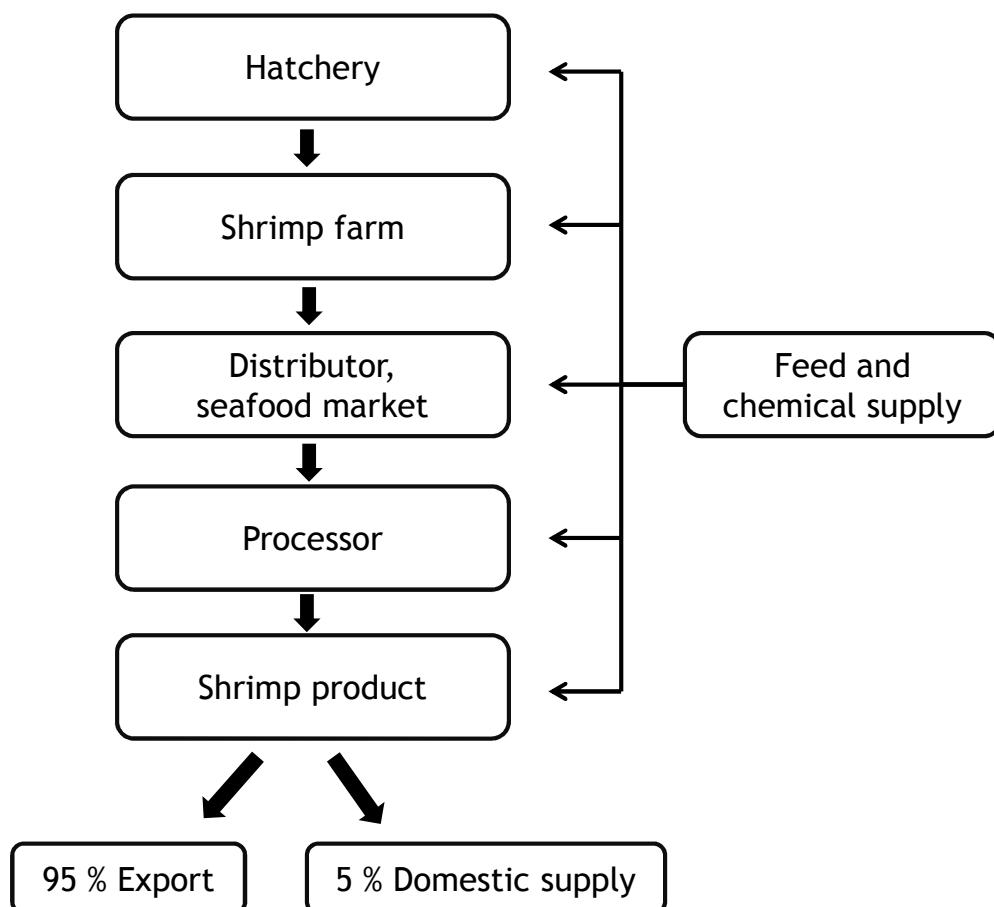
**Figure 1.8.** Schematic of the shrimp production industry in Thailand

**1.1.2.1 Hatcheries**

Approximately 4,000-5,000 shrimp hatcheries are scattered over the shrimp farming area in the East and South provinces, such as Phuket, Chachengsao and Chonburi (Fig. 1.9). Shrimp hatcheries operate intermittently through the year, and the number of active hatcheries at any one time depends on the prevailing market situation. When the shrimp industry is less profitable this number could be as low as 1,000 (Theptaranon *et al.*, 2005). A field trip to a typical shrimp hatchery, the Chen shrimp hatchery, Phuket province, in the South of Thailand (Figs. 1.10-1.12) was carried out as part of the current project, in order to gain first-hand information about the procedures followed.



**Figure 1.9.** The regions in Thailand with shrimp hatcheries and shrimp farms (indicated by arrows)
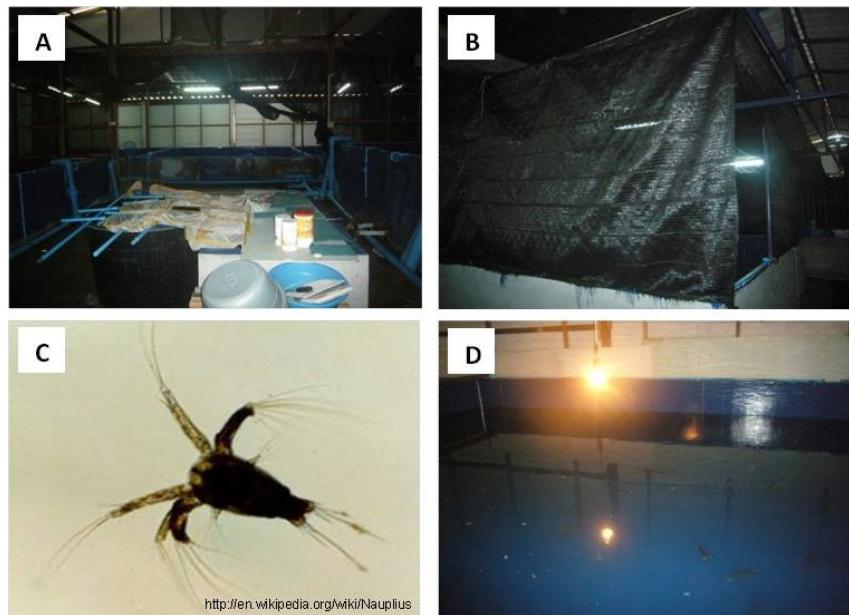
**Figure 1.10.** The Chen shrimp hatchery, Phuket province, Thailand. Pictures represent (A and B) dark atmosphere provided for holding broodstock and nauplius, (C) nauplius stage of *P. vannamei* and (D) a broodstock pond.
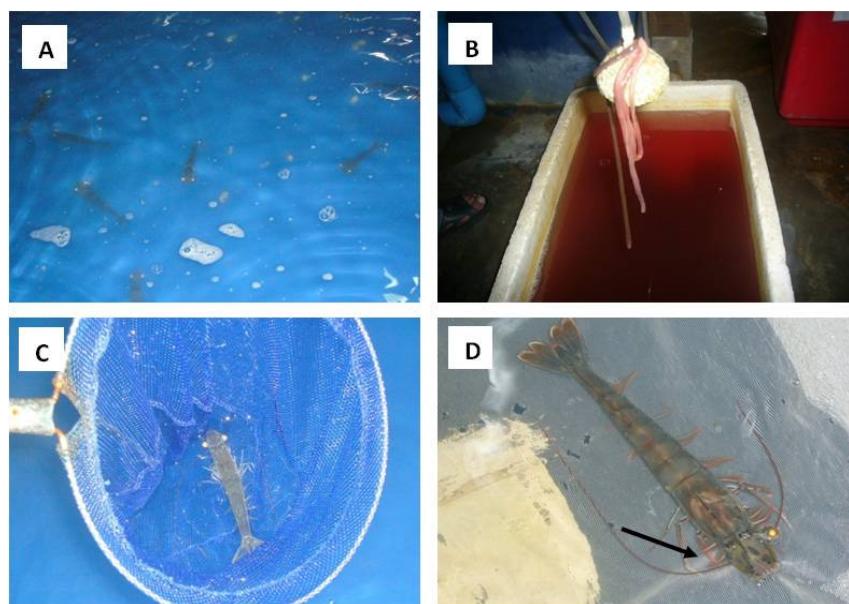


**Figure 1.11.** The Chen shrimp hatchery, Phuket province, Thailand. Pictures represent (A) broodstock pond, (B) sea teredos used as nutritious broodstock feed, (C) Florida broodstock and (D) a female broodstock with one eye removed (arrow indication).

**Figure 1.12.** The Chen shrimp hatchery, Phuket province, Thailand. Pictures represent (A and B) post larvae culture tanks and (C) post larvae at stage of 10 days and (D) 13 days.

In a hatchery, the broodstocks of black tiger shrimps, *P. monodon*, and white shrimps, *P. vannamei*, are obtained differently. Black tiger shrimp broodstocks are caught by fishing boat either from the Andaman Sea or the Gulf of Thailand, whereas white shrimp broodstocks are imported from the USA, mostly from Hawaii and Florida. The broodstocks are fed on sea worms to provide a high level of nutrition. To increase shrimp productivity, one eye of the females is removed (Fig. 1.11D), thus affecting their endocrine balance, leading to an acceleration of egg production. When they are in reproductive condition, male and female broodstocks are put together in a pond for fertilization. After that, the females are separated for laying eggs. After the offspring are released there are four main developmental stages of shrimp progeny: nauplius (2 days), zoea (4-5 days), mysis (3-4 days) and post-larvae (10-15 days), respectively. Post-larval shrimps will be purchased by farmers at a price that depends on their size and according to the period of the post-larval stage from day 8 to day 15. Some

hatcheries have certified specific pathogen-free (SPF) shrimp stocks available. To determine this, pre-stage larval samples will be sent to a local research institute for shrimp pathogen examination by DNA diagnostic methods such as polymerase chain reaction (PCR). As described above, the critical shrimp pathogens in Thailand are various viruses, notably white spot syndrome virus (WSSV), yellow head virus (YHV), hepatopancreatic virus (HPV), monodon baculovirus (MBV), and infectious hypodermal and haematopoietic necrosis virus (IHHNV) (Flegel, 2006; Morakot *et al.*, 2008). However, the presence of seafood-borne bacterial pathogens such as *V. parahaemolyticus* in shrimp culture has not been examined to the same extent. Therefore, bacterial examination of shrimp culture and animal feed can be expected to contribute significantly to the understanding of microbial threats to shrimp production. Previous studies reported a role of *V. parahaemolyticus* in bacterial diseases (e.g. vibriosis) in shrimp (Brock & Lightner, 1990; Jayasree *et al.*, 2006; Nash *et al.*, 1992). However, unlike *Vibrio harveyi*, a major bacterial pathogen causing vibriosis in shrimp and marine fish (Austin & Zhang, 2006), the importance and virulence of *V. parahaemolyticus* to shrimp have not been defined.

### 1.1.2.2 Shrimp farms

Post larval shrimps are grown on at a shrimp farm. Generally, shrimp farms can be divided into four types: extensive, semi-intensive, intensive, and super-intensive (Dennis & Bob, 1992). Extensive farms are constructed as low tide level ponds in order to allow wild post larvae to flow into the farmed area when a water gate is opened. Then the post larvae are grown to a market size in the farm. Since the shrimps feed on natural organisms and are surrounded by naturally-growing vegetation under uncontrolled conditions, production on

extensive farms is at a low intensity. As a result, each farm can typically produce only 50-500 kilograms per hectare per year, which is the lowest yield of any of the farming types.

Semi-intensive operations are conducted at the high tide level, with approximately 2-30 hectares of farming area, which is carefully designed to grow out a stock of 100,000-300,000 post larvae per hectare. Farmers create an adequate food supply and fertile conditions in the pond by additional feeding and aeration. With such an equipped system, the yield range is ten times higher than in the extensive farm operations. Most of the shrimp farms in South America and China are operated as such semi-intensive systems.

Intensive shrimp farming is conducted in a small area either indoors or outdoors, with high stocking densities of more than 300,000 post larvae per hectare. It is well-managed by heavy feeding, waste removal and an aeration system. The shrimp stock is released to a nursery pond then transferred to the second nearby pond on an around-the-clock system when the stocks approach the juvenile stage. Sub-adult and adult shrimp will subsequently be transferred into other ponds. Each pond has a different feed supply and aeration rate to suit a certain developmental stage of the shrimp stock. Efficient harvesting and a pond cleaning technique enable the farms to have year-round production. Intensive farming is extensively used for shrimp production in tropical countries such as Thailand, Indonesia and Taiwan. However, super-intensive farming, which is characterised by a greater control of the resource management such as the water circulating system (no water exchange) and enclosure in greenhouses so they are more bio-secure and ecologically friendly than intensive farms, can produce the highest yield of 100,000 kilogram per hectare per year.

On a field trip in Thailand, the operation of an intensive shrimp farm and a super-intensive shrimp farm was observed by visiting the Jutha shrimp farm, Pang-nga province (Fig. 1.13), in the South of Thailand and the Roiphet Chareon Pokphan Food farm, Trat province, on the East coast of Thailand (Fig. 1.14). The Jutha shrimp farm is an intensive white shrimp farm that has operated to supply local fish markets and a small seafood factory. In contrast, the Roiphet Chareon Pokphan Food farm, abbreviated hereafter to the Roiphet Farm, is a super-intensive white shrimp farm which is operated by one of the largest agro-product companies in Thailand named Chareon Pokphan Thailand (www.cpthailand.com). The Roiphet Farm is a bio-secure, environmentally-friendly, closed system farm with an investment of over one billion Thai baht (£20 million). Approximately 600,000 kilograms of high quality white shrimp are produced per year, with a certified food safety procedure for developed markets such as the European Union (European commission, 2007). All output is supplied to its own processing plant as part of an integrated shrimp production system within the one company.

### 1.1.2.3 Distributors

Shrimps are dispatched from farms either directly to the fish markets or to seafood processing plants. The shrimps are topped with ice during transportation and remain in the ice until they are presented to purchasers in a local market. Live shrimps are transported in a tank equipped with an oxygen supply (Fig. 1.15).
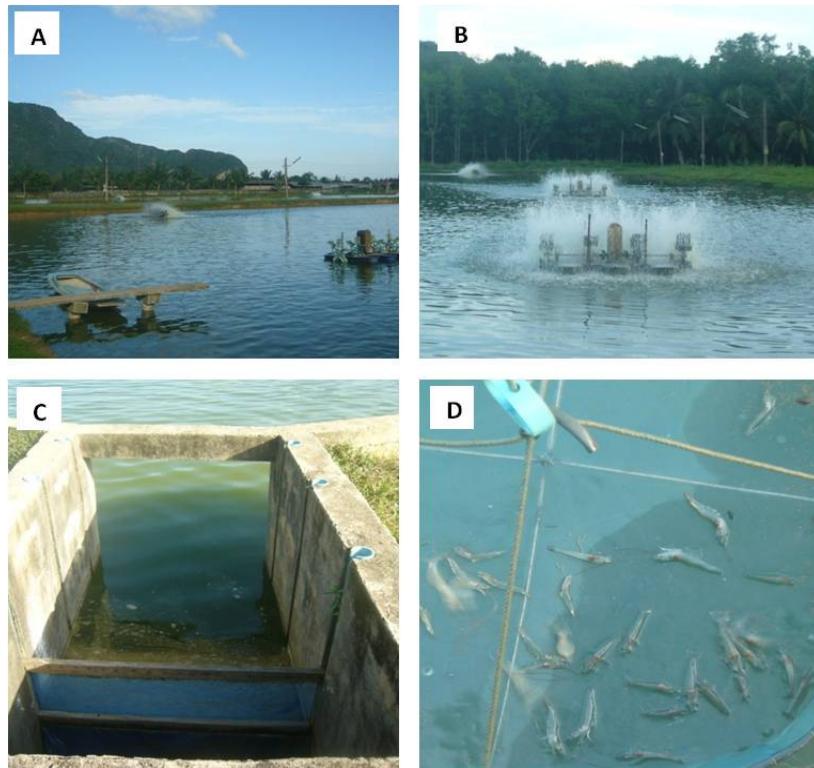
**Figure 1.13.** Intensive shrimp farm, the Jutha shrimp farm, Pang-nga province, Thailand. Picture represents (A and B) shrimp pond equipped with aerating system, (C) water gate used to drain the water from shrimp growing pond before starting the new batch of shrimp culture and (D) farmed shrimps collected by fishing net.

Thailand has tropical weather and the air temperature is 30-37$^{o}$C throughout the year. Thus, temperature is a critical factor for microbial control. Without precise temperature control the long distances between the farms and the markets can generate more risk of microbial growth in shrimp meat. In addition, shrimps for sale at local markets tend to have more risk of microbial contamination as they will not be washed with disinfectant solution (chlorine) after arriving, whereas the seafood manufacturers do disinfect the shrimps before passing products to a processing line.

**Figure 1.14.** Super-intensive-closed system shrimp farm, the Roiphet Chareon Pokphan Food farm, Trat province, Thailand. Pictures represent (A) hygienic practice by washing vehicle wheels with antibiotic solution before entering the farm, (B) circulating water used in the farm is detached from soil by nylon sheets in order to prevent microbial contamination and (C and D) farming area with the closed system for optimal shrimp growing conditions such as temperature, humidity, oxygen, pH and salinity of water.

### 1.1.2.4 Shrimp post-harvest treatments and transportation

After shrimps arrive at a processing plant they are sized and sent through the processing line (Fig. 1.16). Factors such as the quality of the washing water, hygiene and sanitation measures, temperature during processing and distribution, temperature of storage and handling and cross-contamination affect product safety in terms of microbial contamination (WHO & FAO, 2011).

**Figure 1.15.** Transportation of post harvest shrimp at the Charoen Pokphan shrimp industry, Trat province, Thailand. (A and B) Farmed shrimps are transported to processing factory in containers by trucks. (C) Dead shrimps are preserved in ice before processing and (D) live shrimps are supplied with oxygen in the sea water before processing.



**Figure 1.16.** Shrimp handling practice by workers at the shrimp processing plant, the Charoen Pokphan shrimp industry, Trat province, Thailand. After transportation, (A and B) shrimps are drained for 20 min and (C and D) the healthy shrimps are sized manually for processing. These processes are done at ambient temperature.

To control the freshness of seafood material, most of the factories are located near to the shrimp farming area. As an example, the Charoen Pokphand Foods seafood plant, which is established under the CPF company group, operates in Rayong province at 100 kilometres distance from the company's own farm, the Roiphet shrimp farm, Trat province. Also the Kingfisher seafood processing plant (www.kingfisher.co.nz), one of the visited seafood plants on the field trip, is located in the Samutprakarn province of central Thailand where shrimp farms are widely operated.

Chemical and feed supply companies are involved in shrimp aquaculture by producing biological and chemical feeds with effective formulations. Feed additives are important for particular purposes, for example carotenoid-enriched feed supplement can enhance the orange colour of shrimps after cooking. Also feed supplements containing probiotics introduce beneficial bacteria to the shrimp in culture pond ecosystems (Dalmin *et al.*, 2001). Aquaculture equipment can be designed to minimize the environmental effects generated from shrimp farms.

Information from a field trip study of the shrimp (*P. vannamei*) industry in Thailand and a review of shrimp aquaculture provides an understanding of seafood production in the country. This knowledge provides a background for studying the risk of contamination by foodborne pathogenic bacteria in the seafood production chain, which is the aim of the research reported in this thesis. The main focus was on *V. parahaemolyticus,* a seafood-borne pathogenic bacterium which is a major cause of gastroenteritis worldwide.

## 1.1.3 V. parahaemolyticus infection in respect to shrimp aquaculture and the supply chain

At the pre-harvest and harvesting stages, the crucial parameters that influence *V. parahaemolyticus* density are water temperature and salinity, air temperature, tide, and plankton (Codex, 2003; Kumazawa *et al.*, 1999; Sarkar *et al.*, 1985). Since this bacterium is more abundant in regions having warm water temperatures, the geographical location and seasonal factors can be indicators of the level of *V. parahaemolyticus* at harvest. In temperate climates such as in the USA, water and air temperature at harvesting time are the major factors influencing the initial level of pathogenic *V. parahaemolyticus* in oysters (FDA, 2005). However, the seasonal prevalence of *V. parahaemolyticus* is not significantly different in tropical countries, including Thailand. Thus, temperature control during transportation is likely to be a critical factor influencing the growth of *V. parahaemolyticus* in the Thai seafood production chain.

After harvesting, intervention strategies such as minimising the period between harvesting and chilling can help to reduce the level and prevent the growth of *V. parahaemolyticus*. Furthermore, the harvesting method used in different fishing areas also affects the level of *V. parahaemolyticus* after harvest (FDA, 2005). For example, within the Gulf coast of the USA, Louisiana has higher predicted numbers of illnesses compared with the other states in this region. This is because the harvesting boats in Louisiana are typically on the water for longer, and thus the seafood is kept without refrigeration for prolonged periods. However, *V. parahaemolyticus* comprises both non-pathogenic and pathogenic strains (as discussed in section 1.2.4 of this chapter). When considering risk, it is

appropriate to emphasise levels of the pathogenic strains of *V. parahaemolyticus* as this is the actual causative agent of the illness from this bacterium. For example, the incidence of pathogenic *V. parahaemolyticus* in the Pacific Northwest is higher than in the Gulf coast of USA, thus the at-harvest control criterion based on total *V. parahaemolyticus* in the Pacific Northwest should be more stringent than that from the Gulf coast (FDA, 2005). The level of *V. parahaemolyticus* at the point of consumption has been evaluated for oysters in the USA by the level of pathogenic strains associated with typical serving portions (FDA, 2005). Nevertheless, this evaluation may vary depending on seafood species, culture of consumption, and serving size in each particular region.

A previous study detected *V. parahaemolyticus* in healthy workers who work in shrimp farms in the South of Thailand (Assavanig *et al.*, 2008) and in the workers in a seafood processing plant in central Bangkok (Athajariya, 2004). The latter study showed that virulence genes (*tdh* and *trh*) (see section 1.2.4 of this chapter) of *V. parahaemolyticus* isolates were determined by multiplex PCR. Both the *tdh* and *trh* genes ($tdh^+/trh^+$) were found in 4.8% of the isolates, 25.3% contained only *tdh* ($tdh^+/trh^-$), 4.8% contained only *trh* ($tdh^-/trh^+$) and 65.1% had neither virulence gene ($tdh^-/trh^-$). These results indicate that potential virulent strains were detected from healthy carriers who presented no symptoms of gastroenteritis. A study of this suppressive condition of pathogenic *V. parahaemolyticus* in these carriers is required to show whether factors such as human immunization and alternative pathogenic forms can contribute to the survival of *V. parahaemolyticus* in carriers. However, these studies indicate the possibility that human carriers could be a source of bacterial transmission both between and from shrimp aquaculture sites. As the observation from the field

trip study in Thailand, farmers in locally-operated farms such as the Jutha shrimp farm may be at more risk of receiving *V. parahaemolyticus* infection than those in the large commercial-scale farms such as the Roiphet Farm. This is because the advanced equipment in super-intensive farms enables the farmers to manage the shrimp culture system without having much human contact with the environment, while farmers in locally-operated farms have more chance of handling cultured shrimps directly, thus increasing the risk of contamination.

According to the Codex (2002) discussion paper on risk management strategies for *Vibrio* spp. in seafood, more information about seafood transportation is needed to develop further food microbial risk strategies. For example, studies on the growth and survival of pathogenic *V. parahaemolyticus* in shrimp at various temperatures can be used to determine a critical control point for shrimp transportation. Moreover, examination of the samples for bacteria at different steps in the shrimp production process, such as fresh/frozen shrimp meat, water from shrimp farms, and stool sample from workers in seafood factories (bacterial carriers) are useful to study strain variation and molecular epidemiology of *V. parahaemolyticus* in the shrimp production chain. Closing these data gaps could enhance the quality control scheme, which is the strength of the Thai shrimp industry in the highly competitive global industry of shrimp exportation.

## 1.1.4 Seafood safety and risk assessment of V. parahaemolyticus

In Thailand, food safety schemes have been implemented both within the farms and in the seafood processing factories. Good Aquaculture Practice (GAP) is a minimal requirement for shrimp farm management. Under the GAP scheme the farms are assessed in terms of hygiene practice, regulation of antibiotic usage

and environmental practice laws (Department of Fisheries, Thailand, 2007). In addition, a Code of Conduct (CoC) has been adapted from GAP practice, which also covers social responsibility, the involvement of all stakeholders in the production chain and complete traceability of the products. Compliance with this CoC is a full requirement for farm management, harvesting and processing for premium grade seafood. In 2007, most shrimp farms operated in Thailand were approved by GAP, whereas only 274 farms (1-2%) were certified by both GAP and CoC (Anonymous, 2007).

For post-harvest handling and seafood processing, a Good Manufacturing Practice (GMP) scheme has been implemented to maintain product quality control. Moreover, Hazard Analysis and Critical Control Points (HACCP) is an effective approach for food safety inspection, including bacterial examination, and for protecting public health. However, microbiological assessments of seafood products differ depending on the particular seafood purchasers. For example, the European Union (EU) requires a maximum recommended count for *V. parahaemolyticus* of $10^3$ most probable number (MPN) per gram ($g^{-1}$) in cooked molluscs and shellfish, whereas the USA requires $10^4$ MPN $g^{-1}$ maximum for cooked crustacean products (Anonymous, 2009). The country where raw seafood is most widely consumed, Japan, requires zero MPN $g^{-1}$ detection of *V. parahaemolyticus* in raw seafood products including fish, crustaceans, molluscs bivalves, etc (Anonymous, 2009). In addition, food safety control in international food trading was considered by the Codex Alimentarius Commission, abbreviated hereafter to the Codex. The scheme launched by the Codex requires hygienic practices to be followed at all points in the production chain, e.g. seafood needs to be stored below $10^{\circ}C$ throughout the distribution, and in addition fish and shellfish have to be washed with disinfected potable water (Codex, 2003).

According to the Codex discussion on risk management strategies for *Vibrio* spp. in seafood, the major potential causes of *V. parahaemolyticus* infection have been identified as pathogen uptake by fish and shellfish from environmental waters, exposure of the bacteria at the time of harvesting, and improper handling practices after harvest (Codex, 2003).

## 1.2 Characteristics of *V. parahaemolyticus*

*V. parahaemolyticus* infection from contaminated undercooked seafood has been a public health problem as mentioned in section 1.1. Understanding the background of *V. parahaemolyticus* is necessary for epidemiological and molecular evolutionary studies of this organism. Detailed characteristics including historical background, classification and taxonomy, colony morphology and virulence factors of *V. parahaemolyticus* are discussed in this section. In particular, the properties and functions of the major virulence factors, the haemolysins and type three secretion systems (TTSSs), are extensively described for a better understanding of the virulence mechanism of *V. parahaemolyticus*.

### *1.2.1 Historical background V. parahaemolyticus*

*V. parahaemolyticus* was first identified in 1950 from patients presenting with gastroenteritis in Osaka, Japan. The illness was due to the consumption of undercooked salted sardines, called Shirasu (Fujino *et al.*, 1953). The bacterium was also isolated in 1953 as a mixed infection with *Proteus morganii* from stools and intestinal contents of patients. From this isolation, it was first named *Pasteurella parahaemolyticus*. Subsequently, this bacterium was isolated from stool samples of patients with food poisoning in an outbreak at Yokohama National Hospital in 1958 (Takikawa, 1958). In 1960, *Oceanomonas*

*parahaemolyticus,* a halophillic and glucose fermentative bacterium, was isolated from humans and also from the marine environment. However, the Japanese Ministry of Health and Welfare indicated that *P. parahaemolyticus* and *O. parahaemolyticus* are the same organism according to morphological, cultural and chemical analyses. From the report of the International Symposium of *V. parahaemolyticus*, Tokyo in 1974, the organism was reclassified into the genus *Vibrio* and named *Vibrio parahaemolyticus* (Fujino *et al.*,1974).

## 1.2.2 Classification and taxonomy

The family *Vibrionaceae* was first described in 1965 (Janda *et al.*, 1988). The organisms included in this family are generally rod shaped and have an appearance similar to that of other Gram-negative bacteria residing in aquatic habitats. The genus *Vibrio* is considered to be the most species of Gram-negative bacteria residing in aquatic habitats.

*V. parahaemolyticus* belongs to the Genus *Vibrio*, the Family *Vibrionaceae*, the Order *Vibrionales*, the Class *Gamma Proteobacteria*, and the Phylum *Proteobacteria* (Farmer & Janda, 2004) (Fig. 1.17). Of 34 important *Vibrio* species described by Janda *et al.* (1988), one third of these species are known to be human pathogens (Table 1.2). A number of non-human pathogenic species, including *Vibrio anguillarum, Vibrio fischeri* and *Vibrio harveyi* are pathogens of marine fish and shellfish species. Phylogenetic relationships of bacteria within the genus *Vibrio* have been determined by different schemes. Tian *et al.* (2008) proposed that sequence analysis of the *gyrB* gene is more suitable for determining the phylogenetic relationships of the *Vibrios* than sequence analysis of the 16S rRNA gene. The phylogenetic relationships of *Vibrios* and related species based on the maximum likelihood method using multilocus nucleotide

sequences including *ftsZ, gyrB, mreB, pyrH, recA, rpoA* and *topA* was analyzed by Thompson *et al.* (2009) (Fig. 1.18). These authors compared phylogenetic analyses using multilocus nucleotide sequences and other schemes including 16S rRNA, supertrees, average amino acid identity, genomic signatures and genome BLAST atlas, and suggested that a combination of different bioinformatic tools will enable the most accurate species identification and understanding of the genomic taxonomy of *Vibrio* species (Thompson *et al.*, 2009).

Kingdom *Bacteria*

└── Phylum *Proteobacteria*

          └── Class *Gamma Proteobacteria*

                    └── Order *Vibrionales*

                              └── Family *Vibrionaceae*

                                      └── Genus *Vibrio*

                                            └── Species *Vibrio parahaemolyticus*
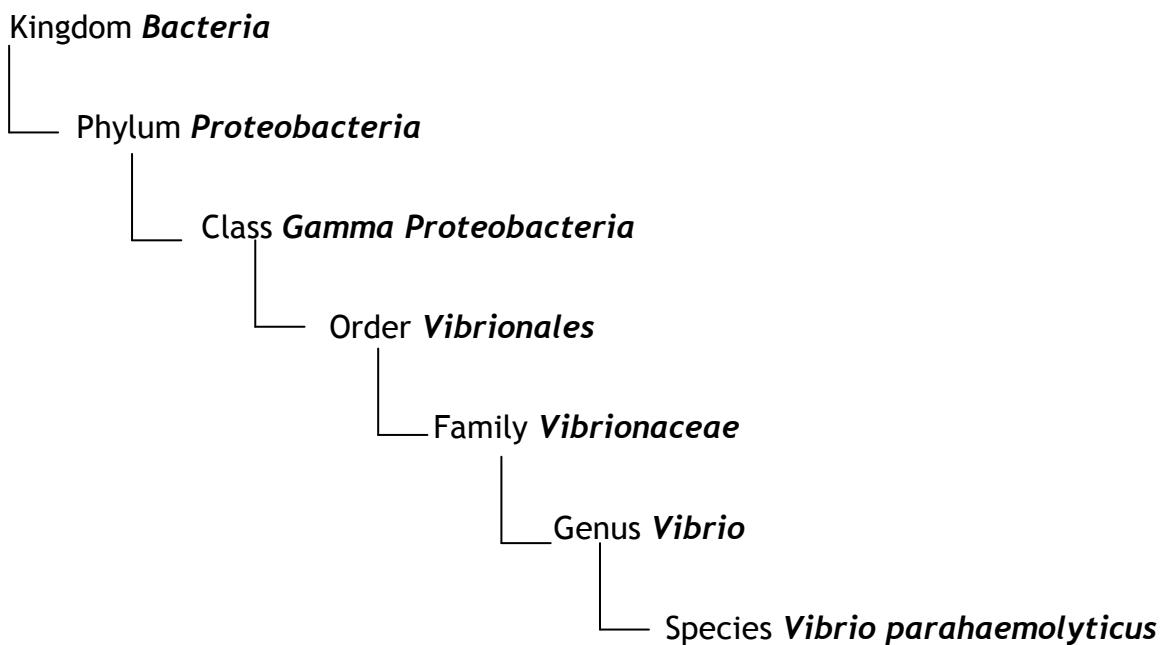
**Figure 1.17.** Classification of *V. parahaemolyticus* described by Madigan *et al. (*2005)

**Table 1.2.** Composition of the genus *Vibrio*. Table adapted from Janda *et al*. (1988)

_____

| Human pathogens | Non-human pathogens |
|---|---|

_____

| Human pathogens | Non-human pathogens |
|---|---|
| *Vibrio alginolyticus* | *Vibrio aestuarianus* |
| *Vibrio chorelae* | *Vibrio anguillarum* |
| *Vibrio cincinnatiensis* | *Vibrio campbellii* |
| *Vibrio damsela* | *Vibrio carchariae* |
| *Vibrio fluvialis* | *Vibrio costicola* |
| *Vibrio furnissii* | *Vibrio diazotrophicus* |
| *Vibrio hollisae* | *Vibrio fischeri* |
| *Vibrio metschnikovii* | *Vibrio gazogenes* |
| *Vibrio mimicus* | *Vibrio harveyi* |
| *Vibrio parahaemolyticus* | *Vibrio logei* |
| *Vibrio vulnificus* | *Vibrio marinus* |
| | *Vibrio mediterranei* |
| | *Vibrio natriegens* |
| | *Vibrio nereis* |
| | *Vibrio nigripulchritudo* |
| | *Vibrio ordalii* |
| | *Vibrio orientalis* |
| | *Vibrio pelagius* |
| | *Vibrio proteolyticus* |
| | *Vibrio psychroerythrus* |
| | *Vibrio salmonicida* |
| | *Vibrio splendidus* |
| | *Vibrio tubiashii* |

_____
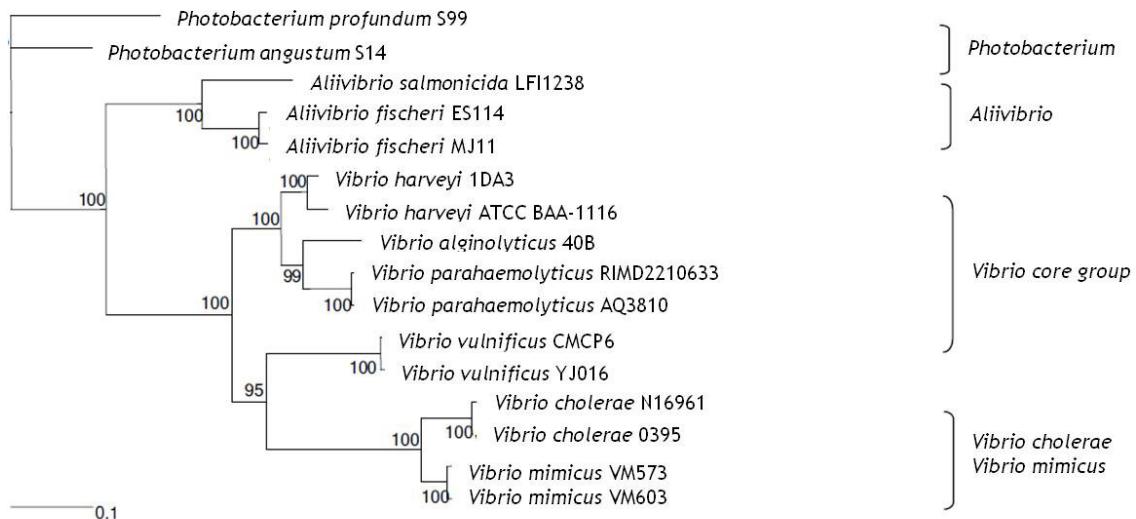
**Figure 1.18.** Phylogenetic tree of *Vibrios* and related species based on the maximum likelihood method using multilocus nucleotide sequences including *ftsZ, gyrB, mreB, pyrH, recA, rpoA* and *topA*. Multilocus sequence tree representing four distinct groups, *Photobacterium* spp., *Aliivibrio* spp., *Vibrio* core group, and *Vibrio cholerae - Vibrio mimicus* group. Figure adapted from Thompson *et al*. (2009)

## *1.2.3 Cell and colony morphology*

*V. parahaemolyticus* is a halophilic, Gram-negative rod-shaped bacterium 0.5-0.8 x 1.4-2.6 µm in size. The optimal growth conditions of *V. parahaemolyticus* are 35-37$^{o}$C, pH 7.5-8.0 and approximately 0.5 M NaCl (Joseph *et al.*, 1982). The colony morphology of *V. parahaemolyticus* is variable. Multiple colony morphotypes can occur in colony descendants from a single isolate. Moreover, the colony types can switch reversibly from translucent (TR) to opaque (OP). The switching mechanism is believed to be a response to specific environmental conditions (McCarter, 1999). *V. parahaemolyticus* are highly competent in biofilm formation although the biofilm structures are developed differently in TR and OP strains (Enos-berlage *et al.*, 2005). In biofilms of TR strains, tall pillars are loosely interspersed with open channels whereas the biofilms of OP strains are more uniform, dense and lack such channels. Biofilm formation of *V. parahaemolyticus* is regulated by the chitin-regulated pilus (ChiRP) and mannose-sensitive hemagglutinin (MSHA) pilus (Shime-Hattori *et al.*, 2006).

*V. parahaemolyticus* possesses multiple cell types, an adaptation for survival under different circumstances. In a liquid environment, free-living organisms called swimmer cells exist which have a single polar flagellum. Growth on a surface or in a viscous environment induces cell differentiation into swarmer cells. The swarmer cells possess peritrichous flagella which are well-adapted to produce movement in a highly viscous environment (McCarter, 1999).

Metabolic adaptation of *V. parahaemolyticus* enables the organism to survive under stressful conditions. Jiang & Chai (1996) found that the morphology of *V. parahaemolyticus* changes from rod-shaped to spheroid after one week of starvation at 3.5$^{o}$C. These spheroid-shaped cells survived but were unable to

grow in growth media, and were hence designated as viable but non culturable (VBNC) cells. The authors suggested that resuscitation of the VBNC cells can occur when the temperature is increased or favourable conditions return.

## 1.2.4 Virulence factors of V. parahaemolyticus

Virulence factors of *V. parahaemolyticus* include haemolysins, virulence genes in pathogenicity islands (PIs), type three secretion systems (TTSSs), colonizing factors and outer membrane proteins (OMPs). The haemolysins, PIs-related genes and TTSSs are well-known virulence factors in this species and details of these are discussed below. Adherence of *V. parahaemolyticus* to host cells is associated with colonizing factors including pili (Nakasone & Iwanaga, 1990; Nakasone *et al.*, 2000), capsular polysaccharide (CPS) (Hsieh *et al.*, 2003) and OMPs. OMPs have significant roles in *V. parahaemolyticus* virulence. They are involved in adherence as well as other activities including resistance to stressful conditions (e.g. bile acid in human intestine) and iron uptake. Thus, OMPs contribute to survival of *V. parahaemolyticus* within the human host. In the present study, comparative outer membrane proteomic analysis was used to compare OMP expression in isolates from different origins, including clinical and environmental sources. Basic information on the structure and classification of OMPs in Gram-negative bacteria are discussed separately in section 1.5 of this chapter and the roles of OMPs in relation to *V. parahaemolyticus* virulence are discussed in chapter 4.

## 1.2.4.1 Haemolysins

The strains isolated from diarrhoeal faeces of patients with gastroenteritis are mostly haemolytic, whereas isolates from the environment are usually non-haemolytic. Haemolysis of *V. parahaemolyticus* is visualized by the lysis of human or rabbit erythrocytes on Wagatsuma agar (Chun *et al.*, 1975). This haemolysis has been termed the 'Kanagawa Phenomenon' after the original discoverers, the Kanagawa Prefectural Public Health Laboratory, Japan.

Kanagawa Phenomenon positive strains (KP-positive) produce a thermostable direct haemolysin (TDH). Another type of haemolysin, thermostable direct haemolysin-related haemolysin (TRH), has been detected in clinical Kanagawa negative strains (KP-negative) (Honda *et al.*, 1988; Janda *et al.*, 1988; Miyamoto *et al.*, 1969). Honda & Iida (1993) demonstrated that TDH and TRH are capable of causing pathogenesis of *V. parahaemolyticus*. A comparison of the properties of TDH and TRH is shown in Table 1.3. Although TDH and TRH are the most well-researched haemolysins of *V. parahaemolyticus,* a thermolabile or lecithin-dependent haemolysin (LDH) and a heat-stable haemolysin (ð-VPH) have also been described in this organism (Taniguchi *et al.*, 1986, 1990).

**Table 1.3.** Comparison of TDH and TRH toxins of *V. parahaemolyticus*. Adapted from Honda & Iida (1993)

| Property | TDH | TRH |
|---|---|---|
| **Molecular weight:** | | |
| - Holo toxin | 46,000 | 47,000 |
| - Subunit | 23,000 | 23,000 |
| **PI** | 4.9 | 4.6 |
| **Heat stability** | Stable at 100°C | Labile at 60°C |
| **Antigenicity** | Related but not identical to that of TRH | Related but not identical to that of TDH |
| **Amino acid sequence** | 67% homology to amino acid sequence of TRH | 67% homology to amino acid sequence of TDH |
| **Biological activity:** | | |
| - Haemolytic activity | Rabbit, human > calf, sheep > horse | Calf, sheep > rabbit, human > horse |
| -Lethal activity (mouse) | cardiotoxicity | cardiotoxicity |
| -Fluid accumulation in rabbit ileal loop (RIL) | 250µg/loop | 100µg/loop |

### *1.2.4.1.1 Thermostable direct haemolysin* (TDH)

TDH is a haemolysin produced by KP-positive strains. Purification and characterization of TDH from *V. parahaemolyticus* cultures was carried out by Sakurai *et al*. (1973). The haemolysin was considered as a thermostable form since it was not inactivated after heating at 100°C for 10 min. The purified TDH was thought to be a protein as it was almost completely destroyed by proteinases such as pepsin, trypsin, alpha chymotrypsin and nagarse. Analysis by gel filtration indicated that TDH has a molecular mass of approximately 118 kilodalton (kDa). The TDH contains a large number of acidic amino acids, resulting in a relatively low isoelectric point of pH 4-5. This haemolysin exhibited high haemolytic activity on erythrocytes of rats, dogs, mice, and monkeys, moderate haemolytic activity on erythrocytes of humans, rabbits,

guinea pigs and chickens, slight haemolytic activity on erythrocytes of sheep and no haemolytic activity on erythrocytes of horses (Zen-Yoji *et al.*, 1971). Crystal structure determination and fuctional characterization of TDH revealed that attachment of the TDH molecule, via the phospholipid bilayer of the targeted cell membrane, allows water molecules to permeate the cell freely through the centre pore of the TDH structure (Yanagihara *et al.*, 2010).

Nashibushi & Kaper (1990) identified two *tdh* chromosomal gene copies, *tdh1* and *tdh2,* from clinical KP-positive *V. parahaemolyticus* strains. These two genes were characterized and assigned as *tdhS* (*tdh1*) and *tdhA* (*tdh2*) by Iida *et al*. (1990). Purification and characterization of these two genes suggested that *tdhA* is the structural gene for TDH and is primarily responsible for the haemolytic phenotype whereas *tdhS* contributes relatively little to extracellular TDH production (Nishibuchi & Kaper, 1990). Furthermore, Nishibushi & Kaper (1990) also identified another two *tdh* gene copies, *tdh3,* a chromosomal-borne gene, and *tdh4,* a plasmid-borne gene, from clinical KP-negative *V. parahaemolyticus*. The *tdh3* and *tdh4* genes are likely to be structural genes encoding new haemolysins, TDH/I and TDH/II, respectively, and these haemolysins were closely similar, but not identical to, TDH (Honda *et al.*, 1991; Nagayama *et al.*, 1995). Although it was still unclear about the expression of *tdh3* and *tdh4* at the RNA transcriptional level, TDH/I and TDH/II were capable of inducing fluid accumulation in ligated rabbit intestine, suggesting that TDH/I and TDH/II participate in enterotoxicity in KP-negative *V. parahaemolyticus* (Honda *et al.*, 1991; Nagayama *et al.*, 1995). Since *tdhA* is the structural gene for TDH, most authors have used *tdhA* primers for *tdh* gene detection in *V. parahaemolyticus* isolates. The expression of *tdh* is regulated by the virulence gene regulator protein ToxRS (Lin *et al.*, 1993).

### *1.2.4.1.2 Thermostable direct haemolysin-related haemolysin* (TRH)

The KP-negative strains of *V. parahaemolyticus* that are capable of causing gastroenteritis were first isolated from travellers in the Republic of Maldives (Honda *et al.*, 1987). Characterization of a new haemolysin that is similar to TDH but different in some physical properties was carried out by Honda *et al.* (1988). This new toxin was named the thermostable direct haemolysin-related haemolysin (TRH). TRH was shown to be a protein with an isoelectric point of pH 4.6 and immunological similarity to TDH. However, unlike TDH, TRH was labile on heat treatment at 60°C for 10 min, and showed differed lytic activity against erythrocytes of various animals compared with the activity caused by TDH (Table 1.3). Sequence variation of *trh* in different *V. parahaemolyticus* strains was examined by Kishishita *et al.* (1992). A haemolysin gene (*trh2*) that shared 84% sequence similarity to *trh* (subsequently named *trh1*), was identified in this study. Amplification primers used for *trh1* have generally been used for *trh* detection in *V. parahaemolyticus* isolates (Honda *et al.*, 1991; Kishishita *et al.*, 1992).

Suthienkul *et al.* (2005) demonstrated that possession of the *trh* gene was associated with urease activity. Four hundred and ninety eight clinical isolates from diarrhoea patients in Thailand were examined for the presence of *trh* and for urease activity. These authors found that all urease-positive strains possessed the *trh* gene. Conversely, the urease–negative strains did not contain the *trh* gene, suggesting that the urease-positive phenotype could be a putative virulence marker of *V. parahaemolyticus*. However, the urease gene cluster of *V. parahaemolyticus* is not involved in the regulation and expression of *tdh* and *trh* (Nakaguchi *et al.*, 2003). Possession of virulence genes in clinical KP-positive and KP-negative *V. parahaemolyticus* is demonstrated in Fig. 1.19.

### 1.2.4.1.3 Lecithin-dependent haemolysin (LDH)

All *V. parahaemolyticus* possess the lecithin-dependent haemolysin gene (*ldh*), and in fact the *ldh* gene has been used as a species-specific marker for *V. parahaemolyticus*. The nucleotide sequence of *ldh* has no homology with that of *tdh*, and that it is thermolabile with a nucleotide sequence of 1.5 Kilobase (kb) in length (Taniguchi *et al.,* 1986). The preprotein and the mature protein consists of 418 and 398 amino acids, with molecular weights 47.5 KDa and 45.3 KDa, respectively. The GC content of *ldh* is 47.6%, which is almost the same as the *V. parahaemolyticus* genome.

### 1.2.4.1.4 Heat-stable haemolysin (ð-VPH)

Tanigushi *et al*. (1990) identified an additional thermostable haemolysin (ð-VPH) gene from a KP-negative *V. parahaemolyticus* strain, the nucleotide and amino acid sequences of which had no homology with those of *tdh* and *ldh* of *V. parahaemolyticus*. The ð-VPH encoding gene in *V. parahaemolyticus* and related species was found to be present in all examined *V. parahaemolyticus* strains, and also in one strain of *V. damsela* (Tanigushi *et al.,* 1990).
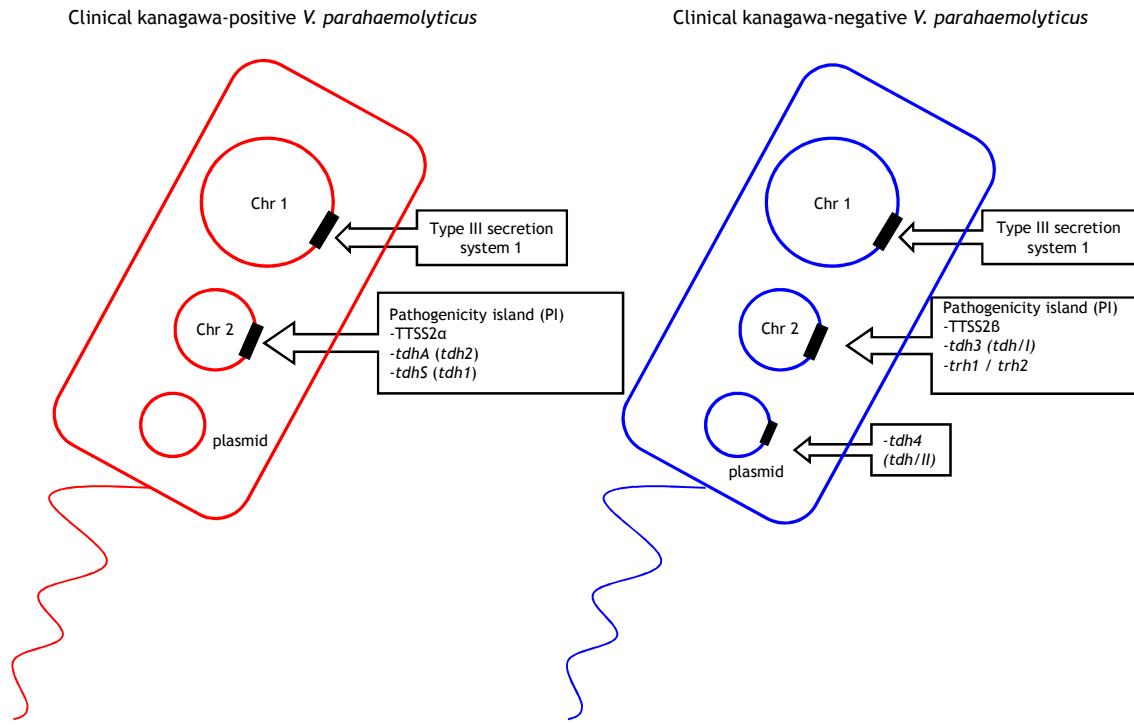
**Figure 1.19.** Possession of virulence genes of clinical KP-positive and KP-negative *V. parahaemolyticus*. Clinical KP-positive strains contain genes associated with type III secretion system 1 in chromosome 1, genes associated with type III secretion system 2 and two copies of *tdh* genes, *tdhA* (*tdh2*) and *tdhS* (*tdh1*), in chromosome 2. Clinical KP-negative strains contain genes associated with type III secretion system 1 in chromosome 1, genes associated with type III secretion system 2, *tdh3* (*tdh/I*), *trh1* or *trh2* in chromosome 2, and *tdh4* (*tdh/II*) in the plasmid.

## 1.2.4.2 Pathogenicity islands (PIs)

A genomic island is a mobile genetic element that can be transferred across bacterial strains or species. Genomic islands that contain virulence-related genes and some antibiotic resistance genes are classified as pathogenicity islands (PIs). A PI can be used as a marker in the molecular diagnostics of pathogenic identity in bacteria (Oelschlaeger & Hacker, 2004). PIs also play an important role in the evolution of bacterial virulence via the process of horizontal gene transfer (HGT) (Dobrindt *et al.*, 2004).

Seven PIs in *V. parahaemolyticus*, namely VpaI1 - VpaI7 with size ranging from 10 kb to 81 kb, were identified in the *V. parahaemolyticus* genome using a bioinformatic approach (Hurley *et al.*, 2006). Examination of these VPals in 235 *V. parahaemolyticus* isolates from China indicated that VpaI-1 and VpaI-5 genes were specifically correlated with pandemic O3:K6 strains, whereas VpaI-7 and TTSS2 were associated with *tdh*-positive strain (Chao *et al.*, 2009a). Comparative genomic analysis using microarrays of pandemic, non-pandemic, and environmental *V. parahaemolyticus* identified the genes that are specifically present in pandemic strains (Izutsu *et al.*, 2008). These genes include 65 genes located in 11 chromosome regions. The authors suggested that evolution of pandemic strains occurred via multiple genetic events, including insertions of several large gene clusters. Moreover, a comparison of the genomes of pathogenic and non-pathogenic strains in that study showed that the nucleotide sequences of genes localized in 80 kb-pathogenicity island are conserved among KP-positive but not in KP-negative strains. This result indicated a strong association between the region of 80 kb-pathogenicity island and pathogenicity of *V. parahaemolyticus*.

## 1.2.4.3 Type three secretion systems (TTSSs)

A type III secretion system (TTSS) is a set of approximately 20 genes that are encoded together within a PI region. The TTSS is a mechanism that enables Gram-negative bacteria to secrete and inject virulence-related proteins into eukaryotic host cells via a needle-like structure (Fig. 1.20). TTSSs have been found in several pathogenic Gram-negative bacteria including *Yersinia spp.*, *Shigella spp.*, *Salmonella spp.*, *Vibrio spp.*, *Pseudomonas aeruginosa* and enteropathogenic *Escherichia coli* (EPEC) (Hueck, 1998).

The discovery of two type III secretion systems, type three secretion system 1 (TTSS1) and type three secretion system 2 (TTSS2), in *V. parahaemolyticus* was first made by Makino *et al*. (2003). The *V. parahaemolyticus* genome consists of two circular chromosomes of 3,288,558 bp and 1,877,212 bp. The entire genome contains 4,832 open reading frames (ORFs). The TTSSI operon is located in chromosome 1, whereas the TTSS2 operon is part of a 80 kb-PI and is located in chromosome 2. The sequence and organisation of the TTSS1-encoding genes are most similar to those of *Yersinia* species, whereas the TTSS2 gene cassette is not similar to any particular TTSS of any other species (Makino *et al*., 2003). However, the TTSS2-associated region in *V. parahaemolyticus* contains several virulence-related genes, including homologues of the *E. coli* cytotoxic necrotising factor, the *Pseudomonas* exoenzyme T and genes presented in the PI of *V. cholerae*. According to examination of TTSSs from various strains of *V. parahaemolyticus* by Makino *et al*. (2003), TTSS1 was detected in all tested strains whereas TTSS2 was found only in clinical KP-positive strains. The G+C content of the *V. parahaemolyticus* PI is lower (39.8%) than the average G+C content of the genome (45.4%), indicating that recent lateral transfer may have occurred in this region.
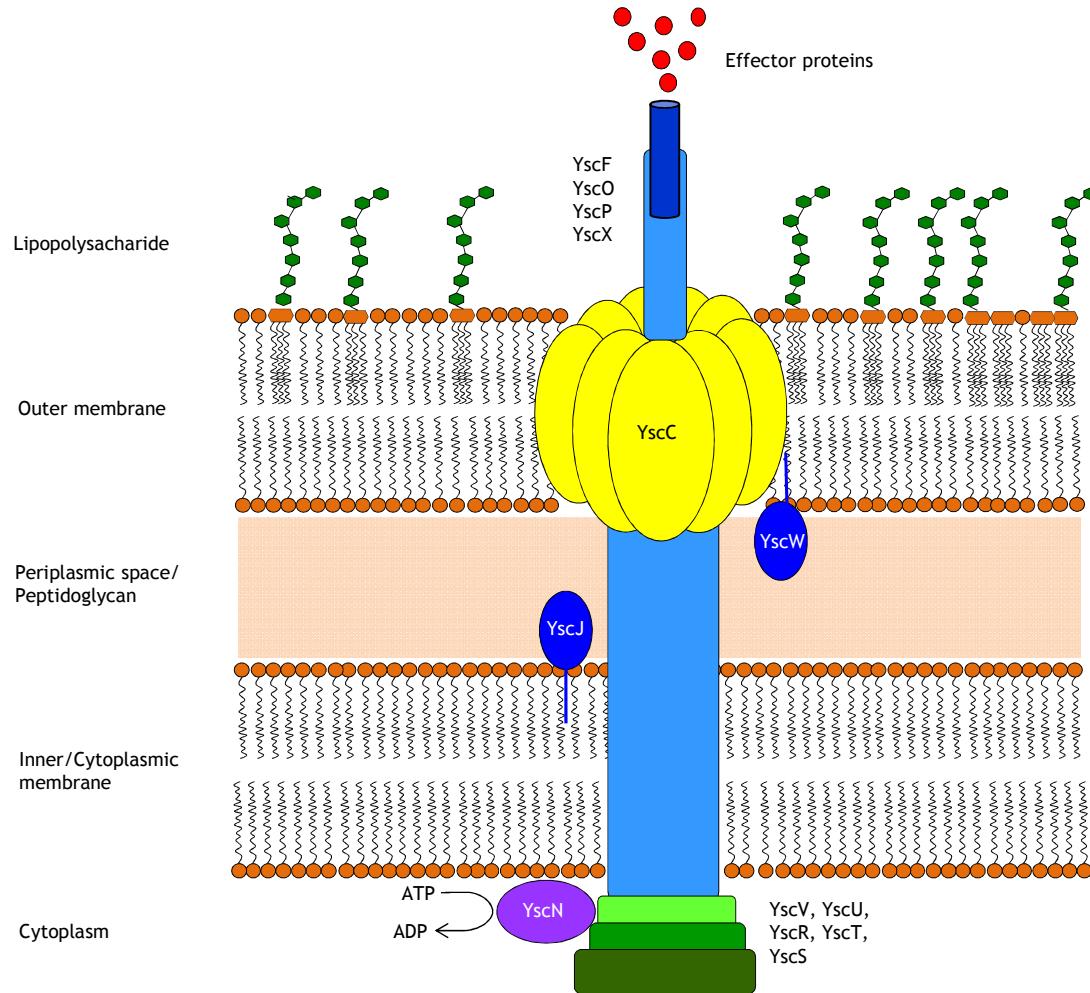
**Figure 1.20.** Schematic representation of the type III secretion system of *Yersinia spp.* YscF, YscO, YscP and YscX are protein components of the needle structure; YscC is the integral ring protein embedded in the outer membrane; YscW and YscJ are lipoproteins that anchor to the outer and inner membrane, respectively; YscV, YscU, YscR, YscT and YscS are protein components of the basal body which provide the contact surface to the cytoplasm; YscN is the ATPase that enables energy utilization for the secretion mechanism.

Functional characterization of putative *V. parahaemolyticus* TTSS1 and 2 were determined by disruption of TTSS1 involving genes, *vcrD1, vscC1,* and *vscN1* and TTSS2 involving genes, *vcrD2, vscC2,* and *vscN2* (Park *et al.,* 2004). The results showed that TTSS1 genes are associated with cytotoxicity whereas those of TTSS2 are associated with enterotoxicity of the host cell. Furthermore, VopD, a virulence-related protein encoded in TTSS1 with homology to YopD in *Yersinia spp.* and VopP, a protein encoded in TTSS2 with homology to YopP in *Yersinia spp.,* were found to be secreted by *V. parahaemolyticus* TTSS1 and TTSS2, respectively (Park *et al.,* 2004). The results indicated that two TTSSs in *V. parahaemolyticus* are responsible for distinct protein secretions. However, Meador *et al.* (2007) suggested that possession of TTSS2 may not be associated with pandemic strains of *V. parahaemolyticus* since they can also be found in *tdh*-negative strains. Conversely, some *tdh*-positive strains did not carry TTSS2. The authors suggested that TTSS2 may be acquired without the surrounding PI that contains two copies of *tdh* genes, or the *tdh* genes were mobile or lost from the PI. The roles of TTSS1 in cytotoxicity and TTSS2 in enterotoxicity were also demonstrated by Hiyoshi *et al.* (2010). In this study, the bacterial pathogenicity contributed by TTSS1 and TTSS2 in relation to the role of *tdh* were determined. The authors suggested that TTSS1 together with TDH may have an additional effect on virulence in mice. Only TTSS2, but not TTSS1 and TDH, is a major contributor to *V. parahaemolyticus*-induced enterotoxicity in a rabbit model. Moreover, secretion of TDH in a manner independent of both TTSS1 and TTSS2 was also demonstrated in this study. Microarray analysis of a TTSS1 deletion mutant also indicated that cell apoptosis requires a functional TTSS1 and showed that TTSS1-dependent translocon proteins were associated with host cell death (Bhattacharjee *et al.,* 2005).

Characterization and functional analysis of *V. parahaemolyticus* TTSS1- and TTSS2-associated proteins have been widely studied over the last few years. An effector protein secreted by TTSS1, VP1686, was found to be responsible for induction of macrophage apoptosis in the infected host (Bhattacharjee *et al.*, 2006). Also, VP1680 and VP1659 were determined to be important effector proteins secreted by TTSS1 and also showed a contribution to the cytotoxicity in Hela cells (Ono *et al.*, 2006; Zhou *et al.*, 2010a). Furthermore, roles for VP1680 in the activation of mitogen-activated protein kinases (MAPK), signalling and interleukin (IL) 8 in host cells were reported (Matlawska-Wasowska *et al.*, 2010; Shimohata *et al.*, 2011). Subsequently, an associated chaperone, VecA (VP1682), of VP1680 was identified (Akeda *et al.*, 2009). A functional study indicated that VecA contributes not only to VP1680 secretion but also to translocation of VP1680 into the host cells.

TTSS2-associated proteins have been characterized by several studies. Two translocon proteins, VopB2 and VopD2, of *V. parahaemolyticus* TTSS2 have been found to play a critical role in TTSS2-dependent enterotoxicity (Kodama *et al.*, 2008). Both translocon proteins were found to be not necessary for *V. parahaemolyticus* TTSS2 effector secretion, but were necessary for effector translocation. They are localized in the host cell membranes and are required for pore formation. VopT was found to be secreted and translocated into the host cell via *V. parahaemolyticus* TTSS2 (Kodama *et al.*, 2007). It is an effector protein that is similar to the ADP-ribosyltransferase effector domain of two effectors proteins, ExoT and ExoS, secreted by *Pseudomonas aeruginosa*. Although a previous study (Park *et al.*, 2004) indicated that TTSS1 is associated with host cell cytotoxicity, the result from this study showed that VopT is partly responsible for cytotoxicity in the host cell. Okada *et al.* (2009) identified a

novel TTSS2 in *trh*-positive *V. parahaemolyticus*. Although TTSS2 from *trh*-positive is closely similar to TTSS2 in *tdh*-positive strains, it belongs to a distinct lineage. TTSS2 from *tdh*-positive and *trh*-positive strains were named as TTSS2α and TTSS2β, respectively (Fig. 1.19). These two distinct TTSSs were also found in the other *Vibrio* species including *V. cholerae* non-O1/non-O139 and *V. mimicus* (Okada *et al.*, 2009, 2010). Distributions of TTSS2β and TTSS2α across different species indicate interspecies HGT of TTSS2 gene clusters.

Genes located in the TTSS2 region, including *vscC2, vopP,* and *vopA/P* and in the VPaI, including *vopC* and *VPA1376,* were detected in environmental *V. parahaemolyticus* isolated from Italy (Caburlotto *et al.*, 2009). It was found that *vscC2* and *vopP* could occur either together or separately, indicating that these two genes may have been acquired independently even though they are located in the same region. Subsequently, the potential virulence of environmental *V. parahaemolyticus* carrying virulence-related genes including *vopT* and *vopB2* and the other genes involved in the VPaI was studied (Caburlotto *et al.*, 2010). These strains are capable of adhering to human cells and causing cell disruption and loss of membrane integrity. These results indicated that there is a threat of pathogenic *V. parahaemolyticus* in the environment, which constitutes a public health concern. Recently, proteins included in TTSS2, VtrA and VtrB, were found to be involved in enterotoxicity and have a vital role in regulating the expression of VPaI-related genes in *V. parahaemolyticus* (Kodama *et al.*, 2010). Furthermore, bile salt has been determined to be a host-derived inducer for transcription of *vtrB* and for expression of *vtrA*-dependent genes of *V. parahaemolyticus* (Gotoh *et al.*, 2010).

# 1.3 Disease and epidemiology

As described in section 1.2.4, pathogenic *V. parahaemolyticus* are capable of causing disease in humans. The clinical features and symptoms caused by *V. parahaemolyticus* infection are described in this section. The global epidemiology of *V. parahaemolyticus*, including the situations in Asia, Europe, the United States of America (USA), and specially in Thailand, are also reviewed for a better understanding of the emergence of certain pandemic strains in diverse geographic regions.

## *1.3.1 Disease caused by V. parahaemolyticus*

*V. parahaemolyticus* is a seafood-borne bacterial pathogen and is the main cause of travellers' diarrhoea and gastroenteritis worldwide. The illness is due to the consumption of contaminated undercooked seafood particularly shellfish. The incubation period of *V. parahaemolyticus* ranges from 13 to 23 hours (Barker *et al.*, 1974). The clinical symptoms usually start 10-15 h after infection with diarrhoea and abdominal pain. The diarrhoeal stool is watery, mucoid, and often bloody. Patients may also have fever, vomiting, nausea, abdominal cramps, chill and general fatigue. The frequency of diarrhoea is normally less than 10 times a day. In many clinical cases, the diarrhoea will spontaneously resolve after 9-10 days as the *V. parahaemolyticus* infection is self-limiting. However, severe infection requires hospitalization. In rare cases, more than 20 episodes of diarrhoea may occur a day, leading to dehydration, collapse, and cyanosis (Janda *et al.*, 1988). This organism can also cause cardiovascular abnormalities on rare occasions (Honda *et al.*, 1976).

The infective dose of *V. parahaemolyticus* varies from $10^5$ to $10^{10}$ colony forming units (CFUs) per gram ($g^{-1}$). However, the infective dose has been found to be associated with possession of virulence components and the pathogenicity of the infecting strains (Joseph *et al.*, 1982). Volunteers who had ingested at least 2 x $10^5$ to 3 x $10^7$ CFU of haemolytic strains of *V. parahaemolyticus* rapidly developed symptoms of gastroenteritis, whereas individuals who received from 4 x $10^9$ to 1.6 x $10^{10}$ CFU of non-haemolytic strains of *V. parahaemolyticus* did not develop diarrhoeal symptoms.

## 1.3.2 Epidemiology of V. parahaemolyticus

The incidence of *V. parahaemolyticus* has been the cause of sporadic diarrhoeal cases throughout the world, including Asia, Europe and the USA. Outbreaks of the pandemic strain serotype O3:K6 have ocurred in many countries in Asia and subsequently spread to the other parts of the world. The geographical distribution of pathogenic *V. parahaemolyticus* is shown in Fig. 1.21.
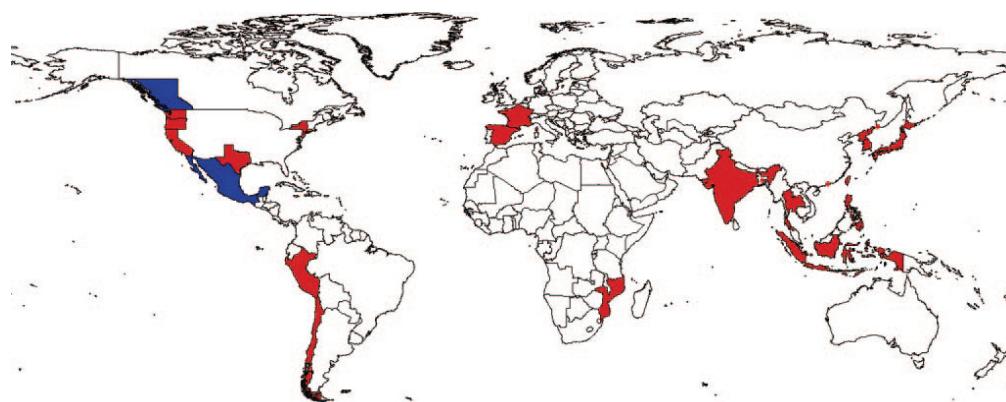


**Figure 1.21.** Global dissemination of *V. parahaemolyticus*. Red represents area where the pandemic *V. parahaemolyticus* strain has spread. Dark blue represents areas where outbreaks of *V. parahaemolyticus* have occurred or presence in the environment but the pandemic status of strains remains unclear. Figure adapted from Nair *et al.* (2007)

## 1.3.2.1 Asia

*V. parahaemolyticus* was first isolated and recognized as a food poisoning bacterium in Japan in 1950 (Fujino *et al.*, 1953). The Infectious Disease Surveillance Centre (IDSC, Japan) identified *V. parahaemolyticus* as the leading cause of food poisoning in Japan during 1996 to 1998 (Su & Liu, 2007). From 1992 until the present time, food-borne illness caused by this organism has been reported in many Asian countries including India, Bangladesh, China, Taiwan, Korea, Vietnam and Thailand. The pandemic O3:K6 serovar first emerged in Calcutta, India in 1996 (Okuda *et al.*, 1997). All pandemic O3:K6 strains from this study possessed *tdh*. These strains accounted for 50-80 % of the strains isolated from gastroenteritis patients during February-August in 1996 in Calcutta. Since it had not been previously identified during *V. parahaemolyticus* surveillance in Calcutta, it was identified as a new pandemic clone. According to these collective data, the outbreak in Calcutta was believed to be the epidemiological origin of the O3:K6 pandemic strain (Nair *et al.*, 2007). In fact, the first O3:K6 isolate was found in 1995 from travellers in Japan who were returning from countries in South East Asia. A molecular typing study showed that strains isolated from travellers in Japan between 1982 and 1993, which is the period before the outbreak had occurred, were distinct from the O3:K6 strains isolated in Calcutta (Nair *et al.*, 2007). However, the isolates from travellers who had returned from South East Asia to Japan between 1995 and 1996 were indistinguishable from the O3:K6 strains isolated in Calcutta in 1996. Thus, Nair *et al.* (2007) suggested that the pandemic O3:K6 clone not only emerged from India but also became the prevalent clone throughout South East Asia.

**1.3.2.2 Europe**

Sporadic outbreaks of diarrhoea due to *V. parahaemolyticus* have been reported in some European countries particularly in France, Spain and Italy. In France, the prevalence of the pandemic serovar O3:K6 was reported by Quilici *et al.* (2005) from coastal areas during 1997-2004. In addition, there was a serious outbreak associated with the consumption of shrimps imported to France from Asia in 1997 (Su & Liu, 2007). Quilici *et al.* (2005) suggested that the clone causing the outbreak might have been transported to France in ballast water discharged from cargo ships entering the European coastal area.

In Spain, *tdh*-positive *V. parahaemolyticus* strains have been identified from faecal samples of gastroenteritis patients. The disease was associated with raw oyster consumption between August and September 1999 (Lozano-Leon *et al.*, 2003). The results from this study indicated that raw oysters and other shellfish are vehicles for the transmission of *V. parahaemolyticus* infection in Europe. The authors also reported the presence of pathogenic *V. parahaemolyticus*, *tdh*-positive strains in molluscs harvested from European waters. In the summer of 2007, pandemic *V. parahaemolyticus* O3:K6 strains were identified in faecal samples of diarrhoeal patients in Italy (Ottaviani *et al.*, 2008). In this study, another toxigenic *V. parahaemolyticus* serovar O1:KUT and other potential pandemic strains were also isolated from local shellfish and seawater from the Adriatic Sea, and it was suggestged that the illness was due to the consumption of fresh shellfish from local sellers.

In the United Kingdom (UK), *V. parahaemolyticus* has been found routinely at low levels (30%) in environmental samples, including shellfish and estuarine water (Wagley *et al.*, 2008). Although over 10% of these environmental isolates

were *tdh*-possitive, pulse field gel electrophorysis (PFGE) analysis showed that none of the isolates from shellfish were clonally related to clinically-derived strains or the pandemic O3:K6 serovar. However, the authors found that clinical isolates from the UK share close clonal similarity with the pandemic O3:K6 strain responsible for outbreaks in Asia.

## 1.3.2.3 The Americas

The geographic distribution of *V. parahaemolyticus* causing infection in North America has been reported on the West coast, Gulf coast, and Pacific sea coast regions of the USA and in British Columbia (Canada) (Anonymous, 1997; Barker *et al.*, 1974; Daniels *et al.*, 2000; DePaola *et al.*, 2000; Lawrence *et al.*, 1979; McLaughlin *et al.*, 2005; Molenda *et al.*, 1972; Nolan *et al.*, 1984). The earliest outbreaks occurred along the East coast and Gulf area including Maryland, Louisiana and the Gulf of Mexico. The pandemic area has subsequently been expanded in the USA to the Pacific Northwest coast, including Washington, Oregon and California, and to British Columbia in Canada.

The first documented outbreaks of *V. parahaemolyticus* gastroenteritis in the USA were reported in Maryland in 1971 (Molenda *et al.*, 1972). Strains of serotypes O4:K11 and O3:K30 were isolated from the stool samples of the affected patients. Steamed crab and crab salad prepared from canned crabmeat were suspected as the cause of the illness in these outbreaks. The case studies of the Louisiana outbreak in 1972 and the outbreaks on two Caribbean cruise ships during 1974-1975, indicated that they were attributable to failures of the shrimp boiling process and to seafood contamination from the internal seawater system (Barker *et al.*, 1974; Lawrence *et al.*, 1979). However, the epidemic

strains isolated from the cruise ships were not identical. This incidence was explained by the fact that the ship had cruised through different territorial waters which contained different regional species, and that these local microorganisms (some of which were pathogenic strains) possibly contaminated the water system on the cruise ship (Lawrence *et al.*, 1979). Certainly, cases of gastroenteritis were also reported in the Pacific Northwest during late summer in 1981 (Nolan *et al.*, 1984).

Pandemic *V. parahaemolyticus* serotype O3:K6 was first recovered in the USA from patients with gastroenteritis who had consumed oysters harvested from Galveston Bay, Gulf of Mexico in 1998 (Daniels *et al.*, 2000). All clinical cases identified during this outbreak were of serotype O3:K6. However, the clinical strains causing this outbreak were not detected from oysters growing in the same region. Although it is not clear how the O3:K6 strain emerged in Galveston Bay, the authors suggested that ballast water from a cargo ship entering the Gulf of Mexico could have introduced this outbreak strain to the Americas. Furthermore, elevated seawater temperatures during El Nino years, including 1998, is also considered to promote a favourable environment for the multiplication and dissemination of this organism.

Small outbreaks during July-September 1998 were reported in Connecticut, New Jersey, and New York as a consequence of raw oyster consumption (Daniels *et al.*, 2000). In the summer of 2004 passengers on board a cruise ship in Alaska developed gastroenteritis after eating raw oysters produced from Alaska (McLaughlin *et al.*, 2005). The incidence of *V. parahaemolyticus* gastroenteritis outbreaks in the USA show that raw oysters are strongly implicated as a vehicle of transmission of the infection.

In South America, pandemic isolates of *V. parahaemolyticus* from Chile during 1998-2004 were analyzed by Gonzalez-Escalona *et al.* (2005). Most clinical isolates belonged to the pandemic clonal complex, predominantly of the O3:K6 serotype. This finding indicated that the pandemic clone that had emerged in Calcutta, India in 1996 had spread to the South American continent.

## 1.3.2.4 Prevalence of *V. parahaemolyticus* in Thailand

Atthasampunna *et al.* (1974) studied the occurrence of *V. parahaemolyticus* infection in Thailand and reported that the incidence of diarrhoea caused by this organism was apparently lower in the cooler months, November 1970 to February 1971, when the sea water temperature ranged from $25.0^{c}$C to $27.9^{o}$C. The incidence of *V. parahaemolyticus* infection in these cooler months was approximately 3.2 % of all diarrhoeal cases. The incidence of infection in the warmer months, during September 1971 when the sea water temperature ranged from $28.4^{o}$C to $30.4^{o}$C, was up to 22.5% of all diarrhoeal cases. However, the authors also suggested that factors other than sea water temperature also affected *V. parahaemolyticus* incidence.

*V. parahaemolyticus* was identified to be a cause of travellers' diarrhoea in Bangkok during 1978 to 1979 (Sriratanaban & Reinprayoon, 1982). Rectal swabs and stool samples from patients with diarrhoea who had stayed in a deluxe international hotel and hotel employees in Bangkok were examined. Although the infections were particularly high in June and July, the monthly incidence of *V. parahaemolyticus* infection was unclear. However, the study showed that the incidence of *V. parahaemolyticus* diarrhoea was significantly higher in hotel guests (31%) than in hotel employees (15%).

A survey of food poisoning-related cases by the Department of Epidemiology, Ministry of Public Heath, Thailand, 1992-2001 revealed that illnesses due to food poisoning were predominantly caused by *V. parahaemolyticus,* followed by *Salmonella spp., Staphylococcus aureus* and *Clostridium perfringen* respectively (Fig. 1.22). Since then gastroenteritis cases caused by *V. paraheamolyticus* were sporadically recovered and this organism has been identified as the leading cause of food poisoning cases in Thailand.
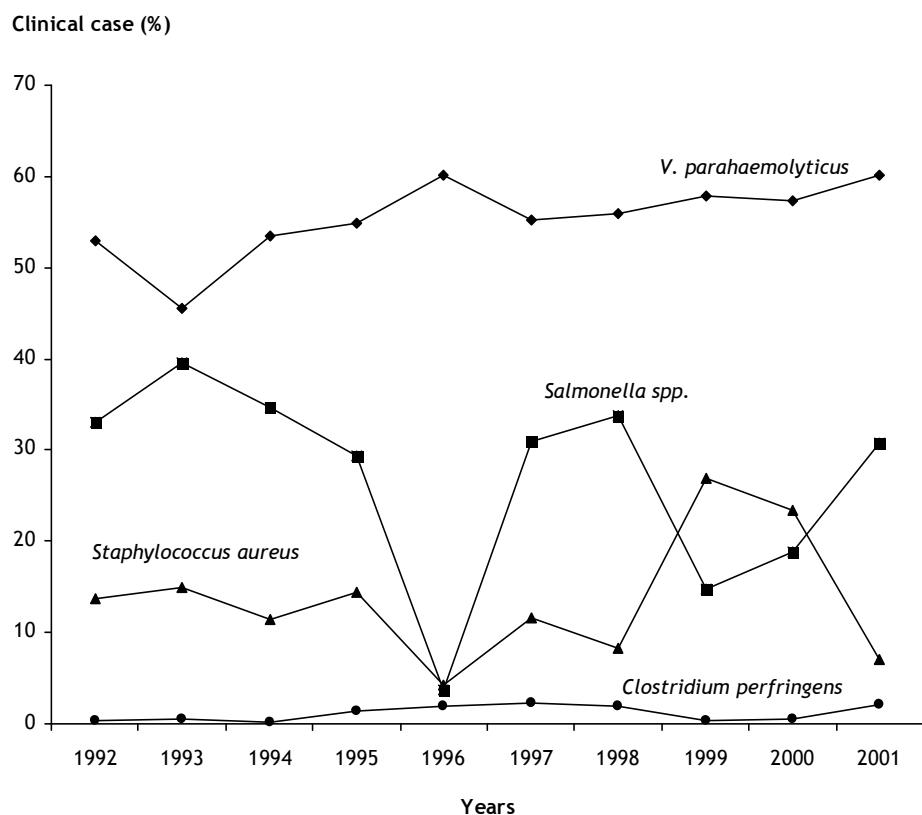


**Figure 1.22.** Occurrence of food poisoning-related cases reported to the Department of Epidemiology, Office of The Permanent of Secretary for Public Heath, Ministry of Public Heath, Thailand, 1992-2001.

The pandemic strain of *V. paraheamolyticus,* serovar O3:K6, that emerged from India in 1996, was found in both clinical and environmental samples in Songkla province, South of Thailand during 1998-1999 (Vuddhakul *et al.*, 2000). This study suggested that pandemic *V. paraheamolyticus* occurred in both clinical and environmental isolates from the same geographical area in Thailand. Prevalence and serodiversity of the pandemic O3:K6 clone among the clinical strains of *V. parahaemolyticus* isolated in Southern Thailand were subsequently examined by Laohaprertthisan *et al.* (2003). Among the clinical strains isolated in 1999, serovar O3:K6 was determined to be a dominant serovar followed by three other virulent serovars, O1:K25, O1:K41 and O4:K12. The epidemiology of the pandemic strain in Thailand continued, with the dominance of O3:K6 serovar during 2001 and 2002 (Serichantalergs *et al.*, 2007). Moreover, a further virulent serovar O3:K46 isolated in Southern Thailand was identified as a new pandemic strain, additional to the three previously recognized pandemic serovars, O3:K6, O1:K25 and O1:KUT (Serichantalergs *et al.*, 2007).

The examination of *V. parahaemolyticus* in seafood from the south of Thailand revealed that the illnesses caused by this bacterium have been associated most strongly with the consumption of bloody clams (*Anadara granosa*) (WHO & FAO, 2011). However, other commercially important seafood species including white shrimp and crab meat are also considered to be potential sources of *V. parahaemolyticus* infection in Thailand.

## 1.4 Molecular typing

Several molecular typing techniques have been applied in order to characterize *V. parahaemolyticus.* Traditionally, serotyping has been extensively used for

epidemiological studies of this organism. However, the limitations of serotyping, such as the availability of commercial antisera and serotypic conversion of this organism, are problematic for using this method. Furthermore, the serotyping scheme of *V. parahaemolyticus* was established based on clinical strains (Sakazaki, 1992) that may not cover the serotypes of environmental strains. Thus, other molecular typing techniques based on nucleotide sequences have been developed for more definitive characterization. This section discusses the typing methods that have been applied to *V. parahaemolyticus*, including serotyping and genotypic-based methods, with the emphasis on multilocus sequence typing (MLST), the technique chosen for use in the present study.

## 1.4.1 Serotyping

There are three antigenic components that can be recognized among strains of *Vibrio* species. The antigenic components comprise the flagella or H antigen, the somatic or O antigen, and the capsular or K antigen. Serological typing of *V. parahaemolyticus* is based on the O and K antigens (Sakazaki, 1992).

### 1.4.1.1 H antigen

Bacteria in the genus *Vibrio* have unique flagellar (H) antigens within each species (Tassin *et al.*, 1983). The H-antigen is unable to be used for serotyping of *V. parahaemolyticus* because it does not exhibit inter-strain variation and is not present in non-motile strains. However, an agglutination assay of the H-antigen can be useful for *Vibrio* species-specific determination in cases where the species have different flagellar types, polar flagella and lateral flagella.

### 1.4.1.2 O antigen

The somatic (O) antigen or lipopolysaccharide (LPS) of *V. parahaemolyticus* is thermostable and not destroyed by treatment with 50% ethanol and n-HCL solution at $37^{o}C$ for 24 h (Sakazaki, 1992). The O antigen is classified into 11 groups. This antigen is inagglutinable in the living state by homologous O-antiserum because of the presence of masking antigens. O-antigen agglutination will occur only when the culture is heated to $100^{o}C$ ($121^{o}$ C in some cases) for 2 h and washed with saline prior to the agglutination test.

### 1.4.1.3 K antigen

The capsular (K) antigen is thermolabile and its agglutination ability can be destroyed by heating to $100°C$ for 1 to 2 h (Joseph *et al.*, 1982). It can be classified into 41 groups (Sakazaki, 1992). Agglutination of the K-antigen can occur by homologous K-antiserum in living cultures without heating. The universal antigenic scheme used for serotype determination of *V. parahaemolyticus* is shown in Table 1.4.

**Table 1.4.** Serotyping scheme of *V. parahaemolyticus* described by Sakazaki (1992)

| O-antigen | K-antigen |
|---|---|
| 1 | 1, 25, 26, 32, 38, 41, 56, 58, 64, 69 |
| 2 | 3, 28 |
| 3 | 4, 5, 6, 7, 29, 30, 31, 33, 37, 43, 45, 48, 54, 57, 58, 59, 65, |
| 4 | 4, 8, 9, 10, 11, 12, 13, 34, 42, 49, 53, 55, 63, 67 |
| 5 | 15, 17, 30, 47, 60, 61, 68 |
| 6 | 18, 46 |
| 7 | 19 |
| 8 | 20, 21, 22, 39, 70 |
| 9 | 23, 44 |
| 10 | 19, 24, 52, 66, 71 |
| 11 | 36, 40, 46, 50, 51, 61 |
| **Total**     11 | 41 serotypes |

The first isolate of *V. parahaemolyticus*, in Japan 1950, was classified as serotype O1:K1. The predominance of serotypes varies across geographical locations and with the date of isolation. To date, there is no report about significant correlations between the serotype and the virulence of clinical strains. However, environmental *V. parahaemolyticus* isolates are frequently untypeable (Joseph *et al.*, 1982). The pandemic *V. parahaemolyticus* serotype O3:K6 emerged in Calcutta, India in 1996. Other pandemic serotypes that were subsequently found are O1:KUT (K antigen untypable), O4:K68, and O1:K25 (Chowdhury *et al.*, 2004).

However, the serotyping method is limited by the availability of commercial antisera, and inconsistent due to serotypic conversion. Other molecular typing

methods based on genotyping identification have been developed and are commonly used for *V. parahaemolyticus*.

## 1.4.2 Genotypic identification

Applications of different molecular typing methods based on genotypic identification of the *Vibrios* were reviewed by Thomson *et al*. (2004). These methods include amplified fragment length polymorphism (AFLP), fluorescence *in situ* hybridization (FISH), amplified ribosomal DNA restriction analysis (ARDRA), random amplified polymorphic DNA (RAPD), repetitive extragenic palindromes (REP), restriction fragment length polymorphism (RFLP), multilocus enzyme electrophoresis (MLEE) and multilocus sequence typing (MLST). Among these techniques, MLST has higher discriminatory power than the other techniques and has been recognised as a very useful tool for species delineation. More details of MLST applications are separately discussed in section 1.4.3 of this chapter. However, several genotypic techniques such as ribotyping, PCR-based techniques, DNA hybridization, and nucleotide sequence analysis have also been developed for *V. parahaemolyticus* over the last two decades.

Molecular typing of the *V. parahaemolyticus* haemolysin genes *tdh* and *trh* was carried out by Suthienkul *et al*. (1996) using RFLP. In this study, 137 *V. parahaemolyticus* isolates from diarrhoeal patients in Thailand were analyzed. As a result, the *Hind*III restriction fragment patterns of *tdh* and *trh* grouped these isolates into five and four types, respectively.

Random amplified polymorphic DNA (RAPD) has been applied to study molecular types of 308 clinical *V. parahaemolyticus* from Taiwan (Wong *et al*., 1999). The results demonstrated that RAPD makes it possible to differentiate strains from

the same serovar. However, this method generated variable band intensity and lacked reproducibility of certain minor bands. Moreover, evaluation of molecular typing methods for *V. parahaemolyticus* by Wong (2003) suggested that the discriminatory ability of RAPD was less than that of PFGE and ribotying.

Chowdhury *et al.,* (2000) studied pulsed field gel electrophoresis (PFGE) by using *Not*I restriction enzyme to characterize *V. parahaemolyticus* pandemic strain serotypes O3:K6, O4:K68 and O1:KUT and nonpandemic strains isolated from different countries including India, Japan, Thailand, Taiwan, Laos, Singapore, Maldives and the USA during 1995 and 1999 (Chowdhury *et al.*, 2000). From this study, *Not*I restriction fragments showed considerable polymorphism between pandemic and non-pandemic strains of various serotypes. Moreover, the PFGE profiles of pandemic serotype O4:K68 and O1:KUT strains isolated from 1997 were closely similar to the pattern obtained with the outbreak O3:K6 strains. The authors suggested that the O4:K68 and O1:KUT strains most likely originated from the new O3:K6 pandemic clone. Subsequently, close genetic relationships among isolates in the O3:K6 clonal group, including O4:K68 and O1:KUT, were confirmed by *EcoR*I ribotyping and *tdh* sequencing (Yeung *et al.*, 2002). However, the results from this study suggested that ribotyping and *tdh* sequencing were less discriminatory than PFGE.

Since the pandemic *V. parahaemolyticus* O3:K6 strains possess a unique *toxRS* nucleotide sequence (*toxRS* is a gene operon that encodes *tdh* regulatory protein, ToxRS), which is distinguishable from that of non-pandemic strains (Matsumoto *et al.*, 2000), epidemiological characterization of pandemic *V. parahaemolyticus* serotype O3:K6 is determined by specific *toxRS*-targeted PCR, also known as gene specific PCR (GS-PCR). Furthermore, all pandemic O3:K6

strains harbour pO3K6, a plasmid containing the *orf8* gene that is acquired from bacteriophage f237 (Nasu *et al.*, 2000). Thus, detection of the *orf8* gene is also used to identify pandemic *V. parahaemolyticus* O3:K6 strains.

Three PCR methods using specific primers have been applied for typing *V. parahaemolyticus* (Wong & Lin, 2001). The primers were designed for specific sequences, namely the ribosomal gene spacer sequence (RS), the repetitive extragenic palindromic sequence (REP) and the enterobacterial repetitive intergenic consensus sequence (ERIC). Typing patterns and clustering analysis indicated that these methods facilitate differentiation of *V. parahaemolyticus* from other species including *Escherichia coli, V. cholerae*, and *V. vulnificus* as well as subspecies typing within *V. parahaemolyticus* strains. Although these three PCR methods were proposed to be suitable for rapid typing of *V. parahaemolyticus*, REP-PCR was the most preferable because it produced the greater reproducibility of fingerprints.

## 1.4.3 Multilocus sequence typing (MLST)

### 1.4.3.1 Background of MLST

Mutilocus sequencing typing (MLST) is a molecular typing method based on comparative nucleotide sequence analysis. It was proposed by Maiden in 1998 as a portable approach to determine clones within populations of pathogenic microorganisms (Maiden *et al.*, 1998). In this method, the nucleotide sequences of fragments of seven housekeeping enzyme genes are compared. A database is complied which can be conveniently accessed via the internet (http://www.pubmlst.org). This approach enables the exchange of molecular

typing data between laboratories that is necessary for global epidemiological studies.

MLST analysis was developed from an earlier molecular typing technique, multilocus enzyme electrophoresis (MLEE). MLEE assesses genetic variation by measuring the phenotype of housekeeping enzymes by means of gel electrophoresis (Selander *et al.*, 1986). A biochemically stainable metabolic enzyme creates different electrophoretic migrations of proteins due to variant alleles at the respective loci. The electrophoretic variants (electromorphs) therefore indicate the genotypes of the examined isolates. However, the disadvantage of MLEE is that the electromorph data are of relatively low resolution. Only genetic changes that alter the electric properties of proteins are detected in the MLEE scheme.

In contrast, nucleotide sequence data from MLST is more effective for bacterial typing as it provides high resolution of genetic discrimination. Since the MLST scheme was proposed, a number of reviews describing its applications for molecular epidemiological studies of pathogenic bacteria have been published (Cooper & Feil, 2004; Feil & Enright, 2004; Feil, 2004; Maiden *et al.*, 1998; Smith *et al.*, 2000; Turner & Feil, 2007; Urwin & Maiden, 2003).

### 1.4.3.2 Considerations of MLST schemes

There are three main considerations in designing a MLST scheme. First, the choice of the strain collection; second, the choice of the genetic loci to be characterized; and third, the design of primers for PCR amplification and nucleotide sequencing (Maiden, 2006; Urwin & Maiden, 2003).

A diverse collection of isolates with existing typing information or epidemiological data is required for MLST analysis. The bacterial collection should represent bacteria from diverse sources of isolation, and approximately 100 isolates are statistically sufficient and recommended (Maiden, 2006).

Seven loci are considered as a minimum number of examined loci for routine typing criteria, since they provide sufficient resolution for the reliable identification, at a reasonable cost and in a relatively short time (Maiden, 2006). However, a larger number of loci would be preferred for studies of population genetics. Allele fragments examined for MLST are usually 400-600 bp in length, because this length is reliably read on a single run of the gel-based automatic sequencing instruments available when the MLST scheme was developed, the mid 1990s (Maiden, 2006). Fragmented housekeeping genes, encoding fundamental metabolic function, are targeted for bacterial typing. The conserved function of these genes enables sufficient discrimination of variant strains without having bias from diversifying selection among bacterial population (Maiden, 2006).

For MLST primer design, a nested system is advisable as it can eliminate false amplification, particularly for highly diverse bacteria, resulting in higher quality of nucleotide data production. A nested strategy requires two sets of primers, one for DNA fragment amplification (PCR) and the other (known as internal primers) for nucleotide sequencing. The sequencing primers are designed within the amplified fragment. Furthermore, primers should be adjusted (if possible) to have the same annealing temperature, so that they can be applied to all the amplification reactions of the seven housekeeping gene fragments (Maiden, 2006).

### 1.4.3.3 MLST data analysis

#### *1.4.3.3.1 Data collection*

Fragments of housekeeping genes are amplified by PCR and the PCR products of these fragments are sequenced using internal primers, alternatively named nested primers. Nucleotide sequences of housekeeping gene fragments are then assigned allele numbers and these allele numbers are in turn used to create sequence types (ST) for each strain. A flow chart of the MLST scheme is shown in Fig. 1.23.

#### *1.4.3.3.2 Data analysis*

Nucleotide sequence data of seven housekeeping gene fragments are used to generate an allelic profile (AP) or sequence type (ST) for each isolate. The ST is a unique combination of housekeeping gene alleles, each of which is represented by an allelic number, alternatively known as allelic type. By this procedure, the ST of each isolate contains seven numbers representing seven allele types of the housekeeping enzyme genes. A set of seven allele types defines an allelic profile (AP) of an individual strain. This AP represents one unique ST of that individual strain. Assignment of STs from MLST data is demonstrated in Fig. 1.24.
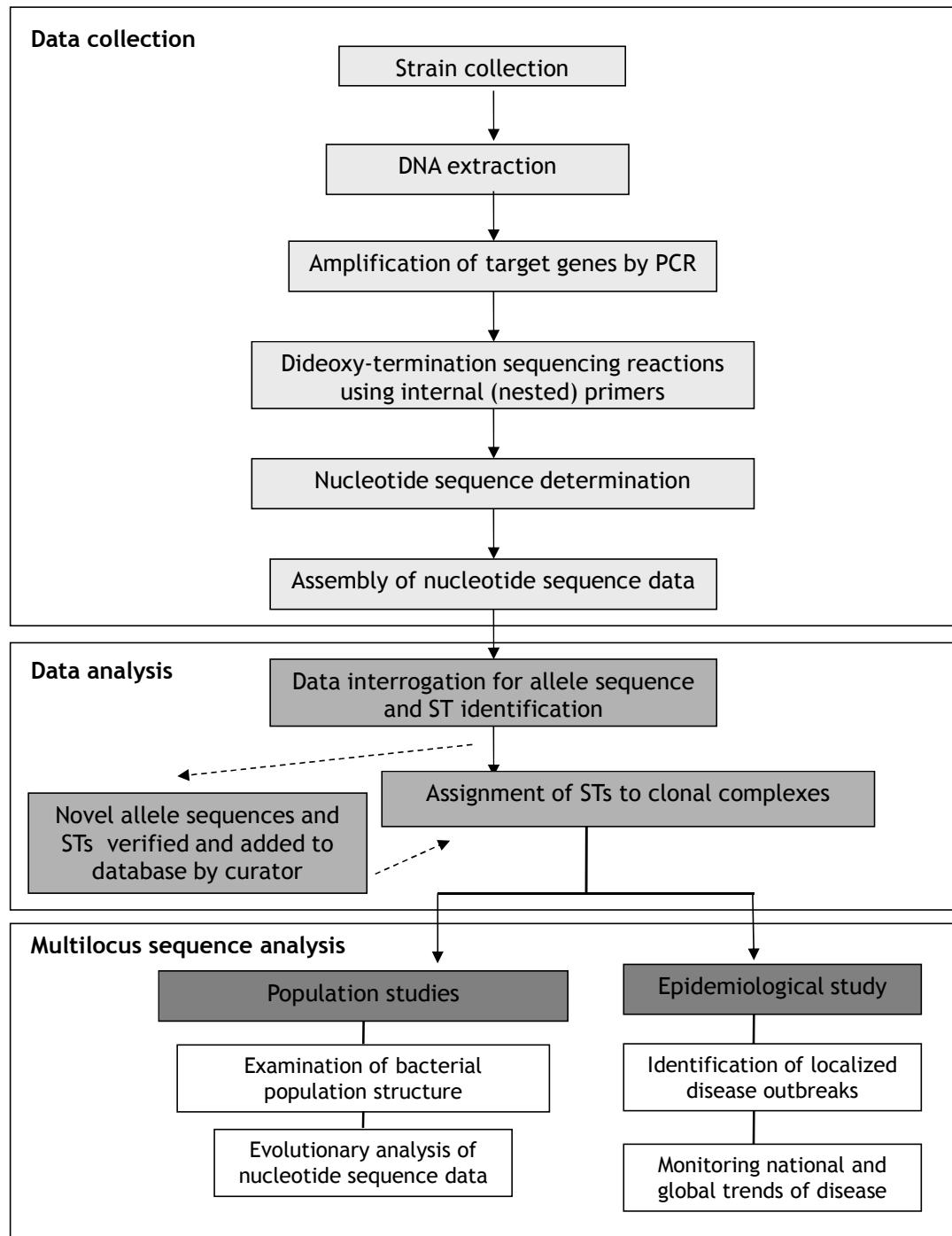
**Figure 1.23.** Flow chart representing the multilocus sequence typing (MLST) scheme, involving data collection, data analysis, and sequence analysis. Diagram adapted from Urwin *et al.* (2003)
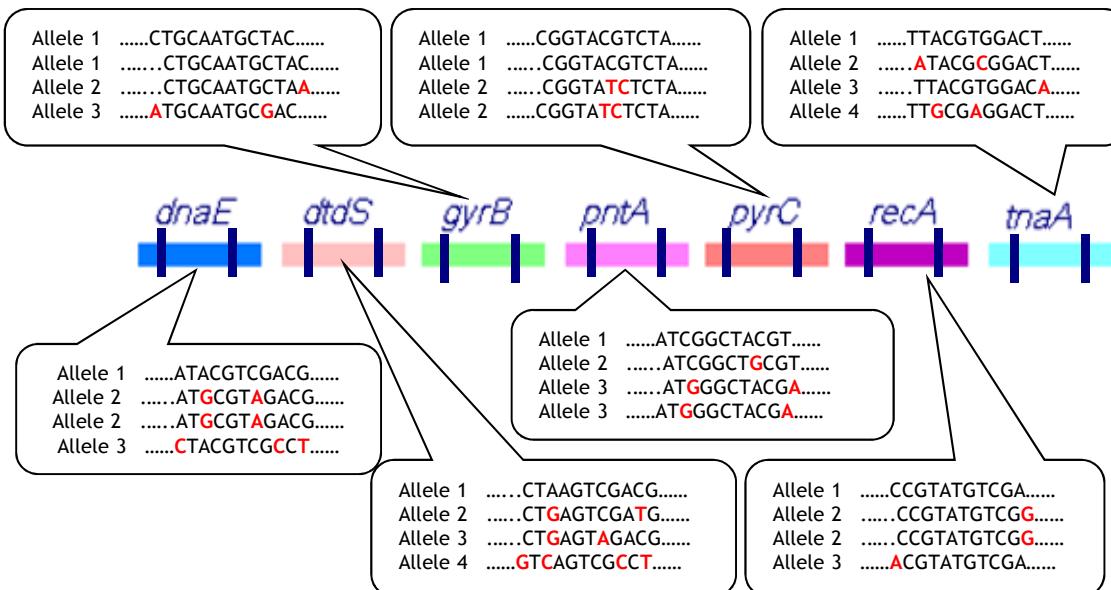
(A)

| Sequence type (ST) | *dnaE* | *dtdS* | *gyrB* | *pntA* | *pyrC* | *recA* | *tnaA* |
|---|---|---|---|---|---|---|---|
| ST-1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| ST-2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 |
| ST-3 | 2 | 3 | 2 | 2 | 1 | 2 | 3 |
| ST-4 | 3 | 4 | 3 | 3 | 2 | 3 | 4 |

(B)

**Figure 1.24.** The diagram represents MLST data interpretation. Seven housekeeping genes, *dnaE, dtdS, gyrB, pntA, pyrC, recA,* and *tnaA* used for the MLST scheme of *V. parahaemolyticus,* are presented as an example. (A) Allelic types, e.g. alleles 1, 2, 3, etc., are assigned for the unique nucleotide sequences of individual gene fragments. (B) Unique combinations of seven allelic types creates the ST, e.g. ST1, 2, 3, etc., for an individual isolate.

The ST of examined isolates can be compared with an existing ST in an online database, or recorded as a novel ST if it contains a new combination of allelic types or a nucleotide sequence that has not previously been submitted to the database. Genetic relationships of representative isolates can be obtained either from concatenated sequences of the seven housekeeping gene fragments or from the AP of individual isolates. Phylogenetic distance is determined by concatenated sequences of examined isolates, whereas the AP data are useful for analysis of clonal relationships among isolates by using an evolutionary analysis tool such as eBURST (Feil *et al.*, 2004). To date, several bioinformatic tools using different algorithms such as Split decomposition, Bayesians clustering analysis, etc. have been developed to exploit MLST data for epidemiological and population genetic studies of pathogenic bacteria.

### 1.4.3.4 MLST applications to pathogenic bacteria

MLST has been widely used to study the molecular evolution and population structure of both Gram-positive and Gram-negative pathogenic bacteria, including *Streptococcus pneumoniae* (Enright *et al.*, 2000), *Staphylococcus epidermidis* (Miragaia *et al.*, 2007), *Streptococcus aureus* (Enright *et al.*, 2000), *Streptococcus oralis* (Do *et al.*, 2009), *Clostridium difficile* (Lemée, *et al.*, 2005), *Clostridium botulinum* (Jacobson *et al.*, 2008), *Bacillus cereus* (Priest *et al.*, 2004), *Listeria monocytogenes* (den Bakker *et al.*, 2008), *Neisseria meningitidis* (Didelot *et al.*, 2009; Jolley *et al.*, 2000; Maiden *et al.*, 1998), *Campylobacter coli* (Dingle *et al.*, 2005; Miller *et al.*, 2006), *Campylobacter jejuni* (Dingle *et al.*, 2005; de Haan *et al.*, 2010), *Salmonella enterica* (Octavia & Lan, 2006), *Haemophilus parasuis* (Olvera *et al.*, 2006), *Lactobacillus casei*

(Cai *et al.*, 2007) and *Yersinia pseudotuberculosis* (Ch'ng *et al.*, 2011). MLST studies for *Vibrio* species are reviewed in the Chapter 2.

The MLST scheme was first applied to *N. menigitidis,* a Gram-negative pathogen (Maiden *et al.*, 1998). MLST analysis indicated that homologous recombination is a main driving force in the diversification of the *N. menigitidis* genome and also showed that invasive disease-causing strains are more clonal than asymptomatic strains (Jolley *et al.*, 2000). *C. coli* and *C. jejuni* are examples of bacteria for which the MLST scheme has proved to be useful for determining associations between bacterial strains and various animal hosts. Sequence types of *C. coli* that are more prevalent in certain animal hosts, including cattle and poultry, were identified using MLST (Miller *et al.*, 2006). *C. coli* strains isolated from swine represented high genotypic diversity, whereas those isolates from cattle were relatively clonal. Recently, close genetic associations of chicken and human *C. coli* isolates from the UK and European countries were established by MLST, and comparative genomic hybridization also suggested that the majority of *C. coli* human infections arise from chickens (Lang *et al.*, 2010). Furthermore, the isolates recovered from turkeys represent evidence of interspecies HGT between *C. coli* and *C. jejuni.* Bayesian analysis of population structure (BAPS) of MLST data yielded a similar overlapping percentage of human disease *C. jejuni* genotypes in both genetic clusters of bovine and poultry isolates in Finland (de Haan *et al.*, 2010). This suggests that bovines and poultry are equally important as reservoirs for *C. jejuni* infections in human. In *Salmonella spp.*, subspecies diversity in *S. enterica* subspecies I (Octavia & Lan, 2006) and serovar Newport (Sangal *et al.*, 2010) were identified by MLST analysis. Three distinct lineages of *S. enterica* serovar Newport strains each represented associations of human isolates with geographical sources, animal hosts and antibiotic

susceptibility. Recently, MLST was suggested to be a replacement for serotyping in classification of *Salmonella spp.* (Achtman *et al.*, 2012). This is due to the failure of the serovar designation to recognize natural evolutionary grouping in *Salmonella spp.*, and may be misleading about the disease potential of certain strains including *S. enteric*. MLST application to the population of *H. parasuis*, a swine pathogenic bacterium of the family *Pasteurellaceae*, revealed that strains of nasal origin (putative non-virulent) were genetically distinct from those isolated from clinical lesions (putative virulent) (Olvera *et al.*, 2006). Interspecies horizontal gene transfer of housekeeping genes was also detected in *H. parasuis*. Isolates containing gene fragments from other species were represented by largely diverged lineages in the phylogenetic tree of *H. parasuis* housekeeping gene sequences. The population structure of a bacterium causing human enteric disease, *Y. pseudotuberculosis*, by MLST showed geographical restriction of the strains from distinct clusters (Ch'ng *et al.*, 2011). The phylogenetic tree of seven housekeeping gene sequences of *Y. pseudotuberculosis* represents two main clusters, cluster A represents strains distributed worldwide from four different continents and cluster B represents strains isolated from Far Eastern countries. Moreover, evidence of the sporadic gain and loss of virulence genes in *Y. pseudotuberculosis* of the same ST was obtained, indicating instability of virulence factors in this species.

The MLST approach is not only applicable to molecular evolutionary analysis of Gram-negative bacteria, but is also capable of resolving evolutionary analysis of Gram-positive bacteria. MLST provides an online tool for assigning clonal complexes of methicillin-resistant *S. aureus* (MRSA) and methicillin-susceptable *S. aureus* (MSSA), since this method is able to determine distinct clonal complexes of these strains (Enright *et al.*, 2000). Furthermore, MLST analysis

66

revealed a genotypic association between *S. aureus* and different animal hosts (Smyth *et al.*, 2009). *S. aureus* isolates recovered from chickens and rabbits were genotypically more similar to those of human-associated strains than to the ruminant-associated genotypes. The MLST technique is able to identify evidence of inter- and intra-species recombination among *S. oralis, S. mitis* and *S. pseudopneumoniae* (Do *et al.*, 2009). Use of the MLST scheme suggested the probable co-evolution of housekeeping genes and virulence-related genes of *C. difficile* due to their congruent phylogenetic topologies (Lemee *et al.*, 2005). In contrast, the evolution of housekeeping genes and botulinum neurotoxin encoding genes (BoNT) is not related in *C. botulinum* (Jacobson *et al.*, 2008). In pathogenic bacterium causing listeriosis, *L. monocytogenes*, the internal/external branch length of a phylogram constructed with recombination correction and Tajima's D test from MLST data has determined a bottleneck effect in the ancestral population (den Bakker *et al.*, 2008).

MLST was applied in a study of niche specificity in industrially-important lactic acid bacteria (LAB) including *L. casei*. The *L. casei* were isolated from cheese from different geographical locations, from the human gastrointestinal tract and from plant materials (Cai *et al.*, 2007). The phylogenetic MLST tree of these *L. casei* isolates was resolved into three clusters, each of which represented isolates from cheese, silage, and various other origins, including the human gastrointestinal tract. Relatively low intragenic polymorphisms in the strains isolated from cheese and silage indicated less diversity within isolates from the same ecological niches. Moreover, analysis of intragenic polymorphisms within the cheese strains indicated that environmental selective pressure for these strains is related more to ecological factors than to geographical regions.

# 1.5 Outer membrane proteomics of Gram-negative bacteria

In the present study, the utilization of MLST has provided an evolutionary framework for investigating the expression of virulence factors in *V. parahaemolyticus*. Since bacterial outer membrane proteins (OMPs) are involved in host adaptation, and some of them serve as virulence factors (as discussed earlier in section 1.2.4 of this chapter), comparative outer membrane proteomics was used to identify and analyze expression of OMPs in selected isolates of *V. parahaemolyticus*. This approach has been used to compare the expression of OMPs in isolates from diverse sources, including clinical and environmental, and its outcome also provides a basis for epidemiological research of *V. parahaemolyticus.* For a better understanding of the outer membrane proteomic analysis used in the present study, this section includes background information on OMP structure and classification in Gram-negative bacteria, OMP analysis by bioinformatic and proteomic approaches and the identification of OMPs by gel-based methods and mass-spectrometry.

## *1.5.1 Structure and classification of the Gram-negative bacterial outer membrane*

A number of comprehensive reviews have been published of our current understanding of the biogenesis of the Gram-negative bacterial outer membrane (Bos *et al.*, 2007; Costerton *et al.*, 1974; Koebnik *et al.*, 2000; Ruiz *et al.*, 2006). The cell envelope of Gram-negative bacteria consists of two membranes, the inner membrane and the outer membrane (Fig. 1.25).
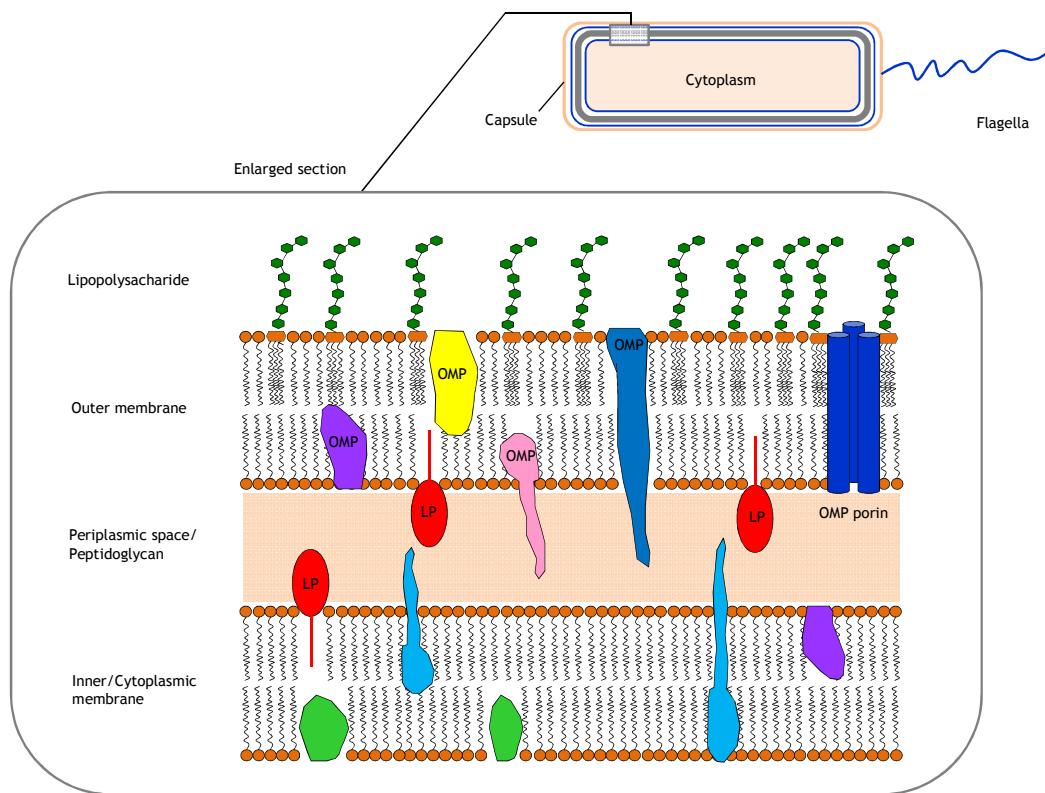
**Figure 1.25.** Structure of the cell envelope of Gram-negative bacteria representing two layers of membranes; the outer membrane and inner membrane. Integral outer membrane proteins (OMPs) such as porins are embedded in the outer membrane. Some OMPs locate across the periplasmic space. Lipopoproteins are present in both outer membrane and inner membrane by attachment of their lipid chains.

These two membranes are separated by periplasm that contains the peptidoglycan layer. The inner membrane comprises a symmetrical phospholipid bilayer, whereas the outer membrane comprises an asymmetrical bilayer containing phospholipid as an inner leaflet and lipopolysachharide (LPS) as an outer leaflet. The properties of the integral proteins embedded in the inner membrane differ from those in the outer membrane: the integral inner membrane proteins fold in the form of α–helices, whereas the integral outer membrane proteins (OMPs) fold into anti-parallel ß-barrels (Koebnik *et al.*, 2000). Integral OMPs are synthesized in the cytoplasm in a form of unfolded ß-strands with a N-terminal signal sequence and are transported through the inner

membrane to the periplasm by translocation machineries (Sec system) (De Keyzer *et al.*, 2003). After transportation across the inner membrane, the OMPs are accessible in the periplasm and are allocated to the outer membrane by a periplasmic chaperone (Eppens *et al.*, 1997). The unfolded ß-stranded OMPs are folded into their ß-barrel structures and integrated into the outer membrane by the ß-barrel assembly machinery (BAM) complex (Hagan *et al.*, 2011). The BAM complex consists of the integral OMP BamA (also known as Omp85 or yaeT) and outer membrane lipoproteins BamB, BamC, BamD and BamE. Outer membrane lipoproteins are transported and positioned in the outer membrane by the localized lipoprotein (Lol) transport machinery, which consists of inner membrane proteins LolC, LolD and LolE, periplasmic chaperone LolA and OMP recepter LolB (Tokuda & Matsuyama, 2004).

OMPs can be classified according to various features such as structure, location, and function. For example, six families of OMPs, namely the OmpA membrane domain, the OmpX protein, phospholipase A, general porins, substrate-specific porins and the TonB-dependent iron siderophore transporters, have been classified according to their atomic structure (Koebnik *et al.*, 2000). Two groups of OMPs including transmembrane and peripheral membrane types are classified according to their subcellular location. This review classifies the OMPs of Gram-negative bacteria based on their major molecular functions including porin, receptor-mediated transport and secretion functions.

### 1.5.1.1 Porins

Porins are integral OMPs, with a ß-barrel structure containing a pore that allows diffusion of small (<700 Da) hydrophilic molecules (Koebnik *et al.*, 2000). The

structure of a porins was first characterized for the OmpF protein, a major porin in *E. coli* (Cowan *et al.*, 1992). The porin consists of a homotrimeric structure of identical subunits. Each subnunit consists of a 16-stranded anti-parallel ß-barrel structure containing a pore in the middle. Porin proteins are divided into two groups; general porins and substrate-specific porins. General porins allow passive diffusion of molecules through the outer membrane, whereas substrate-specific porins may require cellular energy generated by the proton motive force from the inner membrane for active transport. Major porins differ among bacterial species. The well-studied general porins in *E. coli* include OmpF (Cowan *et al.*, 1992), OmpC (Baslé *et al.*, 2006) and PhoE (Korteland *et al.*, 1982). In *V. cholerae,* the porins OmpU and OmpT have equivalent functional roles to the *E. coli* OmpF and OmpC, respectively  (Chakrabarti *et al.*, 1996; Li *et al.*, 2000). Examples of substrate-specific porins include LamB (maltose channel), BtuB (vitamin B12 channel) and ChiP (chitin channel). The LamB protein is responsible for permeation of maltosaccharide and behaves as a maltose-inducible OMP in many bacteria (Lång & Ferenci, 1995). The BtuB protein is a substrate-specific OMP that binds to vitamin B12, allowing active translocation of vitamin B12 (cyanocobalamin) across the outer membrane to the periplasmic space (Aufrere *et al.*, 1986). The energy required to drive the active translocation process involved in substrate uptake is derived from a protein that transfers cellular energy from the inner membrane to the outer membrane, and in the case of BtuB this energy is provided by the TonB system (see section 1.5.1.2 of this chapter). The ChiP protein is a chitin-binding protein that is found, for example, in the marine bacterium *V. furnissii*  (Park *et al.*, 2000). The expression of ChiP is induced by chitin products, e.g. chito-oligopolysaccharide, which is a main component of the crustacean exoskeleton.

## 1.5.1.2 Receptor-mediated transporters

The OMPs in this group include a specific receptor for the substrate (e.g. iron) and require energy for active transport. In the human host, the majority of iron is combined with metalloproteins such as haemoglobin, myoglobin, catalase, and cytochrome *c* (Wooldridge & Williams, 1993). Consequently, the availability of free iron is not sufficient for bacterial growth in the tissues and body fluid of the host. Bacterial pathogens have evolved mechanisms to successfully compete for the iron within the host. Thus, iron-binding OMPs are essential components of the iron uptake mechanisms of bacteria. Several OMPs are involved in iron uptake in various bacterial species (Clarke *et al.*, 2001). A large number of Gram-negative bacterial species contain OMPs that are involved in the uptake of siderophores, iron chelating compounds produced by these microorganisms (Koebnik *et al.*, 2000; Neilands, 1995). The siderophore complex produced by *V. parahaemolyticus* was named vibrioferrin (VF) (Amin *et al.*, 2009). Transportation of iron compounds across the outer membrane requires energy provided by the proton motive force generated in the inner membrane. This energy is transmitted from the inner membrane by the TonB protein complex to the high affinity iron receptors in the outer membrane (Moeck & Coulton, 1998). In general, Gram-negative bacteria, including *E. coli,* possess only one TonB system (Moeck & Coulton, 1998), although multiple TonB systems have been identified in the *Vibrio* species (Kuehl & Crosa, 2011; Kustusch *et al.*, 2011). In particular, *V. parahaemolyticus* contains three TonB systems, namely TonB1, TonB2 and TonB3. The presence of three TonB systems suggests a more complicated energy transduction system for iron uptake and other transport systems in this organism. FhuA is a well-researched ferrichrome-iron receptor in *E.coli* (Ferguson, 1998) and is also found in other Gram-negative bacteria

including *V. parahaemolyticus* (Funahashi *et al.*, 2009). In addition, the porin BtuB is a TonB-dependent protein, requiring energy for vitamin B12 uptake.

### 1.5.1.3 Secretion

The protein secretion systems of Gram-negative bacteria require secretion pathways to allow intracellular proteins to pass though the inner membrane, periplasm and outer membrane. Six different secretion systems (types I – VI) have been identified in Gram-negative bacteria to date (Thanassi & Hultgren, 2000). The Sec system is a set of inner membrane proteins that facilitates the transport of secreted proteins from the cytoplasm across the inner membrane (Economou, 1999). Four secretion systems (types II, V, IV and VI) utilize the Sec system for protein translocation, whereas another two secretion systems (types I and III) are Sec-independent. The secretion pathways of the six secretion systems are illustrated in Fig. 1.26.

For the Sec-dependent secretion systems, substrates are manipulated by different secretion system machineries after crossing the inner membrane via the Sec pathway.
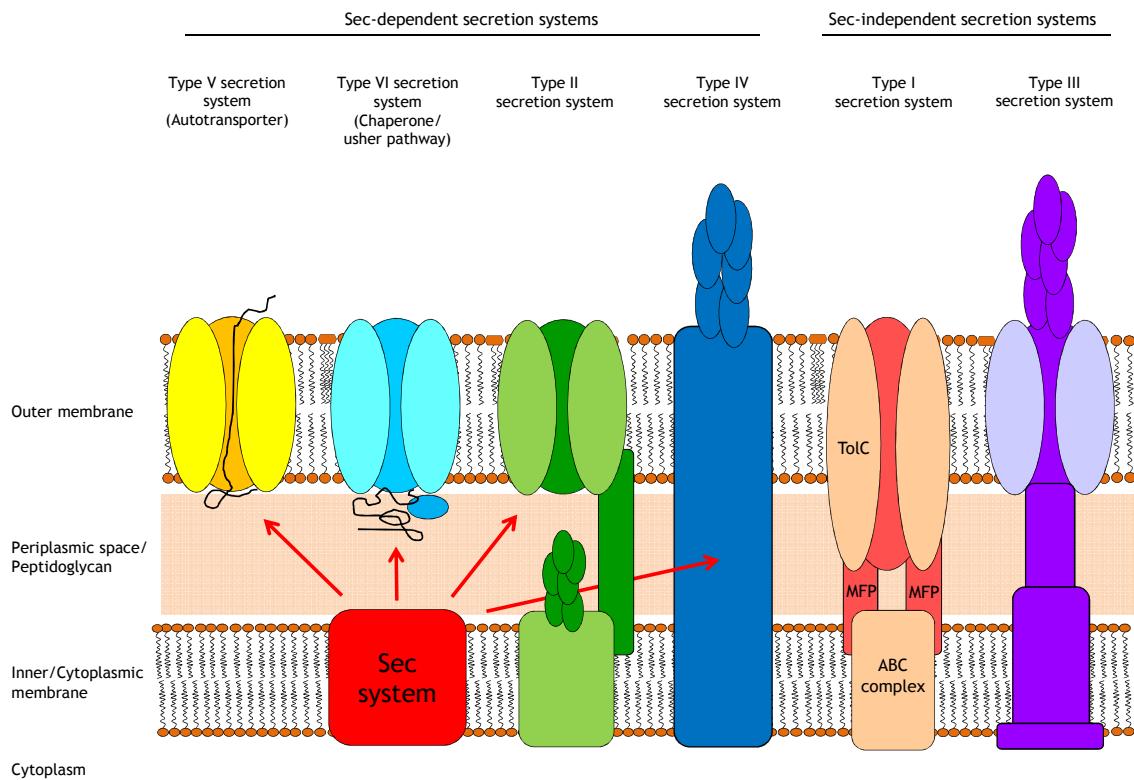
**Figure 1.26.** The six protein secretion systems in Gram-negative bacteria. The type II, type IV, type V and type VI secretion systems are Sec-dependent mechanisms that translocate substrates from the cytoplasm across the inner membrane via the Sec system. The type I and type III secretion systems are Sec-independent mechanisms that translocate intracellular substrates across the inner and outer membranes without involvement of the Sec system and intermediate chaperones. Figure adapted from Büttner & Bonas (2002).

The **type V secretion system** is the most simple, since it acts as an autotransporter, not requiring accessory factors for protein transition from the periplasm to the outer membrane. The **type VI secretion system**, also known as the chaperone/usher pathway, mediates transport of proteins from the periplasm to the bacterial cell surface by a periplasmic chaperone and OMP usher. This system is also involved in pilus subunit transport (Stathopoulos *et al.*, 2000). The **type II secretion system** is responsible for secretion of extra cellular enzymes and toxins (Stathopoulos *et al.*, 2000). This system contains a more

complex pathway than those of the type V and VI secretion systems since it requires 12-16 accessory proteins (Gsp complex) to mediate transport of protein molecules through the periplasm to the outer membrane surface. The **type IV secretion system** contains a channel protein spanning both the inner and outer membranes. This system is involved in DNA export from the bacterial cell to another bacterial cell or to a eukaryotic host cell as well as toxin secretion (e.g. pertussis toxin in *Bordetella pertussis*) and pilus formation (Burns, 1999).

For the Sec-independent secretion systems, substrates are secreted directly from the cytoplasm across the outer membrane without the inner membrane Sec pathway or a periplasm intermediate. The **type I secretion system**, also known as the ATP-binding cassette (ABC) protein exporter, is responsible for the export of several molecules, including toxins and enzymes (protease and lipase), in Gram-negative bacteria (Binet *et al.*, 1997). It consists of three components, an inner membrane ABC complex, a periplasm component or membrane fusion protein (MFP), and the TolC OMP. The TolC OMP is a transmembrane protein that plays an important role in the export of diverse molecules and in the control of multidrug efflux in the bacterial cell (Koronakis *et al.*, 2004). A TolC homologue is ubiquitously found and conserved among Gram-negative bacteria, including *E. coli* and *Vibrio* species (Andersen *et al.*, 2000; Paulsen *et al.*, 1997). TolC comprises a trimeric 12-stranded α/ß barrel which is integrated into the outer membrane and periplasm. The ß-barrel is embedded in the outer membrane with an open exit to the extracellular medium, whereas the α-helical barrel traverses the periplasmic space. A single channel formed by TolC allows chemical substances to pass though the cell envelop in a selective manner. TolC contributes to bacterial antibiotic resistance by regulating multidrug efflux activity. The **type III secretion system**, also known as an injectisome, is capable

of delivering effector proteins (virulence factors) from the bacterial cytoplasm into the cytosol of target eukaryotic cells (i.e. animal and plant cells) (Cornelis, 2006; Hueck, 1998). This system consists of approximately 20 proteins spanning the inner membrane, periplasm and outer membrane; the majority of these proteins are also associated with the flagellar basal body (Hueck, 1998). Several pathogenic Gram-negative bacteria including *E. coli* (EPEC), *Salmonella typhimurium, Shigella flexneri, Yersinia enterocolitica* and *V. parahaemolyticus* (see section 1.2.4.3 of this chapter) secrete virulence factors into host cells via the type III secretion machineries. The secreted virulence factors are capable of inducing various biochemical reactions in host cells such as inflammation, cytotoxicity, apoptosis, etc. (Marlovits & Stebbins, 2010).

## 1.5.2 Bioinformatic tools for discrimination of OMPs in different subcellular compartments

The OMP-encoding genes in the genome can be predicted by using a bioinformatic approach (Gromiha & Suwa, 2006; Gromiha, 2005; Jackups *et al.*, 2006; Juncker *et al.*, 2003). The OMP predictive tools are able to predict the OMPs encoded by the genome by determining subcellular localization (Gardy *et al.*, 2005; Imai *et al.*, 2008; Yu *et al.*, 2004) , ß-barrel conformation (Bagos *et al.*, 2004; Berven *et al.*, 2004; Garrow *et al.*, 2005; Ou *et al.*, 2008), and lipoproteins composition (Berven *et al.*, 2006; Juncker *et al.*, 2003) from the amino acid sequences of total open reading frames in the genome. Integration of predicted proteins by such software will generate a list of putative OMPs from a given bacterial genome.  Validation of these predictive tools for protein prediction was described by E. Komon *et al.* (2012).

## 1.5.2.1 Subcellular localization predictors

Gram-negative bacteria have five major subcellular localization sites: the cytoplasm, the inner membrane, the outer membrane, the periplasm, and the extracellular space. Predictive tools for this group are able to predict subcellular localization of the given protein sequences. Three software packages, CELLO (Yu *et al.*, 2004), PSORTb v.2.0 (Gardy *et al.*, 2005), and SoSUI-GramN (Imai *et al.*, 2008), are reliable tools for prediction of protein subcellular localization(E-komon *et al.*, 2012). CELLO utilizes a single module of the support vector machine (SVM) based on *n*-peptide composition to predict subcellular localization, whereas PSORTb v.2.0 uses multimodular SVM with different modules to examine specific location sites. Since the physico-chemical properties of the proteins in extracellular, the outer membrane, the periplasm, and the cytoplasm are less hydrophobic than the proteins in the inner membrane, SOSUI-GramN has been developed by using physicochemical parameters of the N- and C-terminal signal sequences and total amino acid sequences. Application of the SoSUI-GramN tool provides improved accuracy for predictions of extracellular proteins, compared with other predictive tools including CELLO and PSORTb v.2.0.

## 1.5.2.2 ß-barrel predictors

Integral membrane proteins are divided into two types, the α-helix and the ß-barrel. The α-helical proteins are present and more abundant in the cytoplasmic or inner membrane, whereas the ß-barrel proteins are located in the outer membrane in Gram-negative bacteria (and in chloroplasts and mitrochondria of eukaryotic cells) (Schulz, 2002). The group members of ß-barrel proteins contain membrane-spanning segments formed by antiparallel ß-strands. These structures

generate a channel in a barrel formation, which spans the outer membrane (Schulz, 2002). Several predictive tools have been developed to determine the OMPs of Gram-negative bacteria by classification of ß-barrel structures. A Markov Chain Model for Beta Barrels (MCMBB) utilizes a Hidden Morkov Model to predict transmembrane ß-strands of outer membrane of Gram-negative bacteria (Bagos *et al.*, 2004). This model considers only amino acid sequences and captures the structural characteristics of the transmembrane ß-strands of the outer membrane. Furthermore, the Hidden Morkov Model enables discrimination of the OMPs from the water-soluble proteins, which also form a ß-barrel structure, resulting in a more precise predictive result. The ß-barrel outer membrane protein predictor (BOMP) predicts integral ß-barrel protein based on two separate components. The first component is a recognition of the common C-terminal pattern of ß-barrel integral proteins, whereas the second component evaluates an integral ß-barrel score for the amino acid sequence by considering the sequences containing stretches of typical amino acids for transmembrane ß-strands (Berven *et al.*, 2004). The other predictive tools have been developed by applying different algorithms. TMB-Hunt utilizes a modified *k*-nearest neighbour (*k*-NN) algorithm to discriminate protein sequences as transmembrane ß-barrel or non-transmembrane ß-barrel from the entire amino acid sequence (Garrow *et al.*, 2005). A rigorous cross-validation procedure, including evaluation of differentially weighted amino acids, evolutionary information and calibration of the predictive scoring is incorporated with the *k*-NN algorithm for more accurate prediction. A more recently developed tool, TMBETADISC-RBF uses a predictive method based on radial basis function (RBF) networks (Ou *et al.*, 2008). This algorithm has been widely used for several bioinformatic applications, such as prediction of the cleavage sites in proteins (Yang & Thomson, 2005), inner residue contacts (Zhang & Huang, 2004), etc. This program also includes the

position specific scoring matrix (PSSM) profiles generated by a position-specific interative basic local alignment search tool (PSI-BLAST), and a non-redundant protein database for more robust discrimination.

### 1.5.2.3 Lipoprotein predictors

In Gram-negative bacteria, lipoproteins occur in both the inner and outer membranes. One feature characteristic of lipoproteins is a signal sequence that is covalently linked to the lipid chain (Hayashi & Wu, 1990). The signal sequence of the lipoprotein is cleaved by signal peptidase II (SPaseII). Lipoprotein signal peptides are somewhat similar to the signal peptides of secreted proteins which are cleaved by signal peptidase I (SPaseI) (Hayashi & Wu, 1990; Juncker *et al*., 2003). Lipoproteins can be identified from the peptides cleaved by SPaseII. Their final destination, either the inner membrane or the outer membrane, is determined by a single amino acid at the second amino acid at residue, position +2 (Seydel *et al*., 1999; Yamaguchi *et al*., 1988). Lipoproteins are terminally anchored to the outer membrane when the position +2 is serine (Ser) and to the inner membrane when position +2 is substituted by aspartic acid (Asp). Alteration of position +2 from Asp to Ser enables relocalization of an inner membrane lipoprotein to the outer membrane and *vice versa*. Lipop (Juncker *et al*., 2003) employs the hidden Markov model (HMM) to distinguish between proteins SPaseII-cleaved proteins, SPaseI-cleaved proteins, cytoplasmic proteins and transmembrane proteins. The predicted SPaseII-cleaved proteins will be subsequently categorized to an outer or inner membrane localization according to the amino acid at position +2. Lipo (Berven *et al*., 2006) was developed to analyse entire predicted proteomes and provides a list of recognised lipoproteins

categorised according to the similarity of their lipo-box to those of known Gram-negative lipoproteins in the database (http://www.bioinfo.no/tools/lipo).

## 1.5.3 Identification of OMPs

### 1.5.3.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Gel electrophoresis is widely used for protein identification since it allows the molecular weights of polypeptides in mixtures of protein to be determined. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is the most commonly used gel electrophoresis technique for protein analysis (Garfin, 2003). Important features of SDS-PAGE are its technical simplicity, reliability, and reproducibility. Polymerization of polyacrylamide gels occurs by copolymerization of acrylamide and N,N'-methylenebisacrylamide (bis). Gel formation is catalyzed by ammonium persulphate (APS) and N,N,N'N'-tetramethylethylene diamine (TEMED). The pore size of the gel can be altered by changing the concentration of polyacrylamide. During protein electropholysis, proteins move through the polyacrylamide pores according to both their size and electrical charge, but typically small polypeptides migrate from the cation to anion electrode at a greater rate than larger polypeptides. Consequently, protein profiles are produced based on the separation of different polypeptides according to their molecular weight. The molecular masses of the proteins can be estimated by comparing the positions of the protein bands with those of proteins of known sizes (in a molecular marker mixture). Further proteomic study of protein fractions is carried out by excising the protein bands from the gel followed by trypsin digestion to elute the peptides from the gel matrix.

Samples containing digested peptides are subjected to reverse phase liquid chromatography with tandem mass spectrophotometry (LC-MS/MS) to identify the composition of the proteins within each band.

One disadvantage of SDS-PAGE is that it requires proteins to be first denatured from constituent polypeptide chains. Thus, this method is unable to provide information about protein properties such as biological activity and antigenicity. Furthermore, polypeptides with similar molecular weights may not be distinguished since they will lie very close to each other in the gel.

A comparison of SDS-PAGE with other electrophoretic methods for protein analysis such as discontinuous polyacrylamide gel electrophorysis (DISC-PAGE), pore gradient electrophorysis (PGE), isoelectric focusing (IEF) and two-dimentional gel electrophorysis (2D-PAGE) is provided by Chiou *et al.* (1999). Although more advanced technologies for bacterial surface proteomic studies have been developed, SDS-PAGE still has advantages over the other techniques due to its efficiency in solubilizing integral membrane protein and its technical ease of use in the laboratory (Cordwell, 2006).

### 1.5.3.2 Mass spectrometry (MS)-based outer membrane proteomics

A comprehensive review of the principles and applications of mass-spectometry-based proteomics for the analysis of complex protein samples is provided by Aebersold *et al.* (2003). A mass spectrometer consists of an ion source, a mass analyzer that measures the mass-to-charge ratio ($m/z$) of the ionized peptide mixture, and the detector that recognize the number of ions at each $m/z$ value. The proteins can be ionized by different methods such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). ESI ionizes the

analytes from the peptide mixture solution and is subsequently coupled to liquid-based separation tools such as chromatography and electrophoresis. MALDI ionizes the samples from a dry, crystalline matrix via laser pulses. Integrated liquid chromatography ESI-MS systems (LC-MS) have been applied for analysis of complex protein mixtures, whereas MALDI-MS has been used most commonly for analysis of simple peptide mixtures. Four main types of mass analyser have been developed for protein identification: the ion trap, time-of-flight (TOF), quadrupole and Fourier transform ion cyclotron (FT-MS) analysers. The ion trap analyser has been used extensively in proteomic studies due to its robustness, sensitivity and reasonable cost. By the ion trap system, the ionized analytes are captured for a certain time interval and subsequently subjected to MS, normal mass spectrometry, or MS/MS, tandem mass spectrometry analysis. Protein determination is carried out by a search engine program (MASCOT) that uses the protein sequences from mass spectrometry data to identify the proteins from a primary sequence database.

## 1.6 Aim and objectives of research

The main aim of the research reported in this thesis was to study the molecular evolution and epidemiology of *V. parahaemolyticus* isolated from clinical samples, human carriers, frozen shrimp, farmed-shrimp, seafood, and water in Thailand in order to determine the source of infection and evolutionary relationships of *V. parahaemolyticus* isolates in seafood production. To meet this aim three different research objectives were pursued, each of which used an appropriate molecular technique. First, multilocus sequence typing (MLST) analysis of seven housekeeping genes was used to determine the genetic relationships and population structure of 102 representative *V. parahaemolyticus*

from various sources (Chapter 2). The MLST scheme used in the present study was adapted from an existing scheme but improved by using nested primers. Multiplex PCR was also developed for detection and DNA sequencing of the seven housekeeping genes. This MLST analysis of 101 *V. parahaemolyticus* provided a framework to select representative isolates for studying nucleotide variation of selected virulence genes and comparing outer membrane proteomes.

Second, the distribution and nucleotide variation of the haemolysin genes (*tdhA, tdhS, trh1* and *trh2*) and TTSS1-related genes (*vcrD1, vscC1* and *VP1680*) of isolates recovered from clinical samples, human carriers and seafood in Thailand were analyzed by PCR and DNA sequencing (Chapter 3). Furthermore, the nucleotide sequences of TTSS1-related gene fragments of selected Thai clinical, human carrier, and seafood isolates were compared with those of pathogenic isolates of worldwide distribution to determine genetic relationships of virulence-related genes between Thai and worldwide *V. parahaemolyticus* isolates.

Third, the OMPs of pandemic *V. parahaemolyticus* serotype O3:K6 isolate RIMD2210633 were predicted from the genome sequence by using a bioinformatic approach (Chapter 4). Comparative OMP analyses of eight representative *V. parahaemolyticus* isolates from various sources, including clinical samples, human carriers, seafood, shrimp tissue, and water in Thailand was performed using SDS-PAGE and mass spectrometry.

# 2. MOLECULAR EVOLUTIONARY RELATIONSHIPS OF *V. PARAHAEMOLYTICUS* ISOLATES BY MULTILOCUS SEQUENCE TYPING (MLST)

## 2.1 Introduction

The public health and commercial burden associated with *V. parahaemolyticus* contamination is very high in Thailand due to the wide consumption of seafood. Clinical isolates from Thailand typically correspond to the pandemic serovar O3:K6, which was responsible for the Indian pandemic in 1996 (Matsumoto *et al.*, 2000; Nair *et al.*, 2007; Nasu *et al.*, 2000; Okuda *et al.*, 1997; Vuddhakul *et al.*, 2000), as well as variants of this clone (O1:KUT, O1:K25) (Chowdhury *et al.*, 2000, 2004) and the novel serovar O3:K46 (Serichantalergs *et al.*, 2007). Although a number of molecular approaches have been used to study the epidemiology of *V. parahaemolyticus* in Thailand (Bhoopong *et al.*, 2007; Laohaprertthisan et al., 2003; Serichantalergs *et al.*, 2007; Suthienkul *et al.*, 1996; Vuddhakul *et al.*, 2000; Wootipoom *et al.*, 2007), these studies have not generated detailed sequence-based molecular evolutionary data.

Multilocus sequence typing (MLST) is an important tool for molecular epidemiology and population genetic studies of bacterial pathogens (Cooper & Feil, 2004; Maiden *et al.*, 1998; Maiden, 2006; Turner & Feil, 2007; Urwin & Maiden, 2003). The utilization of housekeeping genes encoding core metabolic enzymes means that the data are unlikely to be impacted by strong positive selection (Maiden, 2006). A successful MLST scheme for *V. parahaemolyticus* has been established by González-Escalona *et al.* (González-Escalona *et al.*, 2008).

Global strains were demonstrated to be genetically diverse and possessed a weakly clonal population structure containing three major clonal complexes CC3, CC34 and CC36. The major clonal complex CC3 comprises pandemic strains of worldwide distribution whereas the clonal complexes CC34 and CC36 consist of strains isolated from the Gulf and Pacific coasts of the USA, respectively. A MLST study of *V. parahaemolyticus* strains isolated from the South Eastern Chinese coast revealed high genetic diversity among strains from a single geographical area (Yu *et al.*, 2011) and a large proportion of clinical strains were associated with the pandemic CC3 identified by González-Escalona *et al.* (2008). The pandemic O3:K6 clone corresponds to clonal complex CC3, and is thought to have originated from an environmental non-pathogenic O3:K6 strain by horizontal gene transfer (Chao *et al.*, 2011). Yan *et al.* (Yan *et al.*, 2011) developed an extended MLST scheme using nine housekeeping genes [four of which are the same loci used in the previous scheme (González-Escalona *et al.*, 2008)] and the haemolysin gene (*tl*) to investigate strains isolated from Asian countries and the USA. These authors identified three major clonal complexes, CC1, CC2 and CC3, representing clinical O3:K6 strains isolated before 1996, pandemic O3:K6 strains isolated in 1996 and in subsequent years, and non-clinical strains, respectively. Furthermore, MLST analysis of *V. parahaemolyticus* isolates from Great Bay Estuary of New Hampshire, USA, suggested high-levels of genetic diversity among environmental isolates recovered from the same region (Ellis *et al.*, 2012). This high genetic diversity is likely to increase in warmer seasons. With the exception of *pyrC,* these authors used housekeeping genes following the existing MLST scheme (González-Escalona *et al.*, 2008) and the virulence-related genes *gacA, toxR* and *vppC* were also included. No significant difference in the level of recombination between housekeeping and virulence genes was observed by these authors. There is no evidence of linkage between

sequence type and serotype in *V. parahaemolyticus*, indicating that serotype switching by recombination has been common in this species (Chao *et al.*, 2011; Chowdhury *et al.*, 2004; González-Escalona *et al.*, 2008). Overall, these studies have confirmed that MLST presents a powerful means to determine the role of recombination in the diversification of natural populations of *V. parahaemolyticus* and for understanding the processes leading to the emergence and spread of clinically relevant strains.

Although undercooked seafood has been identified as a source of *V. parahaemolyticus* infection (Fujino *et al.*, 1953; Barker *et al.*, 1974), the relative likelihood of contamination from different settings (e.g. the natural marine environment, aquacultural sources or the market place) has not been established. However, such evidence is essential to guide future intervention strategies and minimize both the risk to the consumer and the cost to the producer. This study determines the extent to which MLST can help to address this issue, by applying this technique to a strain collection recovered from different epidemiological sources associated with the seafood industry in Thailand. Isolates were obtained from clinical samples, human carriers (healthy workers in a seafood factory), fresh seafood (oysters, bloody clams, crab meat, mussels and white shrimps), frozen shrimp, fresh-farmed shrimp tissue, and shrimp-farm water. The data confirm a highly diverse population, with very limited evidence of concordance between ST and epidemiological source. The data also confirm high rates of recombination, particularly at *recA*, and a novel approach to clustering the isolates on the basis of amino acid sequences is presented. Furthermore, MLST analysis was also applied to nine *V. parahaemolyticus* isolates from European countries. These isolates include clinical and environmental isolates from the UK and Norway. The genetic

relationship of Thai and European isolates was compared by using MLST analyses. This chapter also includes evaluation of *Taq* polymerase enzyme kits from various manufacturers and optimization of PCR conditions, e.g. annealing temperature, magnesium concentration and different primer combination. Multiplex PCR was also developed for amplifying the seven housekeeping gene fragments used in the study.

## 2.2 Materials and methods

### 2.2.1 Bacterial strains and growth conditions

*V. parahaemolyticus* type strain NCTC 10903 (ATCC 17802$^T$) was obtained from the National Collection of Type Cultures (NCTC), Health Protection Agency (HPA), UK. A total of 119 *V. parahaemolyticus* isolates were provided by Prof Orasa Suthienkul, Department of Microbiology, Faculty of Public Health, Mahidol University, Thailand. These isolates were obtained from six categories of sources as follow. Clinical isolates (n=20) were recovered from gastroenteritis patients, human carrier isolates (n=20) were recovered from the faeces of healthy workers in a seafood plant, fresh seafood isolates (n=20) were recovered from various seafood products at local markets in central Thailand, frozen shrimp isolates (n=20) were recovered from frozen shrimp at a processing factory, shrimp tissue isolates (n=20) were recovered from fresh shrimp at two intensive shrimp farms (farms 1 and 2) in southern Thailand, and water isolates (n=19) were recovered from water samples at the same two shrimp farms. Location of shrimp farms, 1 and 2, and protocols of sampling methods for isolates from shrimp tissue and water are demonstrated in Fig. A1-5 Appendix 1.

Of 119 Thai *V. parahaemolyticus* isolates, eight showed colony morphology variation (e.g. transparent and opaque) on the plate after first subculture and were subcultured onto a second plate. For example, opaque and transparent colonies were separately subcultured to two different plates and were renamed as A and B. Each A and B strain was used for separate sequencing analysis. Thus, the total number of strains examined in this study was 128 including eight extra strains representing variable colony morphology and one type strain from Japan. Off these 128 isolates, 27 failed to be amplified one or more housekeeping genes. Thus, total of 101 isolates were used for the MLST analysis.

Details of the isolates examined by MLST are provided in Table 2.1. Furthermore, nine isolates from European countries were kindly provided by Dr. Rachel Rangdale, Centre of Environment, Fisheries, and Aquaculture Science (Cefas), Weymouth, United Kingdom. These isolates included clinical isolates from the UK (n=3) and Norway (n=3) and environmental isolates from the UK (n=3). Details of these European isolates are provided in Table 2.2. The isolates were stored at -80$^{o}$C in 50% (v/v) glycerol in Tryptone Soya Broth (TSB, Oxoid) with 3% (w/v) NaCl and subcultured on Tryptone Soya Agar (TSA, Oxoid) with 3% (w/v) NaCl by overnight aerobic incubation at 37$^{o}$C. For preparation of DNA, a few colonies were inoculated into 10 ml volumes of TSB and grown aerobically overnight at 37$^{o}$C at 120 rpm.

## *2.2.2 Preparation of chromosomal DNA*

Cells from 1.0 ml of overnight cultures were harvested by centrifugation for 1 min at 13,000 x *g* and washed once in sterile distilled H$_2$O. DNA was prepared with InstaGene Matrix (Bio-Rad) according to the manufacturer's instructions and stored at -20$^{o}$C.

**Table 2.1.** Properties of the 102 *V. parahaemolyticus* isolates used in the MLST study

| Isolate | Source of isolation | Year of isolation | Serotype | tdh[a] | trh[a] | ST[b] | Allelic profile[b] |
|---|---|---|---|---|---|---|---|
| | | | | | | | *dnaE gyrB recA dtdS pntA pyrC tnaA* |
| VP2 | Food poisoning (Type strain) | 1950 | O1:K1 | - | - | 1 | 5, 52, 27, 13, 17, 25, 10 |
| VP4 | Frozen shrimp from processing plant A | April 1999 | O3:K20 | - | - | 229 | 109, 136, 25, 121, 83, 107, 83 |
| VP6 | Frozen shrimp from processing plant A | April 1999 | O10:K66 | - | - | 230 | 110, 144, 166, 35, 18, 108, 86 |
| VP8 | Frozen shrimp from processing plant A | April 1999 | O3:KUT | - | - | 231 | 111, 17, 3, 123, 85, 37, 87 |
| VP12 | Frozen shrimp from processing plant A | April 1999 | O1:KUT | - | - | 241 | 112, 143, 25, 120, 26, 109, 81 |
| VP14 | Frozen shrimp from processing plant A | April 1999 | O2:K3 | - | - | 232 | 98, 131, 30, 32, 77, 11, 82 |
| VP16 | Frozen shrimp from processing plant A | April 1999 | O3:K20 | - | - | 233 | 109, 136, 114, 121, 83, 107, 83 |
| VP20 | Frozen shrimp from processing plant A | April 1999 | O1:K25 | - | - | 242 | 113, 145, 61, 70, 28, 11, 26 |
| VP22 | Frozen shrimp from processing plant A | April 1999 | O1:K20 | - | - | 234 | 5, 84, 115, 74, 84, 26, 84 |
| VP24 | Frozen shrimp from processing plant A | April 1999 | O9:K44 | - | - | 235 | 10, 69, 27, 76, 46, 65, 29 |
| VP26 | Frozen shrimp from processing plant A | April 1999 | O7:K19 | - | - | 236 | 114, 100, 61, 122, 66, 54, 85 |
| VP28 | Frozen shrimp from processing plant A | April 1999 | O5:K17 | - | - | 237 | 115, 43, 25, 108, 71, 73, 62 |
| VP30 | Frozen shrimp from processing plant A | April 1999 | O3:K20 | - | - | 233 | 109, 136, 114, 121, 83, 107, 83 |
| VP32 | Frozen shrimp from processing plant A | April 1999 | O2:KUT | - | - | 243 | 12, 146, 117, 124, 28, 10, 54 |
| VP36 | Frozen shrimp from processing plant A | April 1999 | O10:KUT | - | - | 238 | 109, 136, 114, 121, 21, 107, 83 |
| VP38 | Frozen shrimp from processing plant A | April 1999 | O3:K20 | - | - | 239 | 6, 6, 3, 17, 11, 26, 23 |
| VP40 | Frozen shrimp from processing plant A | April 1999 | O11:K40 | - | - | 240 | 31, 147, 62, 84, 4, 45, 88 |
| VP44 | Water from shrimp farm 1, pond A | January 2008 | O9:K23 | - | - | 244 | 19, 74, 61, 68, 86, 11, 26 |
| VP46 | Water from shrimp farm 1, pond A | January 2008 | O1:K38 | - | - | 244 | 19, 74, 61, 68, 86, 11, 26 |
| VP48 | Water from shrimp farm 1, pond B | January 2008 | O9:K24 | - | - | 244 | 19, 74, 61, 68, 86, 11, 26 |
| VP50 | Water from shrimp farm 1, pond B | January 2008 | O7:K52 | - | - | 244 | 19, 74, 61, 68, 86, 11, 26 |
| VP52 | Water from shrimp farm 1, pond B | January 2008 | O1:KUT | - | - | 244 | 19, 74, 61, 68, 86, 11, 26 |
| VP54 | Water from shrimp farm 1, pond B | January 2008 | O1:KUT | - | - | 244 | 19, 74, 61, 68, 86, 11, 26 |
| VP56 | Water from shrimp farm 1, pond B | January 2008 | O2:KUT | - | - | 245 | 49, 148, 25, 125, 60, 110, 89 |

Table 2.1. (continued)

| Isolate | Source of isolation | Year of isolation | Serotype | *tdh*[a] | *trh*[a] | ST[b] | Allelic profile[b] |
|---------|--------------------|--------------------|----------|-----------|-----------|--------|--------------------|
| | | | | | | | *dnaE gyrB recA dtdS pntA pyrC tnaA* |
| VP58 | Water from shrimp farm 1, pond B | January 2008 | O7:KUT | - | - | 239 | 6, 6, 3, 17, 11, 26, 23 |
| VP60 | Water from shrimp farm 1, pond B | January 2008 | O2:K3 | - | - | 246 | 33, 87, 24, 5, 10, 5, 1 |
| VP62 | Water from shrimp farm 1, pond C | January 2008 | O1:KUT | - | - | 247 | 116, 149, 72, 76, 45, 62, 26 |
| VP64 | Water from shrimp farm 1, pond C | January 2008 | OUT:KUT | - | - | 248 | 5, 88, 61, 19, 18, 3, 90 |
| VP66 | Water from shrimp farm 1, pond C | January 2008 | O1:K26 | - | - | 248 | 5, 88, 61, 19, 18, 3, 90 |
| VP72 | Water from shrimp farm 2, pond A | January 2008 | O9:K44 | - | - | 249 | 3, 151, 25, 29, 61, 11, 62 |
| VP74 | Water from shrimp farm 2, pond A | January 2008 | O2:K3 | - | - | 180 | 89, 105, 15, 89, 57, 11, 57 |
| VP76 | Water from shrimp farm 2, pond A | January 2008 | O7:K52 | - | - | 249 | 3, 151, 25, 29, 61,11, 62 |
| VP80 | Water from shrimp farm 2, pond B | January 2008 | O6:K46 | - | - | 250 | 118, 87, 119, 117, 54, 111, 91 |
| VP84 | Shrimp hepatopancreas farm 2, pond B | August 2007 | O10:K71 | - | - | 251 | 119, 152, 120, 29, 23, 11, 61 |
| VP86 | Shrimp hepatopancreas farm 2, pond B | August 2007 | O4:K63 | - | - | 251 | 119, 152, 120, 29, 23, 11, 61 |
| VP88 | Shrimp hepatopancreas farm 2, pond B | August 2007 | O9:K23 | - | - | 251 | 119, 152, 120, 29, 23, 11, 61 |
| VP90 | Shrimp hepatopancreas farm 2, pond B | August 2007 | O2:K28 | - | - | 251 | 119, 152, 120, 29, 23, 11, 61 |
| VP94 | Shrimp muscle farm 2, pond B | August 2007 | O1:KUT | - | - | 251 | 119, 152, 120, 29, 23, 11, 61 |
| VP96 | Shrimp muscle farm 2, pond B | August 2007 | O10:K52 | - | - | 251 | 119, 152, 120, 29, 23, 11, 61 |
| VP98 | Shrimp muscle farm 2, pond B | August 2007 | O2:KUT | - | - | 251 | 119, 152, 120, 29, 23, 11, 61 |
| VP100 | Shrimp intestine farm, 1 pond B | August 2007 | O1:KUT | - | - | 246 | 33, 87, 24, 5, 10, 5, 1 |
| VP102 | Shrimp intestine  farm, 1 pond B | August 2007 | O1:K1 | - | - | 246 | 33, 87, 24, 5, 10, 5, 1 |
| VP104 | Shrimp intestine farm, 2 pond B | August 2007 | O9:K23 | - | - | 251 | 119, 152, 120, 29, 23, 11, 61 |
| VP106 | Shrimp intestine farm, 2 pond B | August 2007 | O1:KUT | - | - | 251 | 119, 152, 120, 29, 23, 11, 61 |
| VP108 | Shrimp internal body farm, 1 pond B | August 2007 | O2:K3 | - | - | 246 | 33, 87, 24, 5, 10, 5, 1 |
| VP110 | Shrimp internal body  farm, 1 pond B | August 2007 | O1:K56 | - | - | 246 | 33, 87, 24, 5, 10, 5, 1 |
| VP112 | Shrimp internal body farm, 1 pond B | August 2007 | O9:K44 | - | - | 246 | 33, 87, 24, 5, 10, 5, 1 |
| VP114 | Internal body farm, 1 pond B | August 2007 | O7:KUT | - | - | 246 | 33, 87, 24, 5, 10, 5, 1 |
| VP116 | Shrimp shell farm, 1 pond B | August 2007 | OUT:KUT | - | - | 246 | 33, 87, 24, 5, 10, 5, 1 |
| VP118 | Shrimp shell farm, 1 pond B | August 2007 | OUT:KUT | - | - | 246 | 33, 87, 24, 5, 10, 5, 1 |
| VP122 | Shrimp shell farm, 2 pond B | August 2007 | OUT:KUT | - | - | 251 | 119, 152, 120, 29, 23, 11, 61 |

Table 2.1. (continued)

| Isolate | Source of isolation | Year of isolation | Serotype | *tdh*[a] | *trh*[a] | ST[b] | Allelic profile[b] *dnaE gyrB recA dtdS pntA pyrC tnaA* |
|---------|---------------------|-------------------|----------|----------|----------|-------|-------------------------------------------------------|
| VP126 | Healthy carrier from seafood plant B | November 2003 | O10:KUT | - | - | 252 | 36, 153, 121, 126, 27, 112, 17 |
| VP128 | Healthy carrier from seafood plant B | November 2003 | O9:K23 | - | - | 253 | 35, 154, 25, 50, 73, 35, 23 |
| VP130 | Healthy carrier from seafood plant B | November 2002 | O1:KUT | - | - | 254 | 112, 155, 122, 19, 87, 96, 92 |
| VP132 | Healthy carrier from seafood plant B | August 2003 | O3:K46 | + | - | 3 | 3, 4, 19, 4, 29, 4, 22 |
| VP136 | Healthy carrier from seafood plant B | October 2003 | O1:KUT | + | + | 199 | 22, 28, 17, 13, 8, 19, 14 |
| VP138 | Healthy carrier from seafood plant B | November 2003 | O11:K5 | + | + | 255 | 12, 156, 123, 127, 19, 12, 47 |
| VP140 | Healthy carrier from seafood plant B | August 2003 | O4:KUT | + | - | 62 | 19, 4, 88, 2, 34, 18, 23 |
| VP142 | Healthy carrier from seafood plant B | July 2002 | OUT:KUT | - | - | 256 | 42, 157, 59, 128, 6, 113, 57 |
| VP144 | Healthy carrier from seafood plant B | July 2002 | O4:K63 | + | - | 257 | 120, 158, 89, 129, 26, 114, 93 |
| VP146 | Healthy carrier from seafood plant B | July 2002 | O1:K25 | - | - | 258 | 121, 95, 124, 130, 26, 115, 12 |
| VP148 | Healthy carrier from seafood plant B | August 2003 | O1:K41 | - | - | 256 | 42, 157, 59, 128, 6, 113, 57 |
| VP150 | Healthy carrier from seafood plant B | October 2003 | O1:KUT | + | - | 259 | 122, 25, 125, 131, 21, 11, 66 |
| VP152 | Healthy carrier from seafood plant B | November 2002 | O8:K21 | - | - | 259 | 122, 25, 125, 131, 21, 11, 66 |
| VP154 | Healthy carrier from seafood plant B | October 2003 | O4:KUT | + | - | 68 | 41, 40, 36, 41, 36, 39, 32 |
| VP156 | Healthy carrier from seafood plant B | November 2002 | O1:K1 | - | - | 260 | 3, 82, 126, 69, 30, 7, 23 |
| VP158 | Healthy carrier from seafood plant B | August 2003 | O1:KUT | + | - | 3 | 3, 4, 19, 4, 29, 4, 22 |
| VP160 | Healthy carrier from seafood plant B | November 2003 | O11:K40 | - | - | 261 | 123, 100, 127, 50, 66, 11, 31 |
| VP162 | Healthy carrier from seafood plant B | November 2003 | O1:K12 | + | + | 255 | 12, 156, 123, 127, 19, 12, 47 |
| VP164 | Clinical sample from a hospital patient | September 1990 | O1:K58 | + | - | 17 | 13, 10, 19, 27, 28, 27, 21 |
| VP166 | Clinical sample from a hospital patient | August 1990 | O1:K1 | + | + | 83 | 5, 52, 27, 13, 17, 25, 40 |
| VP168 | Clinical sample from a hospital patient | August 1990 | O3:KUT | - | + | 66 | 42, 25, 3, 40, 35, 38, 31 |
| VP170 | Clinical sample from a hospital patient | Febuary 1991 | O3:KUT | + | + | 262 | 105, 156, 123, 127, 19, 12, 47 |
| VP172 | Clinical sample from a hospital patient | August 1990 | O1:K1 | + | + | 83 | 5, 52, 27, 13, 17, 25, 40 |
| VP174 | Clinical sample from a hospital patient | Febuary 1991 | O3:KUT | + | + | 263 | 50, 55, 48, 52, 23, 53, 47 |
| VP176 | Clinical sample from a hospital patient | May 1990 | O1:K1 | + | + | 264 | 5, 52, 27, 132, 17, 25, 40 |
| VP178 | Clinical sample from a hospital patient | January 1991 | O1:K69 | + | + | 262 | 105, 156, 123, 127, 19, 12, 47 |
| VP180 | Clinical sample from a hospital patient | September 1990 | O8:K22 | + | - | 262 | 105, 156, 123, 127, 19, 12, 47 |

Table 2.1. (continued)

| Isolate | Source of isolation | Year of isolation | Serotype | *tdh*[a] | *trh*[a] | ST[b] | Allelic profile[b] |
|---|---|---|---|---|---|---|---|
| | | | | | | | *dnaE gyrB recA dtdS pntA pyrC tnaA* |
| VP182 | Clinical sample from a hospital patient | April 1990 | O1:K69 | - | + | 262 | 105, 156, 123, 127, 19, 12, 47 |
| VP184 | Clinical sample from a hospital patient | April 1990 | O4:K11 | + | - | 262 | 105, 156, 123, 127, 19, 12, 47 |
| VP188 | Clinical sample from a hospital patient | January 1991 | O1:K69 | + | + | 262 | 105, 156, 123, 127, 19, 12, 47 |
| VP190 | Clinical sample from a hospital patient | July 1990 | O4:K10 | + | - | 265 | 11, 48, 107, 48, 26, 48, 26 |
| VP194 | Clinical sample from a hospital patient | September 1990 | OUT:KUT | + | - | 83 | 5, 52, 27, 13, 17, 25, 40 |
| VP200 | Clinical sample from a hospital patient | September 1990 | O4:K8 | + | - | 189 | 11, 48, 3, 48, 26, 48, 26 |
| VP204 | Fresh oysters from market | March 2003 | O1:K64 | - | - | 267 | 126, 161, 43, 19, 54, 10, 26 |
| VP206 | Fresh oysters from market | March 2003 | O2:K3 | - | - | 268 | 127, 67, 24, 133, 10, 5, 95 |
| VP208 | Fresh oysters from market | March 2003 | OUT:KUT | - | - | 269 | 128, 162, 128, 134, 1, 11, 94 |
| VP210 | Fresh bloody clams from market | December 2003 | O7:KUT | - | - | 270 | 129, 151, 98, 12, 88, 117, 26 |
| VP212 | Fresh bloody clams from market | December 2003 | O10:K19 | - | - | 271 | 3, 134, 81, 76, 26, 11, 26 |
| VP214 | Fresh bloody clams from market | December 2003 | O1:KUT | - | - | 272 | 60, 116, 91, 135, 50, 118, 2 |
| VP216 | Boiled crab meat from market | October 2002 | O2:KUT | - | - | 273 | 130, 58, 113, 69, 89, 119, 23 |
| VP218 | Boiled crab meat from market | October 2002 | O1:KUT | - | - | 274 | 69, 92, 69, 27, 54, 71, 24 |
| VP220 | Boiled crab meat from market | October 2002 | OUT:KUT | - | - | 275 | 31, 163, 129, 19, 36, 120, 96 |
| VP222 | Boiled mussels from market | July 2003 | NA | - | - | 276 | 131, 147, 60, 136, 90, 27, 23 |
| VP224 | Boiled mussels from market | July 2003 | O10:KUT | - | - | 276 | 131, 147, 60, 136, 90, 27, 23 |
| VP226 | Boiled mussels from market | July 2003 | OUT:KUT | - | - | 277 | 3, 29, 98, 67, 26, 121, 33 |
| VP228 | Fresh shrimp from market | June 2003 | O1:KUT | - | - | 278 | 103, 164, 130, 137, 50, 122, 57 |
| VP230 | Fresh shrimp from market | June 2003 | O3:K58 | - | - | 279 | 132, 165, 25, 110, 4, 116, 12 |
| VP232 | Fresh shrimp from market | June 2003 | O5:KUT | - | - | 278 | 103, 164, 130, 137, 50, 122, 57 |
| VP234 | Fresh shrimp from market | June 2003 | O5:KUT | - | - | 114 | 55, 15, 31, 55, 18, 58, 46 |
| VP236 | Fresh shrimp from market | June 2003 | O1:K69 | - | - | 280 | 28, 144, 3, 138, 26, 123, 97 |
| VP238 | Fresh shrimp from market | June 2003 | O10:K52 | - | - | 281 | 133, 67, 4, 79, 43, 63, 23 |

[a] Results were obtained by Prof Orasa Suthienkul, Mahidol University, Thailand. [b] Results were obtained by MLST analysis from the present study.

**Table 2.2.** Properties of European *V. parahaemolyticus* nine isolates

| Isolate | Source of isolation | Serotype | *tdh*[a] | *trh*[a] | ST[b] | Allelic profile[b] |
|---|---|---|---|---|---|---|
| VP 244 | Oyster, UK | O5:K17 | + | - | 79 | 35, 43, 38, 21, 31, 35, 37 |
| VP 246 | Oyster , UK | N/A | + | - | 79 | 35, 43, 38, 21, 31, 35, 37 |
| VP 248 | Patient (Food poisoning), UK | O3:K4 | + | - | 331 | 147, 181, 127, 69, 26, 18, 23 |
| VP 250 | Clinical, Norway | O6:KUT | - | - | 346 | 45, 45, 143, 7, 14, 46, 36 |
| VP 252 | Clinical, Norway | O3:K6 | + | - | 3 | 3, 4, 19, 4, 29, 4, 22 |
| VP 254 | Clinical, Norway | O4:K9 | + | + | 34 | 20, 25, 15, 13, 7, 11, 5 |
| VP 258 | Clinical, UK | O3:K6 | + | N/A | 3 | 3, 4, 19, 4, 29, 4, 22 |
| VP 260 | Clinical, UK | O3:K6 | + | - | 3 | 3, 4, 19, 4, 29, 4, 22 |
| VP 262 | Chinese mitten crab, UK (Thames) | O1:KUT | + | - | 347 | 103, 186, 31, 78, 2, 144, 26 |

[a] Results were obtained by Dr. Rachel Rangdale, Cefas, UK. [b] Results were obtained by MLST analysis from the present study.

## *2.2.3 Optimization of DNA polymerase kits*

*OmpA* of *Mannheimia haemolytica* (PH2) was amplified for DNA polymerase kit optimization. Thirteen *Taq* DNA polymerase kits from the following manufacturers were evaluated: Invitrogen, New England Biolabs (NEBs), Novagen, Promega, Roche and Thermo Scientific (Table 2.3).  The *ompA* fragments were amplified by PCR from genomic DNA of *M. haemolytica* isolate PH2, using the following forward and reverse primers: 5'-AAGTTCTGTTTCAGGGCCCGCAAGCTAACACTTTCTACGCAGG-3' and 5'-ATGGTCTAGAAA GCTTTAACCTTGACCGAAACGGTATG-3'. PCRs were performed in 50 µl reaction mixes according to the manufacturers' instructions for each *Taq* DNA polymerase kit (product codes are shown in Table 2.3) with 12.5 pmol µl$^{-1}$ of each forward and reverse primer. Amplification was carried out in GeneAmp PCR System 9700 Thermo Cycler (Applied Biosystems) using 30 cycles of the following amplification conditions:  denaturation at

94°C for 45 s, annealing at 56°C for 45 s, and extension at 72°C for 2 min. An initial denaturation step of 94°C for 2 min was used and a final extension step at 72°C for 10 min. dNTPs (GE health care) were used at a final concentration of 1.25 mM and 4 μl of the reagent was used for a 50 μl PCR reaction. The PCR products were confirmed by electrophoresis in a 1% (w/v) agarose gel and visualised with 0.004% (v/v) SybrSafe (Invitrogen). Amplicon size was assured using a 1 Kb DNA ladder (Invitrogen).

**Table 2.3.** Details of DNA polymerase kits and cost calculation

| No. | DNA polymerase kit (product code) | Price (£/number of unit) | Price/ unit (£) | Unit used/ Reaction | Cost/ Reaction (£) | Total cost/700 reactions (£) |
|---|---|---|---|---|---|---|
| 1 | Invitrogen native (18038-018) | 120/500 | 0.24 | 1.00 | 0.24 | 168.00 |
| 2 | Invitrogen recombinant (10342-020) | 118/500 | 0.22 | 1.00 | 0.22 | 154.00 |
| 3 | Invitrogen *Pfx* (11708-013) | 338/500 | 0.68 | 1.00 | 0.68 | 473.20 |
| 4 | NEBs standard (M0273G) | 39/400 | 0.10 | 1.25 | 0.12 | 85.31 |
| 5 | NEBs LongAmp (M0323G) | 65/500 | 0.13 | 5.00 | 0.65 | 455.00 |
| 6 | NEBs Phire (F-120S) | 328/1000 reactions | ND | 1.00 | 0.33 | 229.60 |
| 7 | NEBs Phusion (F-530) | 250/500 | 0.50 | 1.00 | 0.50 | 350.00 |
| 8 | Novagen (71676-3) | 88/200 reactions | ND | 1.25 | 0.44 | 308.00 |
| 9 | Promega GoTaq (M3172) | 98/500 | 0.20 | 1.25 | 0.25 | 171.50 |
| 10 | Promega GoTaq Hotstart (M5002) | 130/500 | 0.26 | 1.25 | 0.33 | 227.50 |
| 11 | Roche (11146165001) | 97/500 | 0.19 | 1.25 | 0.24 | 169.75 |
| 12 | Thermo standard (AB-0192) | 73/250 | 0.29 | 1.25 | 0.37 | 255.50 |
| 13 | Thermo redhot (AB-0406) | 146/500 | 0.29 | 1.25 | 0.37 | 255.50 |

## 2.2.4 Primer design and PCR amplifications of seven housekeeping enzyme genes (dnaE, gyrB, recA, dtdS, pntA, pyrC, and tnaA)

Selection of the seven loci analyzed by MLST was based on the previously published MLST scheme for *V. parahaemolyticus* (González-Escalona *et al.*, 2008). For chromosome I, the housekeeping genes used were *recA* (RecA protein), *dnaE* (DNA polymerase III, alpha subunit), and *gyrB* (DNA gyrase, subunitB). For chromosome II, the housekeeping genes used were *dtdS* (threonine 3-dehydrogenase), *pntA* (transhydrogenase, alpha subunit), *pyrC* (dihydro-oratase), and *tnaA* (tryophanase). Because nested amplification gives more accurate sequencing results and is recommended for MLST (Maiden, 2006), two sets of primers, PCR and sequencing primers, were used for amplification and sequencing of the seven gene fragments. New PCR primers (located upstream of the existing PCR / sequencing primers) were designed using Primer Designer version 2 (Scientific and Educational software). In addition, new shorter sequencing primers were designed based on those of the previous *V. parahaemolyticus* MLST study using the same software. All primers were designed to a length of 18 nucleotides except *gyrB*-F2 which contains 17 nucleotides. The primers were diluted to 12.5 pmol μl$^{-1}$ for PCR reactions and 2 pmol μl$^{-1}$ for sequencing reactions. The nucleotide sequences of the primers used for PCR and DNA sequencing are provided in Table 2.4. Nucleotide sequences and position of PCR and sequencing primers used for each gene are showed in Figs. 2.1-2.7. In some isolates, gene fragments could not be amplified by these primers, in which case additional primers were designed and provided in Table 2.5. PCR fragments containing partial segments of *dnaE* (776 bp), *gyrB* (758 bp), *recA* (932 bp), *dtdS* (572 bp), *pntA* (676 bp), *pyrC* (681 bp), and *tnaA* (600 bp)  were amplified from chromosomal DNA by using a *Taq* polymerase kit (Platinum *Pfx* DNA Polymerase, Invitrogen) according to the manufacturer's instructions. Each PCR reaction (50 μl) consisted of 5 μl PCR buffer, 1.5 μl [1.5 mM millimolar (mM)

(w/v)] MgSO$_4$, 4 µl [1.25 mM (v/v)] dNTPs, 2 µl DNA template, 4 µl (12.5 pmol µl$^{-1}$) of each forward and reverse primer, 29.5 µl dH$_2$O and 0.2 µl *Taq* polymerase enzyme. PCRs were carried out in a GeneAmp PCR System 9700 Thermo Cycler (Applied Biosystems) using 30 cycles of the following amplification conditions: denaturation at 94°C for 45 s, annealing at 59°C for 45 s, and extension at 72°C for 2 min. An initial denaturation step of 94°C for 2 min was used and a final extension step at 72°C for 10 min (Table 2.6). However, in some cases (e.g. for *dnaE, gyrB, dtdS,* and *pyrC*) improved results were obtained by varying the annealing temperatures between 55 to 60°C. The expected size of the PCR products was confirmed by electrophoresis in a 1% (w/v) agarose gel incorporating 0.004% (v/v) SybrSafe (Invitrogen). DNA was purified with a Qiaquick PCR purification kit (Qiagen) and finally eluted in 30-50 µl sterile distilled H$_2$O and stored at -20°C.

**Table 2.4.** Nucleotide sequences of PCR and sequencing primers designed for DNA amplification and sequencing of seven housekeeping gene fragments used for MLST of *V. parahaemolyticus*.

| Gene | Primer | PCR/Sequencing | Base position | Sequence (5'-3') [a] |
|------|--------|----------------|---------------|---------------------|
| *dnaE* | dnaE-F3 | PCR | 639-656 | CGA GAT TCG TGT TGC GAT |
| | dnaE-R1 | PCR | 1414-1397 | CTA GCG TCA TAC CCG GAT |
| | dnaE-F2 | Sequencing | 735-752 | AAT GTG TGA GCT GTT TGC |
| | dnaE-R2 | Sequencing | 1329-1312 | ACG GAT TAC CGC TTT CGC |
| *gyrB* | gyrB-F1 | PCR | 582-599 | GTT CTT GAA CTC AGG CGT |
| | gyrB-R1 | PCR | 1339-1322 | GTG GTA GGA TTG CCT GAT |
| | gyrB-F2 | Sequencing | 655-671 | GAA GGT GGT ATT CAA GC |
| | gyrB-R2 | Sequencing | 1281-1264 | GTC ACC CTC CAC AAT GTA |
| *recA* | recA-F1 | PCR | 75-92 | CAT GCG CCT TGG TGA TAA |
| | recA-R3 | PCR | 1006-989 | CAG GTG CTT CTG GTT GAG |
| | recA-F2 | Sequencing | 111-128 | AAC CAT TTC AAC GGG TTC |
| | recA-R2 | Sequencing | 877-860 | TGT AGC TGT ACC AAG CAC |
| *dtdS* | dtdS-F1 | PCR | 47-64 | GGA TGA CCG AAG TAG ACA |
| | dtdS-R1 | PCR | 618-601 | AGC AAG CTC TAG ACG GTA |
| | dtdS-F2 | Sequencing | 75-92 | TGG CCA TAA CGA CAT TCT |
| | dtdS-R2 | Sequencing | 571-554 | GAG CAC CAA CGT GTT TAG |
| *pntA* | pntA-F1 | PCR | 600-617 | TGA CGT TCG TCC AGA AGT |
| | pntA-R3 | PCR | 1275-1258 | TAC CGA TGC AAT CCA AGC |
| | pntA-F2 | Sequencing | 671-688 | AAG ACT CTG GTT CTG GTG |
| | pntA-R2 | Sequencing | 1158-1141 | TTG AGG CTG AGC CGA TAC |
| *pyrC* | pyrC-F1 | PCR | 194-211 | GCG AAC AAT TCG AAC CTC |
| | pyrC-R1 | PCR | 874-857 | TTG CGA ACG CTT CCA AGT |
| | pyrC-F2 | Sequencing | 269-286 | CAA CCG GTA AAA TTG TCG |
| | pyrC-R2 | Sequencing | 798-781 | AGT GTA AGA ACC GGC ACA |
| *tnaA* | tnaA-F3 | PCR | 548-565 | TCT GTG CCA TCA TCA CGA |
| | tnaA-R3 | PCR | 1147-1130 | CCT CCA GAT ACA ACG CAT |
| | tnaA-F2 | Sequencing | 591-608 | CCA ACC GGT ATC GAT GGA |
| | tnaA-R2 | Sequencing | 1083-1066 | TAT TTT CGC CGC ATC AAC |

[a] Primers were obtained by Sigma-aldrich, United Kingdom

**Table 2.5.** Nucleotide sequences of additional PCR and sequencing primers designed for amplification and sequencing of seven housekeeping gene fragments of the genes which could not be amplified or sequenced by the primers presented in Table 2.4.

| Gene | Primer | PCR/Sequencing | Base position | Sequence (5'-3') [a] |
|------|--------|----------------|---------------|----------------------|
| *dnaE* | dnaE-F1 | PCR | 677-694 | AAG ATC CAC GCC GAC CAA |
| | dnaE-R3 | PCR | 1470-1453 | TTC CTC ATC GGC CTC ATA |
| *recA* | recA-F3 | PCR | 54-71 | GCA ATT CGG TAA AGG CTC |
| | recA-F4 | PCR | 32-49 | CTG CGC TAG GTC ARA TTG |
| | recA-R1 | PCR | 925-908 | GCA GGT AGT TAC AAG CGT |
| | recA-R4 | PCR | 1032-1015 | TTC TTG CTC AGG CTT CTC |
| *dtdS* | dtdS-F3 | PCR | 23-40 | AGC TAA AGC CTG AAS [b] AAG |
| | dtdS-R3 | PCR | 645-628 | TAC TGC RCG Y[b]GT TAC GCC |
| | dtdS-R5 | Sequencing | 598-580 | CGT TTA CGT CTG TGA TAA C |
| *pntA* | pntA-F3 | PCR | 578-595 | TTG GCG CTA TCG TTC GTG |
| | pntA-F4 | PCR | 625-642 | CAA GTT GAG TCG ATG GGT |
| | pntA-R1 | PCR | 1245-1228 | GGC TGC AAC AAG ACC AAT |
| | pntA-R4 | PCR | 1249-1232 | CAA CGG CTG CAA CAA GAC |
| *pyrC* | pyrC-F7 | Sequencing | 232-249 | GAT AAC ACC ACG CCA GAA |
| *tnaA* | tnaA-F1 | PCR | 567-584 | AGT GAC GTG TAA CAG CTC |
| | tnaA-F4 | Sequencing | 623-640 | TGT ACG AAA TTG CCA CCA |
| | tnaA-R1 | PCR | 1128-1111 | ACA CAA GGC TTG TGC TGG |

[a] Primers were obtained by Sigma-aldrich, United Kingdom

[b] IUB symbol for universal primers:  S = G+C, Y = C+T

**Figure 2.1.** Nucleotide sequence (5' → 3') of *dnaE* of *V. parahaemolyticus* isolate RIMD2210633 (GenBank ID:1189816) showing the positions of the PCR (green; dnaE-F3 and dnaE-R1), sequencing (red; dnaE-F2 and dnaE-R2) and alternative (yellow; dnaE-F1 and dnaE-R3) primers used in the present MLST study.

```
                    20              40              60              80              100
gyrB ATGTCTGAAAATTACGATTCATCGAGTATTAAGGTACTGAAGGGTCTGGATGCGGTACGTAAGCGTCCAGGTATGTACATCGGGGACACGGACGACGGCACCGGCCTTCACCACATG
          120             140             160             180             200             220
gyrB GTTTTCGAGGTGGTGGATAACTCAATTGATGAAGCGTTGGCAGGTCACTGTAAAGACATCGTTGTGACAATTCATGAGGACAACTCGGTTTCCGTTAGCGATGATGGTCGTGGCATT
          240             260             280             300             320             340
gyrB CCAACAGAAATGCACCCAGAAGAAAAAGTATCAGCAGCAGAAGTTATCATGACCGTACTGCACGCTGGTGGTAAGTTCGATGATAACTCGTACAAAGTATCAGGCGGTCTTCACGGC
          360             380             400             420             440             460
gyrB GTGGGTGTTTCGGTAGTAAACGCACTGTCAGAAAAAGTGGTACTAACCATCCATCGTGGCGGTCATATCCACACGCAAACTTACCGTCATGGTGAGCCTGAAGCGCCTCTAGCGGTT
          480             500             520             540             560         580 gyrB-F1
gyrB GTGGGTGATACGGATAAAACCGGTACACAAATTCGTTTCTGGCCAAGTGCAGAAACTTTCTCTAACACTGAATTCCATTACGACATCCTAGCAAAACGTCTGCGTGAGCTATCGTTC
     gyrB-F1                600             620             640        gyrB-F2 660        680             700
gyrB TTGAACTCAGGCGTTTCTATCAAGCTTATTGATGAGCGCGAAGCGGACAAGCAAGACCACTTCATGTATGAAGGTGGTATTCAAGCGTTCGTTCAGCACTTAAACACCAACAAAACA
          720             740             760             780             800
gyrB CCAATCATCGAGAAAATCTTCCACTTCGACTTAGAACGTGAAGACGGTATCTCGGTAGAAGTGGCAATGCAGTGGAACGATGGTTTCCAAGAGAACATCTTCTGTTTCACCAACAAC
          820             840             860             880             900             920
gyrB ATTCCACAGCGCGATGGTGGTACTCACCTTGCTGGTTTCCGTGCGGCACTAACACGTACGCTAAACAGCTTTATGGATAAAGAAGGCTTCTCGAAGAAAGCGAAAACAGCAACGTCA
          940             960             980             1.000           1.020           1.040
gyrB GGCGACGATGCGCGTGAAGGTTTGACAGCCGTTGTTTCGGTAAAAGTGCCTGATCCAAAATTCTCGAGCCAAACCAAAGACAAACTGGTTTCTTCTGAAGTGAAATCAGCGGTTGAA
          1.060           1.080           1.100           1.120           1.140           1.160
gyrB TCGGCGATGGGTGAGAAATTATCTGAGTTCTTGGTCGAAAACCCAAGCGAAGCGAAGATGGTTTGTTCGAAAATCATCGATGCAGCACGTGCACGTGAAGCCGCACGTAAAGCGCGT
          1.180           1.200           1.220           1.240           1.260    gyrB-R2     1.280
gyrB GAAATGACTCGTCGTAAAGGCGCGCTAGACCTAGCAGGCCTACCAGGCAAACTTGCAGACTGTCAGGAAAAAGATCCGGCACTCTCTGAACTATACATTGTGGAGGGTGACTCTGCG
          1.300           1.320  gyrB-R1    1.340           1.360           1.380           1.400
gyrB GGTGGTTCAGCTAAGCAGGGTCGTAATCGTAAGAATCAGGCAATCCTACCACTGAAAGGTAAGATCCTGAACGTAGAAAAAGCACGTTTCGACAAGATGTTGTCTTCGCAAGAAGTT
          1.420           1.440           1.460           1.480           1.500           1.520
gyrB GCAACGCTTATTACAGCACTTGGCTGTGGTATCGGTCGTGACGAGTACAACCCGGACAAACTGCGTTACCACAACATCATCATCATGACCGATGCTGACGTCGATGGTTCGCACATC
          1.540           1.560           1.580           1.600           1.620
gyrB CGTACGCTACTGTTGACTTTCTTCTACCGTCAAATGCCAGAGCTTATCGAGCGTGGCTACGTGTACATTGCTCAGCCACCACTATACAAAGTGAAGAAGGGCAAACAAGAGCAGTAC
          1.640           1.660           1.680           1.700           1.720           1.740
gyrB ATCAAAGATGAAGAAGCGATGAACCAATACCAAGTATCATTGGCATTGGACAACGCATCACTGCACGTAAACGCAGAAGCTCCAGCTCTAGCTGGCGAAGCACTAGAAAAACTGGTT
          1.760           1.780           1.800           1.820           1.840           1.860
gyrB CAACAATACAATGCAGGTATCAAGCTTGCTGATCGTATGAGCCGTCGTTACCCTCGCGCCTTAGTGCATGAGCTTATCTACACATCTCGTCTAACGGCTGAGCAGTGTCATGATGCA
          1.880           1.900           1.920           1.940           1.960           1.980
gyrB GCAGCAGTAGAAGCTTGGACAAAACAACTTGTTGAGCAACTAAACGCGAAAGAAGTGGGTGCAAGCCAATATAGCTACGAAGTGGAGCTGCACGCAGAGCTTGGTTTGAGCCTGCCT
          2.000           2.020           2.040           2.060           2.080           2.100
gyrB AAGATCATCGTTCGTACACACGGTGTGACTCACGAGCATGCGCTAAGTGTTGATTTCCTAAACTCAAAAGAATACGGCAAGCTGGCGGACCTTTCTGAAGTCCTAGACGGTCTTCTA
          2.120           2.140           2.160           2.180           2.200           2.220
gyrB GAAGAGGGCGCGTACATCAAACGTGGTGAACGTACTCTGCCAGTATCAAGCTTTGCTGAAGCACTTGAGTGGTTGGTTAAAGAATCGATGCGTGGTCTAAGCCGTCAGCGTTACAAA
          2.240           2.260           2.280           2.300           2.320           2.340
gyrB GGTCTAGGTGAGATGAACCCAGATCAGCTTTGGGAAACCACAATGGACCCAGAGACGCGTCGCATGATGCAAGTAACGATTGAAGATGCGGTAGGCGCAGACCAATTGTTTACAACG
          2.360           2.380           2.400
gyrB CTAATGGGTGACCAAGTTGAACCTCGTCGTCACTTCATCGAAGAGAACGCACTGAAGGTTGCAAACCTAGACGTGTAA
```

**Figure 2.2.** Nucleotide sequence (5' → 3') of *gyrB* of *V. parahaemolyticus* isolate RIMD2210633 (GenBank ID:1187470) showing the positions of the PCR (green; gyrB-F1 and gyrB-R1) and sequencing (red; gyrB-F2 and gyrB-R2) primers used in the present MLST study.

**Figure 2.3.** Nucleotide sequence (5' → 3') of *recA* of *V. parahaemolyticus* isolate RIMD2210633 (GenBank ID:1190074) showing the positions of the PCR (green; recA-F1 and recA-R3), sequencing (red; recA-F2 and recA-R2) and alternative (yellow; recA-F3, recA-F4, recA-R1 and recA-R4) primers used in the present MLST study.
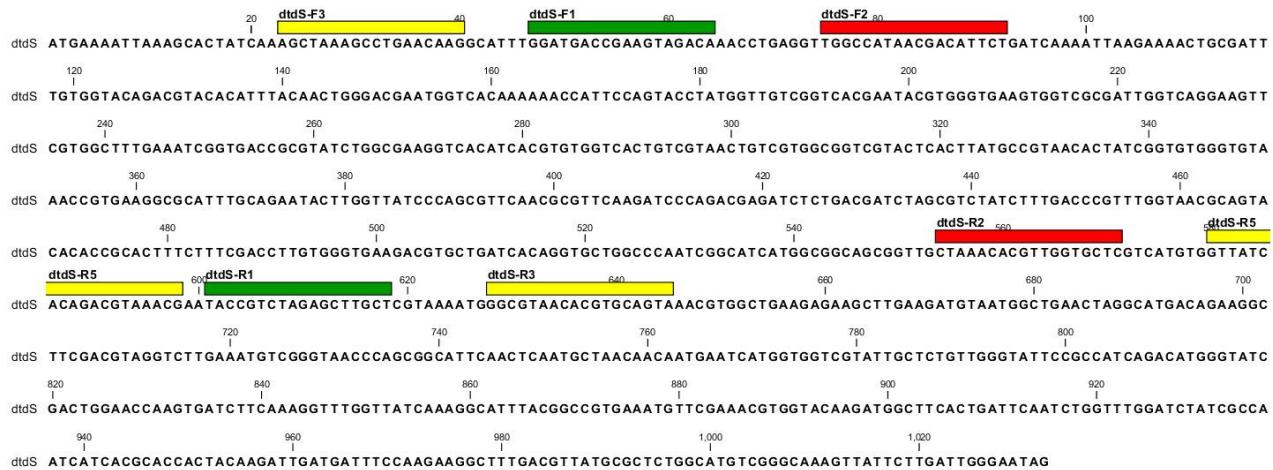


**Figure 2.4.** Nucleotide sequence (5' → 3') of *dtdS* of *V. parahaemolyticus* isolate RIMD2210633 [GenBank ID:BA000032.2 (region 1612798-1613829)] showing the positions of the PCR (green; dtdS-F1 and dtdS-R1), sequencing (red; dtdS-F2 and dtdS-R2) and alternative (yellow; dtdS-F3, dtdS-R3 and dtdS-R5) primers used in the present MLST study.
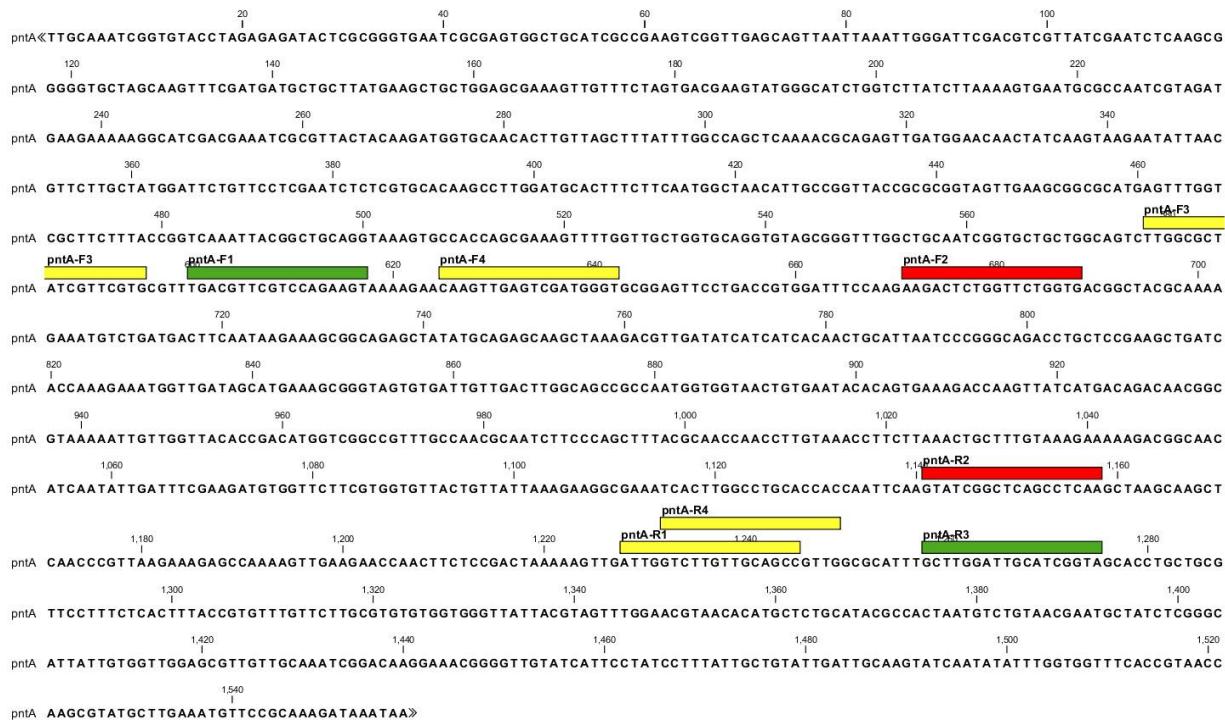
**Figure 2.5.** Nucleotide sequence (5' → 3') of *pntA* of *V. parahaemolyticus* isolate RIMD2210633 (GenBank ID:1191611) showing the positions of the PCR (green; pntA-F1 and pntA-R3), sequencing (red; pntA-F2 and pntA-R2) and alternative (yellow; pntA-F3, pntA-F4, pntA-R1 and pntA-R4) primers used in the present MLST study.

**Figure 2.6.** Nucleotide sequence (5' → 3') of *pyrC* of *V. parahaemolyticus* isolate RIMD2210633 (GenBank ID:1191095) showing the positions of the PCR (green; pyrC-F1 and pyrC-R1), sequencing (red; pyrC-F2 and pyrC-R2) and alternative (yellow; pyrC-F7) primers  used in the present MLST study.
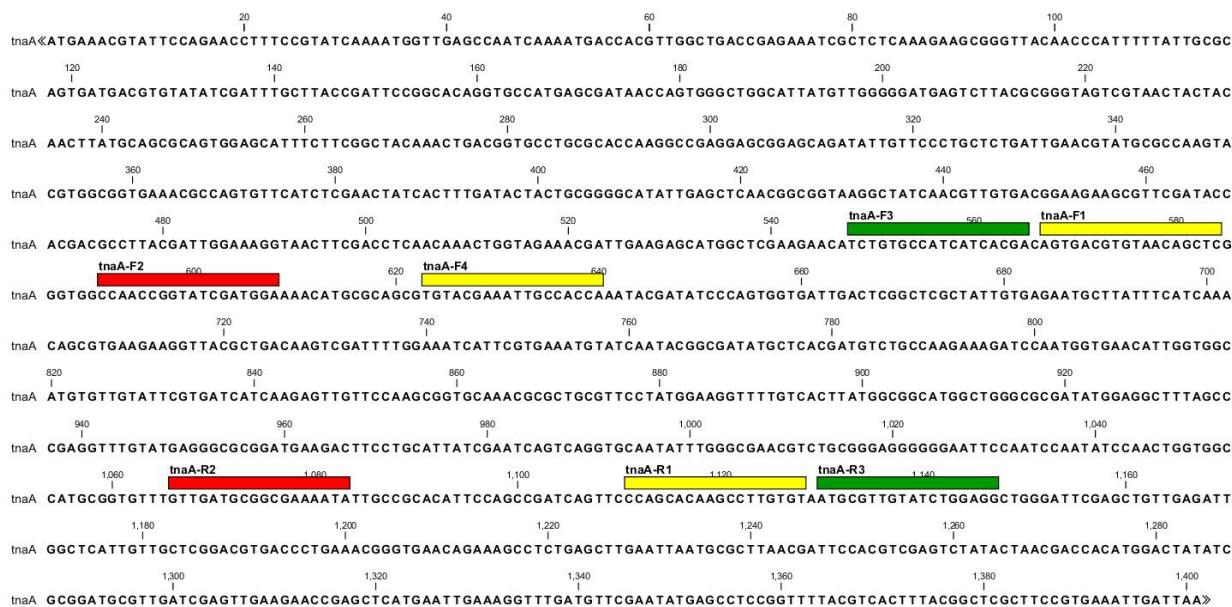
**Figure 2.7.** Nucleotide sequence (5' → 3') of *tnaA* of *V. parahaemolyticus* isolate RIMD2210633 (GenBank ID:1190879) showing the positions of the PCR (green; tnaA-F3 and tnaA–R3), sequencing (red; tnaA-F2 and tnaA-R2) and alternative (yellow; tnaA-F1, tnaA-F4 and tnaA-R1) primers used in the present MLST study.

**Table 2.6.** PCR conditions used for amplification of seven housekeeping gene fragments used for the MLST study of *V. parahaemolyticus*

| PCR process | Temperature |
|---|---|
| Initial denaturation | 94°C 2 min |
| Denaturation | 94°C 45 s |
| Annealing | 59°C 45 s |
| Extension | 72°C 2 min |
| Final extension | 72°C 10 min |

## *2.2.5 Nucleotide sequencing*

Sequencing reactions were performed in 10 µl reaction mixes using the BigDye terminator cycle sequencing kit version 3.1 according to the manufacturer's instructions. Each reaction consisted of 3.5 µl dilution buffer, 3 µl DNA template (approximately 50 ng µl$^{-1}$ DNA), 2 µl of 2 pmol µl$^{-1}$ of forward or reverse sequencing primer, 1 µl dH$_2$O and 0.5 µl of a 1:16 dilution of BigDye. The reactions were carried out in a GeneAmp PCR System 9700 Thermo Cycler (Applied Biosystems) using 25 cycles of the following amplification conditions:  denaturation at 96°C for 10 s, annealing at 50°C for 5 s, and extension at 60°C for 4 min. The DNA fragments were cleaned up by ethanol precipitation and sequenced using an ABI 3730 capillary sequencer (Genepool Sequencing Unit, University of Edinburgh; http://genepool.bio.ed.ac.uk).

## *2.2.6 Nucleotide and population structure analysis from MLST data*

Sequencing data were checked and edited using Lasergene version 5.0 (DNASTAR) sequence analysis software. Nucleotide sequence analyses were conducted with MEGA version 4.0.2. (Tamura *et al.*, 2007), in conjunction with alignment programs written by T.S. Whittam (Michigan State University). Global optimal eBURST (goeBURST) (Francisco *et al.*, 2009) analysis of the 63 STs in the data set was performed using Phyloviz software (http://www.phyloviz.net). Recombination events were detected in concatenated DNA sequences using the RDP version 3.0 software package (Martin *et al.*, 2005). The Bayesian Analysis of Population Structure (BAPS) software version 5.3 (Corander & Marttinen, 2006; Corander & Tang, 2007) was used to infer the population structure by clustering the STs into genetically distinct groups. The "clustering with linked loci" module was employed to approximate the number of genetically distinct groups, i.e., the genetic mixture analysis (Corander & Tang, 2007). Following the recommendations in the BAPS manual

(http://web.abo.fi/fak/mnf/mate/jc/software/BAPS5manual.pdf), several K values (where K is the estimated maximum number of genetically distinct groups) were used to assess how this might affect the results; a range of K values (from 2 to 20) were used and in all cases the results were identical. To test for admixture among the genetic groups identified by BAPS, an admixture analysis was performed (Corander & Marttinen, 2006) with a minimum population size of 5 and the following specifications: the number of iterations used to estimate the admixture coefficient for the individuals was set to 100; the number of reference individuals from each population was 200; the number of iterations used to estimate the admixture coefficient for reference individuals was 20. Only STs having *p*-values less than 0.05 were considered as having "significant" evidence of admixture. BAPS clusters in respect to phylogenetic inference were illustrated by mapping BAPS clusters on to a Neighbour-Joining tree of concatenated sequences.

## 2.2.7 Amino acid sequence type (aaST) designation

The amino acid sequence types (aaSTs) of 348 nucleotide sequence types (STs) from the MLST database (http://pubmlst.org/vparahaemolyticus), including the 63 STs of the Thai isolates from the present study, were assigned according to a program written by Dr. David Aanensen, Department of Infectious Disease Epidemiology, Imperial College London, St. Mary's Hospital Campus, W2 1PG, London.

## 2.2.8 Serotyping of V. parahaemolyticus

Serological identification of *V. parahaemolyticus* was determined by a combination of O and K serotyping. For O antigen serotyping, a sample was prepared by subculturing the bacteria onto TSA + 3% NaCl and incubating at 37$^o$C overnight. Three colonies from the overnight plate were resuspended in 1 ml normal saline [O.85% (w/v)] in a glass test tube. O

The bacterial suspension was autoclaved at 121$^o$C for 60 min. The suspension was transferred to an Eppendorf tube and centrifuged at 12000 × *g* for 20 min. The supernatant was discarded and the cell pellet was resuspended in 1 ml normal saline [O.85% (w/v)] to give a dense suspension for use in the agglutination test. To prepare for the agglutination test, glass slides were divided into 12 equal sections using a wax pencil. One drop of each of 11 O antisera (DENKA SEIKEN, Tokyo, Japan) was added to each compartment, and 10 μl of bacterial suspension was mixed with each antiserum. To test for auto-agglutination, one drop of normal saline instead of O antisera was mixed with the bacterial suspension on the 12$^{th}$ section of the glass slide. The antisera and bacterial cultures were gently mixed on the glass slides until they were homogeneous. A positive agglutination was identified by the formation of fine granules or large aggregates.

For the K antigen agglutination test, the samples were prepared separately from those used for O antigen agglutination. A sample was prepared by subculturing the bacteria onto TSA + 3% NaCl and incubating at 37$^o$C overnight. The colonies from an overnight plate was used for agglutination test. The K antiserum test kit consists of nine polyvalent K antisera and each of them contains different monovalent antisera (DENKA SEIKEN, Tokyo, Japan) (Table A1, Appendix1). Nine polyvalent K antisera were tested with bacterial colonies, which was collected by a loop from an overnight plate, by the slide agglutination test as described above. Monovalent K antisera corresponding to the positive polyvalent K antisera were subsequently tested.

The serotype of *V. parahaemolyticus* was recorded by a combination of both O antisera and K monovalent antisera as shown in Table 1.4.

# 2.3 Results

## *2.3.1 Taq polymerase kit evaluation*

The *ompA* fragments from *M. haemolytica* were amplified by all 13 DNA polymerase kits from the various manufacturers including Invitrogen, NEBs, Novagen, Promega, Roche and Thermo Scientific (Figs. 2.8 and 2.9). It was found that the quality of gene fragment amplification by these enzyme kits was variable.
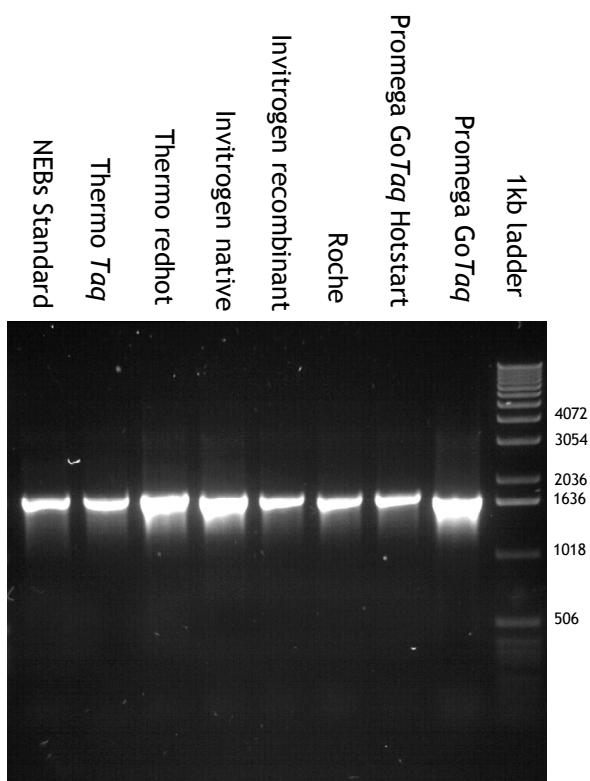


**Figure 2.8.** Agarose gel electrophoresis of *ompA* fragments from *M. haemolytica* by DNA polymerase from various manufacturers
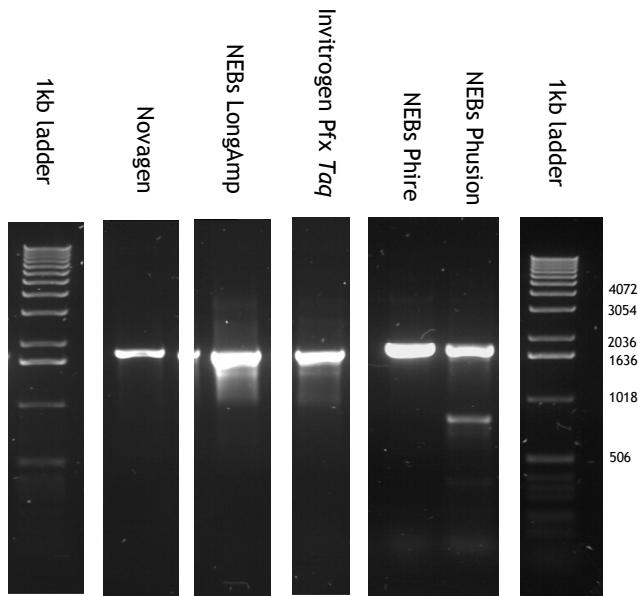
**Figure 2.9.** Agarose gel electrophoresis of *ompA* fragments from *M. haemolytica* by DNA polymerase from various manufacturers

The kits comprise three categories; standard DNA polymerase kits (Invitrogen native, Invitrogen recombinant, NEBs standard, NEBs LongAmp, Novagen, Promega Go*Taq,* Roche and Thermo standard), Hot start DNA polymerase kits (NEBs Phire, Promega Go*Taq* HotStart and Thermo redhot) and high fidelity or proof reading enzyme kits (NEBs Phusion and Invitrogen *Pfx*). The estimated cost for 700 PCR reactions (PCR reactions of seven housekeeping genes in 100 bacterial isolates) and the properties of the kits are compared in Table 2.7. The tested kits were ranked in the Table 2.7 according to estimated cost per 700 reactions. The NEBs standard kit yielded the lowest cost (£85.31) whereas the Invitrogen *Pfx* was the most expensive (£473.00) among all 13 tested kits. The cost of these kits reflected the qualification of the enzyme that the standard *Taq* enzyme kits (NEBs standard, Invitrogen recombinant, Invitrogen native, Roche, Promega Go*Taq*) except Thermo standard  were relatively low cost, whereas the hot start enzyme kits (Promega Go*Taq* Hotstart, NEBs Phire and Thermo redhot) were more expensive. However, the standard *Taq* enzyme kits with specific qualifications such as providing

ready mixed reagent (Novagen) and capability of amplifying long nucleotide sequence (NEBs LongAmp) represented relatively high cost. The proofreading enzyme kit NEBs Phusion was more costly than the other kits except the NEBs LongAmp and Invitrogen *pfx*.

Representative *Taq* kits of different enzyme activities, proofreading ability and starting conditions were selected and the quality of amplification for each kit was compared by gel electrophoresis (Fig. 2.10). All *Taq* kits efficiently amplified *ompA* gene fragments. Since the PCR sequences were to be used for further sequencing analysis, DNA polymerase containing proofreading activity was selected, since it is known to reduce error rates of base substitutions compared with non-proofreading DNA polymerase (Eckert & Kunkel, 1991). Comparatively, between the two high fidelity DNA polymerases, Invitrogen *Pfx* and NEBs phusion, the first produced more specific amplification (Fig. 2.10), even using half the recommended enzyme concentration (0.5 Unit) (data not shown). Thus, the Invitrogen *Pfx* DNA polymerase was selected for the further PCR reactions in this study although it was the most expensive (£473.00/700 reactions) compared with other DNA polymerase kits tested in this study.

**Table 2.7.** DNA polymerase kits cost evaluation

| Cost rank | *Taq* kit | Price per 700 reactions (£) | Advantage | Disadvantage and comments | Enzyme activity |
|---|---|---|---|---|---|
| 1 | NEBs standard | 85.31 | -The most economical price<br>-Good specific amplification | -The band is sometime faint. | Standard *Taq* enzyme |
| 2 | Invitrogen recombinant | 154.00 | Good amplification | -More expensive than standard BioLabs *Taq*<br>-The PCR buffer does not include magnesium solution. | Standard *Taq* enzyme |
| 3 | Invitrogen native | 168.00 | Good amplification | - Expensive<br>-The PCR buffer does not include magnesium solution. | Standard *Taq* enzyme |
| 4 | Roche | 169.75 | -Good amplification<br>-Reference *Taq* kit used for *M. haemolytica* (*ompA* gene) in previous studies. | -The kit does not contain a loading dye. | Standard *Taq* enzyme |
| 5 | Promega Go*Taq* | 171.50 | -Good amplification<br>-The product contains loading dye (Green buffer) | -Produce a big smear band.<br>-The band is sharper by reduce loading volume to 2.5ul | Standard *Taq* enzyme |
| 6 | Promega Go*Taq* Hotstart | 227.50 | - Good amplification<br>-The product contains loading dye | -The amplification ability was performed as same as GoTaq DNA Pol but more expensive | Hot start *Taq* enzyme |
| 7 | NEBs Phire | 229.60 | -Good amplification<br>-Non *Taq* base polymerase with higher power than hot start *Taq* DNA polymerase activity<br>-Fast, high throughput PCR | -Expensive<br>-Require different PCR temperature conditions from the other (denaturing temperature at 98ºC) | Hot start *Taq* enzyme |
| 8 | Thermo standard | 255.50 | -Good amplification | -The kit does not include a loading dye.<br>-The PCR buffer does not include Magnesium. | Standard *Taq* enzyme |
| 9 | Thermo redhot | 255.50 | -Good amplification<br>-Product includes a loading dye | Provided red dye is very light, the colour disappear after gel running. | Hot start *Taq* enzyme |
| 10 | Novagen | 308.00 | -Most specific and the best amplification among the analyzed Taq kits.<br>-Mastermix provided | -Expensive and unnecessary to use a mastermix | Standard *Taq* enzyme |
| 11 | NEBs Phusion | 350.00 | -Good amplification<br>-Proofreading Enzyme, high fidelity<br>-Reasonable price for proof reading enzyme | -Expensive as it is a proofreading enzyme, special for cloning work.<br>- Require different PCR temperature condition from the other (denaturing temperature at 98ºC) | Proofreading *Taq* enzyme |
| 12 | NEBs LongAmp | 455.00 | -Long nucleotide amplification | -Expensive, big smear band<br>-Specially for long nucleotide amplification | Standard *Taq* enzyme |
| 13 | Invitrogen *Pfx Taq* | 473.00 | -Good amplification<br>Proofreading Enzyme, high fidelity | -Expensive but efficient | Proofreading *Taq* enzyme |
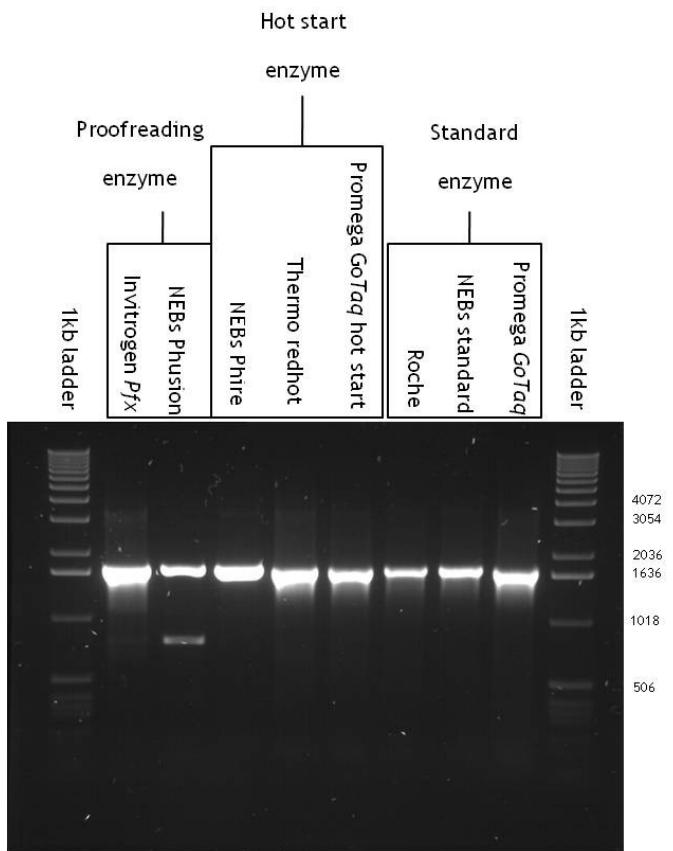
**Figure 2.10.** Agarose gel electrophoresis of *ompA* fragments from *M. haemolytica* by selected DNA polymerase in three categories; proofreading enzyme, hot start enzyme, and standard enzyme

## 2.3.2 Optimization of PCR for seven housekeeping enzyme genes of V. parahaemolyticus

### 2.3.2.1 Effect of different annealing temperatures

Amplification of the *dnaE, dtdS, gyrB, pntA, recA,* and *tnaA* gene fragments of the *V. parahaemolyticus* type strain (VP2) was optimized by using variable annealing temperatures: $56^{\circ}$C, $57^{\circ}$C, $58^{\circ}$C, $59^{\circ}$C, and $60^{\circ}$C. The PCR primers used for these seven gene fragments were dnaE-F1 and dnaE-R1, gyrB-F1 and gyrB-R1, recA-F1 and recA-R1, dtdS-F1 and dtdS-R1, pntA-F1 and pntA-R1, tnaA-F1 and tnaA-R1, respectively. Using an annealing temperature of $56^{\circ}$C, agarose gel electrophoresis of the PCR products of the seven gene fragments showed weak non-specific amplifications, which were observed by additional bands surrounding the primary bands of expected size for *recA* (851bp), *gyrB* (758bp), *pntA* (646bp) and *pyrC* (681bp) fragments (Fig. 2.11). Unwanted DNA bands with higher molecular weight appeared in the amplifications of the *dnaE* (738bp) and *tnaA* (562bp) fragments. The *dtdS* fragment (572bp) was the only gene that was clearly amplified at $56^{\circ}$C. Using $57^{\circ}$C, these non-specific bands still occurred (Fig. 2.12), while at $58^{\circ}$C, amplification of the *recA* and *gyrB* fragments was enhanced although some non-specific bands still remained for *dnaE, pntA*, and *tnaA* (Fig. 2.13). Using $59^{\circ}$C, four out of the seven gene fragments, including *recA, gyrB, dtdS* and *pyrC* were clearly enhanced (Fig. 2.14), although amplifications of *dnaE, pntA,* and *tnaA* required further optimizations. An annealing temperature of $60^{\circ}$C was used, but still did not eliminate the non-specific amplifications of *dnaE, pntA* and *tnaA* (Fig. 2.15).
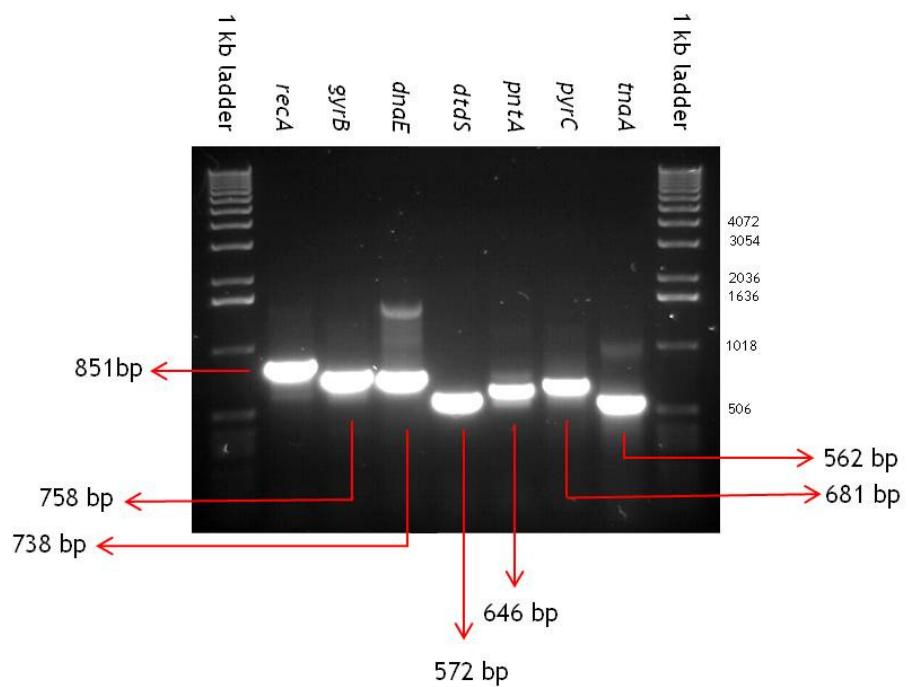
**Figure 2.11.** Agarose gel electrophoresis of *V. parahaemolyticus* (VP2) PCR products corresponding to the PCR primer pairs of seven housekeeping genes at an annealing temperature of 56°C
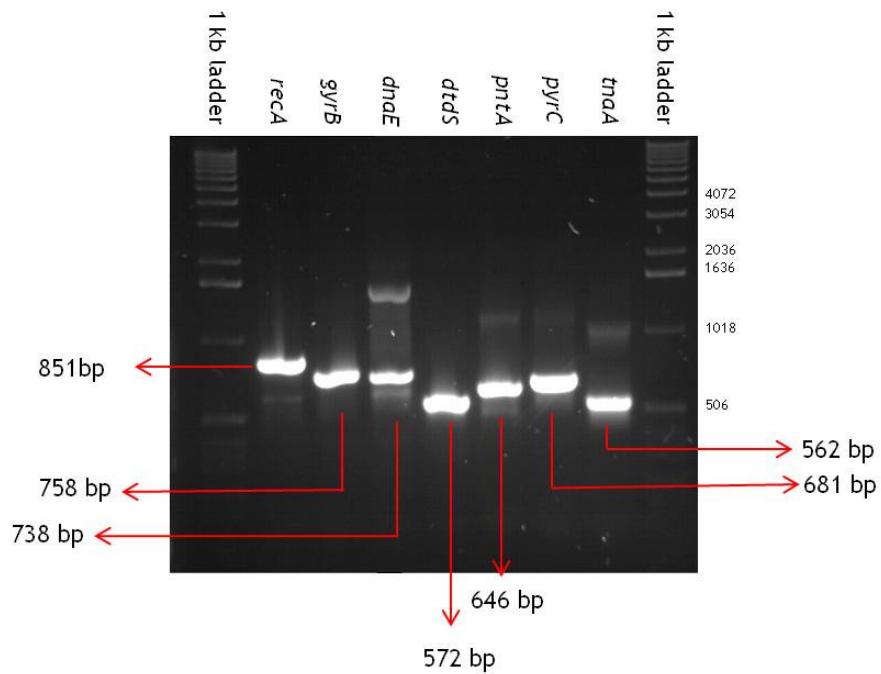


**Figure 2.12.** Agarose gel electrophoresis of *V. parahaemolyticus* (VP2) PCR products corresponding to the PCR primer pairs of seven housekeeping genes at an annealing temperature of 57°C
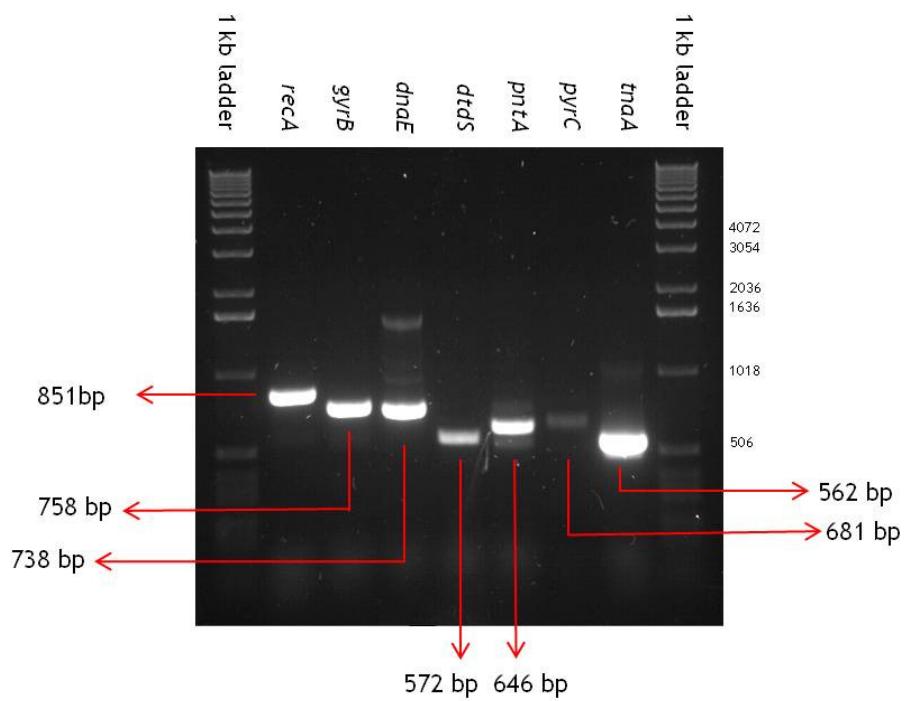
**Figure 2.13.** Agarose gel electrophoresis of *V. parahaemolyticus* (VP2) PCR products corresponding to the PCR primer pairs of seven housekeeping genes at an annealing temperature of 58°C
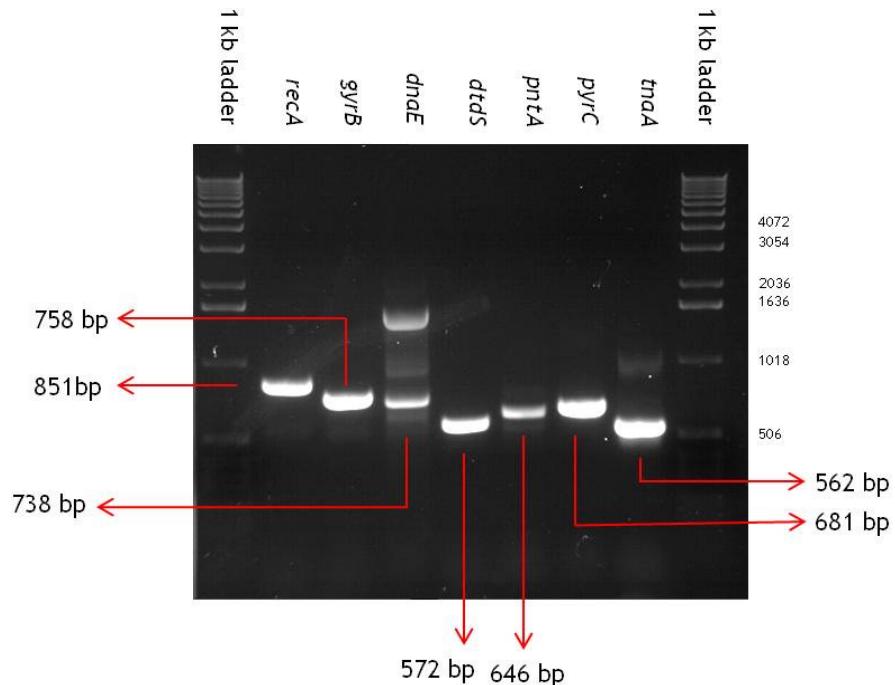


**Figure 2.14.** Agarose gel electrophoresis of *V. parahaemolyticus* (VP2) PCR products corresponding to the PCR primer pairs of seven housekeeping genes at an annealing temperature of 59°C
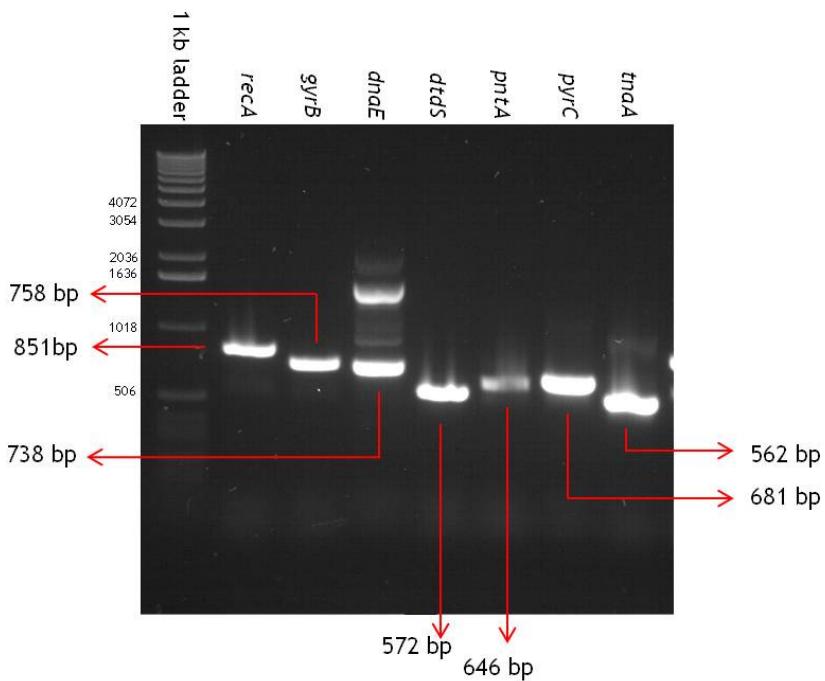
**Figure 2.15.** Agarose gel electrophoresis of *V. parahaemolyticus* (VP2) PCR products corresponding to the PCR primer pairs of seven housekeeping genes at an annealing temperature of 60°C

Among the five different annealing temperatures used for PCR optimization, a temperature of 59°C was selected to be optimal for *gyrB, dtdS, pyrC* amplifications. Although the *recA* fragment of VP2 was successfully amplified at an annealing temperature of 59°C (Fig. 2.14), non-specific bands occurred when this PCR condition was applied to isolates VP166 and VP216 (data not shown). Thus, PCR reactions of *recA, dnaE, pntA*, and *tnaA* fragments were further optimized by applying various magnesium concentrations and different primer pairs.

A range of different magnesium concentrations were applied for PCR optimization of *pntA* and *tnaA* fragments. Since amplification of the *recA* fragment of isolates VP166 and VP216 (data not shown) and the *dnaE* fragment of isolate VP2 (Figs. 2.12-2.15) showed clear non-specific bands, the PCR amplification of these genes were improved by using newly designed primers.

## 2.3.2.2 Effect of different magnesium concentrations

Various magnesium concentrations were used in PCR reactions of *pntA* and *tnaA* fragments. The forward and reverse primers used for *pntA* and *tnaA* amplifications were pntA-F1 and pntA-R1 and tnaA-F1 and tnaA-R1 respectively. The various magnesium concentrations were controlled by adjusting magnesium sulphate (MgSO$_4$), the reagent which is supplied separately with the *Pfx* Invitrogen DNA polymerase kit. The optimal magnesium concentration for *recA, gyrB, dtdS* and *pyrC* was found to be 1.50 mM. Agarose gel electrophoresis showed that *pntA* amplifications were not clearly enhanced by adjusting magnesium concentrations in the range 1.00 mM to 2.50 mM (Fig. 2.16). In contrast, using 1.00 mM magnesium concentration significantly reduced non-specific amplification of the *tnaA* fragment compared to other concentrations in the range, including 1.50 mM (Fig. 2.17).
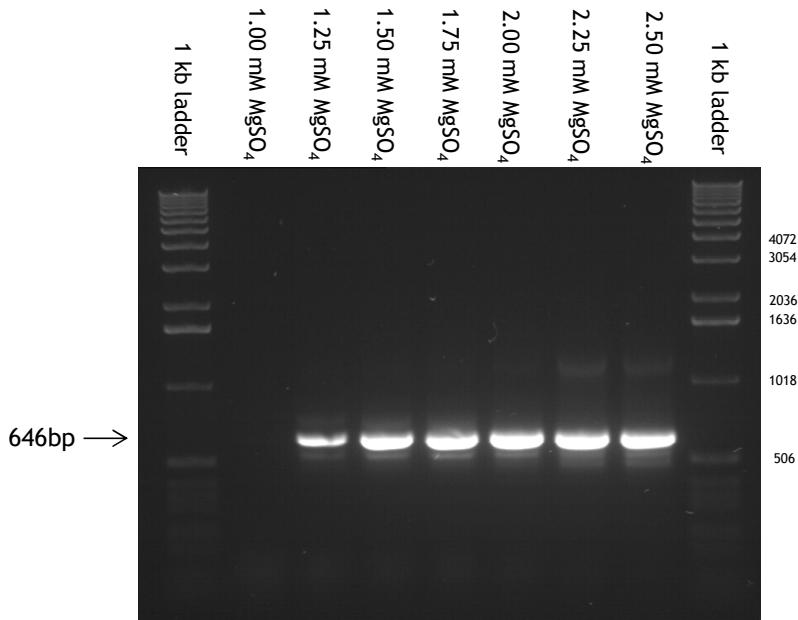


**Figure 2.16.** Agarose gel electrophoresis of *pntA* PCR products of *V. parahaemolyticus* (VP2) using various magnesium concentrations. The PCR reactions were performed using an annealing temperature of 59°C.
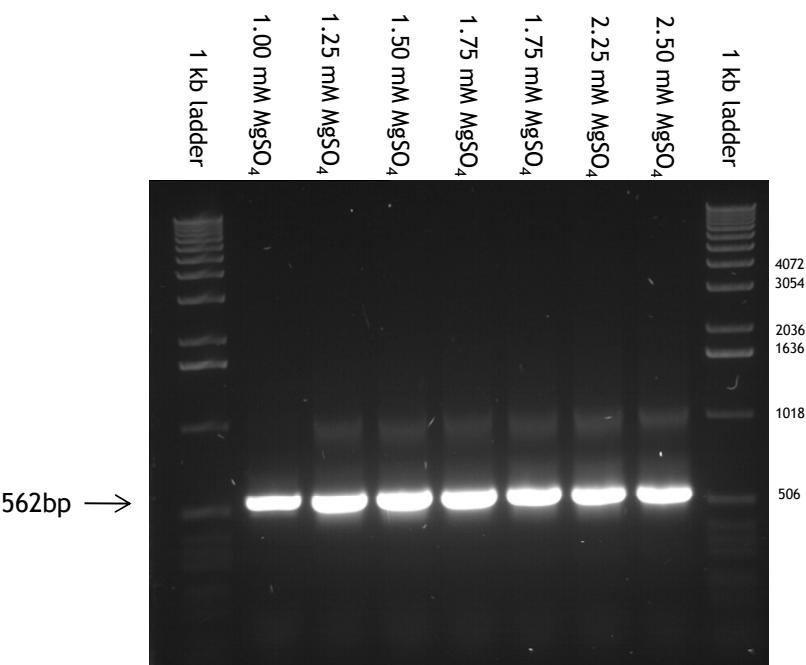
**Figure 2.17.** Agarose gel electrophoresis of *tnaA* PCR products of *V. parahaemolyticus* (VP2) using various magnesium concentrations. The PCR reactions were performed using an annealing temperature of 59°C.

## 2.3.2.3 Effect of different primer pair combinations

Additional PCR primers (Table 2.5) were designed for *dnaE, recA, pntA* and *tnaA*. Agarose gel electrophoresis showed that using the primer pair combinations that were used in previous experiment (Fig. 2.14) and the newly designed primers (Figs. 2.18 and 2.19) at an annealing temperature of 59°C with 1.5 mM magnesium concentration clearly enhanced the quality of the PCR products by showing greater specificity of the targeted DNA fragments. Gel electrophoresis images of PCR products showed that, among tested primer pairs, the best combinations for *dnaE, recA, pntA*, and *tnaA* were dnaE-F3 and dnaE-R1, recA-F1 and recA-R3, pntA-F1 and pntA-R3, and tnaA-F3 and tnaA-R3, respectively (Figs. 2.18 and 2.19) .
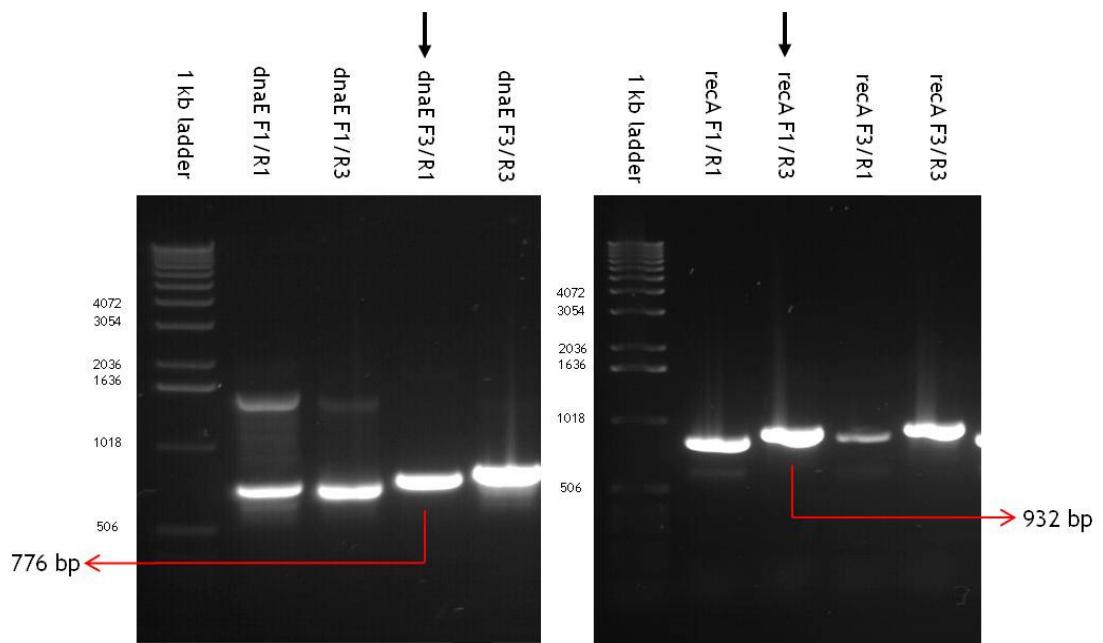
**Figure 2.18.** Agarose gel electrophoresis of *dnaE* and *recA* PCR products of *V. parahaemolyticus* (VP2) corresponding to four primer pair combinations using an annealing temperature of 59°C. Arrows on top indicate the most suitable primer pairs used for gene fragment amplification.
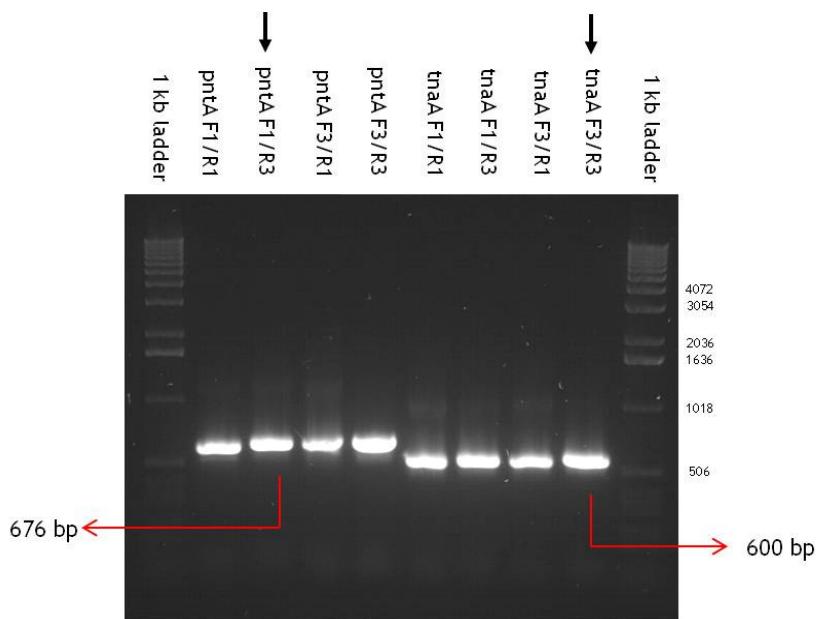


**Figure 2.19.** Agarose gel electrophoresis of *pntA* and *tnaA* PCR products of *V. parahaemolyticus* (VP2) corresponding to four primer pair combinations using an annealing temperature of 59°C. Arrows on top indicate the most suitable primer pairs used for gene fragment amplification.

## 2.3.2.4 Application of optimized PCR condition for seven housekeeping enzyme genes of the other *V. parahaemolyticus* isolates

The seven gene fragments of *dnaE, dtdS, gyrB, pntA, pyrC, recA,* and *tnaA* were amplified under the optimized conditions that had been established: an annealing temperature of 59$^o$C and a magnesium concentration of 1.5 mM. The selected primers used for each gene are shown in Table 2.4. These optimized conditions produced specific amplifications for all seven gene fragments in the *V. parahaemolyticus* type strain (VP2) (Fig. 2.20), as well as the other two *V. parahaemolyticus* strains isolated from a diarrhoeal patient (VP166) and seafood (VP216) (Figs. 2.21 and 2.22). These results therefore indicate that the optimized PCR conditions can potentially be applied to the other *V. parahaemolyticus* isolates in the strain collection.
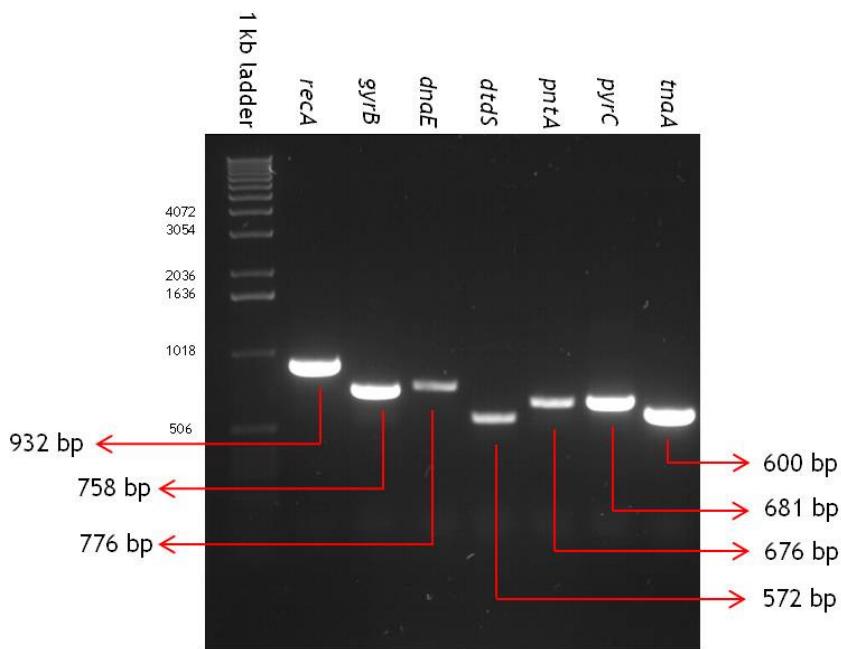


**Figure 2.20.** Agarose gel electrophoresis of *recA, gyrB, dnaE, dtdS, pntA, pyrC* and *tnaA* of the *V. parahemolyticus* type strain (VP2) under the optimized PCR conditions: annealing temperature 59$^o$C and magnesium concentration 1.5 mM.
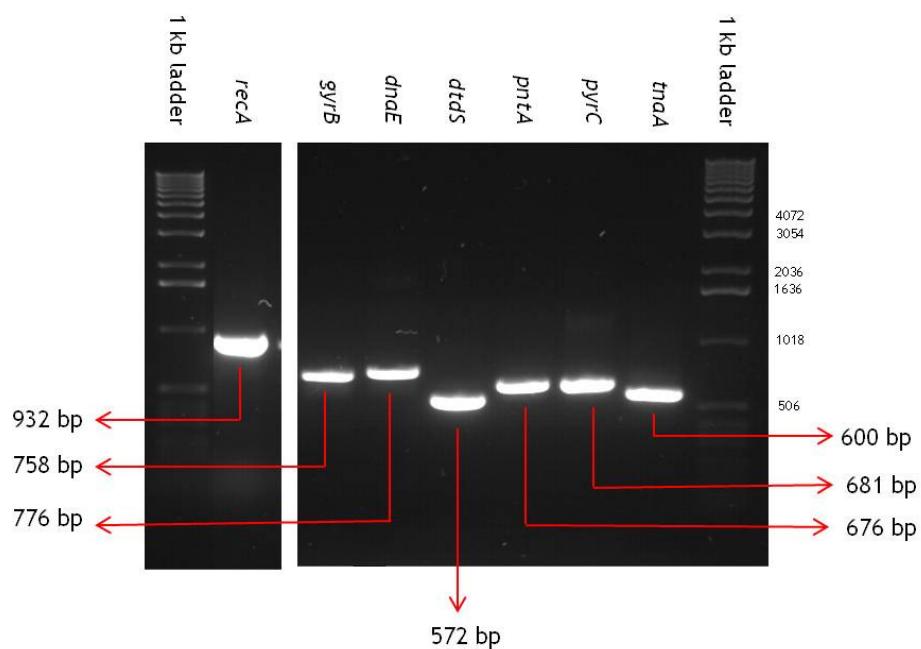
**Figure 2.21.** Agarose gel electrophoresis of *recA, gyrB, dnaE, dtdS, pntA, pyrC* and *tnaA* of *V. parahaemolyticus* clinical isolate (VP166) under the optimized PCR conditions: annealing temperature 59°C and magnesium concentration 1.5 mM.
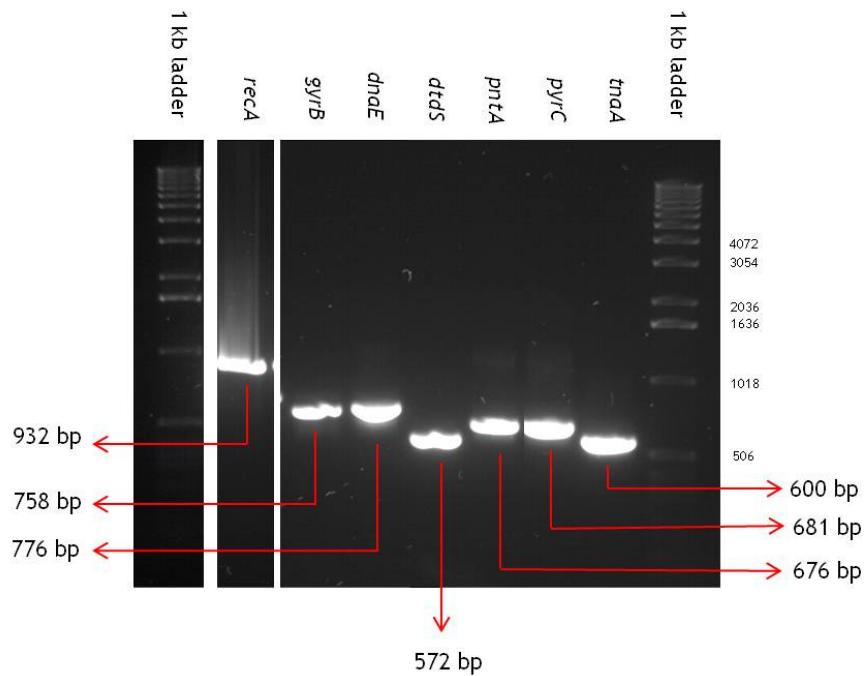


**Figure 2.22.** Agarose gel electrophoresis of *recA, gyrB, dnaE, dtdS, pntA, pyrC* and *tnaA* of *V. parahaemolyticus* seafood isolate (VP216) under the optimized PCR conditions: annealing temperature 59°C and magnesium concentration 1.5 mM.

## *2.3.3 Multiplex PCR*

A multiplex PCR system is a single PCR reaction that uses more than one pair of primers to enable the generation of many amplified gene fragments in one reaction. Application of multiplex PCR is beneficial because it is less time-consuming than a single PCR reaction. From section 2.3.2, seven gene fragments of *dnaE, dtdS, gyrB, pntA, pyrC, recA* and *tnaA* were successfully amplified by specific primers for each individual fragment (Table 2.4), using an annealing temperature of 59°C and magnesium concentration of 1.5 mM.

To develop the multiplex PCR system, the same annealing conditions were used with different primer pair combinations. *V. parahaemolyticus* type strain VP2 was used as a reference strain, and the selection of appropriate primer pairs in each combination was based on their relative individual gene fragment sizes (Fig. 2.23).
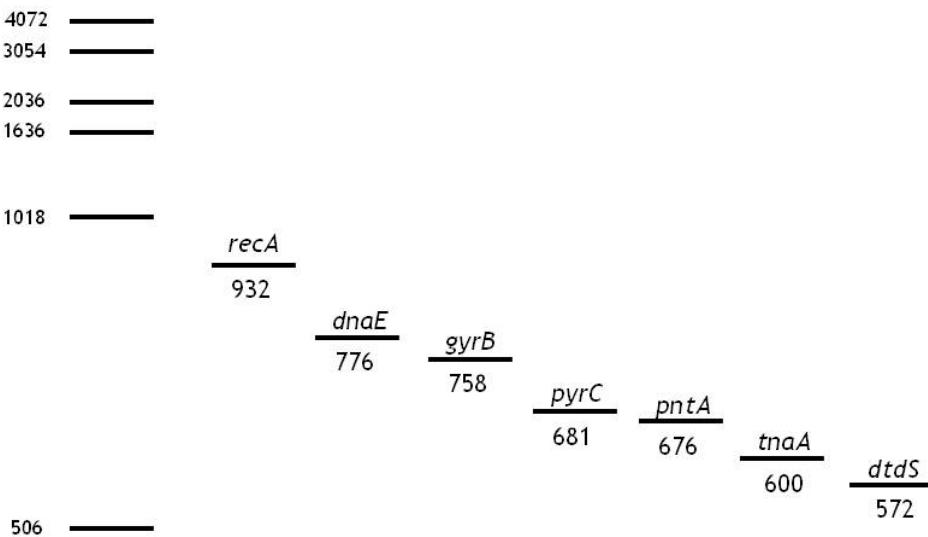


**Figure 2.23.** Fragment sizes (bp) of seven housekeeping genes compared with the molecular weight marker on the left.

## 2.3.3.1 Multiplex PCR reaction with three and four primer pairs

Three and four primers pairs were incorporated into a single PCR reaction and the results are shown in Fig. 2.24. Gel electrophoresis shows that the *gyrB, pntA* and *tnaA* fragments were successfully amplified although their DNA band intensities were not consistent (Fig. 2.24, lane 1). A multiplex reaction using three primer pairs for amplification of *recA, gyrB* and *tnaA* showed that this combination could amplify all three gene fragments (Fig. 2. 24, lane 2). However, the *gyrB* fragment appeared to have lower DNA concentration than the other two, *tnaA* and *recA*. Four primer pairs were used to amplify *recA, dnaE, pyrC* and *dtds* fragments (Fig. 2.24, lane3). The amplification of *recA* and *pyrC* fragments was achieved, but neither *dnaE* nor *dtdS* was successfully amplified. Lastly, a combination of three primers pairs was used to amplify *dnaE, pntA* and *dtdS* fragments (Fig. 2.24, lane 4). As a result, DNA fragments of these three genes were successfully amplified and had similar band intensities although these bands were relatively weak.

## 2.3.3.2 Multiplex PCR reaction by two primer pairs

PCR products from individual gene amplifications of seven housekeeping genes (Fig. 2.25, lanes 1-7) from *V. parahaemolyticus* type strain (VP2) were run comparatively on a gel with multiplex gene amplification using two primer pair combinations (Fig. 2.25, lanes 8-10). Six DNA fragments from multiplex PCR products, *recA and pyrC* (Fig. 2.25, lane 8), *gyrB* and *tnaA* (Fig. 2.25, lane 9), and *dnaE* and *dtdS* (Fig. 2.25, lane 10), were successfully amplified using the optimized PCR conditions.
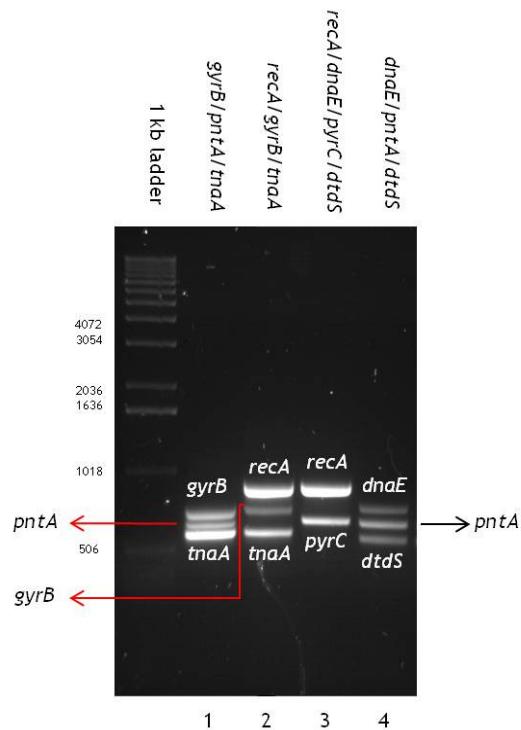
**Figure 2.24.** Gel electrophoresis of multiplex PCR products of seven housekeeping genes for the *V. parahaemolyiticus* type strain (VP2). The PCR annealing temperature was 59°C.
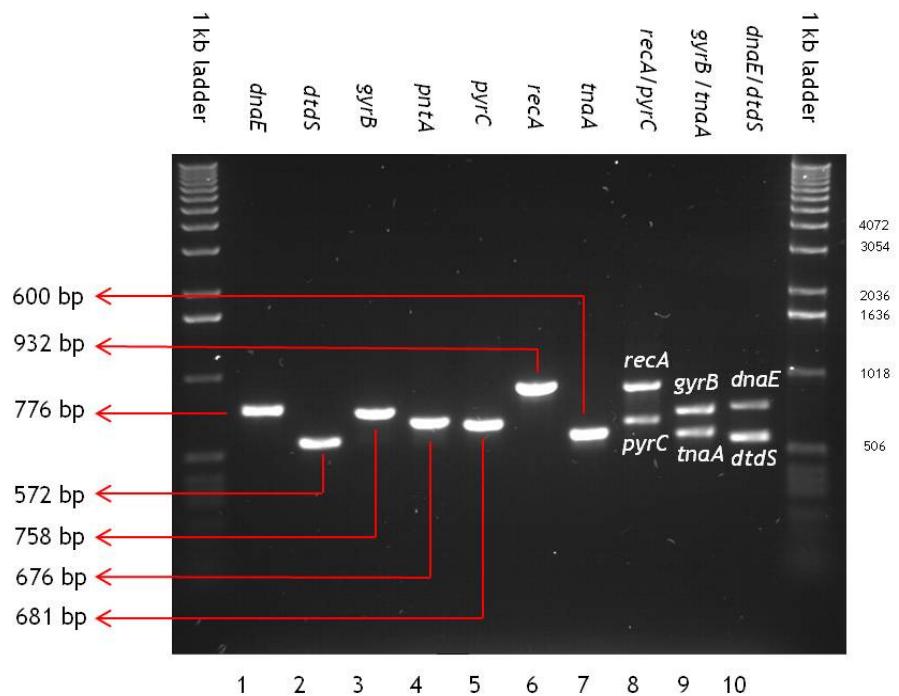


**Figure 2.25.** Comparison of individual and multiplex DNA fragments of seven housekeeping genes for the *V. parahaemolyiticus* type strain (VP2). The PCR annealing temperature was 59°C.

Primer pair combinations of *recA* and *pyrC* and *gyrB* and *tnaA* were applied to representative strains from a clinical sample (isolate VP166) (Fig. 2.26, lane 1-2) and from seafood (isolate VP216) (Fig. 2.26, lane 4-5). Gel electrophoresis showed that these two primer combinations can successfully amplify all four gene fragments, *recA, gyrB, pyrC* and *tnaA* in both VP166 and VP216. A primer combination of *dnaE, pntA* and *dtdS* was also applied to VP166 and VP216 (Fig. 2.26, lane 3 and 6). Unlike the successful amplification in VP2 (Fig. 2.24, lane 4), a primer combination of *dnaE, pntA* and *dtdS* was unable to yield consistent quality of PCR products for VP166 and VP216. However, a primer combination of *dnaE* and *dtdS* was able to amplify the *dnaE* and *dtdS* gene fragments in VP166 and VP216 (Fig. 2.26, lane 7 and 8). These results show that two primer pairs could produce more consistent DNA bands for VP166 (Fig. 2.26, lanes 1-2 and 7) and VP216 (Fig. 2.26, lane 4-5 and 8) than those from the reactions using three primer pairs (Fig. 2.26, lanes 3 and 6). Thus, amplifications of the *pntA* fragment for *V. parahaemolyticus* isolates in this study were carried out separately by a single PCR. Purified PCR products from the duplex system were chosen for sequencing and the sequencing results were compared with the PCR product from a single gene amplification.
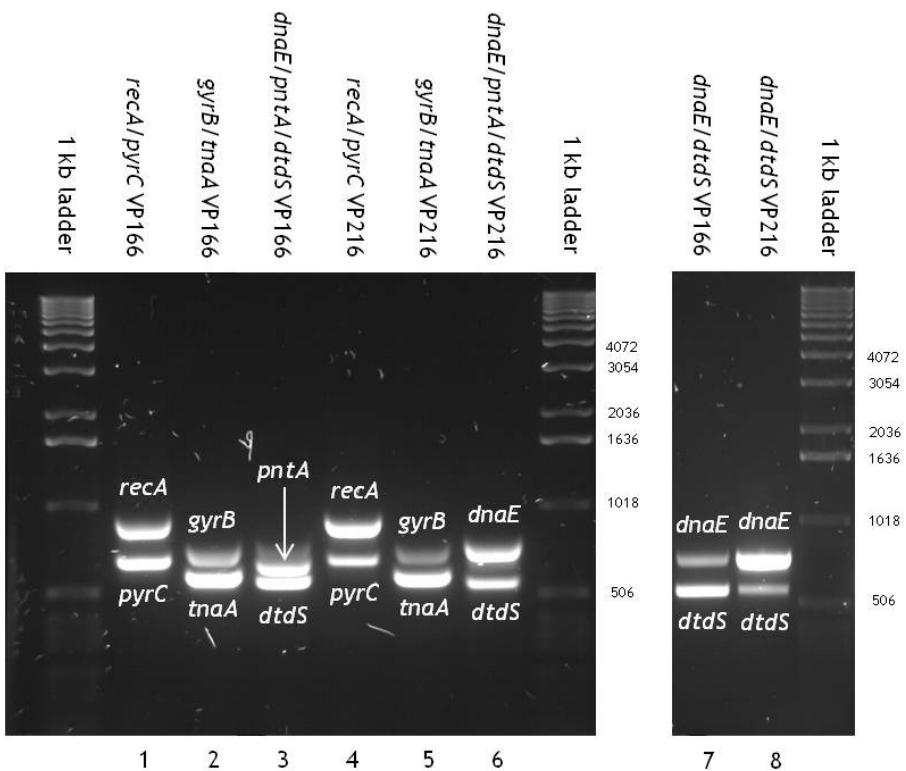
**Figure 2.26.** Gel electrophoresis of multiplex DNA fragments of seven housekeeping genes for the *V. parahaemolyiticus* VP166 and VP216. The PCR annealing temperature was 59°C.

## 2.3.4 Sequencing of multiplex PCR products

Sequencing primers were modified from the primers used in the previous *V. parahaemolyticus* MLST study (González-Escalona *et al.*, 2008) and designed to be located within the PCR primer locations (Figs. 2.1-2.7). Although the sequencing results obtained with multiplex PCR products of *V. parahaemolyticus* type strain (VP2) and the other two references strains (VP166 and VP216) showed some background noise compared with those from the single PCR products, the quality of the sequences were acceptable (data not shown). However, sequencing analysis of other isolates, such as VP10, showed that the results obtained with the single PCR products (Fig. 2.27) yielded consistently better quality compared to the results obtained with the multiplex PCR products (Fig. 2.28).
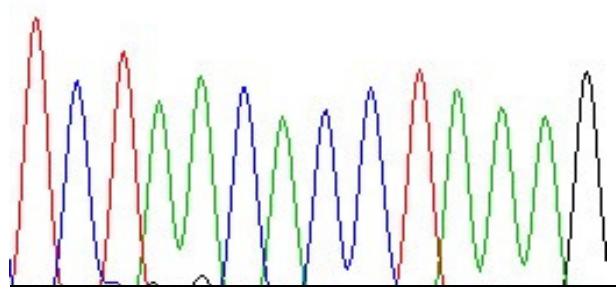
**Figure 2.27.** Sequencing chromatogram obtained from a single PCR product of *recA* for isolate VP10
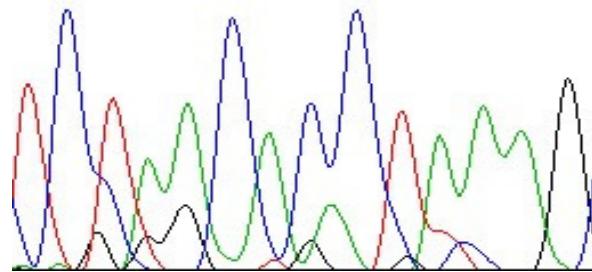


**Figure 2.28.** Sequencing chromatogram obtained from a multiplex PCR product of *recA* for isolate VP10

Due to the inconsistent quality of the sequencing results from multiplex PCR products, the multiplex PCR system was not used for the remaining *V. parahaemolyticus* in the strain collection. Rather, PCR products from single gene amplifications were used for MLST gene sequencing in this study.

## *2.3.5 Optimization of a single PCR for seven housekeeping enzyme genes*

Fragments of seven housekeeping genes from 128 *V. parahaemolyticus* isolates were amplified. The optimized PCR conditions for individual gene fragments were different among these isolates. Although the majority of isolates were amplified by the same optimized condition (annealing temperature 59$^{o}$C, 1.5 mM magnesium concentration and PCR primer pairs provided in Table 2.4), certain genes in some isolates could not be amplified by these conditions. Further optimization of PCR conditions (including the use of various annealing temperatures and design of alternative primers) was necessary to obtain DNA fragments of the seven housekeeping genes for all 128 isolates. Details of PCR optimization for each gene are described below. PCR conditions and primer pairs that were capable of amplifying the seven gene fragments for 128 individual isolates are summarized in Table A2, Appendix 3.

### 2.3.5.1 PCR optimization of *dnaE*

The *dnaE* fragments of 124 *V. parahaemolyticus* isolates were successfully amplified by using primers dnaE-F3 and dnaE-R1 and an annealing temperature of 59$^{o}$C. The *dnaE* fragments of four isolates, VP34A, VP34B, VP206A, and VP206B could not be amplified using this primer pair by these PCR conditions since there were non-specific DNA products being amplified (Fig. 2.29). However, these non-specific bands of VP34A, VP34B, VP206A, and VP206B diminished at the higher annealing temperature of 60$^{o}$C (Fig. 2.30). By such PCR optimization, the *dnaE* fragments of all 128 isolates of *V. parahaemolyticus* were obtained.
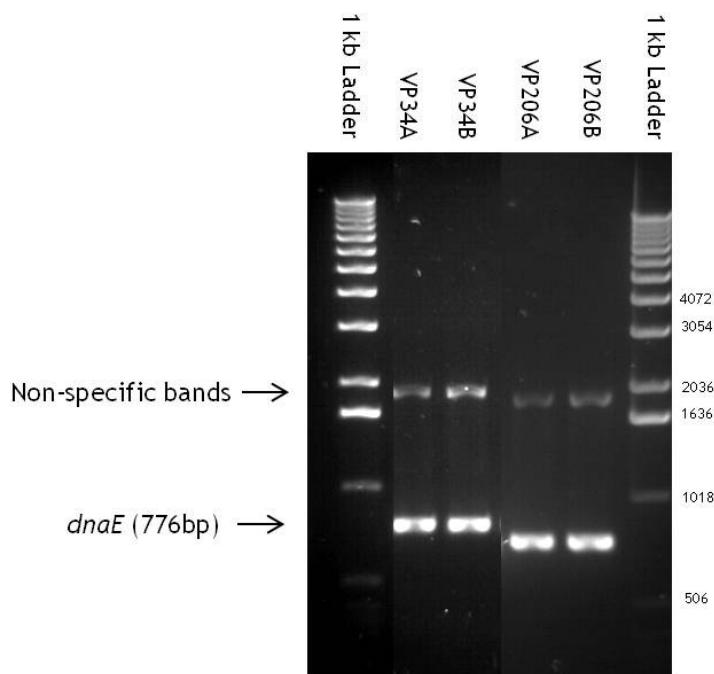
**Figure 2.29.** Gel electrophoresis of *dnaE* fragments from VP34A, VP34B, VP206A, and VP206B. Primers number dnaE-F3 and dnaE-R1 were applied with an annealing temperature of 59C°
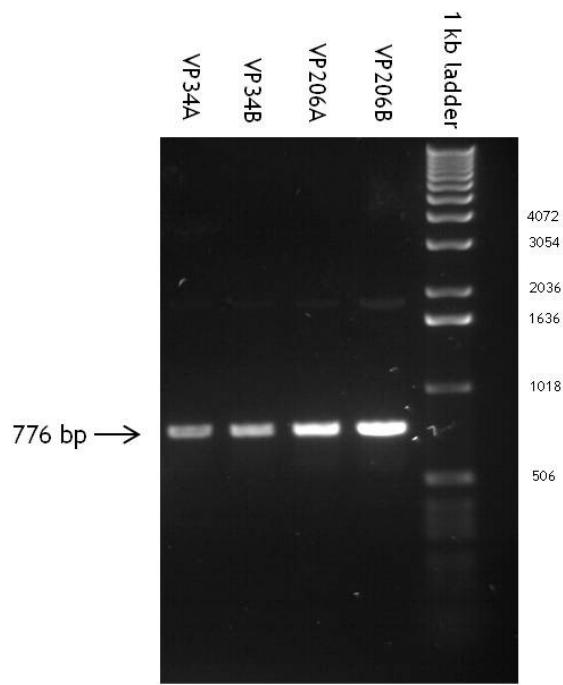


**Figure 2.30.** Gel electrophoresis of *dnaE* fragments from VP34A, VP34B, VP206A, and VP206B. Primers number dnaE-F3 and dnaE-R1 with an annealing temperature of 60C°

## 2.3.5.2 PCR optimization of *gyrB*

The *gyrB* fragments of 121 *V. parahaemolyticus* isolates were successfully amplified by using primers gyrB-F1 and gyrB-R1 with an annealing temperature of $59^{\circ}$C. The *gyrB* fragments of seven isolates, VP44, VP46, VP48, VP50, VP54, VP206A and VP206B could not be amplified using this primer pair by these PCR conditions. Alternative primer combinations of gyrB-F1 and gyrB-R2, gyrB-F2 and gyrB-R2 were applied to isolate VP44. Gel electrophoresis showed that the primers gyrB-F1 and gyrB-R2 were capable of *gyrB* amplification for VP44 at an annealing temperature of $56^{\circ}$C (Fig. 2.31). Theses primers, gyrB-F1 and gyrB-R2, were selected for PCR reactions of the other isolates representing negative PCR reactions, VP46, VP48, VP50, VP54, VP206A and VP206B. As a result, primers gyrB-F1 and gyrB-R2 were able to amplify *gyrB* fragments for VP46, VP48, VP50, VP54, VP206A, and VP206B at the annealing temperature of $56^{\circ}$C (Fig. 2.32). Using the PCR conditions described above, *gyrB* fragments of 128 isolates of *V. parahaemolyticus* were obtained.

## 2.3.5.3 PCR optimization of *recA*

The *recA* fragments of 108 *V. parahaemolyticus* isolates were successfully amplified by using primers recA-F1 and recA-R3 with an annealing temperature of $59^{\circ}$C. The *recA* fragments of 20 isolates, VP18, VP34A, VP34B, VP62, VP78, VP82, VP84, VP86, VP88, VP90, VP92A, VP92B, VP94, VP96, VP98, VP104, VP106, VP122, VP124 and VP190 could not be amplified using this primer pair by these PCR conditions. Additional primers were designed (Fig. 2.3) and various PCR conditions were applied to these isolates in order to obtain *recA* fragments.
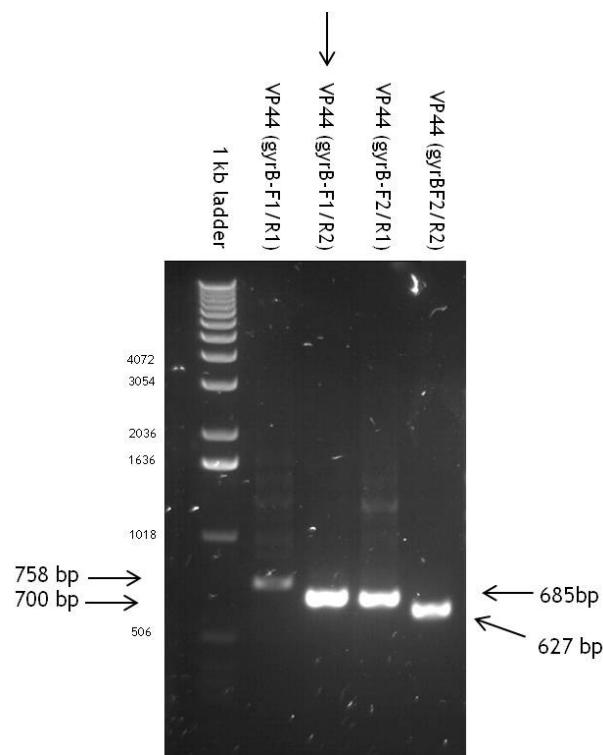
**Figure 2.31.** Gel electrophoresis of *gyrB* from VP44 with four primer combinations of gyrB-F1, gyrB-R1, gyrB-F2 and gyrB-R2 with an annealing temperature of 56°C. Arrow on top indicates the most specific amplification.
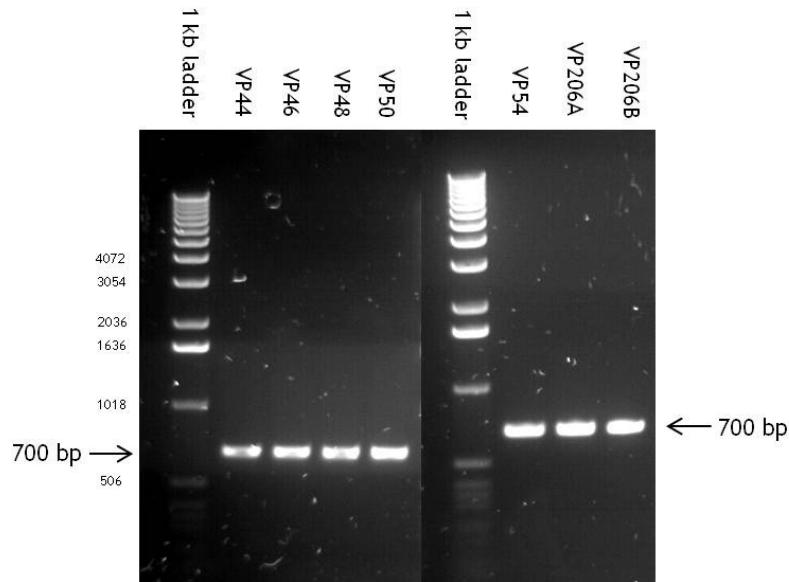


**Figure 2.32.** Gel electrophoresis of *gyrB* from VP44, VP46, VP48, VP50, VP54, VP206A, and VP206B. Primers gyrB-F1 and gyrB-R2 were applied with an annealing temperature of 56°C.

Two primer combinations, primer pairs recA-F3/recA-R2 and recA-F2/recA-R3 were used to amplify *recA* fragments of VP84 and VP94 at an annealing temperature of 57$^o$C (Fig. 2.33). The gene fragments of both VP84 and VP94 were successfully amplified by the primers recA-F3/recA-R2. This primer pair was applied to the other PCR negative isolates, VP18, VP62, VP78, VP86, VP88, VP96, VP98, VP104, VP122, VP124, VP190, VP34A, VP34B, VP82, VP90, VP92A, VP92B and VP106 (Figs. 2.34 - 2.35). As a result, *recA* fragments of VP62, VP86, VP88, VP96, VP98, VP104, VP122, VP124, and VP190 were obtained by these primers whereas those of VP18, VP34A, VP34B, VP78, VP82, VP92A and VP92B were not, as they were represented only by non-specific bands. The *recA* fragments of VP90 and VP106 were amplified but non-specific bands also occurred (Fig. 2.35). However, the same primer pair, recA-F3/recA-R2, was able to specifically amplify *recA* fragments in VP90 and VP106 when the annealing temperature was increased to 58$^o$C (Fig. 2.36).

The alternative primer pair, recA-F1 and recA-R1, was used for isolates VP34A, VP78, VP82 and VP92A at an annealing temperature of 55$^o$C (Fig. 2.37). The results show that the *recA* fragment (851 bp) was not amplified in any of these isolates under these conditions, with only strong non-specific and smearing bands being present (Fig. 2.37).

Newly designed primers, recA-F4 and recA-R4, were used for *recA* amplification of VP18, VP34A, VP78, VP82 and VP92A as well as VP2 as a control (Fig. 2.38). The PCR was performed at an annealing temperature of 56$^o$C. Primers recA-F4 and recA-R4 were able to amplify a *recA* fragment (1001 bp) for VP2, but not for the remaining negative isolates, VP18, VP34A, VP78, VP82 and VP92A (Fig. 2.38).
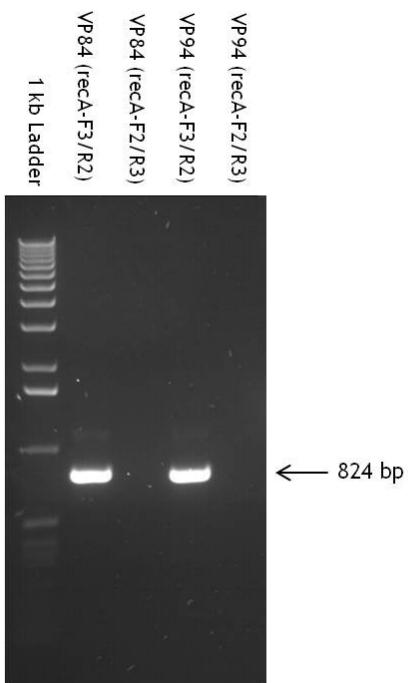
**Figure 2.33.** Gel electrophoresis of *recA* from VP84 and VP94 by using two primer combinations of recA-F2, recA-R3, recA-F3 and recA-R2 with an annealing temperature of 57°C.
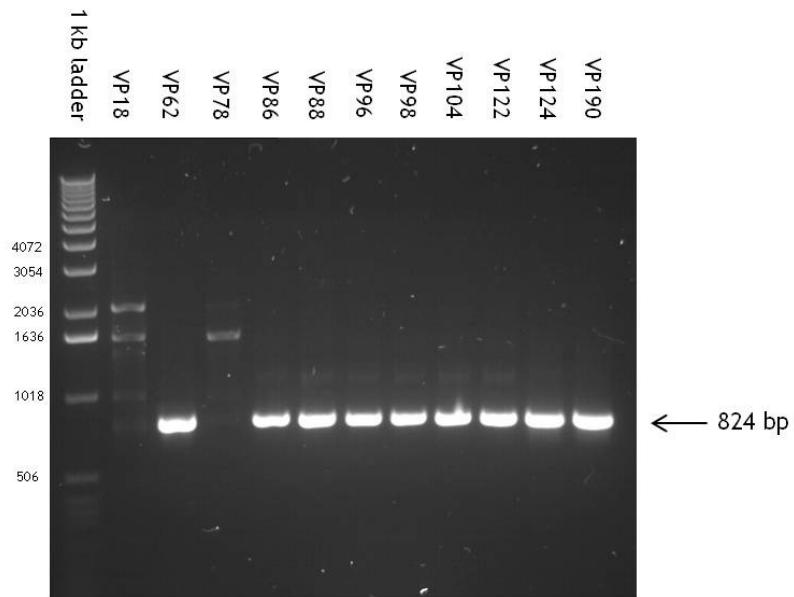


**Figure 2.34.** Gel electrophoresis of *recA* from VP18, VP62, VP78, VP86, VP88, VP96, VP98, VP104, VP122, VP124, and VP190.  Primers recA-F3 and recA-R2 were applied with an annealing temperature of 57°C.
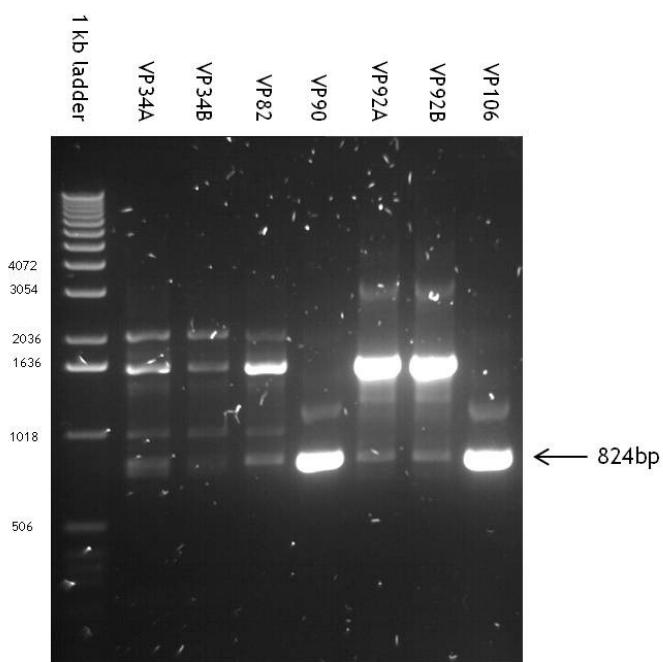
**Figure 2.35.** Gel electrophoresis of *recA* from VP34A, VP34B, VP82, VP90, VP92A, VP92B and VP106. Primers recA-F3 and recA-R2 were applied with an annealing temperature of 57°C.
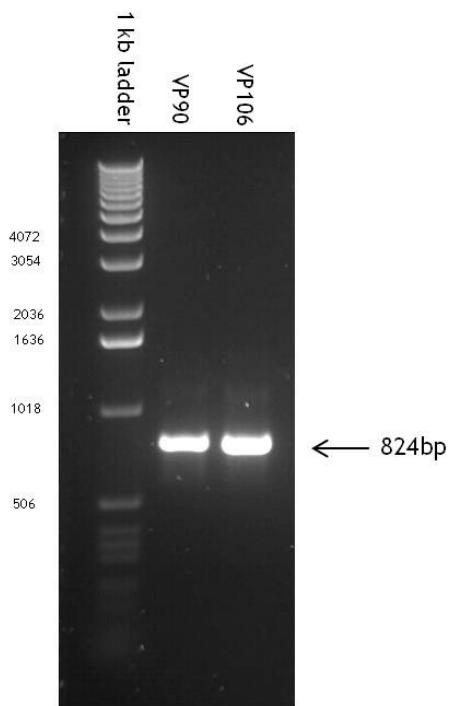


**Figure 2.36.** Gel electrophoresis of *recA* from VP90 and VP106. Primers recA-F3 and recA-R2 were applied with an annealing temperature of 58°C.
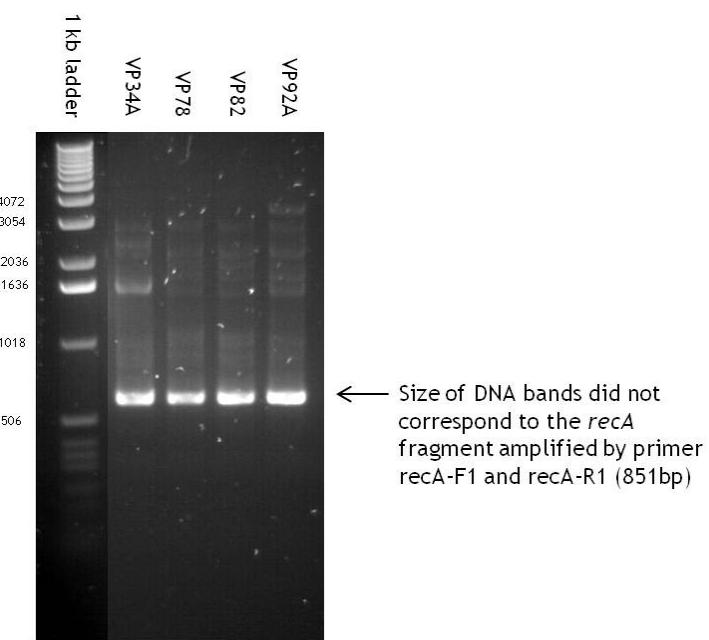
**Figure 2.37.** Gel electrophoresis of *recA* VP34A, VP78, VP82 and VP92A. Primers recA-F1 and recA-R1 were applied with an annealing temperature of 55°C.
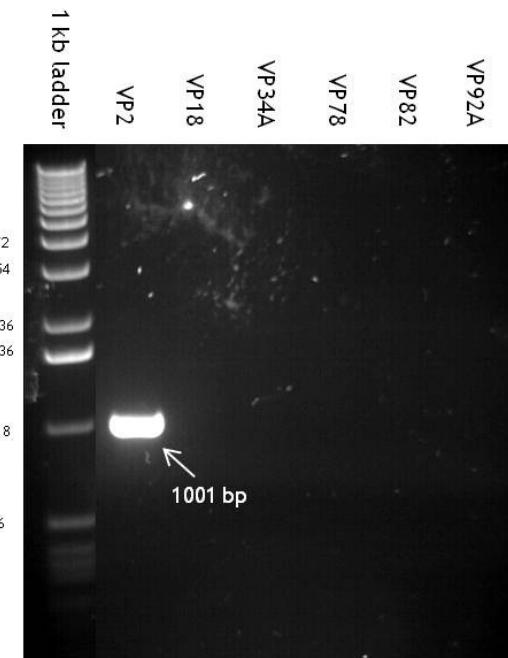


**Figure 2.38.** Gel electrophoresis of *recA* VP2, VP18, VP34A, VP78, VP82 and VP92A. Primers recA-F4 and recA-R4 were applied with an annealing temperature of 56°C.

In summary, the *recA* fragments of 121 *V. parahaemolyticus* isolates were obtained by various PCR conditions with three different primer combinations (recAF1/R1, recAF1/R3 and recAF3/R2) (Table A2, Appendix 3). However, seven isolates, VP18, VP34A, VP34B, VP78, VP82, VP92A and VP92B still gave negative results, despite various different PCR conditions being applied.

### 2.3.5.4 PCR optimization of *dtdS*

The *dtdS* fragments of 107 *V. parahaemolyticus* isolates were successfully amplified by using primers dtdS-F1 and dtdS-R1, at an annealing temperature of 59°C. The *dtdS* fragments of 21 isolates, VP4, VP6, VP16, VP30, VP36, VP40, VP68, VP80, VP142, VP148, VP150, VP152, VP154A, VP154B, VP164, VP196, VP206A, VP206B, VP234, VP208 and VP218 could not be amplified using this primer pair by these PCR conditions. To solve this problem, additional primers (Fig. 2.4) were designed and various different PCR conditions were applied to these isolates.

Primers dtdS-F3 and dtdS-R3 were designed and used in combination with the previously used PCR primer pair dtdS-F1 and dtdS-R1 for the isolate VP4 (Fig. 2.39). The primer combination of dtdS-F3/dtdS-R3 at an annealing temperature of 56°C gave the best *dtdS* amplification for the isolate VP4 although the band was slightly smeared on the gel. The effect of the magnesium concentration on DNA amplification was therefore evaluated. However, magnesium concentrations other than that used previously (1.5 mM) did not improve DNA amplification (Fig. 2.40).
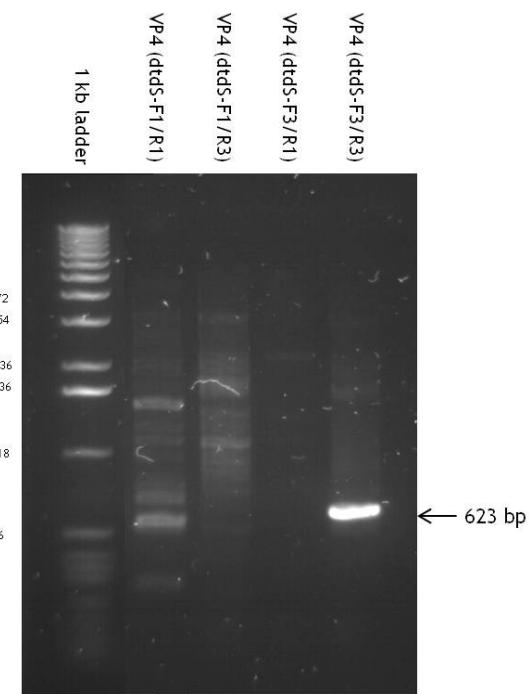
**Figure 2.39.** Gel electrophoresis of *dtdS* fragments from VP4 with four primers combinations of dtdS-F1, dtdS-R1, dtdS-F3, and dtdS-R3. The PCR annealing temperature was 56°C.
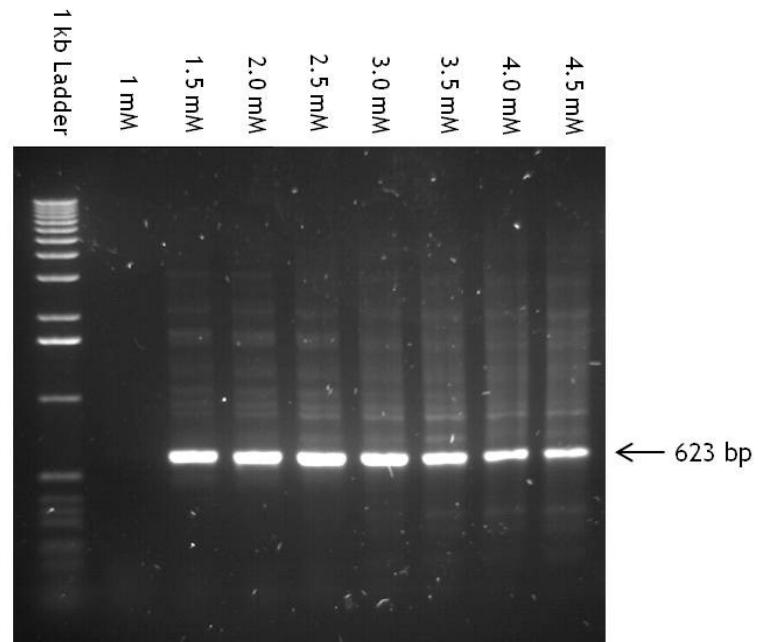


**Figure 2.40.** Gel electrophoresis of *dtdS* from VP16 with variable magnesium concentrations. Primers dtdS-F3 and dtdS-R3 were used with an annealing temperature of 57°C.

The *dtdS* fragments of VP6, VP16, VP30, VP36, VP68, VP142, VP206A and VP234 were amplified by the primers dtdS-F3 and dtdS-R3 at an annealing temperature of 56$^{o}$C (Fig. 2.41). However, the gene fragments of some isolates, namely VP40, VP80, VP148, VP150, VP152, VP154A, VP154B, VP196, VP206B, VP208, VP164 and VP218, could not be amplified under these conditions (data not shown). However, the *dtdS* fragments of these isolates were successfully amplified by decreasing the annealing temperature to 55$^{o}$C (Fig. 2.42). In summary, the *dtds* fragments for all 128 isolates were obtained by using various different PCR conditions (Table A2, Appendix 3).

## 2.3.5.5 PCR optimization of *pntA*

The *pntA* fragments of 122 *V. parahaemolyticus* isolates were successfully amplified by using primers pntA-F1 and pntA-R3 with a PCR annealing temperature of 59$^{o}$C. The *pntA* fragments of six isolates, VP42, VP56, VP130B, VP134, VP140, and VP142 could not be amplified using this primer pair by these PCR conditions. PCR negative isolates, VP56, VP130B, VP140 and VP142 were amplified by alternative primers, pntA-F3 and pntA-R1, at an annealing temperature of 59$^{o}$C (Fig. 2.43).

Primer combinations of pntA-F1/pntA-R3 and pntA-F3/pntA-R1 were used for VP42 and VP134 with a PCR annealing temperature of 55$^{o}$C (Fig. 2.44). Although a primer pair pntA-F1/pntA-R3 was previously applied to these isolates, the previous annealing temperature (59$^{o}$C) used was different.
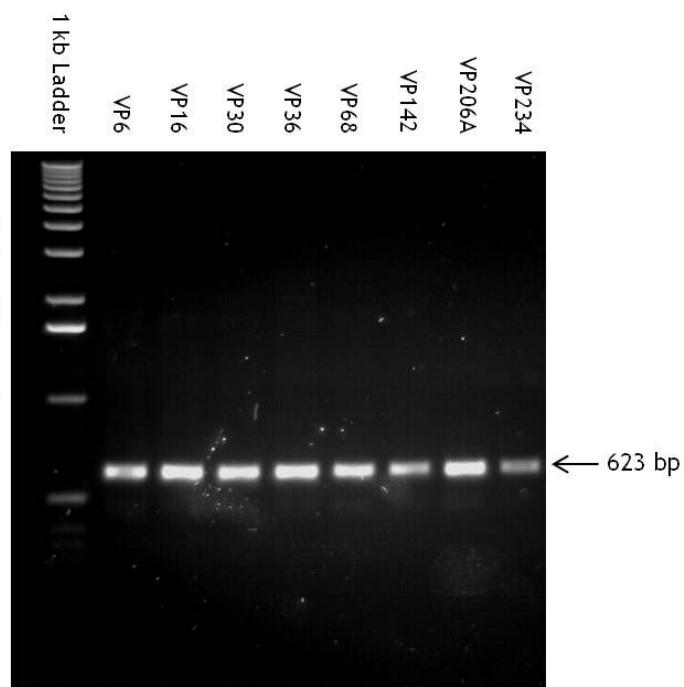
**Figure 2.41.** Gel electrophoresis of *dtdS* VP6, VP16, VP30, VP36, VP68, VP142, VP 206A and VP234. Primers dtdS-F3 and dtdS-R3 were applied with an annealing temperature of 56°C.
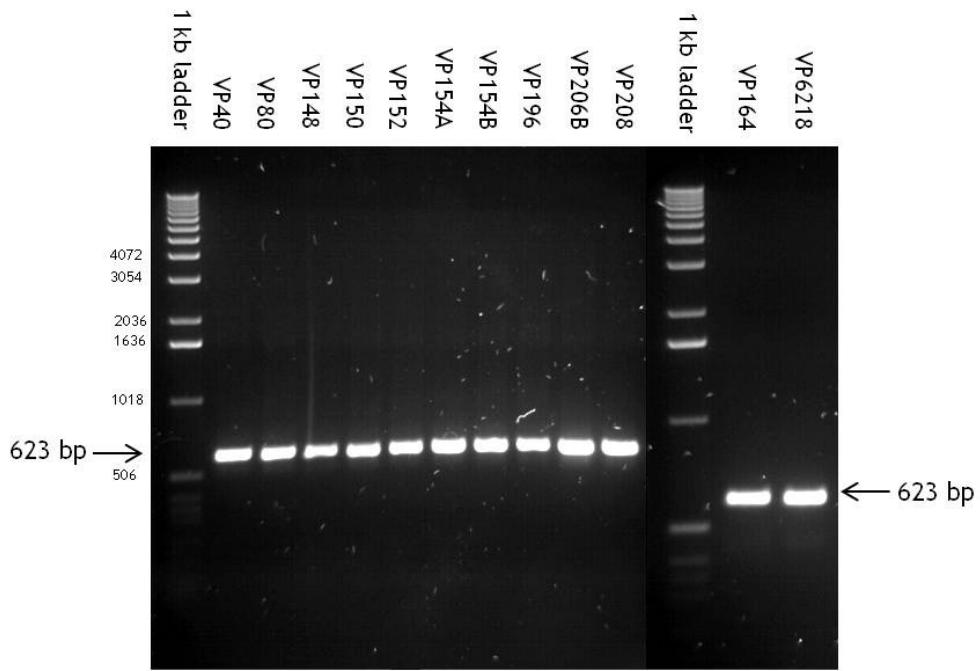


Figure 2.42. Gel electrophoresis of *dtdS* VP40, VP80, VP148, VP150, VP152, VP154A, VP154B, VP196, VP206B, VP208, VP164 and VP218. Primers dtdS-F3 and dtdS-R3 were applied with an annealing temperature of 55°C.
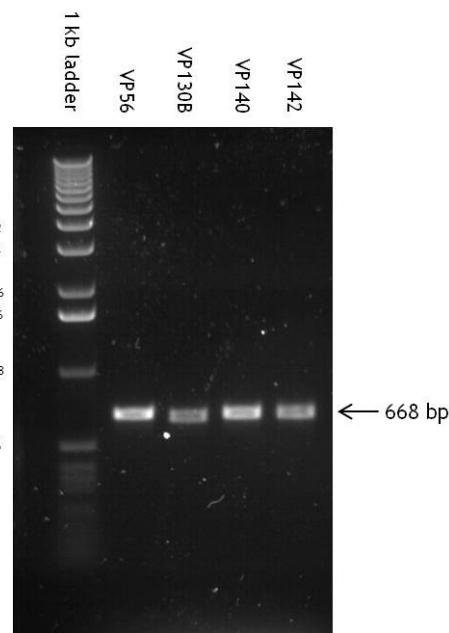
**Figure 2.43.** Gel electrophoresis of *pntA* from VP56, VP130B, VP140 and VP142. Primers pntA-F3 and pntA-R1 were applied at an annealing temperature of 59ºC.
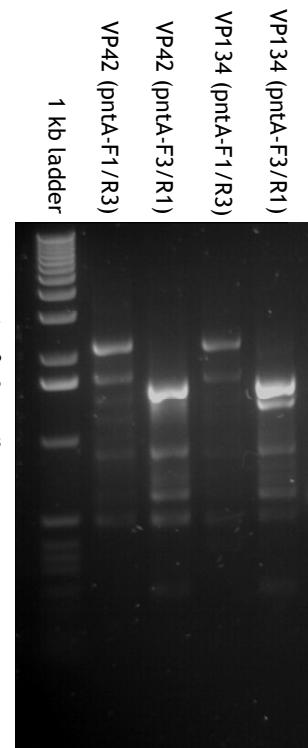


**Figure 2.44.** Gel electrophoresis of *pntA* VP42 and VP134 with four primers combinations. Combination of primers pntA-F1/R3 and pntA-F3/R1 were applied at an annealing temperature of 55ºC.

The results show that these primer combinations were unable to amplify specific *pntA* fragments from VP42 and VP134. Alternative primer combination of pntA-F2 and pntA-R2 was used for VP42 and VP134 with an annealing temperature of 58°C (Fig. 2.45). This condition was still unable to specifically amplify *pntA* fragments of VP42 and VP134. Instead, newly designed PCR primers pntA-F4 and pntA-R4 were used for optimization of these two isolates with a PCR annealing temperature of 56°C (Fig. 2.46). This primer pair was able to amplify the *pntA* fragment of VP2, the isolate used as a control, but not of VP42 and VP134. In summary, *pntA* fragments of 126 *V. parahaemolyticus* isolates were obtained by an annealing temperature of 59°C with different primer combinations (Table A2, Appendix 3). However, two isolates, VP42 and VP134 gave negative results with all PCR conditions used.

## 2.3.5.6 PCR optimization of *pyrC*

The *pyrC* fragments of 121 *V. parahaemolyticus* isolates were successfully amplified by using primers pyrC-F1 and pyrC-R1 with an annealing temperature of 59°C. The *pyrC* fragments of seven isolates, VP32, VP62, VP130A, VP130B, VP226, VP234 and VP236 could not be amplified using this primer pair by these PCR conditions. Four primer combinations of pyrC-F1, pyrC-R1, pyrC-F2, and pyrC-R2 were applied to the PCR negative isolates, VP62 and VP226, at an annealing temperature of 56°C (Fig. 2.47). Gel electrophoresis showed that the primers pyrC-F1 and pyrC-R1, which were used to amplify gene fragments of 121 isolates, gave positive results for both VP62 and VP226 at an annealing temperature of 56°C (Fig. 2.47).
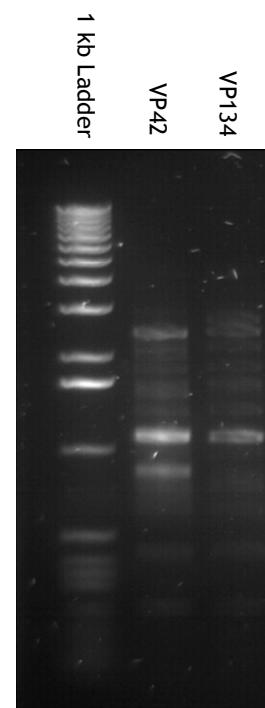
**Figure 2.45.** Gel electrophoresis of *pntA* VP42 and VP134. Primers pntA-F2 and pntA-R2 were applied with an annealing temperature of 58$^o$C.
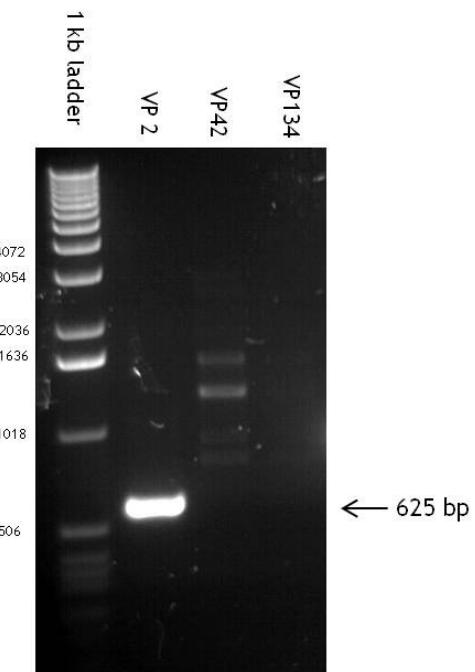


**Figure 2.46.** Gel electrophoresis of *pntA* from VP2, VP42 and VP134. Primers pntA-F4 and pntA-R4 were applied with an annealing temperature of 56$^o$C.
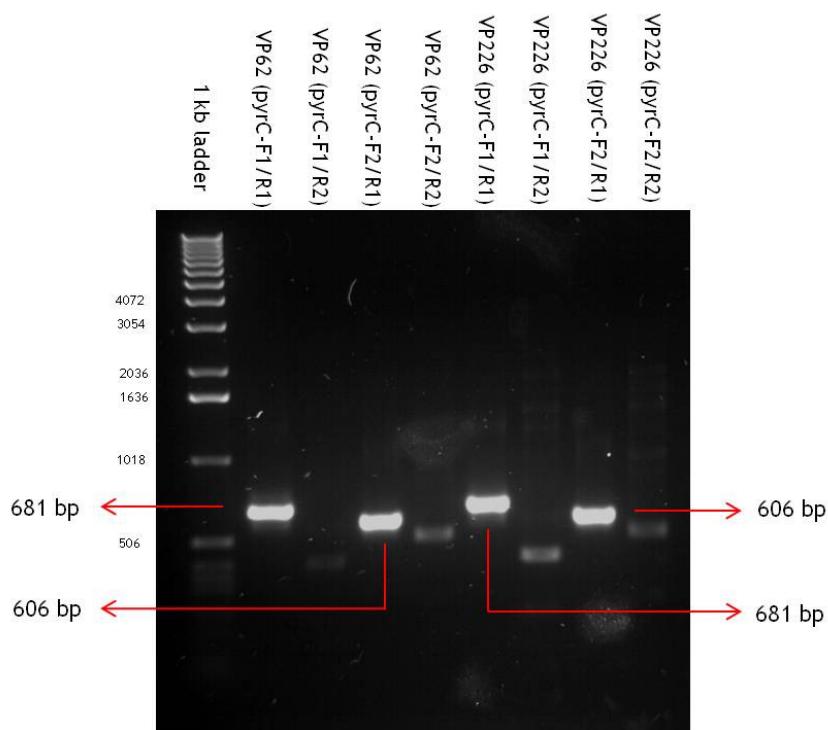
**Figure 2.47.** Gel electrophoresis of *pyrC* from VP62 and VP226 with four primers combinations. Primers pyrC-F1, pyrC-R1, pyrC-F2, and pyrC-R2 were applied with a PCR annealing temperature of 56°C.

This result indicates that, among the 128 isolates of *V. parahaemolyticus, pyrC* fragments cannot be amplified by the same primers under the same PCR conditions, but that the same primers can be used when the annealing temperature is decreased to 56°C. Thus, the primer combination of pyrC-F1 and pyrC-R1 was used for the other PCR negative isolates, VP32, VP130A, VP130B, VP234 and VP236 at an annealing temperature of 56°C (Fig. 2.48). As a result, the *pyrC* fragments of these negative isolates were successfully amplified. In summary, *pyrC* fragments from all 128 isolates were obtained by the same primer pair but using different annealing temperatures (Table A2, Appendix 3).
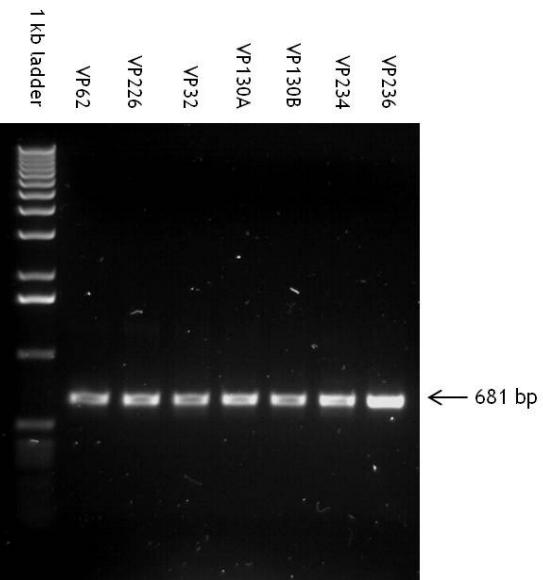
**Figure 2.48.** Gel electrophoresis of *pyrC* from VP62, VP226, VP32, VP130A, VP130B, VP234 and VP236. Primers pyrC-F1 and pyrC-R1 were applied with an annealing temperature of 56°C.

**2.3.5.7 PCR optimization of *tnaA***

The *tnaA* fragments of all 128 *V. parahaemolyticus* isolates were successfully amplified by using primers tnaA-F3 and tnaA-R3 with an annealing temperature at 59°C. Examples of *tnaA* fragment amplifications are shown in Fig. 2.49.
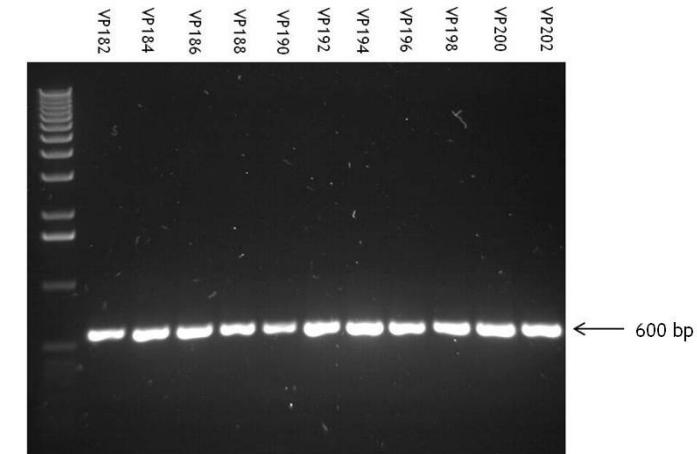


**Figure 2.49.** Gel electrophoresis of *tnaA* from VP182, VP184, VP186, VP188, VP190, VP192, VP194, VP196, VP198, VP200 and VP202. Primers tnaA-F3 and tnaA-R3 were applied with an annealing temperature of 59°C.

## 2.3.6 Sequencing of individual seven housekeeping gene fragments

DNA sequences of seven housekeeping genes of most isolates were obtained by using the sequencing primers provided in Table 2.4. In some cases, alternative primers (Table 2.5) were used to obtain the DNA sequences. According to PCR and sequencing optimization of the 128 *V. parahaemolyticus* isolates, not all of these isolates could be used for the MLST study since some could not be amplified or sequenced for one or more housekeeping genes. A total of 106 isolates were selected for the MLST study because their seven housekeeping gene fragments were successfully sequenced. However, preliminary analysis from the Neighbour-Joining tree of MLST data for these 106 isolates demonstrated that the isolates showing different colony morphologies, that were labelled as A and B isolates, represent the same sequence type (ST) (data not shown). Thus, only the A isolate was chosen for further MLST analyses. Consequently, the numbers of Thai isolates used for the MLST study was decreased to 101. However, *V. parahaemolyticus* type strain ATCC17802[T] (VP2), the strain that was identified as the food poisoning agent in Japan, was incorporated into this MLST study as a reference strain. Thus, the total number of *V. parahaemolyticus* isolates used in the MLST study was 102.

## 2.3.7 Analysis of seven concatenated housekeeping genes sequences of Thai V. parahaemolyticus

### 2.3.7.1 Nucleotide diversity at each locus

The nucleotide sequence data of the seven housekeeping gene fragments for the 102 *V. parahaemolyticus* isolates are summarized in Table 2.8. The number of

alleles observed for each locus ranged from 39 (*pntA*) to 49 (*gyrB*), and the percentage of polymorphic nucleotide sites varied from 7.3% (*tnaA*) to 25.8% (*recA*). The most frequently occurring alleles at each locus were *dnaE119* (10), *gyrB87* and *gyrB152* (10), *recA24*, *recA61* and *recA120* (10), *dtdS29* (12), *pntA23* (11), *pyrC11* (26), and *tnaA26* (13). The $d_N/d_S$ ratios were < 1 for *dnaE, recA, pntA, pyrC,* and *tnaA*; no non-synonymous changes were detected at *gyrB* or *dtdS*. The mean $d_N/d_S$ ratio for the two genes on chromosome I (0.038) was slightly higher than that of the three genes on chromosome II (0.024).

**Table 2.8.** Nucleotide and allelic diversity of MLST loci for 102 *V. parahaemolyticus* isolates

| Chromosome and locus | Fragment size (bp) | No. of alleles | No. of polymorphic nucleotide sites (%) | No. of inferred variable amino acid sites (%) | *dN/dS* ratios |
|---|---|---|---|---|---|
| I | | | | | |
| *dnaE* | 595 | 47 | 44 (7.9%) | 4 (2.2%) | 0.044 |
| *gyrB* | 627 | 49 | 47 (8.0%) | 1 (0.5%) | 0.000 |
| *recA* | 767 | 42 | 187 (25.8%) | 17 (7.0%) | 0.033 |
| II | | | | | |
| *dtdS* | 497 | 48 | 36 (7.9%) | 0 (0.0%) | 0.000 |
| *pntA* | 488 | 39 | 36 (8.4%) | 7 (4.9%) | 0.024 |
| *pyrC* | 530 | 44 | 39 (8.0%) | 10 (6.1%) | 0.026 |
| *tnaA* | 493 | 40 | 31 (7.3%) | 6 (4.3%) | 0.021 |

### 2.3.7.2 Genotypic diversity

Details of the isolates, including source and year of isolation, serotype, presence of the haemolysin-encoding genes, *tdh* and *trh*, sequence type (ST) and allelic profile are presented in Table 2.1. A total of 63 STs were identified among the 102 isolates, and 53 (86%) of these were novel since they had not previously been recorded in the MLST database (http://pubmlst.org/vparahaemolyticus). The high proportion of novel STs in this study illustrates the high degree of

environmental diversity, even on a fairly local scale, and how poorly the current MLST dataset represents this diversity. It should also be noted that the recovery of novel STs was non-random with respect to epidemiological source. A total of 68 isolates were recovered from seafood, frozen shrimp, shrimp tissue and water. These isolates accounted for 40 STs and 39 (98%) of these were novel. In contrast, the 18 isolates recovered from human carriers were represented by 14 STs and 10 (71%) of these were novel. Finally, the 16 isolates recovered from human disease corresponded to 9 STs and only four (44%) of these were novel. Of the 53 novel STs identified in this study, only two were associated with more than one source; ST239 was recovered from a frozen shrimp and from water at farm 2, and ST246 was recovered from a shrimp and from water at the corresponding farm (farm 1).

Four of the clinical STs identified in the present study were also associated with clinical isolates in the MLST database. ST83 (recovered from three patients in our study) represents a common clinical ST in Japan and India, ST189 corresponds to clinical isolates from China, Japan, and India, ST66 represents clinical isolates previously recovered from Mozambique, and ST17 corresponds to clinical isolates from Spain and the USA. Similarly, the four human carriage STs identified in the present study that were also recorded in the MLST database had also been recovered from cases of disease. Two of our carriage isolates, VP132 and VP158, correspond to the pandemic ST3. The other STs associated with human carriage had previously been recovered from clinical cases in China (ST62 and ST199) and Mozambique (ST68). Significantly, none of the clinical and carrier STs were associated with environmental isolates. These data strongly suggest that both clinical and human carrier isolates do not represent a random sample of the reservoir of diversity present in the environment. Rather, a

limited number of genotypes appear to be adapted to human carriage and some of these isolates have the potential to cause disease. Interestingly, the single environmental ST (ST114 from seafood) in our study that was already present in the MLST database was previously recorded from a seafood source in the USA.

### 2.3.7.3 Clonal relationships of *V. parahaemolyticus* population

#### *2.3.7.3.1 Index of Association ($I^S_A$)*

High rates of recombination have previously been demonstrated in *V. parahaemolyticus* (González-Escalona *et al.*, 2008; Yan *et al.*, 2011) and other *Vibrio* species (Byun *et al.*, 1999; Keymer & Boehm, 2011; Thompson *et al.*, 2005). The extent of recombination within natural populations can be determined by calculating the amount of linkage between alleles relative to a null of random association using the standardized Index of Association ($I^S_A$) (Haubold & Hudson, 2000; Smith *et al.*, 1993). When this is calculated over all 102 isolates, $I^S_A$ = 0.5966. However, this decreases to 0.1350 when only the 63 generated STs are considered, indicating that a large proportion of the linkage is accounted for by the expansion of specific STs, that is, the population corresponds to an "epidemic" population structure (Smith *et al.*, 1993). Although the value of 0.1350 still represents a significant departure from linkage equilibrium, this decrease is consistent with high rates of recombination between clones, and the inclusion of all unique STs (348 STs) from the *V. parahaemolyticus* MLST database (http://pubmlst.org/vparahaemolyticus/) results in a similarly low $I^S_A$ value (0.1162).

### 2.3.7.3.2 Analysis of clonal structure

eBURST was used to identify and visualize clonal groups within the nucleotide sequence dataset of seven housekeeping genes for the 102 *V. paraheamolyticus* isolates (Fig. 2. 50). The 63 STs generated were separated into two clonal complexes (CC83 and CC233), two doublets (D1 and D2), and 53 singletons. Clone complex 83 consists of five clinical isolates and includes the type strain; these isolates represent serotypes O1:K1 (4) and OUT:KUT (1). Clone complex CC233 comprises four isolates recovered from frozen shrimp; these isolates represent serotypes O3:K20 (3) and O10:KUT (1). Doublet 1 (D1) corresponds to six clinical isolates representing ST262 and a single-locus variant, ST255, which includes two human carrier isolates (VP138 and VP162). The clinical isolates represent serotypes O1:K69 (3), O4:K11 (1), O8:K22 (1) and O3:KUT (1) and the carrier isolates serotypes O1:K12 and O11:K5. These two genotypes differ only at the *dnaE* locus, but inspection of the variant allele sequences reveals that they differ by 5 polymorphisms in 557 sites (0.9%) which suggests an intra-species recombination event involving the *dnaE* locus. A far more striking example of recombination is provided by doublet 2 (D2). This corresponds to two clinical isolates represented by STs 189 (VP200, O4:K8) and 265 (VP190, O4:K10). Close inspection of the *recA* allele sequences of these two STs reveal that they differ by 138 polymorphisms in 729 sites (18.9%). Such extreme divergence can only be explained by way of an inter-species recombination event. In support of this, the best BLAST score for *recA107* of ST265 corresponded to *Vibrio cincinnatiensis*, although this sequence was still >10% divergent from the query sequence. Clearly, *recA107* of ST265 was likely imported from an as yet unidentified donor species (see Figure A6 in Appendix 3). Of the 53 singleton STs, 45 were represented by only a single isolate.  The singleton ST251, corresponding to the

shrimp tissue isolates from farm 2, was the most frequent ST in the dataset (10 out of 102 isolates). However, these isolates were represented by a range of different serotypes, including O1:KUT (2), O2:K28 (1), O2:KUT (1), O4:K63 (1), O9:K23 (2), O10:K52 (1), O10:K71 (1), and OUT:KUT (1).
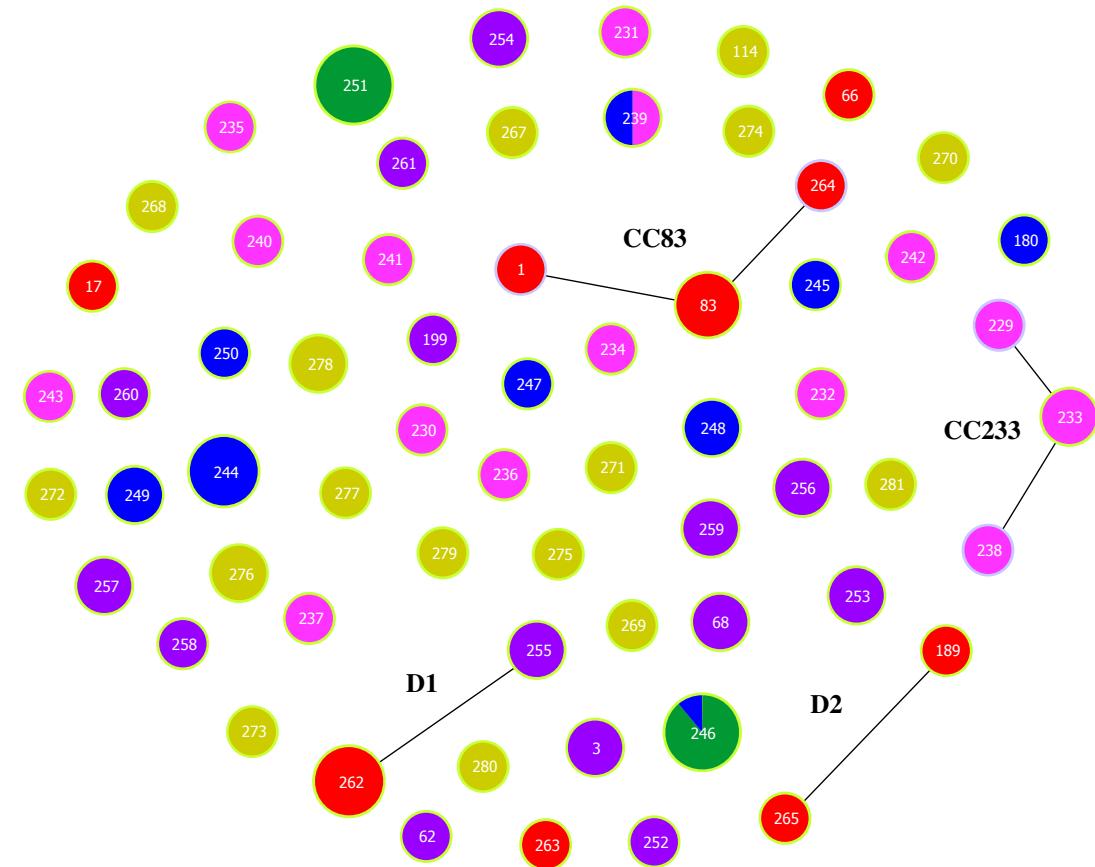


**Figure 2.50.** eBURST analysis of 63 STs of *V. parahaemolyticus*. The analysis is based on allelic profiles of MLST data and displays clusters of linked and individual unrelated STs. Single locus variants (SLVs) are illustrated by linkage lines among the nodes. Colour coding represents the source of isolation of each ST: red = clinical sample; purple = human carrier; yellow = seafood; green = shrimp tissue; pink = frozen shrimp; dark blue = shrimp-farm water. The frequency of each ST is indicated by the size of each node.

Due to the high degree of nucleotide sequence diversity of housekeeping enzyme genes in *V. parahaemolyticus*, amino acid sequence analysis was used to investigate clonal relationships from a wider perspective. Nucleotide sequences were translated into amino acid sequences and amino acid sequence types (aaSTs) were assigned to individual STs (Table A3, Appendix 3). In total, 87 aaSTs were assigned from the 348 STs in the *V. parahaemolyticus* MLST database (http://pubmlst.org/ vparahaemolyticus/) and these were used to perform an eBURST analysis (Fig. 2.51). Two major predicted ancestors, aaST2 and aaST34, were identified. Each of these was represented by a higher proportion of environmental isolates (aaST2 = 76%; aaST34 = 62%) than of clinical isolates (aaST2 = 24%; aaST34 = 38%) regardless of geographical region. However, it is noteworthy to investigate the relationship of pandemic ST3 isolates with respect to other strains based on aaSTs. ST3 corresponds to subgroup founder aaST7 which comprises 18 STs. Unlike aaST2 and aaST34, the majority (92%) of isolates in aaST7 were from clinical sources whereas a much lower proportion (8%) of environmental isolates were present. Of the Thai isolates in the collection, ST3 was associated with only two isolates that were recovered from human carriers, but was not found among the clinical isolates. Thus, only a small number of isolates from human carriers in Thailand, and none from clinical cases, are closely related to clinical strains having a worldwide distribution.

Although amino acid eBURST analysis was capable of demonstrating clonal relationships within the bacterial population more clearly than nucleotide eBURST, there was no clear evidence that clinical strains were associated with a particular epidemiological source in Thailand.
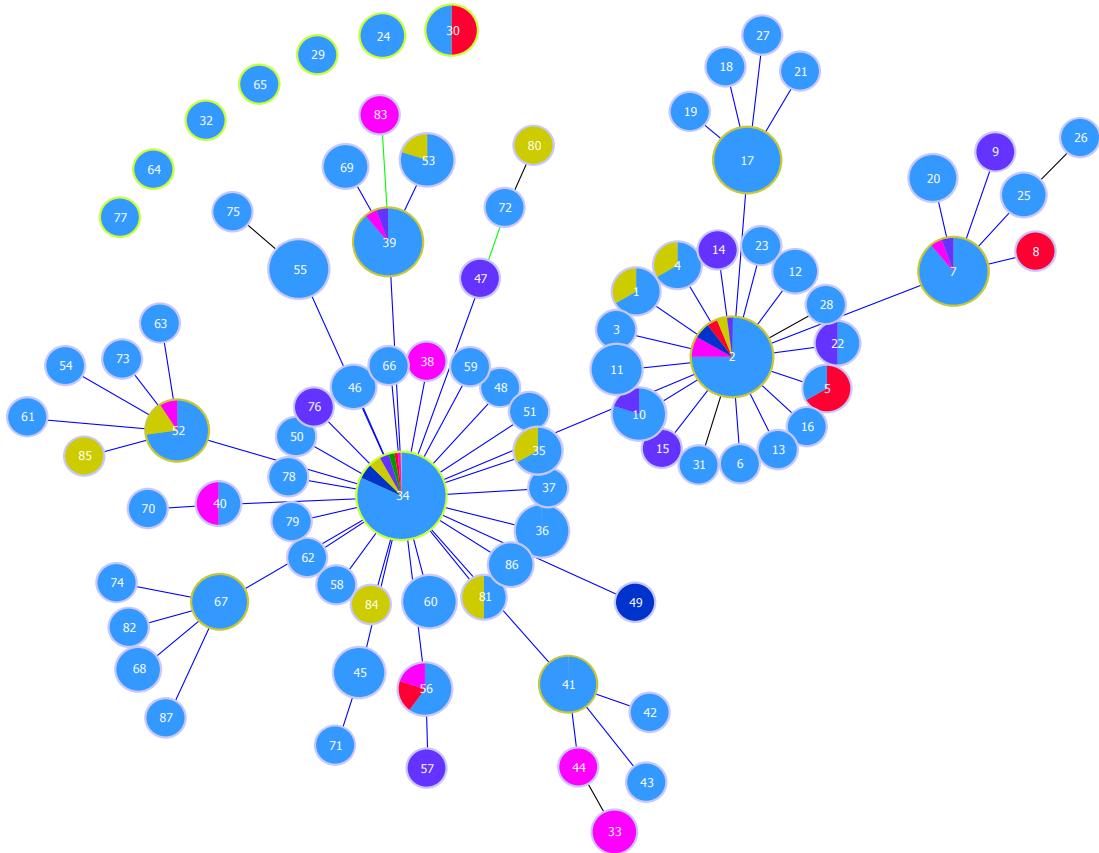
**Figure 2.51.** Population snapshot of 87 aaSTs of *V. parahaemolyticus* which were resolved from 348 STs from the *V. parahaemolyticus* MLST database (http://pubmlst.org/ vparahaemolyticus/). Two predicted founder groups, aaST2 and aaST34, were identified and each was surrounded by a ring of subgroup founders and SLVs. Blue represents isolates recovered from the *V. parahaemolyticus* MLST database while other colours represent the source of isolation of the Thai isolates in the present study: red = clinical samples; purple = human carrier; yellow = seafood; green = shrimp tissue; pink = frozen shrimp; dark blue = shrimp-farm water. The frequency of each aaST is indicated by the size of each node.

## 2.3.7.4 Phylogenetic analysis

A Neighbour-Joining tree representing the concatenated sequences of the seven housekeeping gene fragments in 102 isolates is shown in Fig. 2.52. The phylogenetic tree consists of two major lineages, A and B, which are separated by a relatively large genetic distance and have a high bootstrap value. Lineage A is further sub-divided into two clades, I and II, although these are closely related and have low bootstrap values. The isolates recovered from human carriers, seafood, and frozen shrimp were very diverse and were distributed widely throughout the tree. The short internal nodes and low bootstrap scores evident from the phylogenetic analysis is consistent with a history of frequent recombination. Furthermore, the low bootstrap values indicate that the topology of the tree is poorly supported. However, five clear clusters, 1 to 5, representing isolates of the same and closely related STs (in the case of cluster 2) are apparent within the tree and these are strongly supported by high bootstrap values. With the exception of cluster 1, each cluster corresponds to isolates from a single source.

Cluster 1 is represented by eight isolates recovered from shrimp tissue at farm 1 in August 2007 and a single isolate from water at the same farm in January 2008; cluster 2 corresponds to four clinical isolates obtained from a single hospital in Bangkok in May and August 1990 and the type strain which was associated with a case of food poisoning in Japan in 1950; cluster 3 corresponds to six isolates recovered from shrimp farm water at farm 1 (ponds A and B) in January 2008; cluster 4 corresponds to six clinical isolates originating from the same hospital as those present in cluster 2 between April 1990 and February 1991; and cluster 5
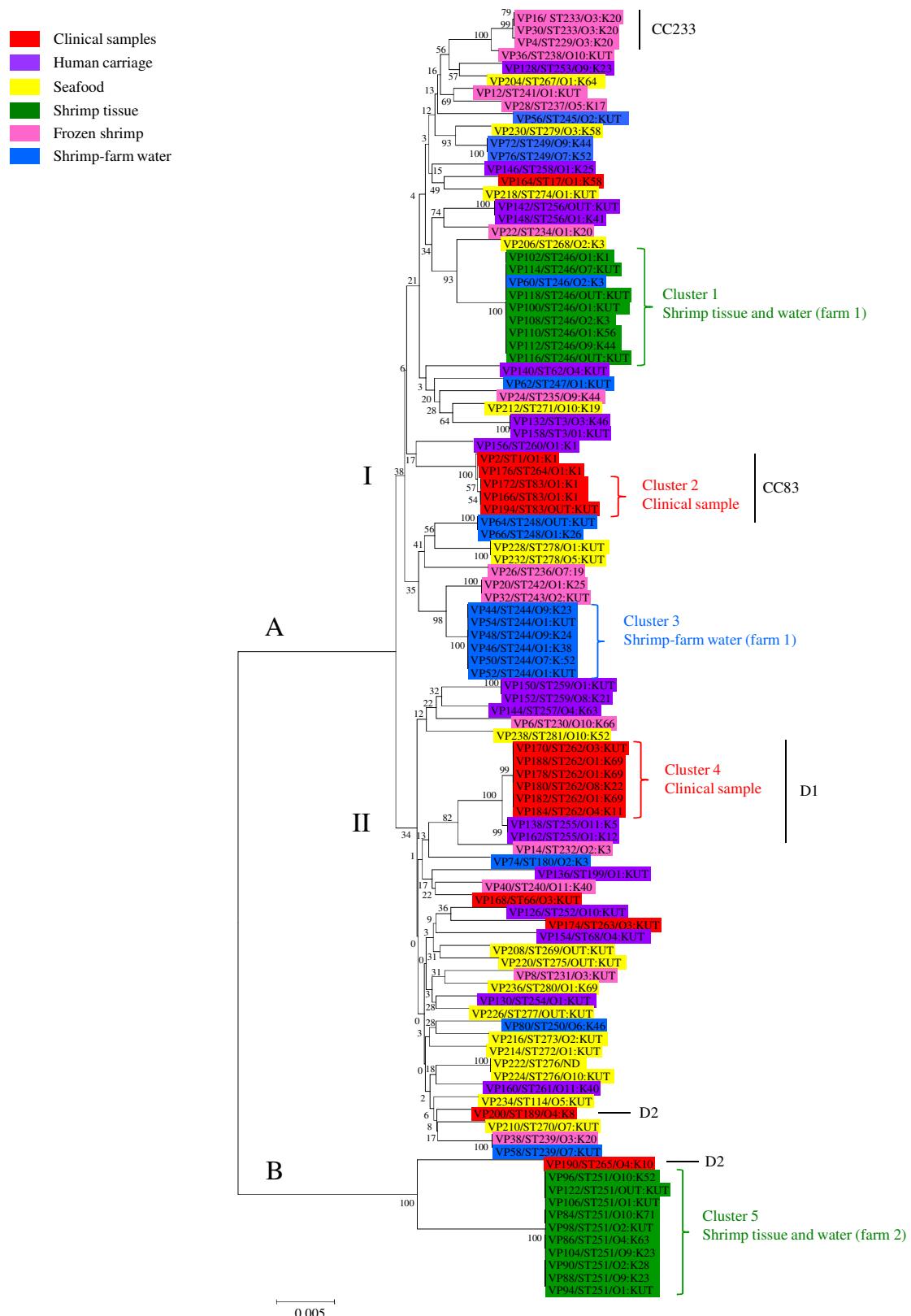
**Figure 2.52.** Neighbour-Joining tree of 102 concatenated sequences of *V. parahaemolyticus* from multiple sources in Thailand.

corresponds to ten isolates recovered from shrimp tissue at farm 2 in August 2007. The contemporaneous recovery of clinical isolates representing distinct clusters, 2 and 4, from patients in the same hospital in Bangkok in 1990/91 clearly indicates that two distinct disease-causing clones were circulating at this time. We also note a close association between these clinical clusters and three isolates from human carriage (VP156, cluster 2; and VP138 and VP162, cluster 4). Different haemolysin gene profiles were also observed among clinical isolates within each of the clusters 2 and 4 (Table 2.1). In cluster 2, isolate VP194 possesses only *tdh* whereas isolates VP166, VP172 and VP176 possess both *tdh* and *trh*; in cluster 4, VP170, VP178, and VP188 contain both *tdh* and *trh* whereas VP180 and VP184 contain only *tdh,* and VP182 contains only *trh*. With the exception of VP194 (OUT:KUT), all isolates in cluster 2 represent serotype O1:K1; in contrast, isolates in cluster 4 represent multiple serotypes (O1:K69, O4:K11, and O8:K22, O3:KUT).

Clusters 1 and 5 represent predominantly shrimp tissue isolates and correspond to isolates recovered from two different farms, 1 and 2, respectively. The existence of these clusters points to very limited diversity within the farms at any given point in time. In support of this, cluster 3 represents isolates recovered from two separate ponds (A and B) at farm 1 in January 2008. However, this cluster is distinct from cluster 1 which represents isolates recovered from shrimp tissue at the same farm five months earlier in August 2007. These observations suggest that the clusters may represent temporal effects resulting from cycles of rapid clonal expansion and replacement within a single farm.

Analysis of the distribution of polymorphic nucleotide sites within the seven gene fragments provides an explanation for the divergence of lineage B (Fig. 2.53). It is clear from Fig. 2.53 that isolates of STs 251 and 265 (lineage B) have highly divergent *recA* alleles. As discussed above, these *recA* alleles have most probably been acquired by horizontal DNA transfer. However, inspection of Fig. 2.53 indicates that intragenic recombination within *recA* has also occurred involving STs from clade I and especially clade II. Clearly, *recA* is having a major influence on the overall branching pattern of the phylogenetic tree and is responsible for the delineation of lineages A and B, as well as clades I and II. Indeed, when *recA* is removed from the concatenated sequences, a Neighbour-Joining tree is recovered which lacks any major lineages (Fig. A7 in Appendix 3). A high level of divergence at the *recA* locus is also apparent within other *V. parahaemolyticus* STs in the MLST database (http://pubmlst.org/ vparahaemolyticus). Distribution of polymorphic nucleotide sites of individual housekeeping genes of isolates in this study are demonstrated by happlot diagrams in Fig. A8-14 in Appendix 3.
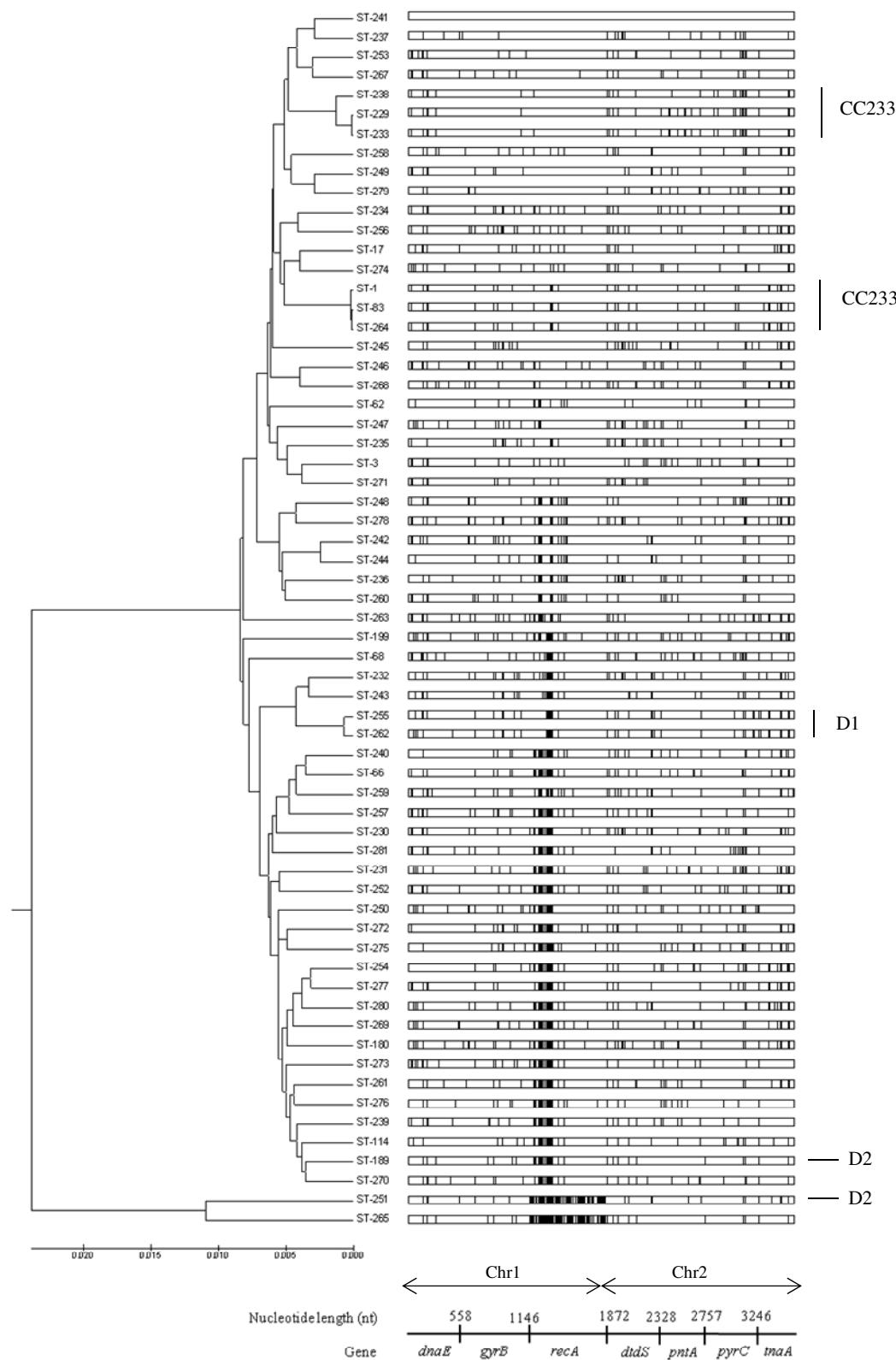
**Figure 2.53.** Distribution of polymorphic nucleotide sites among concatenated sequences of 63 STs. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, ST241. The demarcation and nucleotide lengths of the seven genes are indicated along the bottom scale.

**2.3.7.5 Distribution of polymorphic nucleotide sites within *recA* alleles**

Nucleotide sequence alignments of representative *recA* alleles were examined in further detail to determine the nature and extent of the intragenic recombination events. Visual inspection of the distribution of polymorphic nucleotide sites revealed evidence of multiple intragenic recombinational exchanges among the *recA* alleles of strains isolated from clinical, human carrier, and environmental sources since these alleles had complex mosaic structures (Fig. 2.54). In the region spanning nucleotides 48 to 224 three mosaic segments, A to C, could be identified. Segments A, B, and C comprised four, five and three different nucleotide sequences, respectively, and these could be arranged to give 10 different A, B, C combinations. For example, *recA60* (A1, B1, C1) and *recA116* (A2, B1, C1) share identical B (nt 90-161) and C (nt 162-224) segments but have different A (nt 48-89) segments. Allele *recA17* (A3, B2, C1) shares an identical segment C with *recA60* (A1, B1, C1) and *recA116* (A2, B1, C1) but contains different segments A and B. Alleles *recA116* (A2, B1, C1) and *recA48* (A2, B1, C2) share identical A and B segments but contain very different C segments. Similarly, alleles *recA17* (A3, B2, C1) and *recA88* (A3, B2, C3) share identical A and B segments but contain very different C segments. Alleles *recA36* (A4, B3, C1), *123* (A4, B4, C1), *27* (A4, B4, C2), and *126* (A4, B5, C2) share identical A segments , but contain three segment Bs and two segment Cs in four different combinations.

 Furthermore, *recA36* (A4, B3, C1) shares identical A and B segments with *recA19* (A4, B3, C3) but contains a very different C segment. Alleles *recA27* (A4, B4, C2) and *recA6*0 (A1, B1, C1) differ in all three segments. Significantly, seven of the

(A)

```
                                    1111 1111111111 1111111111 1111111111 1111111112 2222222222 2222222222 2222222333 3333333333 3333333333 3333333333 3444444444 4444444445 5555555555 5555555666 6666666666 666667777
                 111222333 4455556666 7799990000 1112222223 3445556667 7788888990 0011122233 4445666777 7888899000 1111233345 5556667777 7888888999 9000111122 5667788990 0111112334 4577789001 1555666778 899990112
                 9259147049 5814702369 2503692358 1470123492 8173692352 4701369266 7813625817 0692124036 9025817369 2358103981 2343692356 8234789046 9025148903 0581706251 7013692576 9803627062 5457346291 806798576
recA60   TTCTGTTTCA TTCTGTATTA ATAATTACTT TTGACGTGCA TGCAGCTCAG ATAAGATCGC ACTGTAACTG AAGTATACTC AACCAGTTTT TCAATGCCCC AGCTGATCTA GTCTAGCACT GCTACGCAGT ATTGCTTGTC CGCTGTCTAT TACTAATCGG CTAGCCATGA TGCTAGCAA
recA116  .C........ ..T...C... .......... .......... .......... .......... .......... ........T. .......... .......T.. ....T..... .......... .......... .......... T......... ...C...... .C........ .........
recA17   .......... ......C... ..G.AAGTG. GAA.....T. ...T...... .......... .......... .......... .......... .......T.. .....C.... .......... .......... .......... T......... .......... .C........ .........
recA36   .......... ..AAC.C... ..G..CG..C ..A.....T. ...T...... C......... .......... .......... .......... .......T.. .....C.... .......... .......... .......... .......... .......... .C........ .........
recA123  .......... G..AAC.C... ..G..G..C ..A.....T. ...T...... .......... .......... .......... .......... .......T.. .....C.... .......... .......... .......... .......... .......... .C........ .........
recA27   .......... ..AAC.C... ..G..G..C ..A.....T. ...T...GTG. GCT.AG.T.. .......... .......... .......... .......T.. .....C.... .......... .......... .......... .......... .......... .C........ .........
recA126  .......... G..AAC.C... .......... .......T. ...T...GTG. GCT.AG.T.. .......... .......C .........C ........TT .....C.... .......... .......... C......... .C........ .C........ .........
recA48   .C........ ..T...C... .......... .......T. ...T...GTG. GCT.AG.T.. .......... .......... .........C ........T. .....C.... .......... .......... .C........ .C........ .........
recA88   .......... ......C... ..G.AAGTG. GAA.....T. ...T...GTG. GCT.AG.T.. G...AG... .......... .......T .........A .......... .......... .......... .......... .C........ .C........ .........
recA19   .......... ..AAC.C... ..G..CG..C ..A.....T. ...T...GTG. GCT.AG.T.. G...AG... C......... .......... .....C.... .......... .......... .......... .......... .C........ .C........ .........
recA107  A.TCC.GC.T GC.AACT.CT GA.T...GC. ..AGAAGCTG .TT.ATC.TA C.TCAGATAA GTC..GTTAA .TACTC.TGT TGTTGC.GA. ATGTAATT.. ...CTGG.CC A....AGCAA AAGGTTGCCA GACTTCAACT TT.AAC.CGC .TT.GTC..A .CCTAAG.TT CA.CTATGG
recA120  A.TCCCGCTT GC.AACT.C. .A.T...GC. ..AGAAGCTG CTT.A.G.T. C.TCAGATAA GTCT.GTTA. GTA.TCGTAG TGTT...G.. ATGTAATT.. TCACC..AGC TAGCGT.... AAGG.TGCTA GACTTCAAC. TTTAACTCT. .TTCGT.TC. TCCTAAGCTT CATCTATGG

                   <————————> <————————><——————————————————————> <————————————————————>
                        A          B                                      C
```

(B)

| Alleles | Recombinant segments | Mosaic designations | Isolates | ST | Sources |
|---|---|---|---|---|---|
| *recA60* | A1 / B1 / C1 | A1,B1,C1 | VP222, VP224 | 276 | seafood |
| *recA116* | A2 / B1 / C1 | A2,B1,C1 | VP6 | 230 | frozen shrimp |
| *recA17* | A3 / B2 / C1 | A3,B2,C1 | VP136 | 199 | human carriage |
| *recA36* | A4 / B3* / C1 | A4,B3,C1 | VP154 | 68 | human carriage |
| *recA123* | A4 / B4 / C1 | A4,B4,C1 | VP138, VP162, VP170, VP178, VP180, VP182, VP184, VP188 | 255, 262 | clinical, human carriage |
| *recA27* | A4 / B4 / C2 | A4,B4,C2 | VP24, VP176, VP166, VP172, VP194 | 235, 364, 83 | clinical, frozen shrimp |
| *recA126* | A4 / B5 / C2 | A4,B5,C2 | VP156 | 260 | human carriage |
| *recA48* | A2 / B1 / C2 | A4,B1,C2 | VP174 | 263 | clinical |
| *recA88* | A3 / B2 / C3 | A3,B2,C3 | VP140 | 62 | human carriage |
| *recA19* | A4 / B3 / C3 | A4,B3,C3 | VP132, VP158, VP164 | 3, 17 | clinical, human carriage |

**Figure 2.54.** (A) Distribution of polymorphic nucleotide sites among a sample of *recA* alleles. The numbers written vertically above the sequences represent the positions of polymorphic nucleotide sites. The dots represent sites where the nucleotides match those of the top sequence (*recA*60). The boxes indicate regions of sequence identity that represent proposed recombinant segments. (B) Schematic representation of recombinant fragments A, B and C among *recA* alleles corresponding to the nucleotide sequences shown in Fig. 2.54A. Different nucleotide sequences in recombinant segments A, B, and C are represented by A1-4, B1-5, and C1-3, respectively. Segments B3 and B3* differ at only a single nucleotide site. Mosaic designations are represented by different combination of A, B and C. Mosaic alleles have been formed by from one to three separate intragenic recombination events.

mosaic *recA* alleles were present exclusively in isolates from clinical samples or human carriers whereas two alleles were present only in environmental isolates (seafood and frozen shrimp); a single allele was present in isolates from clinical samples and frozen shrimps. These data suggest that recombinational exchanges are occurring more frequently within the human host (i.e. within the intestinal tract) than in the environment.

## 2.3.7.6 Recombination events in housekeeping genes and the role of *recA* in phylogenetic structure

Bayesian clustering analysis (BAPS software) was employed to identify genetically distinct subpopulations. Since ST251 and ST265 (Fig. 2.53) possess divergent *recA* sequences which were likely acquired by horizontal gene transfer from other unidentified species, these STs were removed from the clustering analysis to offset the effects of recombination. As a result, two distinct genetic groups were identified within the bacterial population which comprised 61 STs (Fig. 2.55). The tree represents two clusters each of which contains a mixture of isolates from multiple sources. BAPS admixture analysis was further performed to investigate genetic hybridization between these two clusters (Fig. A15 in Appendix 3). Consequently, six hybrid STs were detected and these are shown as blue lineages in Fig. 2.55. These hybrid STs included ST242 (carrier), ST244 (water), ST263 (clinical), ST68 (carrier), ST255 (carrier) and ST262 (clinical).
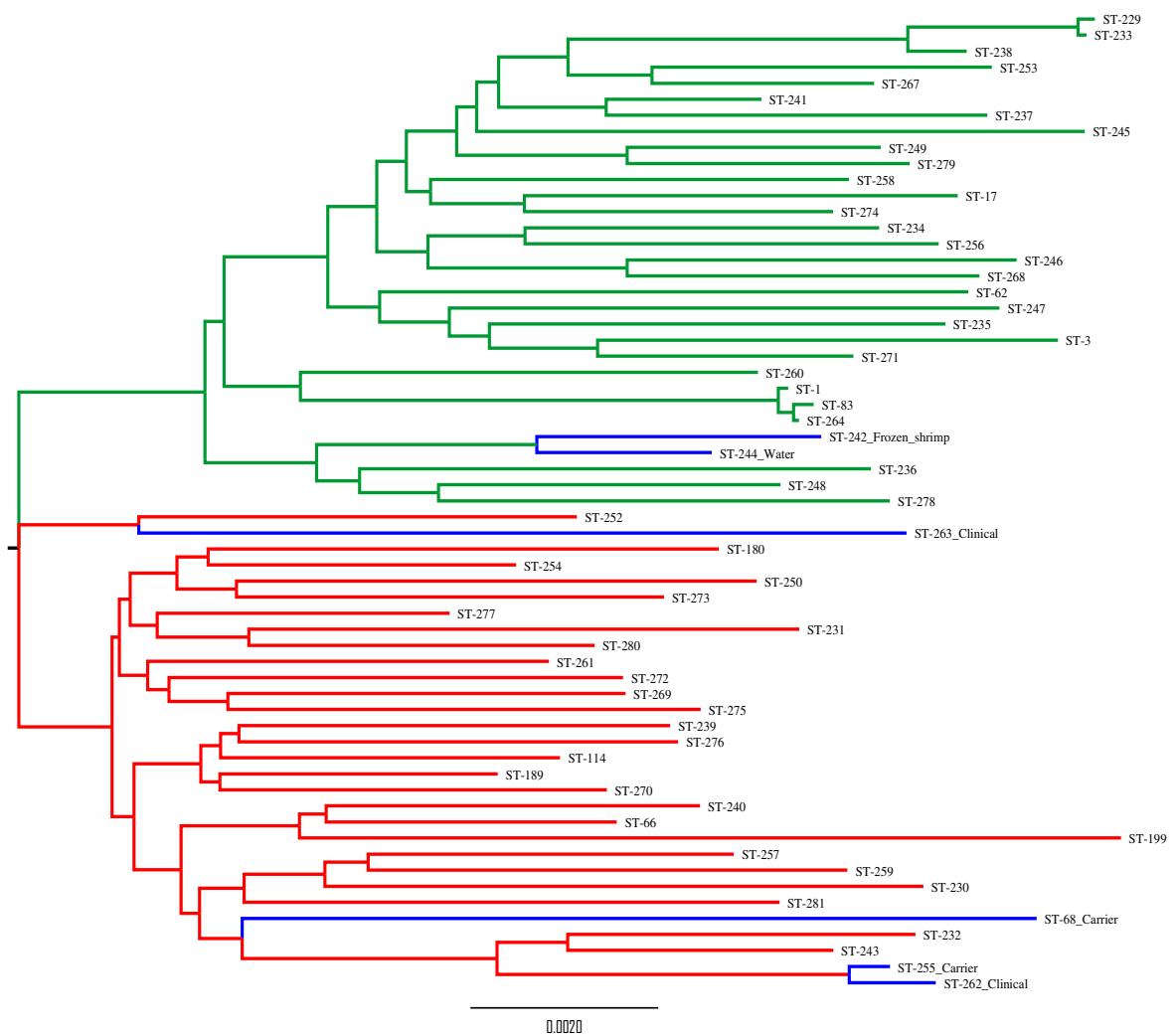
**Figure 2.55.** Neighbour-Joining tree representing BAPS clusters and admixture STs. The lineages representing different STs are coloured according to the BAPS cluster classification. Red and green colours represent STs in distinct population clusters whereas blue represent admixture STs. The population structure was obtained using the admixture model where K = 2.

To investigate whether *recA* substantially influenced the *V. parahaemolyticus* population structure, *recA* sequences were removed from the 63 concatenated sequences of the seven housekeeping genes and the data set was re-analyzed by BAPS. Consequently, BAPS clustering analysis was unable to differentiate the population that contained concatenated sequences excluding *recA* and revealed that all STs were represented in the same genetic group (data not shown). This finding confirmed that the apparent phylogenetic relationships of *V. parahaemolyticus* are strongly affected by *recA*.

Recombination events within the MLST dataset were examined in further detail using RDP3 (Martin *et al.*, 2005). Thirteen unique recombination events and 175 recombination signals were detected by the RDP, GENECONV and MaxChi programs within the RDP3 package. The recombination per mutation rate ($\sigma/\theta$) within the 63 STs was 12.922 indicating that the observed diversity has been driven predominantly by recombination. A graphic representation of the concatenated sequences of the 63 STs representing recombination breakpoints show that a majority of recombination events occur at the position of *recA* (between base positions 1147-1872) (Fig. A16 in Appendix 3). The six hybrid STs predicted by BAPS analysis and described above were further examined using RDP3. The predicted parents of the isolates representing these six STs are shown in Table 2.9.

**Table 2.9.** Strain information of six hybrid STs predicted by BAPS and RDP.

| Predicted recombinants by BAPS and RDP | | Donor | | Recipient | |
|---|---|---|---|---|---|
| ST | Source | ST | Source | ST | Source |
| 262 | Clinical sample | 68 | Human carrier | 279 | Seafood |
| 263 | Clinical sample | 235 | Frozen shrimp | 239 | Water |
| 255 | Human carrier | 68 | Human carrier | 278 | Seafood |
| 242 | Frozen shrimp | 248 | Water | 232 | Frozen shrimp |
| 244 | Water | 248 | Water | Unknown | - |
| 68 | Human carrier | - | - | - | - |

The clinical ST262 and ST263 have arisen as a consequence of recombination between ST68 (human carrier) and ST279 (seafood) and environmental ST235 (frozen shrimp) and ST239 (shrimp-farm water), respectively, whereas the other three hybrid STs, ST255 (human carrier), ST242 (frozen shrimp) and ST244 (water) have arisen as a consequence of recombination between ST68 (human carrier) and ST278 (seafood), ST248 (water) and ST232 (frozen shrimp), and ST248 (water) and unknown donor, respectively. However, RDP was unable to detect a recombination event in ST68.

## 2.3.8 Phylogenetic analysis of Thai *V. parahaemolyticus* isolates based on source of isolation

Happlot diagrams representing distributions of polymorphic nucleotide sites among concatenated sequences of seven housekeeping genes for isolates from the six sources (clinical samples, human carriers, seafood, shrimp tissue, frozen shrimp, and water) are shown in Figs. 2.56-2.61. The Neighbour-Joining tree of the concatenated sequences representing the seven gene fragments of each isolate are represented on the left of the happlot diagrams. The phylogenetic

tree of 17 clinical isolates (Fig. 2.56) represented two main genetic clusters that correspond to the MLST clusters 2 and 4 (Figs. 2.52). The high density of polymorphic sites in VP190 represents interspecies horizontal gene transfer of *recA* as discussed in previous section. High levels of genetic diversity were observed within the isolates from human carriers (Fig. 2.57), seafood (Fig. 2.58), and frozen shrimp (Fig. 2.60). The distribution of polymorphic sites among isolates from these three sources is more abundant compared to those of clinical isolates.

However, isolates from shrimp tissue (Fig. 2.59) and water (Fig. 2.61) represent more clonal populations compared to the isolates from human carriers (Fig. 2.57), seafood (Fig. 2.58), and frozen shrimp (Fig. 2.60). Isolates from shrimp tissue represent only two sequence types (STs). These represent isolates recovered from two different shrimp farms, farms 1 and 2. Among multiple polymorphic sites between two distinct groups of shrimp tissue isolates, a high density of polymorphic sites at *recA* was clearly observed between the isolates from clusters 1 and 5 which correspond to shrimp tissue isolates from farms 1 and 2, respectively. Concatenated sequences of isolates from water (Fig. 2.61) are more diverse than those from clinical and shrimp tissue isolates, although six isolates (VP44, VP46, VP48, VP50, VP52 and VP54) from water at farm 1 share identical STs that represent cluster 3 in the MLST phylogenetic tree (Fig. 2.52).

Analysis of the distribution of polymorphic sites within the concatenated sequences shows a high level of genetic diversity and frequent recombination among isolates from the same source. However, the isolates from seafood, human carriers, and frozen shrimp are relatively more diverse than the isolates from clinical samples, shrimp tissue, and water. Although the density of

polymorphic sites at *recA* in the isolates from human carriers, seafood, frozen shrimp, and water are not as high as those in the clinical isolate VP190 (Fig. 2.56) and the isolates from shrimp tissue cluster 1 (Fig. 2.59), the polymorphic sites of concatenated sequences of isolates from these sources nevertheless occur predominantly in *recA*.
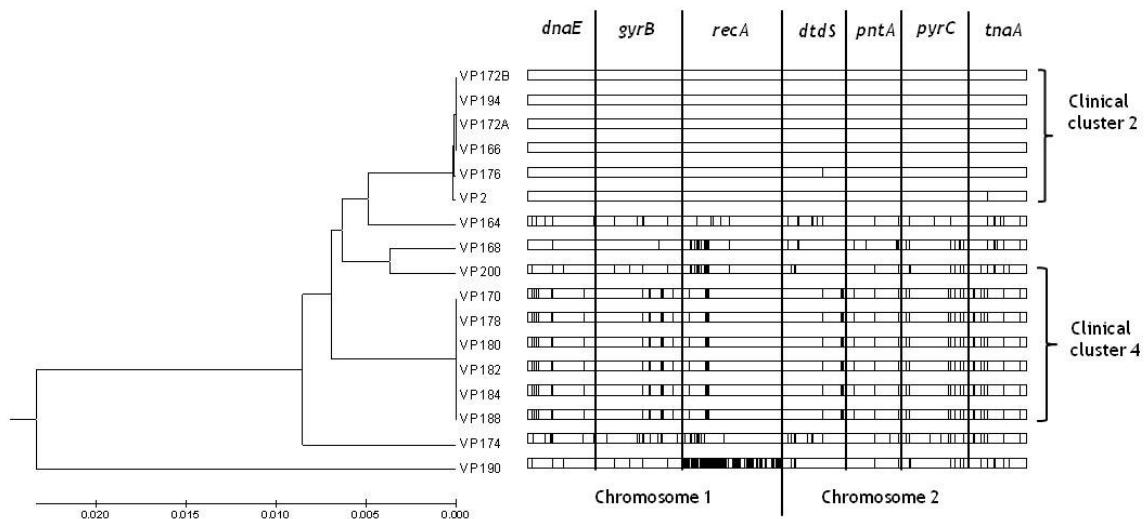


Figure 2.56. Neighbour-Joining tree and distribution of polymorphic nucleotide sites among concatenated sequences of housekeeping genes for 17 clinical isolates. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, VP172B.
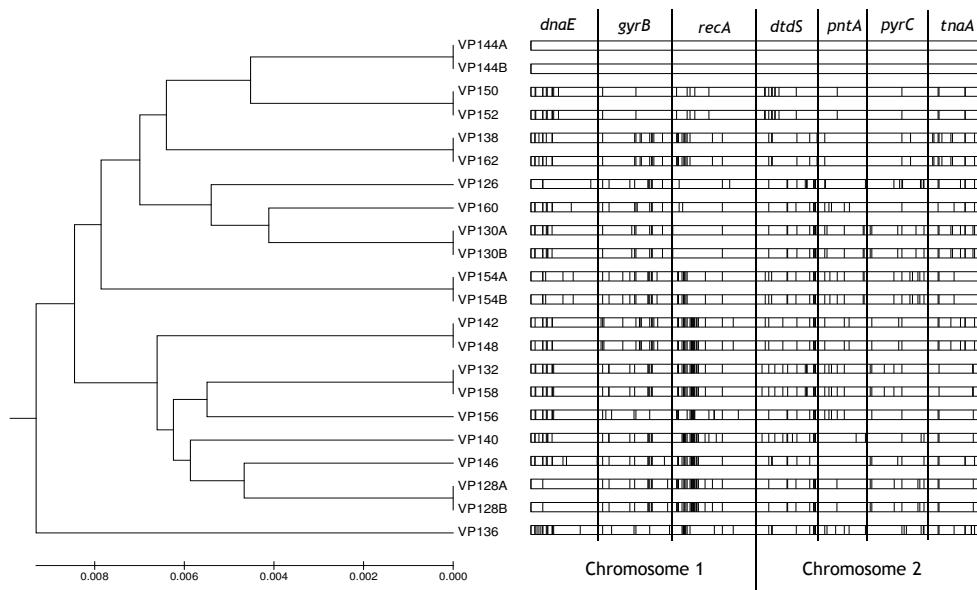
**Figure 2.57.** Neighbour-Joining tree and distribution of polymorphic nucleotide sites among concatenated sequences of housekeeping genes for 22 human carrier isolates. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, VP144A.
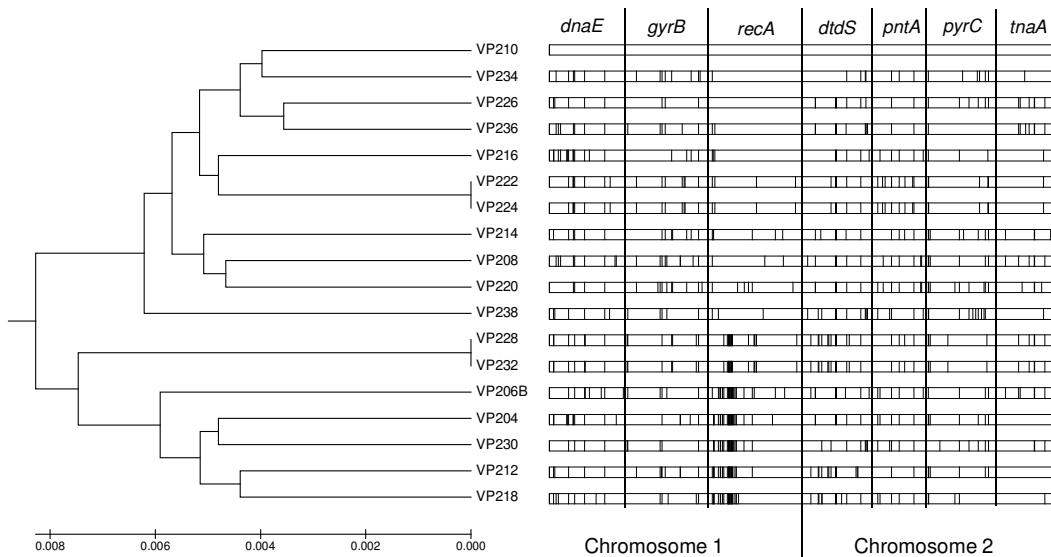


**Figure 2.58.** Neighbour-Joining tree and distribution of polymorphic nucleotide sites among concatenated sequences of housekeeping genes for 18 seafood isolates. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, VP210.
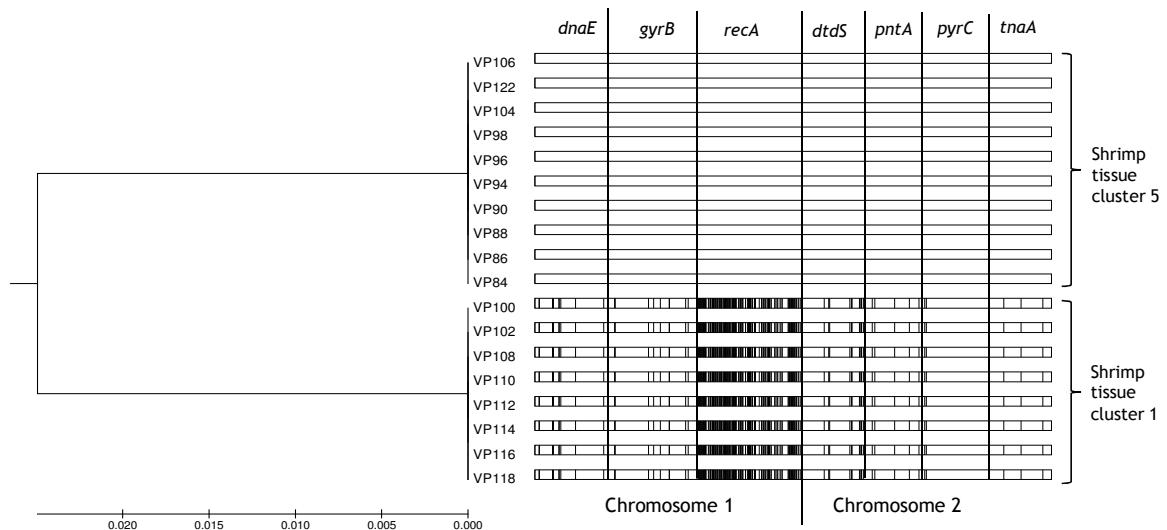
**Figure 2.59.** Neighbour-Joining tree and distribution of polymorphic nucleotide sites among concatenated sequences of housekeeping genes for 18 shrimp tissue isolates. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, VP106.
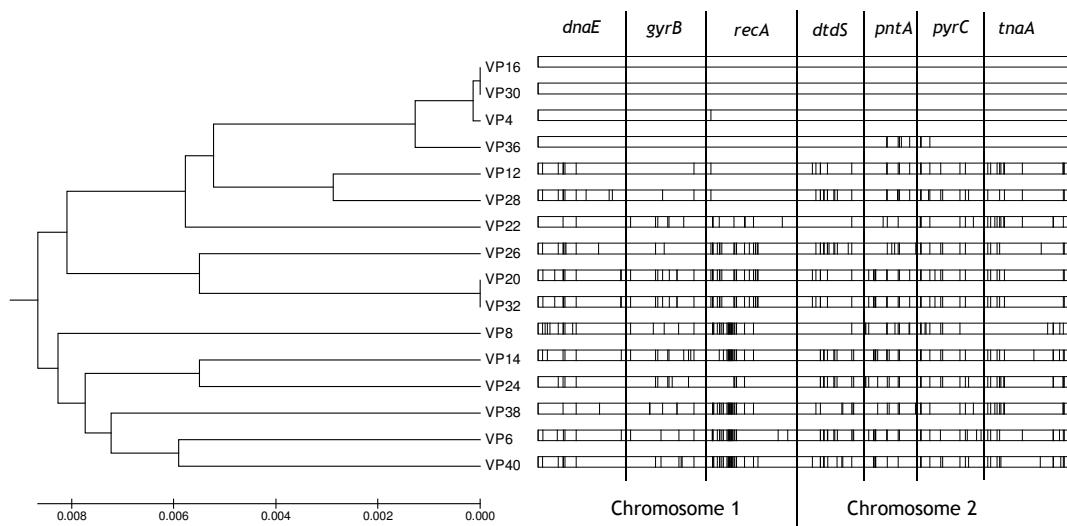


**Figure 2.60.** Neighbour-Joining tree and distribution of polymorphic nucleotide sites among concatenated sequences of housekeeping genes for 16 frozen shrimp isolates. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, VP16.
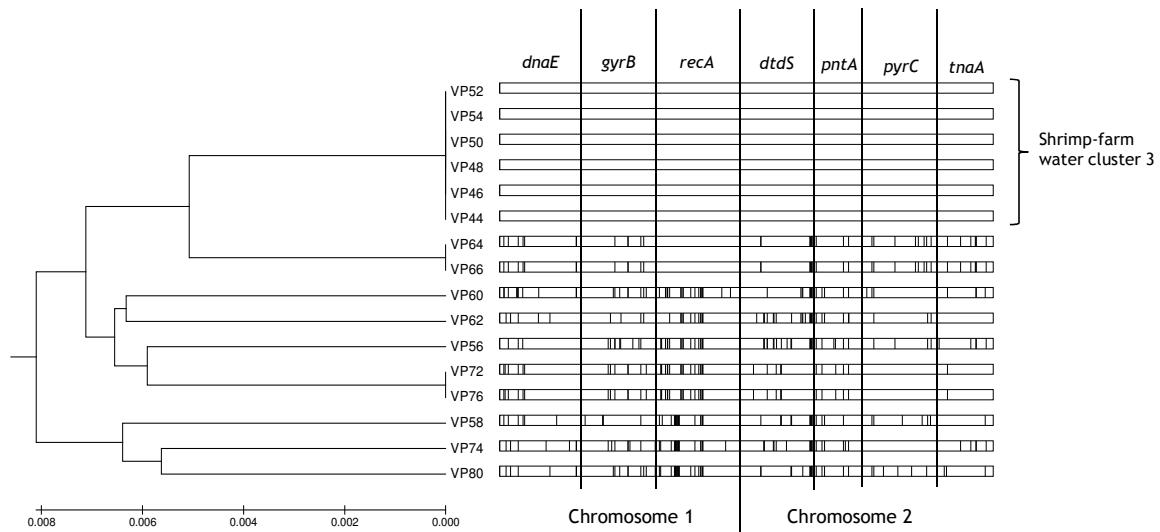
**Figure 2.61.** Neighbour-Joining tree and distribution of polymorphic nucleotide sites among concatenated sequences of housekeeping genes for 16 water isolates. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, VP52.

## 2.3.9 MLST analysis of V. parahaemolyticus isolates from Europe

### 2.3.9.1 Genotypic diversity

Details of European isolates used in the present study, including source, presence of haemolysin-encoding genes (*tdh and trh*), serotype, ST and allelic profile are presented in Table 2.2. A total of six STs were identified among nine isolates and two of these were novel. Novel ST346 and ST347 represent a clinical isolate (VP250) from Norway and an environmental isolate (VP262) from the UK, respectively. Three clinical isolates, VP252, VP258 and VP260, represent ST3, the ST that consists of pandemic *V. parahaemolyticus* O3:K6 and related strains of worldwide distribution (http://pubmlst.org/vparahaemolyticus/). These three European ST3 isolates also share the same O3:K6 serotype. Isolates VP252 and VP260 possessed *tdh* but not *trh* (*tdh⁺/trh⁻*), which is a typical haemolysin gene profile of pandemic strain serotype O3:K6 (Chen *et al.*, 2011). Isolate VP258

possesses *tdh* but no data were available for *trh*. The clinical isolate from Norway, VP254, was resolved to ST34 and this ST also contains other clinical Norwegian isolates that have been submitted to the MLST database (http://pubmlst.org/vparahaemolyticus/). Interestingly, the data from the MLST database show that a number of isolates recovered from seafood (oyster) in the USA also represent ST34. Furthermore, serotypic diversity was observed in the isolates representing ST34. Isolate VP254 and two environmental isolates from the USA are of serotype O4:K9, whereas a clinical isolate from Norway in the database is of serotype O3:KUT. Serotypic information for other 11 USA environmental strains representing ST34 is not provided in the MLST database. It is noteworthy that all isolates in ST34, including Norwegian clinical isolates from this study as well as the Norwegian clinical and the USA environmental isolates from the MLST database, share an identical haemolysin gene profile that is positive for both *tdh* and *trh* ($tdh^+/trh^+$).

Environmental isolates recovered from oyster in the UK, VP244 and VP246, represent ST79. This ST also includes environmental isolates from Norway and the Baltic Sea from the MLST database (http://pubmlst.org/ vparahaemolyticus/). However, haemolysin profiles of the isolates within ST79 are different. Both VP244 and VP246 possess *tdh* but not *trh* ($tdh^+/trh^-$), whereas environmental isolates from Norway and the Baltic Sea in the MLST database possess *trh* but not *tdh* ($tdh^-/trh^+$). Since serotypic data of VP246 and two environmental isolates from Norway and the Baltic Sea in MLST database are not available, it is not possible to assess serotypic diversity within this ST. However, these results do indicate a close genetic relationship between isolates from the UK and Norway since a number of these isolates were resolved to the same STs (ST3 and ST79).

Finally, the ST331 represents a clinical isolate from the UK (VP248) as well as a clinical isolate from China in the MLST database (http://pubmlst.org/vparahaemolyticus/). These two isolates also share an identical haemolysin gene profile, *tdh* positive and *trh* negative (*tdh*+/*trh*-).

**2.3.9.2 Phylogenetic analysis of European *V. parahaemolyticus* isolates**

The Neighbour-Joining tree of nine concatenated sequences of the European *V. parahaemolyticus* shows that the isolates are divided into two main clades I and II (Fig. 2.62). Clinical isolates representing ST3 (VP252, VP258, and VP260) and environmental isolates representing ST79 (VP244 and VP246) are resolved in clade I whereas clinical isolates representing ST34, ST346, and ST331 (VP254, VP250, and VP248) and environmental isolate representing ST347 (VP262) are resolved in clade II. Bootstrap scores of phylogenetic lineages representing clinical isolates (100) and environmental isolates (100) in clade I indicate a genetic distinction between clinical and environmental isolates in this group.
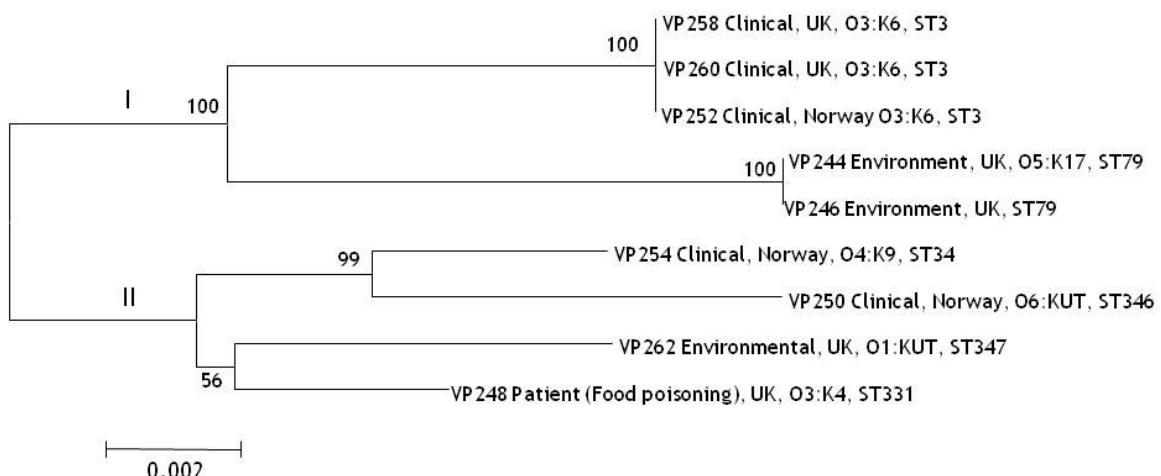


**Figure 2.62.** Neighbour-Joining tree of nine concatenated sequences of *V. parahaemolyticus* from multiple sources in European countries

## 2.3.9.3 Phylogenetic relationships of European and Thai *V. parahaemolyticus* isolates

The genetic relationship of *V. parahaemolyticus* isolated from European countries and Thailand is represented by the Neighbour-joining tree of 111 isolates including 101 Thai isolates, nine European isolates, and one Japanese type strain (VP2) (Fig. 2.63). The UK environmental isolates VP244 and 246 are closely related to an isolate from Thai human carrier, VP140, which represents ST62 and also contains a clinical strain from China (http://pubmlst.org/ vparahaemolyticus/). This result indicates a close genetic relationship among UK environmental strains, a Thai human carrier strain, and a Chinese clinical strain.

Isolates from Thai human carriers (VP132 and VP158) and European clinical isolates (VP252, VP258, and VP260) are resolved to ST3. The result confirms the genetic relatedness of Thai human carrier isolates ST3 and worldwide clinical isolates, including those from the UK and Norway in the present study. A clinical isolate from Norway, VP254, has a close genetic relationship to a Thai clinical isolate VP168. However, these two clinical isolates have a different haemolysin profile. VP254 possesses both *tdh* and *trh* ($tdh^+$/$trh^+$) whereas VP168 possesses only *trh*. An environmental isolate from the UK (VP262) has a close genetic relationship to a Thai human carrier isolate (VP154). Furthermore, the UK clinical isolate VP248 showed a close relationship to Thai human carrier isolate VP160. These results indicate that three Thai human carrier isolates (VP132, VP158 and VP160) have close genetic relationships to clinical isolates from the UK (VP258, VP260 and VP248) and Norway (VP252).
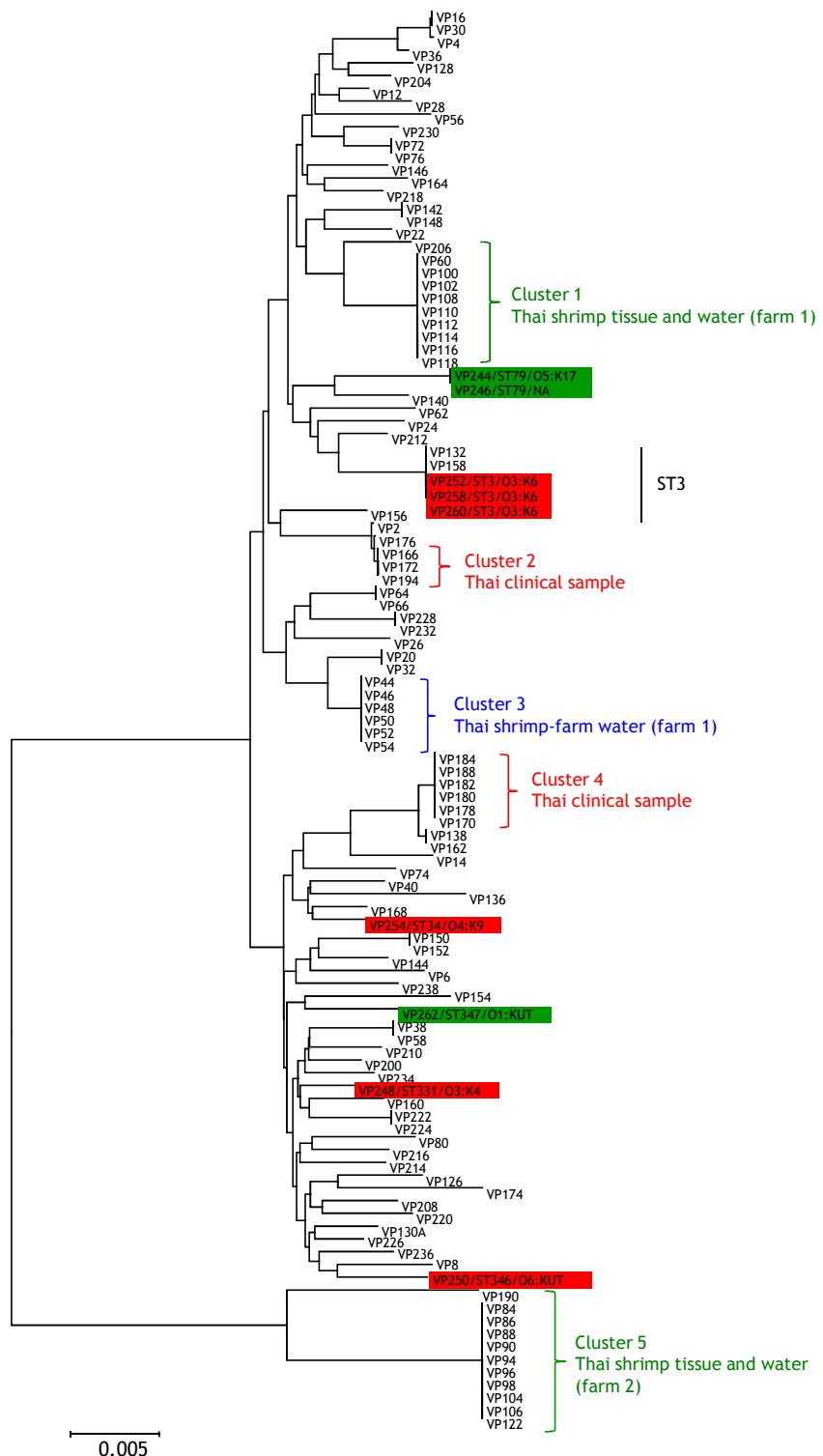
**Figure 2.63.** Neighbour-Joining tree of 111 concatenated sequences of *V. parahaemolyticus* from multiple sources in European countries, Thailand and Japanese type strain (VP2). Five genetic clusters of Thai isolates were indicted in the tree. Isolates with red highlight represent European clinical strains. Isolates with green highlight represent European environmental strains.

Similarly, another two Thai human carrier isolates (VP140 and VP154), are closely related to environmental isolates (VP244, VP246 and VP262) from the UK. Thus, Thai human carrier isolates are genetically related to both the clinical and environmental isolates from European countries. Furthermore, a close genetic relationship between a clinical isolate from Norway (VP250) and a Thai isolate from frozen shrimp (VP8) was demonstrated in this study. Beside the isolates representing ST3, none of the *V. parahaemolyticus* isolates from Europe in this study share the same ST with Thai isolates.

**2.3.9.4 Phylogenetic relationships of *V. parahaemolyticus* isolates from the MLST database in relation to European isolates in this study**

The Neighbour-Joining tree of 348 STs for all *V. parahaemolyticus* in the MLST database (http://pubmlst.org/vparahaemolyticus/) was constructed (Fig. 2.64). The positions of STs from European isolates in this study were identified in the phylogenetic tree. Seven STs representing European isolates were randomly distributed among 348 worldwide STs. Close genetic relationships between clinical and environmental isolates from Norway were also identified. Thus ST34, which represents clinical isolate VP254 from Norway in this study, is closely related to ST77, which represents environmental isolates from the same country in the MLST database. Moreover, ST346 representing a Norwegian clinical isolate (VP250) in this study is closely related to ST81, the ST which represents a clinical isolate also recovered from Norway in the MLST database. Also ST331, which represents an isolate causing food poisoning in the UK, is closely related to the clinical ST344 from China (ST344). These results indicate the genetic relatedness of clinical isolates from different geographical regions. The important ST3 that contains pandemic strains of serotype O3:K6, as well as the other worldwide

clinical isolates including VP252, VP258, and VP260 from this study, has a close genetic relationship with STs containing clinical (ST192) and environmental isolates (ST2, ST266, and ST305) from China in the MLST database.

Environmental isolates from the UK in this study have genetic similarity to isolates in the MLST database recovered from distant geographical regions. Thus, environmental isolate VP262 representing ST347 is closely related to an environmental isolate representing ST10 from Chile. Furthermore, environmental isolates VP244 and VP246 representing ST79 are closely related to an environmental isolate representing ST139 from the USA.
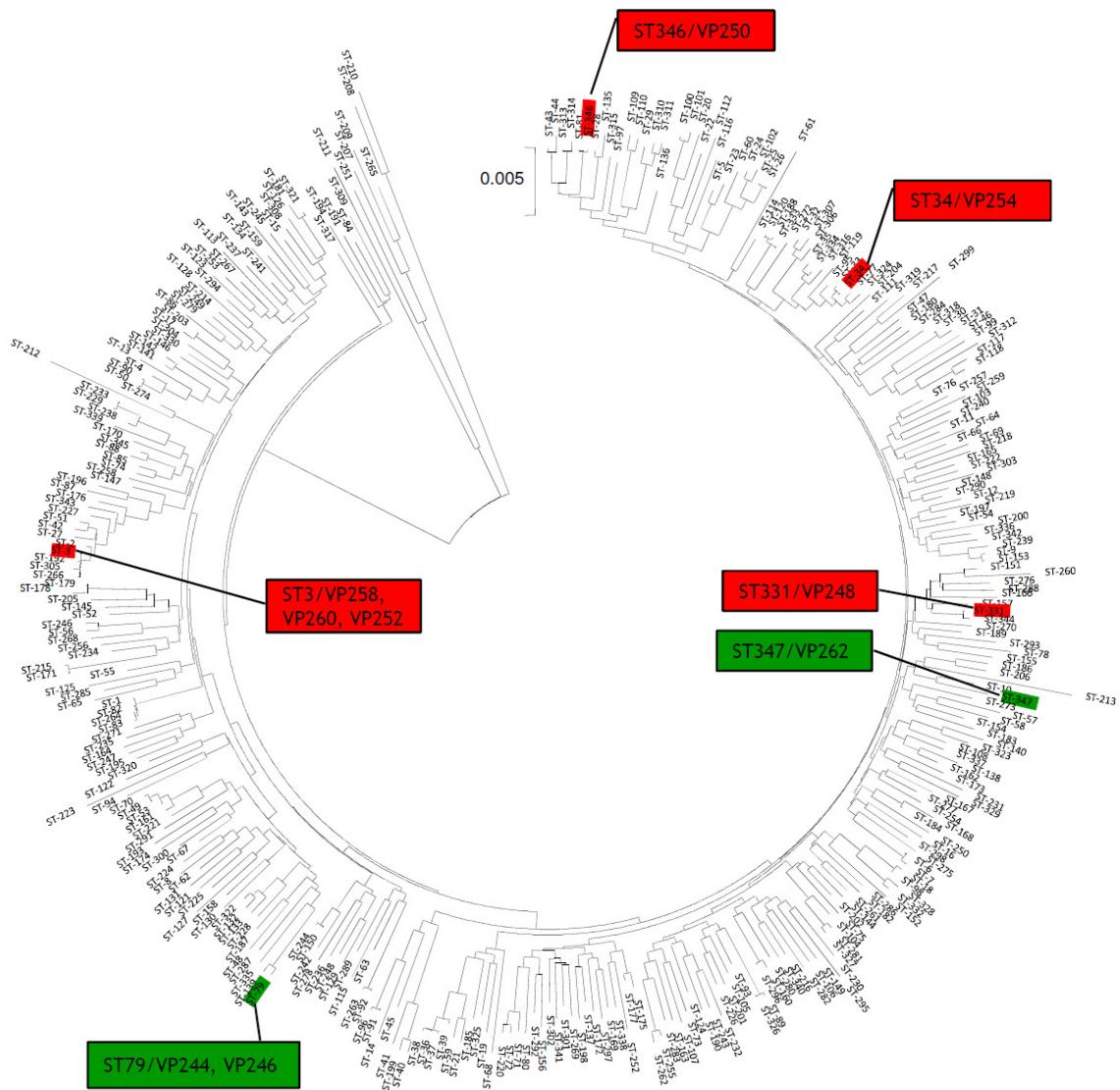
**Figure 2.64.** Circular phylogenetic tree of 348 *V. parahaemolyticus* STs from the MLST database (http://pubmlst.org/vparahaemolyticus/). Seven STs which represent isolates from European countries in this study are highlighted in red (clinical isolates ST3, ST34, ST331, and ST346) and green (environmental isolates ST79 and ST347).

## 2.4 Discussion

Evaluation of *Taq* polymerase kits from different manufacturers demonstrated variation in the quality of the PCR product. Among several types of *Taq* polymerase kits, enzymes with proofreading properties are preferred for PCR amplification in MLST studies because they yield more reliable sequencing results. However, evaluation of *Taq* polymerase enzyme kits is necessary in order to make a compromise between the high cost of proofreading *Taq* enzyme with the large number of required PCR reactions for a MLST study. PCR optimization depends on several parameters such as annealing temperature, magnesium concentration, DNA primers, etc (Gundry & Poulson, 2011). In the present study, there was difficulty in optimizing the PCR for each gene in all strains of *V. parahaemolyticus*. Among the seven housekeeping genes of *V. parahaemolyticus* used in this study, *recA* and *dtdS* were the most problematic for PCR optimization. The results indicated different levels of nucleotide variation for the individual housekeeping genes. Since *V. parahaemolyticus* is diverse organism inhabiting a wide range of environments, including different animal hosts, it is known to have a diverse genetic background. This may contribute to the large variation of housekeeping enzyme gene sequences in this organism. Thus, the housekeeping gene fragments of different strains could not be amplified by the same primer pairs.

In previous research, application of multiplex PCR has been used for the haemolysin genes (*tdh, trh* and *tl*) of *V. parahaemolyticus* (Garrido *et al.*, 2012; Jones *et al.*, 2012a; Rizvi & Bej, 2010; Wang *et al.*, 2011b). Multiplex PCR of housekeeping genes of *Vibrios* species (*gyrB* and *pntA)* was developed for differentiation between two species, for example between *V. parahaemolyticus*

and *V. alginolyticus* and between *V. chorelae* and *V. mimicus* (Teh *et al.*, 2009). There is no report to date of using multiplex PCR for MLST applications in *V. parahaemolyticus*. In the present study, the development of a multiplex PCR system for seven housekeeping genes of *V. parahaemolyticus* yielded successful DNA amplifications. However, the sequencing data obtained from a single PCR product were more consistent than data derived from the multiplex system. Multiple PCR products in the multiplex system may have affected the efficiency of the sequencing reactions, resulting in poor sequencing chromatograms. These results suggest that multiplex PCR of housekeeping enzyme genes for *V. parahaemolyticus* may be beneficial for nucleotide detection but not for a sequencing application.

The phylogenetic relationships of the concatenated sequences indicated that *V. parahaemolyticus* is highly diverse even though the isolates examined were recovered from a single country. Although a high level of recombination was detected amongst our *V. parahaemolyticus* isolates, it was not sufficiently high to lead to linkage equilibrium within the population. However, MLST analyses of the present dataset are in agreement with those of Gonzalez-Escalona *et al.* (González-Escalona *et al.*, 2008) who suggested that *V. parahaemolyticus* has an epidemic population structure where new alleles arise from a highly recombining background. In previous MLST studies of *C. coli* (Lang *et al.*, 2010) and *S. aureus* (Smyth *et al.*, 2009), a close relationship has been demonstrated between human isolates and sources of infection. Phylogenetic analysis of seven housekeeping genes of *S. enterica* serovar Newport also revealed an association of strains with host of origin, epidemiological source, and antibiotic resistance (Sangal *et al.*, 2010). However, in the present MLST study, we were unable to demonstrate a close genetic relatedness between clinical or human carrier

isolates and the epidemiological source of *V. parahaemolyticus*. These results suggest that virulent, and potentially virulent, strains of *V. parahaemolyticus* are not associated with particular sources of seafood production.

Interestingly, five isolates from human carriers represented STs (ST3, 62, 68, and 199) that were identical to those of isolates recovered from clinical sources of worldwide distribution. Although the pandemic ST3 was not detected among the Thai clinical isolates, this ST was represented among isolates from human carriage (VP132 and VP158). These findings provide some evidence of a genetic link between human carriage and clinical isolates. Due to a lack of medical records for individuals whose faecal samples were examined, we were unable to identify whether asymptomatic carriers had developed protective immunity to these strains or whether the carrier strains lacked virulence and the ability to cause disease. However, symptomless carriers could become sources of infection by transmitting potentially pathogenic bacteria to seafood products within the factory or directly to uninfected individuals. Furthermore, ten novel STs were associated with human carrier isolates but not with clinical and environmental isolates (Fig. 2.52). This suggests that the human intestinal tract provides a potential reservoir of unique *V. parahaemolyticus* isolates that are not commonly seen among clinical or environmental isolates. Mixed colonization with other micro-flora in the human intestinal tract may provide the opportunity for horizontal gene transfer to occur among different strains of *V. parahaemolyticus* or between *V. parahaemolyticus* and other species. Horizontal gene transfer of antibiotic resistance genes has been reported in *Enterococcus faecalis,* a commensal bacterium of the gastrointestinal tract of humans and other mammals (Haug *et al.*, 2011; Sparo *et al.*, 2012). Conjugative transposons also play an important role in horizontal gene transfer among the

*Enterobacteriaceae* (Pembroke *et al.*, 2002). Stecher *et al.* (Stecher *et al.*, 2012) demonstrated that conjugative gene transfer between *S. enterica* serovar Typhimurium and *E. coli* can be induced by an inflammatory response and a high density of *E. coli* in the mammalian gut. Therefore, it is highly likely that horizontal gene transfer occurs between *V. parahaemolyticus* strains inhabiting the human gut and other commensal bacteria.

Substantial serotypic diversity was observed among isolates within clusters 1, 3, 4, and 5 but not among isolates of cluster 2 (Fig. 2.52). All of the isolates in clinical cluster 2, with the exception of isolate VP194 (OUT:KUT), as well as the closely related carrier isolate VP156, were of serotype O1:K1. In contrast, clusters 1, 3, 4, and 5 comprise isolates of multiple serotypes. The three environmental clusters, 1, 3 and 5, contain isolates of seven, five, and eight combinations of O and K antigens, respectively, and clinical cluster 4 comprises isolates of serotypes O1:K69, O3:KUT, O4:K11 and O8:K22. These results clearly demonstrate a remarkably high degree of serotypic diversity among environmental *V. parahaemolyticus* isolates, in particular, since isolates in clusters 1, 3 and 5 represent a single ST and were recovered from the same source on the same day. The present study confirms previous findings that multiple serotypes of *V. parahaemolyticus* occur within a single ST, or closely related STs (Chao *et al.*, 2011; Chowdhury *et al.*, 2000, 2004; González-Escalona *et al.*, 2008; Yu *et al.*, 2011). In particular, extensive serotypic diversity has been described in ST3 which represents the pandemic O3:K6 strains as well as non-pandemic and environmental isolates of *V. parahaemolyticus* (Chao *et al.*, 2011; González-Escalona *et al.*, 2008; Yu *et al.*, 2011) (Fig. A17 in Appendix 3). Previous studies have demonstrated that the serotype O4:K68 strain has most likely evolved from the pandemic O3:K6 strain by replacement of both the O and

K antigens as a consequence of recombination events involving the entire O and K antigen-encoding gene clusters (Chen *et al.*, 2011; Okura *et al.*, 2008). In *V. cholerae*, serotype conversion by horizontal gene transfer has been suggested to play an important role in the evolution of *V. cholerae* O139 serotype pandemic strains from *V. cholerae* serotype O1 strains (Bik *et al.*, 1995; Stine *et al.*, 2000). Unlike other *Vibrio* species, including *V. cholerae* and *V. vulnificus*, which have a single region encoding both O and K antigens, the O antigen-encoding region of pandemic O3:K6 *V. parahaemolyticus* is not present in the same location as the K antigen gene cluster (Chen *et al.*, 2010). Therefore, *V. parahaemolyticus* is likely to have greater potential for a larger number of O and K antigen combinations than other *Vibrio* species as a consequence of genetic recombination of the O and K antigen-encoding genes. As suggested by Chen *et al.* (2011), recombinational exchange is the likely explanation for the serotypic diversity present within our *V. parahaemolyticus* isolates. However, the presence of numerous serotypes among isolates of the same ST and recovered from the same environmental source on the same day suggests that recombination is occurring at an exceptionally high rate. There was also evidence of horizontal gene transfer influencing the distribution of the haemolysin genes (*tdh* and *trh*) among isolates of the same, or closely related, STs. Clinical isolates within cluster 2 (ST1, ST83 and ST264) possessed two different haemolysin gene profiles (*tdh*$^+$/*trh*$^+$ and *tdh*$^+$/*trh*$^-$) whereas those within cluster 4 (ST262) possessed three different haemolysin gene profiles (*tdh*$^+$/*trh*$^+$, *tdh*$^+$/*trh*$^-$, and *tdh*$^-$/*trh*$^+$).

Bayesian analysis of *V. parahaemolyticus* isolates belonging to pandemic clonal complexes associated with South America and Asia identified two genetic clusters linked with the geographical history of the isolates in each group

(Ansede-Bermejo *et al.*, 2010). Although this method was unable to discriminate between isolates from different epidemiological sources in Thailand, Bayesian analysis did confirm that the population structure of Thai *V. parahaemolyticus* isolates is strongly influenced by *recA*. To support this finding, a phylogenetic analysis of all 348 STs within the *V. parahaemolyticus* MLST database (http://pubmlst.org/vparahaemolyticus/) identified three main clades that were resolved on the basis of having distinct *recA* sequences (data not shown). Clearly, *recA* has a major influence on the apparent phylogenetic relationships and population structure of *V. parahaemolyticus* based on the current MLST scheme. In the present study, we identified two highly divergent *recA* alleles, *recA107* and *recA120*, that have been acquired by horizontal DNA transfer by isolates representing STs 265 and 251, respectively. Nucleotide blast analysis of the *recA107* and *recA120* alleles were best matched to the *recA* sequences of *V. cincinnatiensis* (83% similarity) and *Vibrio halioticoli* (83%), respectively. *V. cincinnatiensis* has been identified as a human pathogenic bacterium (Brayton *et al.*, 1986) whereas *V. halioticoli* is commonly found in the gut of abalones (Sawabe *et al.*, 1998). It is clear, therefore, that certain *V. parahaemolyticus* strains have acquired highly divergent *recA* alleles by horizontal gene transfer from other *Vibrio* species. Furthermore, the evidence indicates that this has occurred on at least two occasions among clinical and environmental strains. Because *recA* sequences are more discriminatory than 16S rRNA, *recA* has been proposed to be an alternative identification marker in the family *Vibrionaceae* (Thompson *et al.*, 2004). However, incoherence of the phylogenetic tree using *recA* sequences of *V. harveyi* and *Vibrio campbellii* was observed by Thompson *et al.* (Thompson *et al.*, 2007), suggesting that *recA* is unreliable for use as a marker for *Vibrio* species discrimination. High *recA* diversity within *Vibrio* species has also been reported in a number of previous MLST studies (Chowdhury

*et al.*, 2004; González-Escalona *et al.*, 2008; Thompson *et al.*, 2008). Highly divergent *recA* alleles and evidence for frequent recombination at this locus were also observed in previous MLST studies of *V. parahaemolyticus* isolates from the southeastern Chinese coast (Yu *et al.*, 2011) and the Chinese mainland (Chao *et al.*, 2011), respectively. Based on our current analysis of Thai *V. parahaemolyticus* isolates, together with the findings of previous studies, *recA* is clearly not an ideal molecular marker for evolutionary analyses of *V. parahaemolyticus* and other *Vibrio* species.

In addition to potential assortative recombination events involving the entire *recA* gene from other species, intragenic recombination has also played a significant role in *V. parahaemolyticus* evolution since *recA* itself has a complex mosaic structure (Fig. 2.54). The *recA60, recA116, recA17, recA36, recA123, recA27, recA126, recA48, recA88,* and *recA19* alleles have complex combinations of the internal segments A, B, and C indicative of multiple intragenic recombination events. Significantly, these mosaic alleles were present predominantly in clinical or human carrier isolates, suggesting that recombinational exchange has occurred more frequently in the human intestinal tract than in the environment. These findings suggest that the carriage of different strains of *V. parahaemolyticus* within the human intestinal tract is acting as a driving force in the evolution and emergence of new strains of this pathogen; the intestinal tract is providing an environment which stimulates the occurrence of genetic recombination between different strains and species of the genus *Vibrionaceae* (Haley *et al.*, 2010; Okada *et al.*, 2009; Ruwandeepika *et al.*, 2010; Wang *et al.*, 2011a). Notably, intragenic recombination of *recA* has also been reported in *V. cholerae* and *V. mimicus* (Byun *et al.*, 1999; Thompson *et al.*, 2008), suggesting that this gene may represent a hot-spot of

recombination in *Vibrio* species. However, evidence has also been presented to suggest that hybridization between non-virulent environmental strains can lead to the emergence of virulent strains. In this case, the clinical ST263 comprised genotypic fragments of housekeeping genes derived from isolates recovered from water (ST239) and frozen shrimp (ST235) (Table 2.8). The phenomenon that hybrid variants are apparently more pathogenic than existing strains has also been described in *V. vulnificus* (Bisharat *et al.*, 2005). Bayesian analysis demonstrated that *V. vulnificus* biotype 3, an emerging pandemic strain that was detected in Israel in 1996, and has evolved as a consequence of genetic hybridization between existing biotype 1 and 2 strains. Strains of biotype 3 are apparently more pathogenic than those of biotypes 1 and 2. Thus, the present study has demonstrated the possibility that virulent *V. parahaemolyticus* strains could emerge from a background of non-virulent strains by genetic recombination, together with acquisition of virulence genes such as *tdh* and *trh*.

The *dN*/*dS* ratios were calculated separately for genes from the two chromosomes in order to evaluate the degree of purifying selection within each chromosome. It has been documented that chromosome I of *V. parahaemolyticus* contains genes predominantly encoding proteins for basic cell function, whereas chromosome II possesses genes predominantly responsible for bacterial adaptation to environmental change and associated with a pathogenicity island (Makino *et al.*, 2003). Previous research has demonstrated that purifying selection is relatively stronger in chromosome I of *V. parahaemolyticus* compared to chromosome II (Cooper *et al.*, 2010). However, in the present study, the *dN*/*dS* ratio of chromosome II genes was less than that of chromosome I genes, indicating that purifying selection is stronger in chromosome II. This can be explained by a slow purging process of non-synonymous substitutions in

bacterial core genes (Castillo-Ramírez *et al.*, 2011). The period over which the isolates were recovered in this study (ten years) may not be sufficient to allow non-synonymous substitutions in chromosome I to be deleted from the gene pool. Although Cooper *et al.* (2010) suggested that chromosome I has been subject to stronger purifying selection than chromosome II, the existence of the remaining non-synonymous substitutions in the gene pool of chromosome I may result in the higher *dN/dS* ratio compared to that of chromosome II. A lower *dN/dS* ratio in chromosome I compared to chromosome II has been reported for *V. parahaemolyticus* in the previous MLST studies (González-Escalona *et al.*, 2008; Yan *et al.*, 2011), although it remains possible that more striking differences are apparent when larger samples of genes are considered.

In this study, the genetic relatedness of *V. parahaemolyticus* isolated from Thailand was compared with isolates originating from the UK and Norway by MLST. It is interesting that clinical isolates from the UK and Norway (VP248 and VP250, respectively) are closely related to a Thai human carrier isolate (VP160) and a Thai environmental isolate (VP8), respectively. The Thai environmental strain (VP8) was recovered from a commercial frozen shrimp source in a seafood processing factory. These results indicate genetic relatedness between European clinical strains and environmental strains isolated from seafood and a worker who was involved in a seafood factory in Thailand. This study suggests that awareness of pathogenic *V. parahaemolyticus* in exported seafood products from Thailand should be raised, in order to maintain a high standard appropriate to a world class seafood-exporting country. Future research should include a study of the conditions that could trigger the emergence of virulent strains throughout the domestic and international seafood supply chain, in order to avoid potential infection by *V. parahaemolyticus* from seafood products in worldwide markets.

Furthermore, the isolate recovered from Chinese mitten crab in the river Thames, UK, represents the novel ST347. The Chinese mitten crab is a native Asian species that has been introduced to the UK and elsewhere and has significant effects on the ecological systems of European countries particularly the UK and Germany (Bentley, 2011). A previous study found that a Chinese mitten crab from the river Thames harboured *V. parahaemolyticus* possessing the virulence gene *tdh* (Wagley *et al.*, 2009). Together with the results of the present study, the Chinese mitten crab not only has an impact on estuarine water ecology but also is capable of introducing new strains of potentially pathogenic *V. parahaemolyticus* to European countries including the UK.

In conclusion, MLST analyses of Thai *V. parahaemolyticus* isolates from different sources indicated that clinical strains are unrelated to those recovered from seafood and other environmental sources. In addition, the majority of STs represented by human carrier isolates are novel and are not associated with clinical or environmental isolates. These findings suggest that the human intestinal tract serves as a potential reservoir of *V. parahaemolyticus* strains that are mostly different to those commonly associated with infection and environmental sources. However, a small number of STs associated with human carrier isolates were genetically related to clinical strains from both Thailand and from worldwide sources, suggesting that a limited number of pathogenic phenotypes may be positively selected for within the human population. Very high levels of serotypic diversity, presumably due to recombinational exchange, were observed among isolates representing the same ST and recovered from a single source at the same period of time. Extensive recombination was also observed to be affecting the *recA* locus, particularly within clinical and carrier isolates. The preponderance of a large number of mosaic *recA* alleles in clinical

and carrier isolates suggests that many of these horizontal DNA transfer events are occurring within the human intestinal tract. Recombinational exchange clearly plays an important role in the evolution of *V. parahaemolyticus* but the human intestinal tract provides an environment that appears to be driving the emergence of new potentially pathogenic strains.

# 3. DISTRIBUTION AND MOLECULAR EVOLUTIONARY RELATIONSHIPS OF HAEMOLYSIN AND TYPE III SECRETION SYSTEM 1 GENES AMONG *V. PARAHAEMOLYTICUS*

## 3.1 Introduction

Haemolysin genes (*tdh*A, *tdhS, trh1* and *trh2)* and type III secretion system genes are known to encode virulence determinants of *V. parahaemolyticus* as has been described in Chapter 1. In KP-positive *V. parahaemolyticus* strains*, tdhA* and *tdhS* are responsible for haemolytic activity and both of these genes are generally found in single *V. parahaemolyticus* isolates  (Nishibuchi & Kaper, 1990). The *tdhA* and *tdhS* genes are located within VPaI-7 in chromosome 2 (Fig. 3.1A) (Makino *et al.,* 2003). In the VPaI-7 gene cluster, *tdhA* and *tdhS* are surrounded by other virulence-related genes including type three secretion system 2 and *toxR,* the TDH-encoding gene regulator. Genes involved in mobile genetic elements such as integrase- and transposase- encoding genes are also present in a gene cluster containing *tdhA* and *tdhS*. Although the nucleotide sequences of *tdhA* and *tdhS* are very similar (97.2%), *tdhA* is the structural gene mainly responsible (97.0%) for TDH production in *V. parahaemolyticus*, while *tdhS* contributes to this process to only a small extent (0.5-9.4%) (Iida & Yamamoto, 1990; Nishibuchi & Kaper, 1990). For this reason primers specific for *tdhA* have generally been used for *tdh* gene detection in *V. parahaemolyticus* isolates (Tada *et al.*, 1992).

(A)



(B)
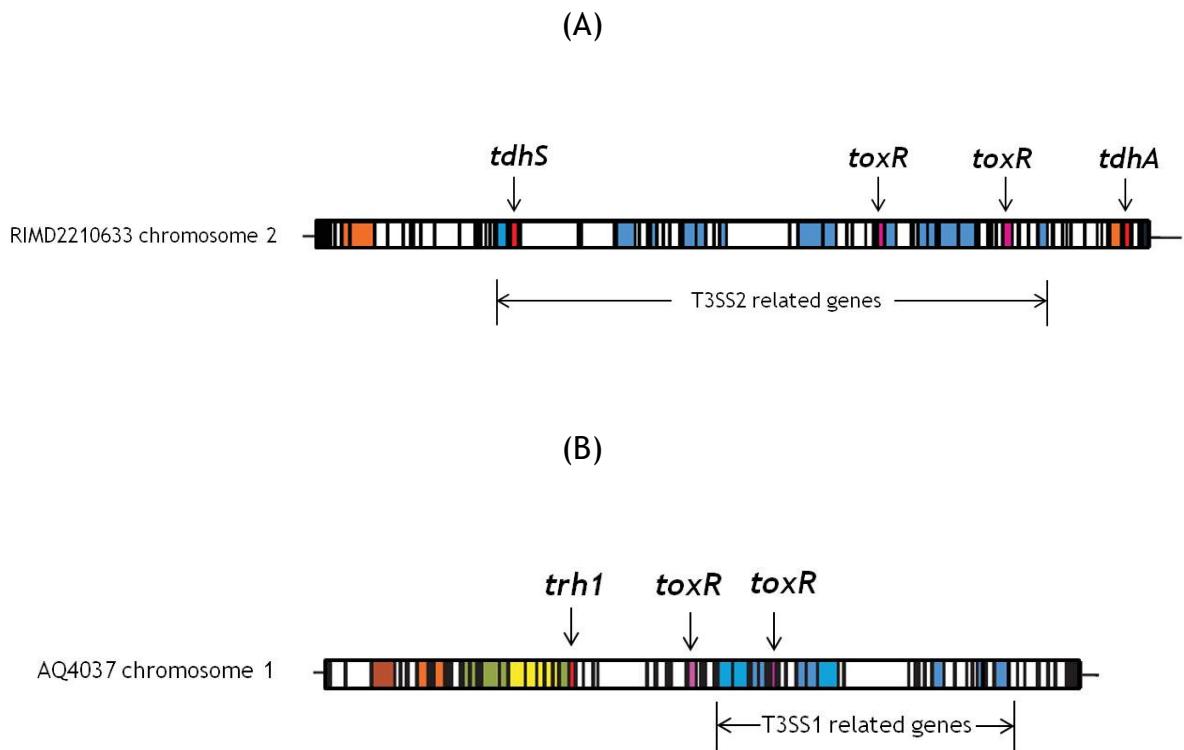


**Figure 3.1.** Gene clusters in the *V. parahaemolyticus* pathogenicity islands (VPaIs) (A) shows locations of *tdhA* and *tdhS* in chromosome 2 of strain RIMD2210633 (*tdh+/trh-*) and (B) shows location of *trh1* in chromosome 1 of strain AQ4037 (*tdh-/trh+*). Boxes indicate origins of replication (ORFs). Colours represent various functional categories including *tdh* and *trh* (red), TTSS-related genes (blue), *toxR* (pink), integrase (brown), transposase (orange), urease-encoding genes (green), nickel-peptide transport-encoding genes (yellow). Open boxes represent genes of other functions. The diagram was adapted from Chen *et al*. (2011).

The *tdh*-related haemolysin (*trh*) is responsible for haemolytic activity of Kanagawa-negative (KP⁻) *V. parahaemolyticus* strains and is known to produce TDH-related haemolysin (TRH) (Honda *et al.*, 1988, 1991). Another haemolysin that is similar to TRH was subsequently identified and characterized by Kishishita *et al.* (1992). The previously identified *trh* was renamed *trh1* and the new haemolysin was designated *trh2* (Kishishita *et al.*, 1992). The *trh1* and *trh2* genes are located in the VPaI of chromosome 1 [Fig.3.1B] (Chen *et al.*, 2011). The similarity of *trh1* and *trh2* (84.0%) is less than that of *tdhA* and *tdhS* (97.2%). Unlike *tdhA* and *tdhS*, a previous study showed that individual *V. parahaemolyticus* isolates are likely to contain either *trh1* or *trh2*, but not both genes (Kishishita *et al.*, 1992). These authors also demonstrated that haemolytic activity caused by *trh2* is weaker than that caused by *trh1*. Amplification primers used for detection of the original *trh* (*trh1*) have been used for general *trh* detection in *V. parahaemolyticus* isolates (Honda *et al.*, 1991; Kishishita *et al.*, 1992).

Type III secretion systems 1 (TTSS1) and 2 (TTSS2) of *V. parahaemolyticus* are responsible for cytotoxicity and enterotoxicity to host cells, respectively (Hiyoshi *et al.*, 2010; Makino *et al.*, 2003; Park *et al.*, 2004). *V. parahaemolyticus* TTSS1 is located in chromosome 1 whereas the TTSS2 gene clusters are located within the VPaI-7 of chromosome 2. (Fig. 3.1) (Makino *et al.*, 2003). TTSS1 is commonly present in pathogenic and non-pathogenic strains of *V. parahaemolyticus*. In contrast, TTSS2 is predominantly recovered from pathogenic strains (Makino *et al.*, 2003). The relatively low G+C content of TTSS2 (39.8%) indicates that this region has been acquired by horizontal gene transfer. In contrast, TTSS1 has an average G+C content (45.4%) which is similar to that of the entire *V. parahaemolyticus* genome, indicating that this region is ancestral and was not

acquired by a recent horizontal DNA transfer event (Makino *et al.*, 2003). Since both pathogenic and non-pathogenic *V. parahaemolyticus* isolates contain TTSS1, analysis of nucleotide sequence variation of TTSS1 from both clinical and environmental isolates will enable us to understand the molecular evolution of this virulence factor in *V. parahaemolyticus.*

The organization of the *V. parahaemolyticus* TTSS1 gene cluster was proposed by Ono *et al*. (2006) and is shown in Fig. 3.2. Among these TTSS1 genes, *vcrD1, vscC1,* and *VP1680* play important roles in an effective TTSS1 and are evenly distributed among the TTSS1 operons of *V. parahaemolyticus.* The TTSS1 genes *vcrD1* and *vscC1* are most similar to *lcrD* and *yscC* of *Yersinia spp.*, respectively (Makino *et al.*, 2003; Park *et al.*, 2004).

The LcrD-family proteins are located in the inner membrane and are also involved in the expression of other TTSS-secreted proteins such as the Yop family (Hueck, 1998; Plano & Straley, 1993). The YscC-family proteins are pore-forming outer membrane proteins that are involved in the virulent phenotype caused by TTSS (Haddix & Straley, 1992; Hueck, 1998; Ochman *et al.*, 1996; Plano & Straley, 1995). The VcrD1 and VscC1 proteins of *V. parahaemolyticus* are involved in cytotoxic activity towards HeLa cells (Hiyoshi *et al.*, 2010; Ono *et al.*, 2006; Park *et al.*, 2004). The VP1680 protein was identified to be a *V. parahaemolyticus* effector protein that is responsible for cytotoxicity and capable of inducing acute apoptosis in HeLa cells (Ono *et al.*, 2006). Furthermore, VP1680 also plays an important role in stimulating inflammation in Caco-2 cells by induction of interleukin (IL)-8 (Shimohata *et al.*, 2011).
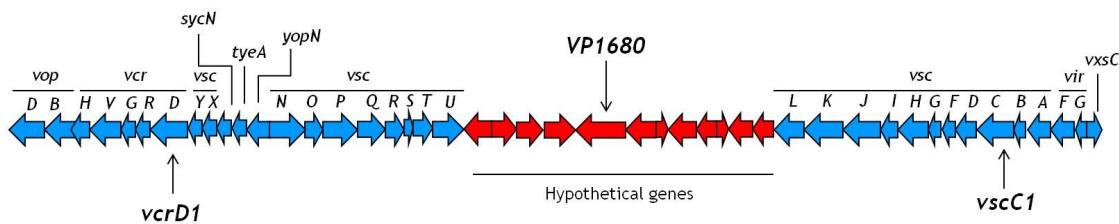
**Figure 3.2.** Organization of the *V. parahaemolyticus* TTSS1 gene cluster in chromosome 1. Blue represents the TTSS apparatus genes that are similar to those of *Yersinia spp*. The *vcrD1* and *vscC1* loci are present within these TTSS apparatus regions. Red represents hypothetical genes including VP1680 that was subsequently identified to be a TTSS1 effector protein. The figure is adapted from Ono *et al*. (2004).

Utilization of bioinformatic approaches to interrogate the virulence gene sequences from *V. parahaemolyticus* strains for which the genome sequences are available (http://www.ncbi.nlm.nih.gov/) makes it possible to study the evolution of virulence genes in a wide range of isolates, including those of the present study and others from other parts of the world. The genome sequences of seven pathogenic *V. parahaemolyticus* isolates, AQ3810, AQ4037, RIMD2210633, Peru466, AN5034, 10329 and K5030, were established by previous studies (Chen *et al.*, 2011; Gonzalez-Escalona *et al.*, 2011). These seven pathogenic isolates were recovered from different geographical regions at different periods of time and possess different haemolysin profiles. Two isolates, AQ3810 (O3:K6, *tdh+/trh-*) and AQ4037 (O3:K6, *tdh-/trh+*), were recovered prior to the Indian pandemic in 1996 and are recognized as pre-pandemic strains. Three isolates, RIMD2210633 (O3:K6, *tdh+/trh-*), Peru466 (O3:K6, *tdh+/trh-*), and AN5034 (O4:K68, *tdh+/trh-*), were recovered during 1996-1998, the period during which the virulent *V. parahaemolyticus* serotype O3:K6 strain emerged originally from India in 1996, and were subsequently isolated from other parts of

the world. These virulent serotype O3:K6 strains (RIMD2210633, Peru466, and AN5034) are recognized as pandemic strains. Isolate 10329 (O4:K12, *tdh+/trh+*) was recovered in 1998 and is recognized as a pathogenic isolate with low infectivity. Isolate K5030 (O3:K6, *tdh+/trh-*) was recovered in 2005 and is recognized as a post-pandemic strain. However, nucleotide sequence diversity of TTSS1-related genes in pre-pandemic, pandemic, and post-pandemic isolates, and in isolates from different sources, including clinical, human carrier, and seafood, remain to be studied.

The haemolysin-encoding genes *tdhA, tdhS, trh1* and *trh2* were selected in the present study to establish the evolutionary relationships of selected virulence genes. The TTSS1 genes *vcrD1, vscC1,* and *VP1680* were also selected because they are capable of causing pathogenicity and can be found in both clinical and environmental isolates. Since the *vcrD1, vscC1* and *VP1680* genes are evenly distributed among TTSS1 operons (Fig. 3.2), the presence of these genes may imply the existence of the entire TTSS1 operon in *V. parahaemolyticus* isolates. In the present study, comparative nucleotide sequence analysis of *vcrD1, vscC1* and *VP1680* gene fragments within seven worldwide pathogenic strains, five Thai isolates from clinical, human carrier, and seafood sources, and a Japanese type strain causing food poisoning, were determined by phylogenetic analysis. Seven worldwide pathogenic isolates were selected due to their epidemiological patterns, which include pre-pandemic isolates (AQ3810 and AQ4037), pandemic isolates (RIMD2210633, Peru466 and AN5034), post-pandemic isolates (K5030), and a pathogenic isolate with low infectivity (10329).

The objective of this chapter was to study the distribution and sequence variation of the virulence genes *tdhA, trh1, trh2, vcrD1, vscC1* and *VP1680*

among Thai *V. parahaemolyticus* isolates from different sources, including clinical samples, human carriers, seafood, and water. Detection of *tdhS* was examined in selected isolates that possessed *tdhA*, and nucleotide sequences of both *tdhA* and *tdhS* fragments in certain isolates were analyzed. Furthermore, nucleotide sequences of the TTSS1 genes in isolates from Thailand were compared to those of clinical strains of worldwide distribution. This knowledge will establish whether environmental isolates from Thailand can serve as reservoirs of these virulence determinant-encoding genes and will lead to a better understanding of the genetic diversity of these genes in *V. parahaemolyticus.*

## 3.2 Materials and methods

### 3.2.1 Bacterial strains and growth conditions

One hundred and two *V. parahaemolyticus* isolates were used for the characterization of virulence genes (Table 2.1). The bacterial culture conditions have been described in Chapter 2.

### 3.2.2 Preparation of chromosomal DNA

Chromosomal DNA was prepared by the method described in Chapter 2. Prepared genomic DNA of 102 *V. parahaemolyticus* isolates was stored at -20$^{\circ}$C.

### 3.2.3 Primer design and PCR amplifications of virulence genes tdhA, tdhS, trh1, trh2, vcrD1, vscC1, and VP1680

The primers used for DNA amplification and sequencing of the *tdhA* gene fragment were the same as those described by Nishibuchi *et al* (1985). The

primers used for DNA amplification and sequencing of the *tdhS, trh1, trh2, vcrD1, vscC1,* and *VP1680* gene fragments were newly designed using Primer Designer version 2 (Scientific and Educational software). The nucleotide sequences of these primers are shown in Table 3.1. The locations of these primers in gene sequences are shown by CLC Genomics Workbench version 3.7.1 (Figs. 3.3-3.7). The primers for *tdhS* gene fragments were designed based on specific nucleotide regions of the *tdhS* gene sequence from *V. parahaemolyticus* isolate RIMD2210633 (Fig. 3.3). The primers for the *trh1* and *trh2* gene fragments were designed based on specific nucleotide regions of the *trh1* and *trh2* gene sequences of *V. parahaemolyticus* strains GCSL28 and M88112, respectively (Fig. 3.4). The primers used for DNA amplification and sequencing of the *vcrD1, vscC1,* and *VP1680* genes were designed based on gene sequences of *V. parahaemolyticus* strain RIMD2210633 (Figs. 3.5-3.7). The primers were diluted to 12.5 pmol $\mu l^{-1}$ for PCR reactions and 2 pmol $\mu l^{-1}$ for sequencing reactions. PCR fragments containing partial segments of *tdhA* (382 bp), *tdhS* (368 bp), *trh1* (345 bp), *trh2* (431 bp), *vcrD1* (1417 bp), *vscC1* (1094 bp), and *VP1680* (716 bp) were amplified from chromosomal DNA by using a *Taq* polymerase kit (*Pfx*, Invitrogen) according to the manufacturer's instructions. PCRs were carried out in a GeneAmp PCR System 9700 Thermo Cycler (Applied Biosystems) using 30 cycles of the following amplification parameters: denaturation at 94°C for 45 s, annealing at 55°C for *tdhA, tdhS, trh1,* and *trh2,* 50°C for *vcrD and vscC1* and 54°C for *VP1680,* for 45 s, and extension at 72°C for 2 min. An initial denaturation step of 94°C for 2 min was used and a final extension step at 72°C for 10 min. However, in some cases of *VP1680* amplification, improved results were obtained by decreasing the annealing temperature to 52°C. The expected size of the PCR products was confirmed by electrophoresis in a 1% w/v agarose gel incorporating 0.004% (v/v) SybrSafe

(Invitrogen). DNA was purified with a Qiaquick PCR purification kit (Qiagen) and finally eluted in 30-50 µl sterile distilled $H_2O$ and stored at -20$^{o}$C.

**Table 3.1.** Nucleotide sequences of PCR and sequencing primers designed for DNA amplification and sequencing of the haemolysin genes *tdhA, tdhS, trh1* and *trh2* and the TTSS1-associated genes *vcrD1, vscC1* and *VP1680* of *V. parahaemolyticus*.

| Gene | Primer | Primer position | Primer direction | Size (bp) | Sequence (5'-3') |
|------|--------|-----------------|------------------|-----------|------------------|
| *tdhA* | tdhA-F | 6-23 | Forward | 382 | GTA CCG ATA TTT TGC AAA |
| | tdhA-R | 387-369 | Reverse | | ATG TTG AAG CTG TAC TTG A |
| *tdhS* | tdhS-F | 65-83 | Forward | 368 | CAT CTG CTT TTG AGC TTC C |
| | tdhS-R | 432-415 | Reverse | | AGA ACC TTC ATC TTC ACC |
| *trh1* | trh1-F | 157-175 | Forward | 345 | CAC CAG TTA ACG CAA TCG |
| | trh1-R | 500-483 | Reverse | | TCC GCT CTC ATA TGC TTC |
| *trh2* | trh2-F | 65-82 | Forward | 431 | CAT CGC GA TTG ATC TGC |
| | trh2-R | 495-478 | Reverse | | CTC ATA TGC CTC GAC AGT |
| *vcrD1* | vcrD-F | 53-70 | Forward | 1417 | AAG ACA TCA TGC TCG CAG |
| | vcrD-R | 1469-1452 | Reverse | | CTC GCA GTC AAC TTC TCT |
| *vscC1* | vscC1-F | 107-124 | Forward | 1094 | AGC TCA ATT GGC CAG AAC |
| | vscC1-R | 1200-1183 | Reverse | | TAG CAC CGC TTC GAC GTT |
| *VP1680* | VP1680-F | 151-168 | Forward | 716 | TCG GTT AGC GAA GGC GTA |
| | VP1680-R | 866-849 | Reverse | | CCG CTG ATA ATG CCA GTA |
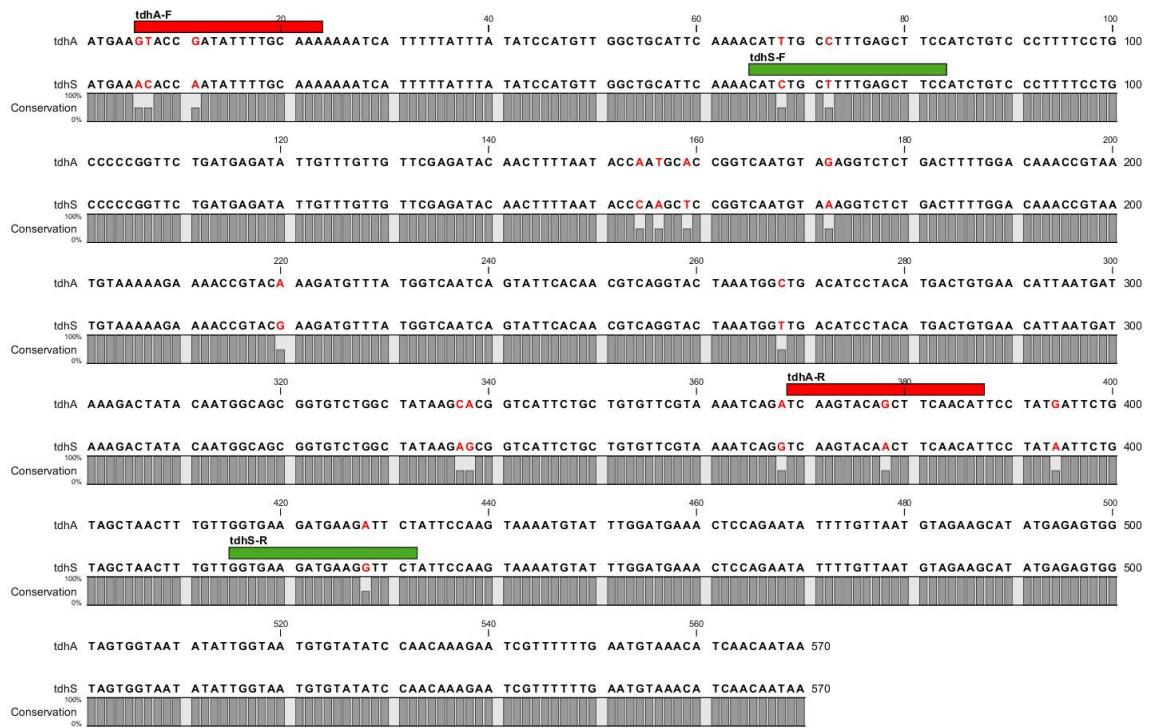
**Figure 3.3.** Nucleotide sequences (5' → 3') of the *tdhA* and *tdhS* genes of *V. parahaemolyticus* isolate RIMD2210633 showing the positions of specific forward and reverse primers of *tdhA* and *tdhS* gene fragments. Red represents the primers for the *tdhA* gene fragment and green represents the primers for the *tdhS* gene fragment. The nucleotide sequences were obtained from http://www.ncbi.nlm.nih.gov/genbank.
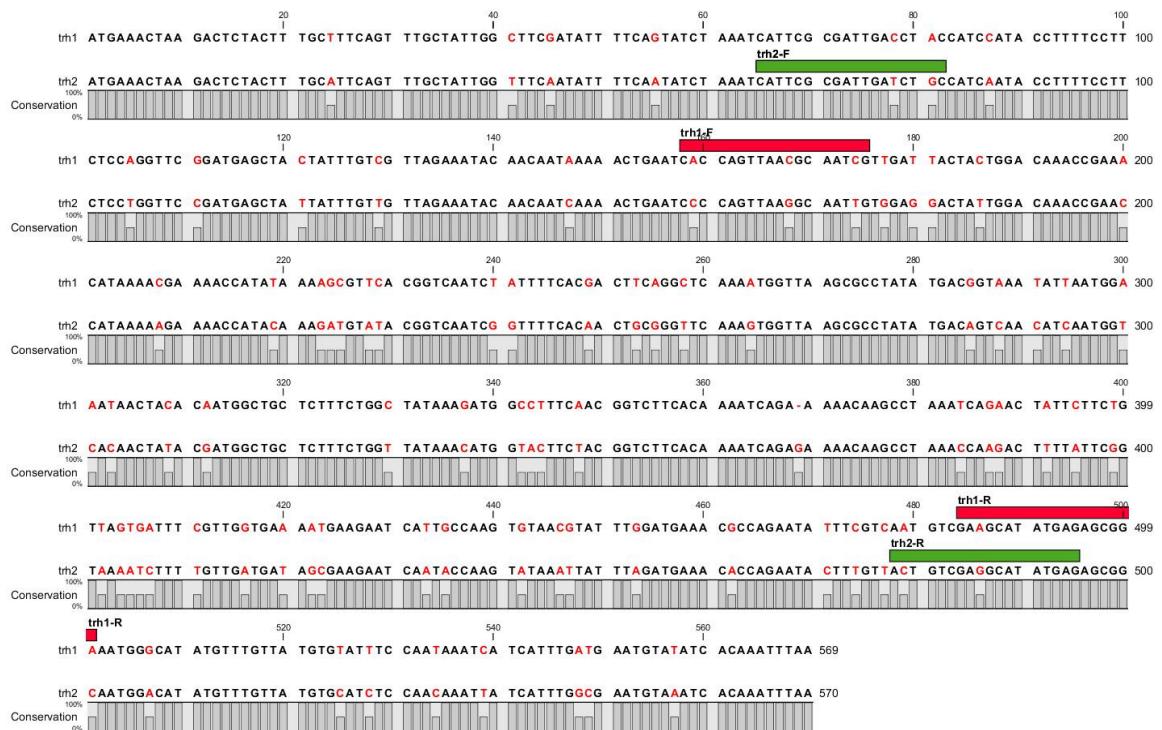
**Figure 3.4.** Nucleotide sequences (5' → 3') of the *trh1* gene of *V. parahaemolyticus* isolate GCSL28 and *trh2* gene from *V. parahaemolyticus* isolate M88112 showing positions of specific forward and reverse primers. Red represents the primers for the *trh1* gene fragment and green represents the primers for the *trh2* gene fragment. The nucleotide sequences were obtained from http://www.ncbi.nlm.nih.gov/genbank.

**Figure 3.5.** Nucleotide sequence (5' → 3') of the *vcrD1* gene of *V. parahaemolyticus* isolate RIMD2210633 showing the positions of forward and reverse primers. The nucleotide sequence was obtained from http://www.ncbi.nlm.nih.gov/genbank.

vscC1-F

vscC1 GTGGTTACCGTAATGCGAACTCTCATGCCAAAAATAGGGCGCATTGCCGCCAAAATGACATTGTGCGCACTGTGTGTCGCTCCGATGTTTTCTGTACAAGCGACAGAGCTCAAT

vscC1-F 120

vscC1 TGGCCAGAACAACCGTTTCGTTACTATGCCGATAACGATTCGCTGAAAGATCTGCTTAACAACTTTGGTGCCAACTATCGAGTATCCGTGTCTGTGAGCGACAAAGTGAATGAC

vscC1 CGAGTGAGTGGTCGATTCACGCCAGAAGATCCAGCGGAATTTCTCGACTACCTAGCGCAAGTTTACAACTTGATGTGGTACTTCGATGGCGCAGTACTGCACGTTTACAAAGCG

vscC1 ACAGAAACGCGCTCCCGATTGCTGCAATTGGAACTGCTTACCGCGCGCGAGCTTCGCAGTACCTTGATTTCAACCGGTGTTTGGGATGCTCGTTATGGTTGGCGCGCTGCCGAA

vscC1 AATAAAGGGCTGGTATATCTAGCTGGGCCGCCGCGTTACGTTGAGTTGGTGGTGCAAACCGCAGAAGCGTTAGAGAGCCGATTGCTGCAAAAGTCCAACTCGACGGATGAGCTG

vscC1 TTTGTGGAACTCATTCCGCTTAAATATGCCTCGGCGACGGACAGAAGTATCTCGTATCGCGACCAAAGCATTACCGTTCCGGGGATTGCTTCGGTATTGAGCCGTGTGGTTGGC

vscC1 GGTGTTCAAACTCAAATTACGGATTCTGCGTCTGTACAAACATCGTCTGTAAATGGATTACCCGCAGAAGCAGCGAAACCTCGTGGCAAAACTGCGTCTGTGCACGGCGGTGCA

vscC1 ACCGTAGAAGCGGAGCCGGGGCTGAATGCGATCATCGTTCGCGCGATACCCAAGCAAGACTGCCGTTGTATCGCAAGCTGGTCGCTCAACTCGATCAGCCACAATCTCGTATCGAG

vscC1 GTTGCGTTATCTATTGTGGACATCAGCGCAAACGATTTACGTCAGCTCGGTGTGGATTGGCGTGCAGGTGTATCGGTTGGAAACAACCGCATTGTCGACATCAAAACAACCGGC

vscC1 GACGTTGATAACGGTGATGTCACATTGGGCAGTGGTCAGAGTTTCAAGTCCTTGTTGGATTCGACCAACTTGAACTACTTGTTGGCGCAAATTCGTTTATTGGAATCGAAAGGC

vscC1-R

vscC1 AGTGCGCAAGTCGTTTCCAGACCAACGCTGCTCACGCAAGAAAACGTCGAAGCGGTGCTAAACAACAGCAGCACGTTTTACGTCAAATTGGTTGGCAAAGAGACGGCGGCTTTG

vscC1 GAAGAAGTGACGTATGGCACCTTGTTGCGTATTGTGCCTCGTATCGTCGGCGACAGATTCGCGACGCGACCAGAGATCAACTTGAGCCTTCACTTAGAAGATGGCGCAAAGATC

vscC1 CCAGATGGTGGTGTGGATGATCTTCCATCAGTACGTAAAACCGAAATCAGTACGTTAGCGACGGTAAAACAAGGGCAAAGCTTACTCATTGGTGGCGTATACCGAGACGAAGTT

vscC1 AGCCATCAGCTACGTAAAGTCCCGCTGTTGGGCGACATTCCGTACCTAGGTGCTTTGTTCCGTAGTAATACCAACACAACGCGCCGAACAGTTCGCATGTTCATTATCGAGCCG

vscC1 AGAATCGTGGTCGATGGGATTGGCGACAGCGTGTTGATAGGTAATGAGCATGATTTGCGCCCGAGTATTGGCCAGCTCAACAACATTTCGAACAACAGCGCAGAATTTAAATCG

vscC1 GTTGTTGAAGTATTCAGCTGTACGTCAAAAACGCAGGCAGAACGTTACCAACAAGACTTACTGTCACAGCAAAAGTCTTCGCTGCTAACTCAATGTCAATTGCCGTCAGGGCAA

vscC1 GTAGGCTGGCGAGTCAAAGTGGCGGAATGTGACCTTTCACAGGCCGAATGTGTTCGCCCTTCAGAGGAGCCATAA

**Figure 3.6.** Nucleotide sequence (5' → 3') of the *vscC1* gene of *V. parahaemolyticus* isolate RIMD2210633 showing the positions of forward and reverse primers. The nucleotide sequence was obtained from http://www.ncbi.nlm.nih.gov/genbank.

**Figure 3.7.** Nucleotide sequence (5' → 3') of the *VP1680* gene of *V. parahaemolyticus* isolate RIMD2210633 showing the positions of forward and reverse primers. The nucleotide sequence was obtained from http://www.ncbi.nlm.nih.gov/genbank.

### 3.2.4 Sequencing

PCR products of *tdhA, tdhS, trh1, trh2, vcrD1, vscC1,* and *VP1680* of selected isolates were sequenced. Sequencing primers were the same as those used for the PCR reactions (Table 3.1). Sequencing was performed as described in Chapter 2. Representative isolates for virulence gene sequencing were selected based on different epidemiological sources including clinical, human carrier, and seafood samples. Properties of isolates selected for virulence gene sequencing are shown in Table 3.2. The information on the presence and absence of the *tdh* and *trh* shown in Table 3.2 was provided by Prof Orasa Suthienkul, Faculty of Public Health, Department of Microbiology, Mahidol University, Thailand or obtained from the online database (http://www.ncbi. nlm.nih.gov). Sequencing data were checked and edited using Lasergene version 5.0 (DNASTAR) sequence analysis software. Nucleotide sequence analyses were conducted with MEGA version 4.0.2. (Tamura *et al.*, 2007).

## 3.3 Results

### 3.3.1 Detection and distribution of haemolysin and TTSS1-related genes

Gene fragments of *tdhA* (VP170)*, tdhS* (VP178)*, trh1* (VP170)*, trh2* (VP166)*, vcrD1* (VP2)*, vscC1* (VP2), and *VP1680* (VP2) were successfully amplified in *V. parahaemolyticus* isolates (Figs. 3.8-3.9). The expected sizes of the PCR products for these gene fragments were visualized by gel electrophoresis.

**Table 3.2.** Details of isolates selected for DNA sequence analysis of *tdhA, tdhS, trh1, trh2, vcrD1, vscC1* and *VP1680* gene fragments

| No. | Isolates | Source | Country | Year | Serotype | *tdh* | *trh* | ST | Reference |
|---|---|---|---|---|---|---|---|---|---|
| 1 | VP2 | Food poisoning agent | Japan | 1950 | O1:K1 | - | - | 1 | Fujino *et al.*, 1974 |
| 2 | VP36 | Frozen shrimp | Thailand | 1999 | O10:KUT | - | - | 238 | This study |
| 3 | VP132 | Human carrier | Thailand | 2003 | O3:K46 | + | - | 3 | This study |
| 4 | VP138 | Human carrier | Thailand | 255 | O11:K5 | + | + | 255 | This study |
| 5 | VP166 | Clinical sample | Thailand | 1990 | O1:K1 | + | + | 83 | This study |
| 6 | VP178 | Clinical sample | Thailand | 1991 | O1:K69 | + | + | 262 | This study |
| 7 | VP216 | Oyster | Thailand | 2002 | O2:KUT | - | - | 273 | This study |
| 8 | RIMD2210633 | Clinical sample | Thailand | 1996 | O3:K6 | + | - | 3 | Makino *et al.*, 2003 |
| 9 | 10329 | Clinical sample | USA | 1998 | O4:K12 | + | + | 36 | Gonzalez-Escalona *et al.*, 2011 |
| 10 | AQ3810 | Clinical sample | Singapore | 1983 | O3:K6 | + | - | - | Boyd *et al.*, 2008 |
| 11 | AQ4037 | Clinical sample | Maldives | 1985 | O3:K6 | - | + | - | Chen *et al.*, 2011 |
| 12 | Peru466 | Clinical sample | Peru | 1996 | O3:K6 | + | - | - | Chen *et al.*, 2011 |
| 13 | AN5034 | Clinical sample | Bangladesh | 1998 | O4:K68 | + | - | - | Chen *et al.*, 2011 |
| 14 | K5030 | Clinical sample | India | 2005 | O3:K6 | + | - | - | Chen *et al.*, 2011 |
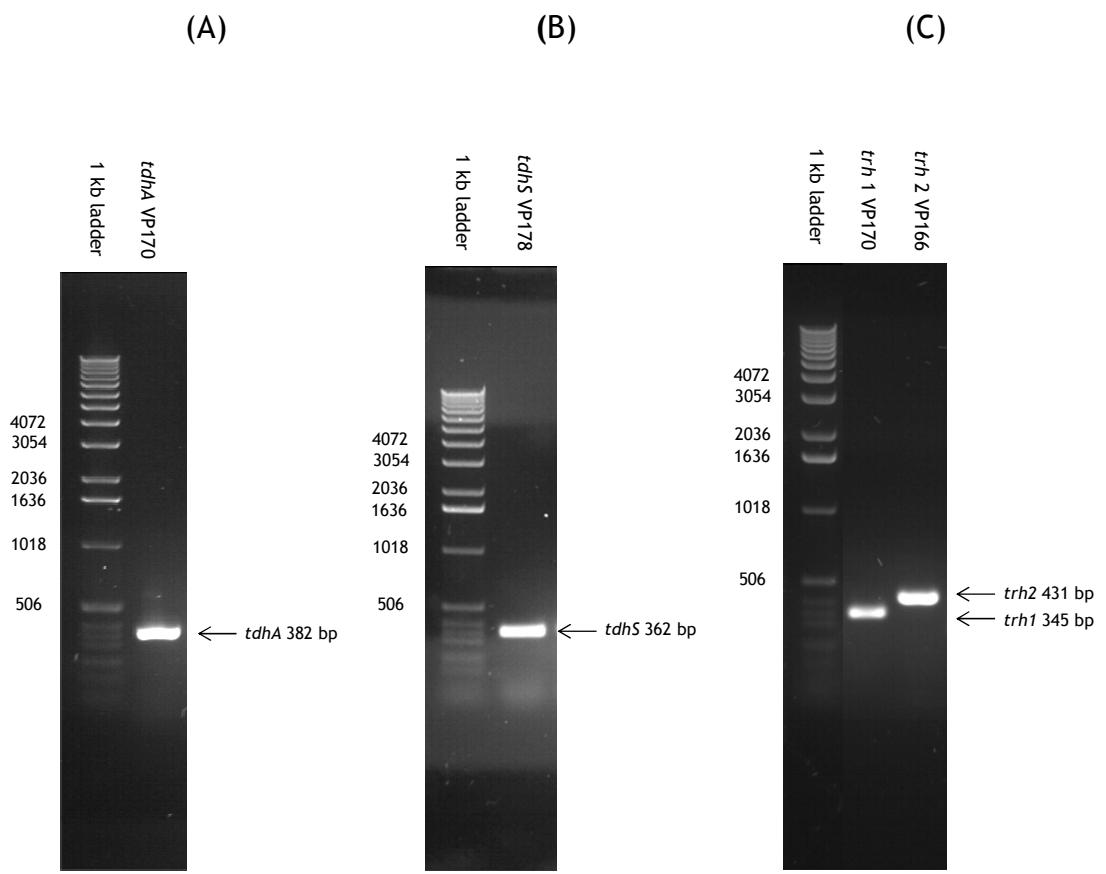
**Figure 3.8.** Agarose gel electrophoresis of (A) *tdhA,* (B) *tdhS* and (C) *trh1* and *trh2* gene fragments. The expected sizes of *tdhA* (382 bp), *tdhS* (362 bp), *trh1* (345 bp) and *trh2* (431 bp) gene fragments were detected in isolates VP170, VP178, VP170 and VP166, respectively.
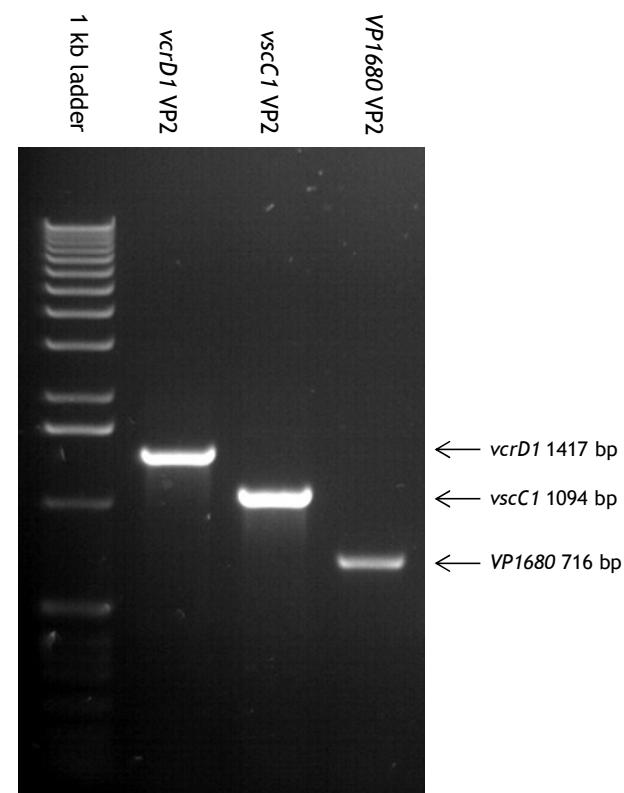
**Figure 3.9.** Agarose gel electrophoresis of *vcrD1, vscC1,* and *VP1680* gene fragments. The expected sizes of *vcrD1* (1417 bp)*, vscC1* (1094 bp) and *VP1680* (716 bp) fragments were detected in isolate VP2.

PCR amplifications of the *tdhA, trh1, trh2, vcrD1, vscC1,* and *VP1680* gene fragments were performed in 102 *V. parahaemolyticus* isolates and the distribution of each gene among these isolates was determined. PCR amplification of the *tdhS* gene fragment was performed in selected isolates and the PCR products of certain isolates were used for sequencing analysis.

The distributions of *tdhA, trh1,* and *trh2* in the 102 *V. parahaemolyticus* isolates were non-random among the isolates from different sources, namely frozen shrimp, water, shrimp tissue, human carriers, clinical samples and seafood in Thailand (Tables 3.3-3.4). The *tdhA* gene was present in 13/16 clinical isolates and in 9/18 human carrier isolates. However, none of the isolates from environmental sources including seafood, shrimp tissue, frozen shrimp, and water in this study were found to contain the *tdhA* gene. Twelve clinical (VP164, VP166, VP170, VP172, VP174, VP176, VP178, VP180, VP184, VP188, VP194, and VP200) and six human carrier (VP132, VP136, VP138, VP140, VP154, and VP162) isolates possessing *tdhA* also contained *tdhS* (Table 3.3). Exceptionally, the clinical isolate VP168 possessed *tdhS* but lacked *tdhA*. Six environmental isolates (VP36, VP44, VP56, VP88, VP204 and VP216) which lacked *tdhA* also lacked *tdhS*. These results indicate that isolates possessing *tdhA* are likely to contain *tdhS*.

The *trh1* gene was detected in 8/16 isolates from clinical samples and in 3/18 isolates from human carriers (Table 3.4). The *trh1* gene was not detected in any isolates from environmental sources including seafood, shrimp tissue, frozen shrimp, and water. The presence of *trh2* among clinical and human carrier isolates was more frequent compared to that of *trh1*.

**Table 3.3.** Presence of haemolysin and TTSS1-related genes among 102 *V. parahaemolyticus* isolates

| Isolate | Source of isolation | ST | Serotype | Haemolysin genes | | | | TTSS1 genes | | |
|---------|---------------------|-----|----------|------|------|------|------|-------|-------|--------|
| | | | | *tdhA* | *tdhS* | *trh1* | *trh2* | *vcrD1* | *vscC1* | *VP1680* |
| VP2 | Food poisoning (Type strain) | 1 | O1:K1 | - | - | - | + | + | + | + |
| VP4 | Frozen shrimp | 229 | O3:K20 | - | ND | - | - | + | + | + |
| VP6 | Frozen shrimp | 230 | O10:K66 | - | ND | - | - | + | + | + |
| VP8 | Frozen shrimp | 231 | O3:KUT | - | ND | - | - | + | + | + |
| VP12 | Frozen shrimp | 241 | O1:KUT | - | ND | - | - | + | + | + |
| VP14 | Frozen shrimp | 232 | O2:K3 | - | ND | - | - | + | + | + |
| VP16 | Frozen shrimp | 233 | O3:K20 | - | ND | - | - | + | + | + |
| VP20 | Frozen shrimp | 242 | O1:K25 | - | ND | - | - | + | + | + |
| VP22 | Frozen shrimp | 234 | O1:K20 | - | ND | - | - | + | + | + |
| VP24 | Frozen shrimp | 235 | O9:K44 | - | ND | - | - | + | + | + |
| VP26 | Frozen shrimp | 236 | O7:K19 | - | ND | - | - | + | + | + |
| VP28 | Frozen shrimp | 237 | O5:K17 | - | ND | - | - | + | + | + |
| VP30 | Frozen shrimp | 233 | O3:K20 | - | ND | - | - | + | + | + |
| VP32 | Frozen shrimp | 243 | O2:KUT | - | ND | - | - | + | + | + |
| VP36 | Frozen shrimp | 238 | O10:KUT | - | - | - | - | + | + | + |
| VP38 | Frozen shrimp | 239 | O3:K20 | - | ND | - | - | + | + | + |
| VP40 | Frozen shrimp | 240 | O11:K40 | - | ND | - | - | + | + | + |
| VP44 | Water from shrimp farm 1/A | 244 | O9:K23 | - | - | - | - | + | + | + |
| VP46 | Water from shrimp farm 1/A | 244 | O1:K38 | - | ND | - | - | + | + | + |
| VP48 | Water from shrimp farm 1/B | 244 | O9:K24 | - | ND | - | - | + | + | + |
| VP50 | Water from shrimp farm 1/B | 244 | O7:K52 | - | ND | - | - | + | + | + |
| VP52 | Water from shrimp farm 1/B | 244 | O1:KUT | - | ND | - | - | + | + | + |

**Table 3.3.** (continued)

| Isolate | Source of isolation | ST | Serotype | Haemolysin genes | | | | TTSS1 genes | | |
| | | | | *tdhA* | *tdhS* | *trh1* | *trh2* | *vcrD1* | *vscC1* | *VP1680* |
|---|---|---|---|---|---|---|---|---|---|---|
| VP54 | Water from shrimp farm 1/B | 244 | O1:KUT | - | ND | - | - | + | + | + |
| VP56 | Water from shrimp farm 1/B | 245 | O2:KUT | - | - | - | - | + | + | + |
| VP58 | Water from shrimp farm 1/B | 239 | O7:KUT | - | ND | - | - | + | + | + |
| VP60 | Water from shrimp farm 1/B | 246 | O2:K3 | - | ND | - | - | + | + | + |
| VP62 | Sediment from shrimp farm 1/C | 247 | O1:KUT | - | ND | - | - | + | + | + |
| VP64 | Sediment from shrimp farm 1/C | 248 | OUT:KUT | - | ND | - | - | + | + | + |
| VP66 | Sediment from shrimp farm 1/C | 248 | O1:K26 | - | ND | - | - | + | + | + |
| VP72 | Water from shrimp farm 2/A | 249 | O9:K44 | - | ND | - | - | + | + | + |
| VP74 | Water from shrimp farm 2/A | 180 | O2:K3 | - | ND | - | - | + | + | + |
| VP76 | Water from shrimp farm 2/A | 249 | O7:K52 | - | ND | - | - | + | + | + |
| VP80 | Water from shrimp farm 2/B | 250 | O6:K46 | - | ND | - | - | + | + | + |
| VP84 | Shrimp tissue farm 2/B | 251 | O10:K71 | - | ND | - | - | + | + | + |
| VP86 | Shrimp tissue farm 2/B | 251 | O4:K63 | - | ND | - | - | + | + | + |
| VP88 | Shrimp tissue farm 2/B | 251 | O9:K23 | - | - | - | - | + | + | + |
| VP90 | Shrimp tissue farm 2/B | 251 | O2:K28 | - | ND | - | - | + | + | + |
| VP94 | Shrimp tissue farm 2/B | 251 | O1:KUT | - | ND | - | - | + | + | + |
| VP96 | Shrimp tissue farm 2/B | 251 | O10:K52 | - | ND | - | - | + | + | + |
| VP98 | Shrimp tissue farm 2/B | 251 | O2:KUT | - | ND | - | - | + | + | + |
| VP100 | Shrimp tissue farm 1/B | 246 | O1:KUT | - | ND | - | - | + | + | + |
| VP102 | Shrimp tissue farm 1/B | 246 | O1:K1 | - | ND | - | - | + | + | + |
| VP104 | Shimp tissue farm 2/B | 251 | O9:K23 | - | ND | - | - | + | + | + |
| VP106 | Shimp tissue farm 2/B | 251 | O1:KUT | - | ND | - | - | + | + | + |
| VP108 | Shrimp tissue farm 1/B | 246 | O2:K3 | - | ND | - | - | + | + | + |
| VP110 | Shrimp tissue farm 1/B | 246 | O1:K56 | - | ND | - | - | + | + | + |

207

**Table 3.3.** (continued)

| Isolate | Source of isolation | ST | Serotype | Haemolysin genes | | | | TTSS1 genes | | |
|---------|---------------------|-----|----------|------|------|------|------|-------|-------|--------|
| | | | | *tdhA* | *tdhS* | *trh1* | *trh2* | *vcrD1* | *vscC1* | *VP1680* |
| VP112 | Shrimp tissue farm 1/B | 246 | O9:K44 | - | ND | - | - | + | + | + |
| VP114 | Shrimp tissue farm 1/B | 246 | O7:KUT | - | ND | - | - | + | + | + |
| VP116 | Shrimp tissue farm 1/B | 246 | OUT:KUT | - | ND | - | - | + | + | + |
| VP118 | Shrimp tissue farm 1/B | 246 | OUT:KUT | - | ND | - | - | + | + | + |
| VP122 | Shrimp tissue farm 2/B | 251 | OUT:KUT | - | ND | - | - | + | + | + |
| VP126 | Carrier in a seafood plant | 252 | O10:KUT | - | ND | - | - | + | + | + |
| VP128 | Carrier in a seafood plant | 253 | O9:K23 | - | ND | - | - | + | + | + |
| VP130 | Carrier in a seafood plant | 254 | O1:KUT | - | ND | - | + | + | + | + |
| VP132 | Carrier in a seafood plant | 3 | O3:K46 | + | + | - | + | + | + | + |
| VP136 | Carrier in a seafood plant | 199 | O1:KUT | + | + | + | - | + | + | + |
| VP138 | Carrier in a seafood plant | 255 | O11:K5 | + | + | + | + | + | + | + |
| VP140 | Carrier in a seafood plant | 62 | O4:KUT | + | + | - | + | + | + | + |
| VP142 | Carrier in a seafood plant | 256 | OUT:KUT | - | ND | - | + | + | + | + |
| VP144 | Carrier in a seafood plant | 257 | O4:K63 | + | ND | - | - | + | + | + |
| VP146 | Carrier in a seafood plant | 258 | O1:K25 | - | ND | - | - | + | + | + |
| VP148 | Carrier in a seafood plant | 256 | O1:K41 | - | ND | - | + | + | + | + |
| VP150 | Carrier in a seafood plant | 259 | O1:KUT | + | ND | - | + | + | + | + |
| VP152 | Carrier in a seafood plant | 259 | O8:K21 | - | ND | - | + | + | + | + |
| VP154 | Carrier in a seafood plant | 68 | O4:KUT | + | + | - | + | + | + | + |
| VP156 | Carrier in a seafood plant | 260 | O1:K1 | - | ND | - | - | + | + | + |
| VP158 | Carrier in a seafood plant | 3 | O1:KUT | + | ND | - | + | + | + | + |
| VP160 | Carrier in a seafood plant | 261 | O11:K40 | - | ND | - | - | + | + | + |
| VP162 | Carrier in a seafood plant | 255 | O1:K12 | + | + | + | + | + | + | + |
| VP164 | Clinical samples | 17 | O1:K58 | + | + | - | + | + | + | + |

**Table 3.3.** (continued)

| Isolate | Source of isolation | ST | Serotype | *tdhA* | *tdhS* | *trh1* | *trh2* | *vcrD1* | *vscC1* | *VP1680* |
|---------|---------------------|-----|----------|--------|--------|--------|--------|---------|---------|----------|
| | | | | Haemolysin genes | | | | TTSS1 genes | | |
| VP166 | Clinical samples | 83 | O1:K1 | + | + | + | + | + | + | + |
| VP168 | Clinical samples | 66 | O3:KUT | - | + | + | - | + | + | + |
| VP170 | Clinical samples | 262 | O3:KUT | + | + | + | + | + | + | + |
| VP172 | Clinical samples | 83 | O1:K1 | + | + | - | + | + | + | + |
| VP174 | Clinical samples | 263 | O3:KUT | + | + | - | + | + | + | + |
| VP176 | Clinical samples | 264 | O1:K1 | + | + | - | + | + | + | + |
| VP178 | Clinical samples | 262 | O1:K69 | + | + | + | + | + | + | + |
| VP180 | Clinical samples | 262 | O8:K22 | + | + | + | + | + | + | + |
| VP182 | Clinical samples | 262 | O1:K69 | - | ND | + | + | + | + | + |
| VP184 | Clinical samples | 262 | O4:K11 | + | + | + | + | + | + | + |
| VP188 | Clinical samples | 262 | O1:K69 | + | + | + | + | + | + | + |
| VP190 | Clinical samples | 265 | O4:K10 | + | ND | - | + | + | + | + |
| VP194 | Clinical samples | 83 | OUT:KUT | + | + | - | + | + | + | + |
| VP200 | Clinical samples | 189 | O4:K8 | + | + | - | + | + | + | + |
| VP204 | Fresh oysters | 267 | O1:K64 | - | - | - | - | + | + | + |
| VP206 | Fresh oysters | 268 | O2:K3 | - | ND | - | - | + | + | + |
| VP208 | Fresh oysters | 269 | OUT:KUT | - | ND | - | - | + | + | + |
| VP210 | Fresh bloody clams | 270 | O7:KUT | - | ND | - | - | + | + | + |
| VP212 | Fresh bloody clams | 271 | O10:K19 | - | ND | - | - | + | + | + |
| VP214 | Fresh bloody clams | 272 | O1:KUT | - | ND | - | - | + | + | + |
| VP216 | Boiled crab meats | 273 | O2:KUT | - | - | - | + | + | + | + |
| VP218 | Boiled crab meats | 274 | O1:KUT | - | ND | - | - | + | + | + |
| VP220 | Boiled crab meats | 275 | OUT:KUT | - | ND | - | - | + | + | + |
| VP222 | Boiled mussels | 276 | NA | - | ND | - | - | + | + | + |

**Table 3.3.** (continued)

| Isolate | Source of isolation | ST | Serotype | Haemolysin genes | | | | TTSS1 genes | | |
|---------|---------------------|-----|----------|------|------|------|------|-------|-------|--------|
| | | | | *tdhA* | *tdhS* | *trh1* | *trh2* | *vcrD1* | *vscC1* | *VP1680* |
| VP224 | Boiled mussels | 276 | O10:KUT | - | ND | - | - | + | + | + |
| VP226 | Boiled mussels | 277 | OUT:KUT | - | ND | - | - | + | + | + |
| VP228 | Fresh shrimp from a plant | 278 | O1:KUT | - | ND | - | - | + | + | + |
| VP230 | Fresh shrimp from a plant | 279 | O3:K58 | - | ND | - | - | + | + | + |
| VP232 | Fresh shrimp from a plant | 278 | O5:KUT | - | ND | - | - | + | + | + |
| VP234 | Fresh shrimp from a plant | 114 | O5:KUT | - | ND | - | - | + | + | + |
| VP236 | Fresh shrimp from a local market | 280 | O1:K69 | - | ND | - | - | + | + | + |
| VP238 | Fresh shrimp from a local market | 281 | O10:K52 | - | ND | - | + | + | + | + |

ND = No data available

**Table 3.4.** Total number of isolates possessing *tdhA, trh1, trh2, vcrD1, vscC1,* and *VP1680* gene fragments in six different sources of isolation

| Source | *tdhA* | *trh1* | *trh2* | *vcrD1* | *vscC1* | *VP1680* |
|---|---|---|---|---|---|---|
| Clinical sample (n=16) | 13 | 8 | 15 | 16 | 16 | 16 |
| Human carrier (n=18) | 9 | 3 | 11 | 18 | 18 | 18 |
| Seafood (n=18) | - | - | 2 | 18 | 18 | 18 |
| Shrimp tissue (n=18) | - | - | - | 18 | 18 | 18 |
| Frozen shrimp (n=16) | - | - | - | 16 | 16 | 16 |
| Water (n=16) | - | - | - | 16 | 16 | 16 |
| Total (n=102) | 22 | 11 | 28 | 102 | 102 | 102 |

The *trh2* gene was present in 15/16 clinical samples and in 11/18 human carrier isolates; it was also present in 2/18 seafood isolates. The TTSS1-related genes, *vcrD1, vscC1,* and *VP1680,* were detected in all 102 isolates examined, although the band intensity was variable across individual isolates.

## 3.3.2 *Comparative nucleotide sequences of haemolysin gene fragments*

Representative isolates for sequencing analysis of *tdhA, tdhS, trh1,* and *trh2* were selected based on isolates from different sources including clinical, human carriers and seafood, and the genetic relationships among 102 isolates from MLST data in Chapter 2. Clinical isolates VP166 (cluster 2) and VP178 (cluster 4) represent isolates from distinct clinical clusters (Fig. 2.52; Chapter 2). Human carrier isolate VP132 and clinical isolate RIMD2210633 serotype O3:K6 (nucleotide sequence derived from GenBank) represent ST3 which contains pandemic *V. parahaemolyticus* causing outbreaks of worldwide distribution.

Human carrier isolate VP138 is closely related to clinical isolates in cluster 4 from the MLST phylogenetic tree (Fig. 2.52; Chapter2). Lastly, isolate VP216 represents a strain isolated from oyster.

### 3.3.2.1 Thermostable direct haemolysins (*tdhA* and *tdhS*)

A Neighbour-Joining tree of the nucleotide sequences of the *tdhA* and *tdhS* gene fragments of two clinical isolates, VP166 and VP178, two human carrier isolates, VP132 and VP138, and the pandemic serotype O3:K6 isolate RIMD2210633 (http://www.ncbi.nlm. nih.gov/) is shown in Fig. 3.10.
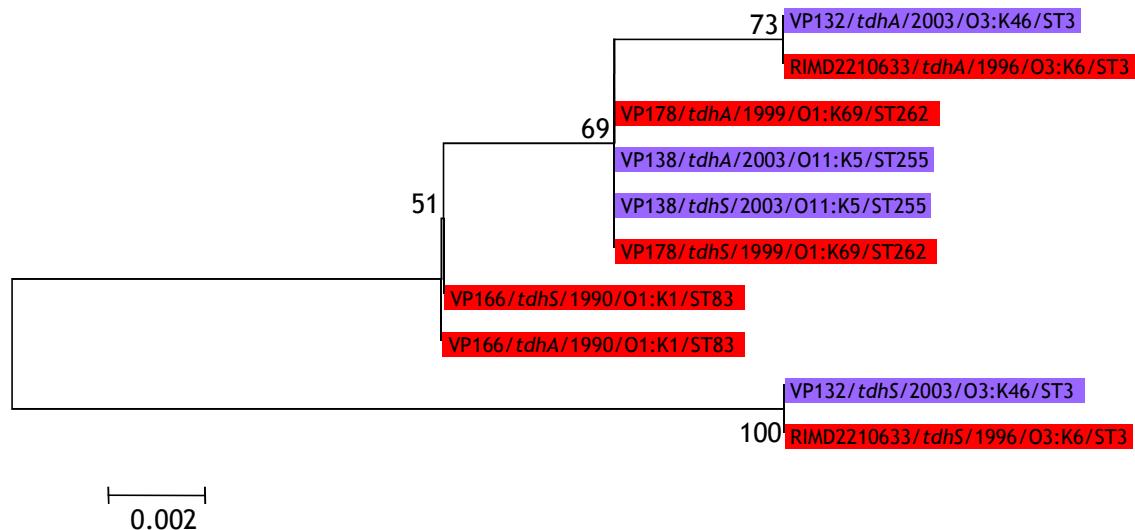


**Figure 3.10.** Neighbour-Joining tree of nucleotide sequences of *tdhA* and *tdhS* gene fragments of four representative *V. parahaemolyticus* isolates from Thailand and pandemic serotype O3:K6 isolate RIMD2210633. Red represents the isolates from clinical samples and purple represents the isolates from human carriers.

The phylogenetic tree of the *tdhA* and *tdhS* gene fragments shows that the human carrier isolate VP132 shares identical *tdhA* and *tdhS* fragments with the pandemic isolate RIMD2210633 serotype O3:K6. The human carrier isolate VP138 shares identical *tdhA* and *tdhS* fragments with the clinical isolate VP178. The nucleotide sequences of the *tdhA* and *tdhS* gene fragments from VP138 and VP178 are identical. According to the results of the housekeeping gene sequence analyses in Chapter 2, the clinical isolate VP178 (ST262) represents a different sequence type (ST) to the human carrier isolate VP138 (ST255) (Fig. 2.52 and Table 2.1). However, the concatenated sequences of the housekeeping genes of VP138 and VP178 differ only at the *dnaE* locus; this difference is probably due to genetic recombination involving *dnaE* (D1, Fig. 2.50; Chapter 2). This suggests that VP138 and VP178 are likely to have the same ancestor but one of them has acquired a *dnaE* fragment from another strain. These two isolates may also have acquired the same *tdhA* and *tdhS* segments via horizontal gene transfer. In the other scenario, identical *tdhA* and *tdhS* genes in VP138 and VP178 may be due to a gene duplication event. The clinical isolate VP166 also possesses identical *tdhA* and *tdhS* gene fragment and the sequences of these two genes differ only at a single polymorphic site to those of the VP138 and VP178.

In contrast to VP138, VP178 and VP166, the nucleotide sequences of *tdhA* and *tdhS* fragments of the pandemic isolate RIMD2210633 and of the human carrier isolate VP132 are relatively different (97.0% similarity) (Fig. 3.10). In addition, both VP132 and RIMD2210633 represent ST3 which represents the pandemic form of *V. parahaemolyticus* that was responsible for the Indian outbreak in 1996 and was subsequently found in the other parts of the world (http://pubmlst.org/vparahaemolyticus/). These results show that the nucleotide sequences of *tdhA* and *tdhS* fragments are identical in isolates VP138, VP178, and VP166 but they

are different in VP132 and the pandemic isolate RIMD2210633. The presence of both *tdhA* and *tdhS* in a single isolate may be a consequence of either gene duplication or independent acquisition from horizontal gene transfer. Furthermore, since *tdhA* and *tdhS* are located in *V. parahaemolyticus* pathogeneticity island (VPaI), the distinct nucleotide sequences of *tdhA* and *tdhS* in VP132 and the RIMD2210633 suggests a higher degree of genetic plasticity within the VPaI of strains representing the pandemic ST3.

### 3.3.2.2 TDH-related haemolysins (*trh1* and *trh2*)

A Neighbour-Joining tree of the nucleotide sequences of the *trh1* (VP178 and VP138) and the *trh2* (VP166, VP216 and VP132) gene fragments is shown in Fig. 3.11. The phylogenetic tree consists of two distinct lineages representing *trh1* and *trh2.* Nucleotide polymorphic sites between *trh1* and *trh2* gene fragments of VP178 and VP166 is 124 (n=334) whereas that of VP138 and VP132 is 123 (n=334). The *trh1* gene fragment of the clinical isolate VP178 is closely related to that of human carrier isolate VP138 (they differ at a single polymorphic site). This indicates that the clinical isolate (VP178) and human carrier isolate (VP138) share a very similar *trh1* although they represent different serotypes and STs (VP178 = O1:K69 and ST262; VP138 = O11:K5 and ST255). Furthermore, the *trh2* gene fragments of clinical isolate VP166, human carrier isolate VP132, and seafood isolate VP216 are also closely related. The *trh2* gene fragment sequences of VP216 and VP132 are identical whereas the *trh2* gene fragment sequence of VP166 differs from those of VP216 and VP132 at a single polymorphic site. The results indicate that the seafood isolate (VP216) shares a very similar *trh2* gene fragment with the clinical (VP166) and human carrier (VP132) isolates although these three isolates comprise multiple serotypes and

STs (VP166 = O1:K1 and ST83; VP216 = O1:KUT and ST273; VP132 = O3:K46 and ST3).
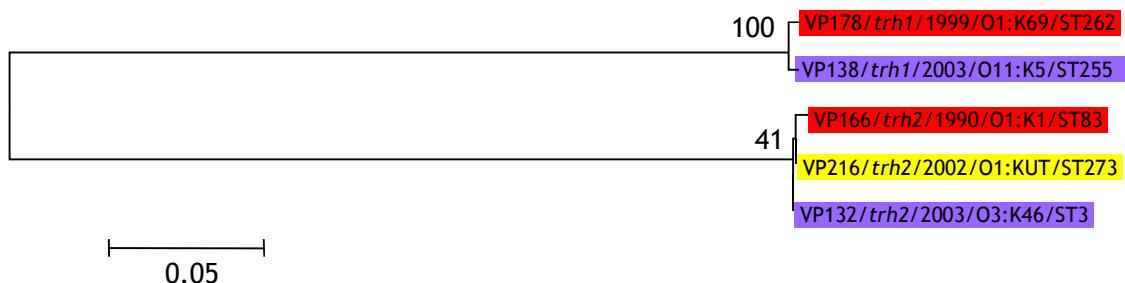


**Figure 3.11.** Neighbour-Joining tree of nucleotide sequences of *trh1* and *trh2* gene fragments of five representative *V. parahaemolyticus* isolates from Thailand. Red represents the isolates from clinical samples, purple represents the isolates from human carriers, and yellow represents an isolate from seafood.

### *3.3.3 Comparative nucleotide sequences of type three secretion system 1 (TTSS1) gene fragments*

Neighbour-Joining trees representing the nucleotide sequences of the *vcrD1, vscC1,* and *VP1680* gene fragments of 13 representative *V. parahaemolyticus* isolates were constructed (Figs. 3.12-3.14). These isolates included five isolates from Thailand (VP36, VP132, VP138, VP178 and VP216), seven pathogenic isolates from worldwide sources (AQ3810, AQ4037, RIMD2210633, Peru466, AN5034, K5030, and 10329) and the type strain (VP2) which was isolated in Japan. The details of these isolates are provided in Table 3.2. The sequences of the *vcrD1, vscC1,* and *VP1680* gene fragments for the seven pathogenic isolates from worldwide sources and the Japanese type strain were obtained from the GenBank (http://www.ncbi.nlm.nih.gov/). The sequences of the *vcrD1, vscC1,*

and *VP1680* gene fragments for the five Thai isolates were determined in the present study.

### 3.3.3.1 *vcrD1*

The phylogenetic tree for *vcrD1* (Fig. 3.12) shows that the nucleotide sequences of the *vcrD1* gene fragments of the isolates AQ3810, AQ4037, RIMD2210633, Peru466, AN5034, K5030, and VP132 were identical. With the exception of the isolates AN5034 (O4:K68) and VP132 (O3:K46), these isolates contain the same serotype O3:K6. All isolates mentioned above shared the same haemolysin profile (*tdh+/trh-*) except AQ4037 (*tdh-/trh+*). The phylogenetic tree for *vcrD1* suggests that human carrier isolate VP132 shares the same *vcrD1* gene with clinical isolates that were responsible for cases of *V. parahaemolyticus* infection. The other human carrier isolate, VP138, shares an identical *vcrD1* gene fragment with the clinical isolate VP178 (Fig. 3.12) although these sequences are very different from those of the main cluster. The *vcrD1* gene fragment of the clinical isolates VP2 and 10329, the seafood isolate VP216 and the frozen shrimp isolate VP36 were unrelated.

However, the *vcrD1* gene fragments of the seafood isolate VP216 and clinical isolate 10329 are more closely related to those of pathogenic isolates AQ3810, AQ4037, RIMD2210633, Peru466, AN5034 and K5030, and to the human carrier isolate VP132, than they are to those of other isolates including the frozen shrimp isolate VP36, the clinical isolates VP2 and VP178, and the human carrier isolate VP138. Interestingly, the seafood isolate VP216 was recovered from oysters in Thailand, while isolate 10329 was recovered from the faecal sample of a patient with "oyster-associated illness" on the west coast of Washington State,

USA (Gonzalez-Escalona *et al.*, 2011). Thus, the *vcrD1* gene fragments from the Thai oyster isolate and the clinical isolate involved in oyster-causing illness in the USA are genetically related. Finally, the *vcrD1* gene fragment of clinical isolate VP2 was most closely related to the frozen shrimp isolate VP36.
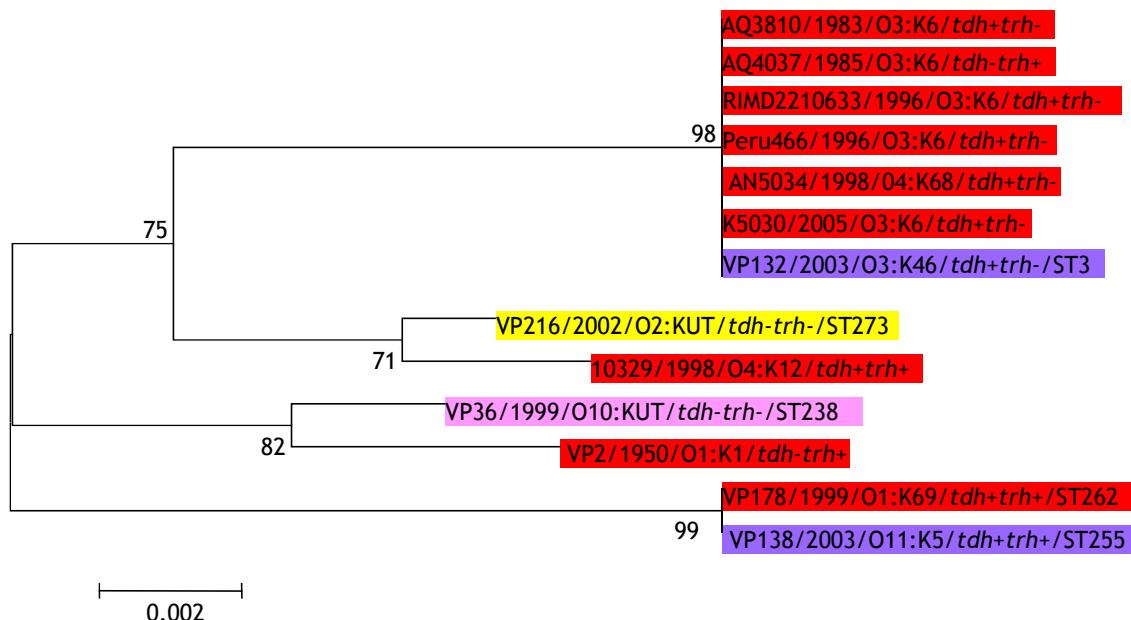


**Figure 3.12.** Neighbour-Joining tree of the nucleotide sequences of *vcrD1* gene fragments of representative *V. parahaemolyticus* isolates including five isolates from Thailand and eight isolates from worldwide sources. Red represents isolates from clinical samples, purple represents isolates from human carriers, yellow represents isolates from seafood, and pink represents isolates from frozen shrimp.

### 3.3.3.2 *vscC1*

The phylogenetic tree for *vscC1* (Fig. 3.13) shows that the nucleotide sequences of the *vscC1* gene fragments of worldwide pathogenic isolates AQ4037, RIMD2210633, Peru466, AN5034, K5030, 10329, and Thai human carrier isolate VP132 are identical. The isolates AQ4037, RIMD2210633, Peru466 and K5030 contain the same serotype O3:K6 whereas the isolates AN5034 (O4:K68), 10329 (O4:K12) and VP132 (O3:K46) are of different serotypes. All isolates mentioned above shared the same haemolysin profile (*tdh*+/*trh*-) except AQ4037 (*tdh*-/*trh*+) and 10329 (*tdh*+/*trh*+). The *vscC1* fragment of isolate AQ3810, a pre-pandemic isolate of serotype O3:K6, is closely related, but not identical, to that of the above isolates. The Thai clinical isolate VP178 and human carrier isolate VP138 share identical *vscC1* gene fragments but these are divergent from those of the above isolates. However, the nucleotide sequences of the *vscC1* gene fragments of the seafood isolate VP216, clinical isolate VP2 and frozen shrimp isolate VP36 are relatively unrelated. The *vscC1* gene fragments of the clinical isolate VP2 and frozen shrimp isolate VP36 are more closely related to those of the clinical isolate VP178 and human carrier isolate VP138 than are those of the seven worldwide pathogenic isolates and Thai human carrier isolate VP132.
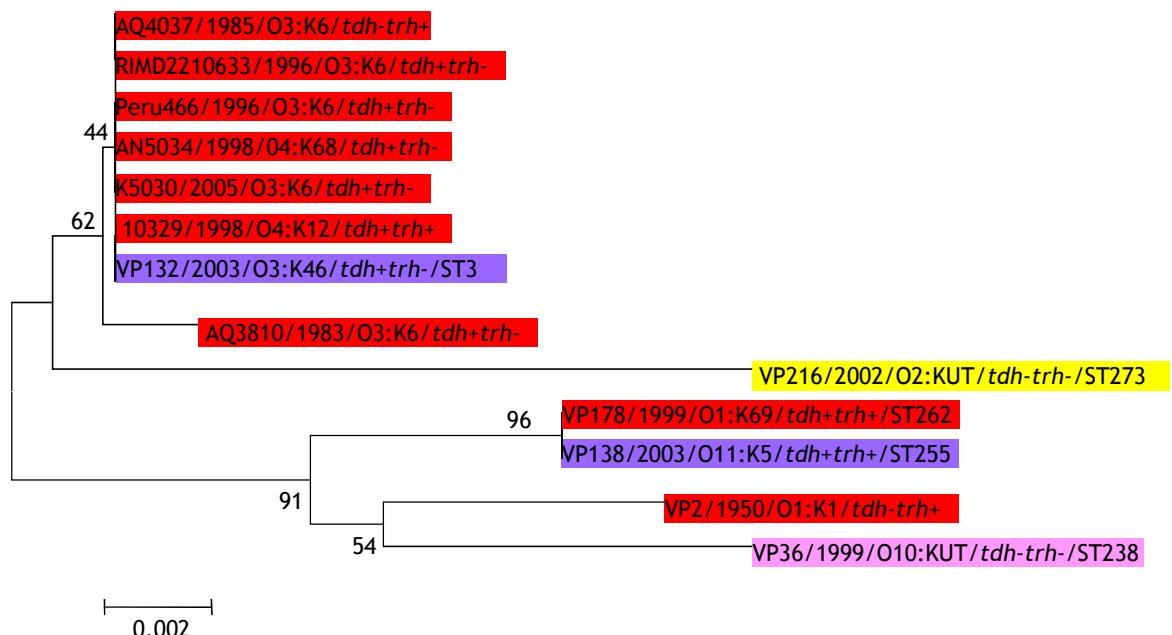
**Figure 3.13.** Neighbour-Joining tree of the nucleotide sequences of *vscC*1 gene fragments of representative *V. parahaemolyticus* isolates including five isolates from Thailand and eight isolates from worldwide sources. Red represents isolates from clinical samples, purple represents isolates from human carriers, yellow represents isolates from seafood, and pink represents isolates from frozen shrimp.

### 3.3.3.3 *VP1680*

The phylogenetic tree for VP1680 (Fig. 3.14) shows that the nucleotide sequences of the *VP1680* gene fragments of isolates AQ3810, AQ4037, RIMD2210633, Peru466, AN5034, K5030 and VP132 are identical. With the exception of the isolates AN5034 (O4:K68) and VP132 (O3:K46), these isolates contain the same serotype O3:K6. All isolates mentioned above share the same haemolysin profile (*tdh+*/*trh-*) except AQ4037 (*tdh-*/*trh+*). The *VP1680* gene fragment of the clinical isolate 10329 is distinct from those of above isolates. The Thai clinical isolate VP178 and the human carrier isolate VP138 share identical *VP1680* gene fragments, although *VP1680* of these two isolates is distinct from those of the other isolates with a high bootstrap score (100). The

*VP1680* gene fragments of the clinical isolate VP2, seafood isolate VP216 and frozen shrimp isolate VP36 are unrelated.
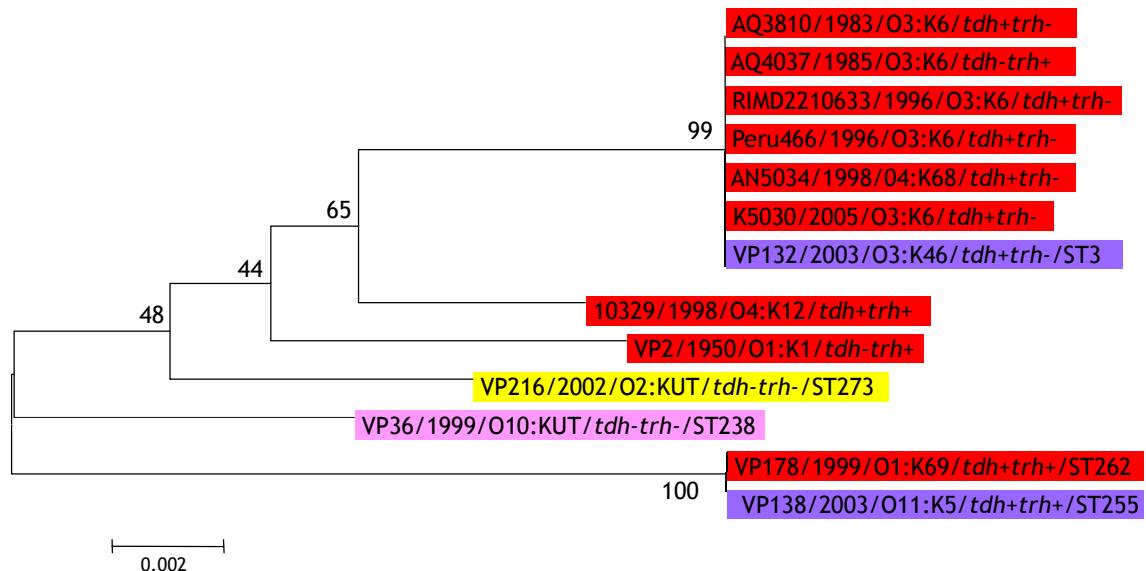


**Figure 3.14.** Neighbour-Joining tree of the nucleotide sequences of *VP1680* gene fragments of representative *V. parahaemolyticus* isolates including five isolates from Thailand and eight isolates from worldwide sources. Red represents isolates from clinical samples, purple represents isolates from human carriers, yellow represents isolates from seafood, and pink represents isolates from frozen shrimp.

## 3.5 Discussion

In the present study, the detection of haemolysin genes, including *tdhA, trh1* and *trh2,* in Thai *V. parahaemolyticus* isolates shows that there is a non-random distribution of *tdhA, trh1,* and *trh2* in isolates from different sources. The *tdhA, trh1* and *trh2* genes were predominantly present in clinical and carrier isolates (Tables 3.3 and 3.4). With the exception of two seafood isolates which possessed *trh2,* none of the environmental isolates from seafood, shrimp tissue, frozen shrimp, and water contained *tdhA, trh1* or *trh2.* Confirmation of PCR negative results can be obtained by applying alternative DNA-based methods such as Southern blot (Southern, 1975) and fluorescence *in situ* hybridization (FISH) (Wagner & Haider, 2012; Wagner *et al.*, 2003; Zwirglmaier, 2005). False positive PCR results can be avoided by preventing contamination of exogenous DNA, that may be a consequence of poor DNA preparation, relative to the examined DNA in the sample whereas false negative PCR results can be avoided by using alternative primers that are specific to the targeted gene.

Most isolates containing *tdhA* in this study also possessed *tdhS*. From previous literature, the distribution of *tdhS* has not been well studied in *V. parahaemolyticus,* presumably because *tdhS* is not as crucial for TDH production as is *tdhA* (Nishibuchi & Kaper, 1990). Detection of *tdhA* and *trh1* in *V. parahaemolyticus* isolates has been widely studied and most authors use the term *tdh* instead of *tdhA, and trh* instead of *trh1.* The presence of *tdh* (*tdhA*) *and trh* (*trh1*) in isolates recovered from seafood and marine samples (seawater, seaweed, sediment, etc.) has been reported from many countries worldwide, including Japan (Hara-kudo *et al.*, 2003; Mahmud *et al.*, 2006), China (Chao *et al.*, 2009b), India (Deepanjali *et al.*, 2005; Pal & Das, 2010; Raghunath *et al.*,

2008), Bangladesh (Alam *et al.*, 2009), Malaysia (Bilung *et al.*, 2005; Sujeewa *et al.*, 2009), Indonesia (Marlina *et al.*, 2007), Thailand (Yamamoto *et al.*, 2008), Turkey (Terzi *et al.*, 2009), Italy (Pinto *et al.*, 2008) , Spain (Roque *et al.*, 2009), England (Wagley *et al.*, 2008, 2009), and the USA (DePaola *et al.*, 2000; Jones *et al.*, 2012b; Parveen *et al.*, 2008; Rizvi & Bej, 2010). In particular, four (12.5%) and two (6.3%) of 32 isolates recovered from bloody clams at a harvesting site in southern Thailand contained *tdh* and *trh,* respectively (Yamamoto *et al.*, 2008). From that study, the prevalence of *tdh* and *trh* from seafood isolates in Thailand was relatively high (12.5% for *tdh* and 6.3% for *trh*). In contrast, *tdh (tdhA)* and *trh* (*trh1*) were not detected in Thai seafood isolates in the present study, even though the sample set was large (n=52). However, seafood isolates examined in the present study were from shrimps, oysters, bloody clams, crab meat, and mussels collected at shrimp farms, seafood processing factories and seafood markets, whereas the isolates examined in the previous study (Yamamoto *et al.*, 2008) were only from wild bloody clams collected at a harvesting site on the sea shore. These results suggest that *V. parahaemolyticus* isolates inhabiting different host species of diverse geographic origins may have different distribution of virulence genes.

To date, the presence of haemolysin genes in *V. parahaemolyticus* isolates recovered from healthy human carriers has not been well studied. The presence of both *tdhA* and *tdhS* in isolates recovered from faecal samples of healthy Thai human carriers indicates the persistence of these virulence determinants in *V. parahaemolyticus* isolates in the human intestinal tract. The *trh1* and *trh2* genes were also detected in isolates from Thai human carriers. Although a previous study (Kishishita *et al.*, 1992) showed that individual *V. parahaemolyticus* isolates possessed either one *trh1* or *trh2,* both *trh1* and *trh2* were detected in

eight clinical (VP162, VP166, VP170, VP178, VP180, VP182, VP184, and VP188) isolates and one human carrier (VP138) isolate in the present study. However, the distribution of *trh2* was consistent with the previous study (Kishishita *et al.*, 1992) in that *trh2* was predominantly present in clinical isolates rather than in environmental isolates. In the present study, we found that *trh2* (15/16) was more prevalent than *trh1* (8/16) in clinical isolates. Similarly, *trh2* (11/18) was also more prevalent than *trh1* (3/18) in human carrier isolates. The *trh2* gene was also detected in two seafood isolates, whereas *trh1* was not detected in any environmental isolates from Thailand. Since the haemolytic activity of *trh2* is relatively weak compared to that of *trh1*, the abundance of *trh2* in the clinical and human carrier isolates may suggest another important role to benefit bacterial survival in the human host, other than haemolytic activity. Evidence of *V. parahaemolyticus trh2* gene transfer among *Vibrio* species, including *V. alginolyticus*, was reported in previous studies (González-Escalona *et al.*, 2006; Xie *et al.*, 2005). This suggests that the *trh2* gene may have a significant role in the evolution of *Vibrio* species.

Phylogenetic analysis showed that the *tdhA* and *tdhS* gene fragments of the pandemic isolate RIMD2210633 and human carrier isolate VP132 are different (only 97.0% similarity) whereas the *tdhA* and *tdhS* fragments from the other isolates examined, including the clinical isolates VP166 and VP178, and human carrier isolate VP138, are identical (Fig. 3.10). Based on the presence of flanking insertion sequences in the *tdh*-encoding region in *V. parahaemolyticus,* it has been proposed that the *tdh* gene is located on a mobile genetic element (Kamruzzaman *et al.*, 2008; Terai, 1991). These insertion sequences facilitate horizontal gene transfer of *tdh* among other *Vibrio* species including non-O1 *V. cholera*, *V. mimicus* and *V. hollisae*; they also play an important role in *tdh*

deletion in *tdh*-negative *V. parahaemolyticus* isolates (Kamruzzaman *et al.*, 2008; Nishibuchi & Kaper, 1995). As a result of the present study, the presence of different *tdhA* or *tdhS* genes in the pandemic isolate RIMD2210633 and human carrier isolate VP132 is more likely due to gene acquisition rather than gene duplication. In contrast, indistinguishable *tdhA* and *tdhS* genes in the clinical isolates VP166 and VP178, and in the human carrier isolate VP138, could be due to gene duplication. Thus, the genetic structure of the pathogenicity island, where both *tdhA* and *tdhS* are located, of isolates representing the pandemic ST3 (RIMD2210633 and VP132) tends to be more dynamic than that of the other Thai isolates such as VP138, VP166 and VP178. Since *tdhS* is responsible for 0.5-9.4% of TDH production (Nishibuchi *et al.*, 1991), acquisition of a certain *tdhS* that is capable of producing a greater amount of TDH may increase the ability of pandemic strains to produce TDH. The identical *tdhA* and *tdhS* genes of Thai human carrier VP132 and pandemic isolate RIMD2210633, as well as the Thai human carrier VP138 and clinical isolate VP178, indicate that the Thai human carriers are capable of harbouring *V. parahaemolyticus* isolates with virulence determinants. Evidence that the human carrier isolates VP132 and VP138 share identical, or very similar, virulence gene sequences with the clinical strains RIMD2210633, VP166 and VP178 has been provided for *trh1, trh2, vcrD2, vscC1,* and *VP1680* (Figs. 3.11-3.14). Furthermore, the *trh2* gene fragment from seafood isolate VP216 is very similar (different at a single polymorphic site) to the *trh2* fragment from the clinical isolate VP166. On the other hand, the isolates VP216 and VP166 are unrelated in the phylogenetic tree based on the seven housekeeping genes (Fig. 2.52). This finding provides evidence that horizontal gene transfer of *trh2* has occurred between clinical (VP166) and environmental (VP216) *V. parahaemolyticus* isolates in Thailand. Comparative phenotypic tests of the virulence factors of clinical, human carrier, and seafood isolates should

be a focus for future research, to allow a better understanding of the function of virulence genes of *V. parahaemolyticus* isolates from the human intestinal tract and from seafood.

The presence of TTSS1-related genes including *vcrD1, vscC1,* and *VP1680* in all 102 isolates confirms the existence of the TTSS1-protein encoding region in *V. parahaemolyticus* isolates from both clinical and environmental sources (Makino *et al.,* 2003). Sequencing analysis of the *vcrD1, vscC1* and *VP1680* gene fragments indicated that the TTSS1-related genes of the five pandemic isolates (AQ4037, RIMD2210633, Peru66, K5030, and AN5034) and a Thai human carrier isolate (VP132) are highly conserved and also distinct from those of the other clinical and environmental isolates examined (Figs. 3.12-3.14). Futhermore, recombination events of these TTSS1 genes were observed in pathogenic strains AQ380 and 10329 (Figs. 3.12-3.14). Previous authors studied the evolutionary relationships of the pandemic *V. parahaemolyticus* serotype O3:K6 strain and its serovariants (e.g. O1:KUT, O1:K25, and O4:K68, etc.) and the role of pathogenicity islands (VPaI-1 to VPaI-7) in the evolution of the pandemic strain (Chen *et al.,* 2011; Han *et al.,* 2008; Hurley *et al.,* 2006). Han *et al.* (2008) suggested that the O3:K6 non-pandemic strains have evolved into the pandemic clone by acquisition of new *toxRS* (a part of VPaI-3) and *tdh* (a part of VPaI-7). These authors also proposed that genetic diversity within the pandemic clone occurs by subsequent acquisitions of VPaI-4 and VPaI-6, serotype conversion and gene deletion (Han *et al.,* 2008). According to these studies, the evolution of the pandemic serotype O3:K6 strain has been caused by considerable genomic flux in the VPaI-regions, including the TDH- and TTSS2-encoding genes. Unlike the TTSS2-encoding genes, comparative genomics of six pre- and post-pandemic *V. parahaemolyticus* isolates (AQ3810, AQ4037, RIMD2210633, Peru466, AN5034,

and K5030) revealed that TTSS1-encoding genes are conserved among these isolates (Chen *et al.*, 2011). However, genetic diversity of TTSS1-associated genes among a wider range of clinical and environmental isolates from different parts of the world has not been explored. The present study has demonstrated that the conserved TTSS1-related genes (*vcrD1, vscC1, and VP1680*) of six worldwide pathogenic *V. parahaemolyticus* isolates differ from those of other isolates including clinical and environmental isolates from Thailand. Nucleotide sequences of *vcrD1* and *VP1680* gene fragments of clinical isolate 10329 differ from those of six worldwide pathogenic *V. parahaemolyticus* isolates (Fig. 3.12 and Fig. 3.14). To some extent, the clinical isolate 10329, which was recovered from a patient with "oyster-associated illness" in Washington State, was determined to infect at much lower dose than the other pathogenic *V. parahaemolyticus* from another area in the USA, although this latter strain possesses both *tdh* and *trh* (Gonzalez-Escalona *et al.*, 2011). Variation in the nucleotide sequences of *vcrD1* and *VP1680* gene fragments between isolate 10329 and the six worldwide pathogenic isolates may contribute to different degrees of virulence caused by these isolates.

The role in pathogenicity of TTSS1-encoding genes in environmental *V. parahaemolyticus* isolates has not been described to date. Genotypic differences in *vcrD1, vscC1* and *VP1680* between environmental and clinical isolates in the present study may correlate with difference in pathogenicity in the individual isolates. The contribution to *V. parahaemolyticus* pathogenicity of the *tdh* and TTSS2-associated genes in environmental isolates has been described in several previous studies. Caburlotto *et al (*2010) demonstrated that environmental *V. parahaemolyticus* isolates possessing TTSS2-associated genes are capable of adhering and causing cell disruption in human cells. Furthermore, Vongxay *et al.*

(2008) showed that clinical *V. parahaemolyticus* isolates harbouring *tdh* exhibited higher cytotoxicity than environmental isolates also containing *tdh*. Studies of the contribution of TTSS1 to pathogenesis in environmental isolates will enable a better understanding of the virulence potential of environmental strains of *V. parahaemolyticus*.

Comparative sequence analysis of *tdh, vcrD1, vscC1* and *VP1680* gene fragments clearly showed that Thai human carrier isolates VP132 and VP138 share identical haemolysin and TTSS1-associated genes with pandemic RIMD2210633 and Thai clinical isolates VP178, respectively. However, the expression of virulence genes at the transcriptional level in isolates from human carriers and clinical samples may differ. Previous studies demonstrated that the expression of *V. parahaemolyticus* TTSS1 was regulated by a homologue of *Pseudomonas aeruginosa* transcriptional factors ExsA, ExsC, and ExsD (Zhou *et al.*, 2008, 2010b). Several factors derived from the human host such as bile acid, NaCl concentration and temperature have been suggested to induce *V. parahaemolyticus* virulence (Gotoh *et al.*, 2010; Mahoney *et al.*, 2010; Osawa & Yamai, 1996; Pace *et al.*, 1997; Whitaker *et al.*, 2010). However, the degree of pathogenicity may vary among *V. parahaemolyticus* strains. For example, virulence traits including haemolysin, protease, motility, biofilm formation and cytotoxicity were induced at $37^{\circ}$C in clinical isolates but not in environmental isolates (Mahoney *et al.*, 2010). *V. parahaemolyticus* is capable of adhering to the human intestinal cells regardless of the presence of *tdh* or *trh,* suggesting that both pathogenic and non-pathogenic isolates are able to colonize the human gut (Gingras & Howard, 1980; Iijima *et al.*, 1981; Reyes *et al.*, 1983; Vongxay *et al.*, 2008). However, the *tdh*-positive isolates show greater adherence to cell lines than do *tdh*-negative isolates (Chakrabarti *et al.*, 1991; Hackney *et al.*,

1980). A recent study showed that OmpU, a major OMP of *V. parahaemolyticus*, is involved in bacterial colonization and prolongs bacterial survival under stressful conditions, including the bile-containing environment of the gut (Whitaker *et al.*, 2012). These authors also demonstrated that high numbers of the pandemic *V. parahaemolyticus* serotype O3:K6 isolate RIMD2210633 were maintained in the mouse intestine after seven days of infection without any signs of pathology (e.g. cell disruption and degradation of epithelial or colonic crypt structure). An explanation for the finding that isolates carrying virulence determinants can survive in the gut of healthy individuals remains to be elucidated. It is possible that there are unknown mechanisms involved in the regulation of virulence-associated genes in *V. parahaemolyticus* isolates inhabiting the human intestinal tract, so that these isolates become asymptomatic in the host. Alternatively, this bacterium may develop mechanisms to survive in the human gut of healthy individuals while the active innate immune system in the body is operating. Mechanisms to survive the gastrointestinal immune response and an ability to gain benefits for bacterial growth during intestinal inflammation have been reported in *Salmonella typhimurium*, an enteric bacterium that can cause acute human gastroenteritis via the *Salmonella*-TTSS1 and 2 (Broz *et al.*, 2012; Thiennimitr *et al.*, 2012). Although proteins encoded by *V. parahaemolyticus* TTSS1 and 2 are capable of inducing inflammatory factors such as mitogen-activated protein kinases (MAPK) (Matlawska-Wasowska *et al.*, 2010) and interleukin-8 (IL-8) (Shimohata *et al.*, 2011),  the interactions of the human intestinal innate immune response with virulence mechanisms of *V. parahaemolyticus* have not been clearly described. A better understanding of the host defence mechanisms of *V. parahaemolyticus* may help to explain the survival of *V. parahaemolyticus* carrying virulence genes in healthy individuals.

In conclusion, virulence-related genes including the haemolysin-encoding genes *tdhA, tdhS, trh1* and *trh2,* and the TTSS1-related genes *vcrD1, vscC2* and *VP1680,* were detected in isolates from healthy individuals who were working at a seafood processing factory. Thai human carrier isolates (VP132 and VP138) share identical nucleotide sequences of the virulence genes *tdhA, vcrD1, vscC2* and *VP1680* with clinical isolates, indicating a potential ability of these isolates to cause disease. Consequently, healthy individuals who are involved in the seafood industry should be considered as a reservoir of potential pathogenic *V. parahaemolyticus* in accordance with seafood safety surveillance. Furthermore, distinct nucleotide sequences of TTSS1-related gene fragments (*vcrD1, vscC2* and *VP1680*) were demonstrated in clinical isolates of worldwide distribution as well as in the Thai human carrier isolate VP132, compared to other Thai isolates of clinical and environmental origins. These findings contribute to our understanding of the epidemiology of potential pathogenic *V. parahaemolyticus* isolates in Thailand and this knowledge can be applied for the development of *V. parahaemolyticus* risk assessment in the seafood production industry.

# 4. COMPARATIVE OUTER MEMBRANE PROTEOMICS OF *V. PARAHAEMOLYTICUS* ISOLATES FROM CLINICAL, HUMAN CARRIER AND ENVIRONMENTAL SOURCES

## 4.1 Introduction

Since *V. parahaemolyticus* has been routinely recovered from a very wide range of sources, including estuarine and sea water, marine plankton, marine animals and from the human body, the organism is clearly capable of adaptation to a wide range of environmental conditions (temperature, osmolarity, nutrient concentration, etc.). However, the mechanisms of host adaptation and the evolution of virulent strains have not been established for this organism. The outer membrane of Gram-negative bacteria plays an important role in adaptation to the external environment, including the host in the case of commensal and pathogenic bacteria. This is because the outer membrane is the outermost layer of the bacterial cell (with the exception of the capsule) and is responsible for bacterial adaptive responses to the conditions encountered (Lin *et al.*, 2002). The outer membrane functions as a selective barrier that protects bacteria from harmful substances. Proteins localized in the outer membrane are essential for maintaining membrane integrity, controlling permeability of chemical substances across the membrane and behaving as virulence factors (Bos *et al.*, 2007; Buchanan, 1999; Costerton *et al.*, 1974; Delcour, 2002; Klebba & Newton, 1998; Koebnik *et al.*, 2000). The outer membrane is composed of inner

and outer leaflets which contain phospholipids and lipopolysaccharide (LPS), respectively (Ruiz *et al.*, 2006).

Since *V. parahaemolyticus* is able to survive in various habitats, the outer membrane is an important factor involved in bacterial adaptation. *V. parahaemolyticus* synthesizes three major surface antigens, namely, LPS or somatic O antigens, capsular polysaccharide or K antigens and flagellar or H antigens (Hsieh *et al.*, 2003). However, little is known about the composition of the outer membrane of *V. parahaemolyticus*. Isolation and characterization of outer membrane proteins (OMPs) from *V. parahaemolyticus* was first reported by Koga & Kawata (1983). This study identified five main OMPs with molecular weights of 44.0, 36.0, 33.5, 26.5, and 22.0 kilodaltons (kDa). The authors also demonstrated that the OMP profiles of *V. parahaemolyticus* altered under different NaCl concentrations. Heterogeneous OMP profiles were observed among *V. parahaemolyticus* with different K-serotypes and there was no association between OMP profile and serotype. A recent study of the outer membrane proteome identified 44 proteins including OmpU, OmpK, OmpA, OmpW, OmpV, TolC and iron-regulated proteins in *V. parahaemolyticus* (Li *et al.*, 2010a). OmpU functions as a major porin protein in *V. cholerae* and is also found in *V. parahaemolyticus* (Chakrabarti *et al.*, 1996; Mao *et al.*, 2007a). The protective role of OmpU for bacterial survival under stressful conditions such as acid- and bile-containing environments, as well as a role in colonization of host cells, have been reported in *V. parahaemolyticus* and *V. cholerae* (Simonet *et al.*, 2003; Sperandio *et al.*, 1995; Whitaker *et al.*, 2012; Wibbenmeyer *et al.*, 2002). OmpK is a channel-forming protein and receptor for the broad-host-range vibriophage KVP40 in *V. parahaemolyticus* (Inoue *et al.*, 1995a). It was also suggested to be a genus-specific antigen which could be used to develop

vaccines against pathogenic *Vibrio* species including *V. alginolyticus, V. vulnificus, V. parahaemolyticus, V. fluvialis, V. mimicus, and V. harveyi* (Li *et al.*, 2010b, c; Ningqiu *et al.*, 2008; Qian *et al.*, 2008). OmpA is a heat-modifiable integral protein that is generally present in the outer membranes of Gram-negative bacteria (Beher *et al.*, 1980). Beside its role in maintaining cell shape (Sonntag *et al.*, 1978), OmpA is immunogenic and has been used in programmes for developing vaccines against *Salmonella spp.* (Jeannin *et al.*, 2002; Lee *et al.*, 2010; Puohiniemi *et al.*, 1990). In *Vibrio* species, an OmpA-like protein has been detected in *V. cholerae*, and two and four OmpA orthologues have been identified in *V. alginolyticus* and *V. parahaemolyticus*, respectively (Alm, 1986; Li *et al.*, 2010a).

A number of OMPs involved in nutrient transport and osmoregulation in *V. parahaemolyticus* have been described in previous studies (Abdallah *et al.*, 2009a; Bhattacharya *et al.*, 2000; Koronakis *et al.*, 2004; Qian *et al.*, 2007; Xu *et al.*, 2004; Yang *et al.*, 2010). Magnesium transport from the environment across the outer membrane of *V. parahaemolyticus* is facilitated by a 40 KDa OMP (Bhattacharya *et al.*, 2000). The OmpW and OmpV proteins are important for marine bacteria as they are osmotic stress responsive OMPs. Expression of OmpW and OmpV varies with changing NaCl concentrations in *V. parahaemolyticus* (Qian *et al.*, 2007; Xu *et al.*, 2004). The OMP profiles of *V. parahaemolyticus* are altered under stressful conditions such as a low salt environment and exposure to gamma radiation (Abdallah *et al.*, 2009a, b; Yang *et al.*, 2010). TolC is an outer membrane efflux protein that allows export of a variety of substrates (Koronakis *et al.*, 2004). TolC family proteins are ubiquitous among Gram-negative bacteria. In *V. parahaemolyticus,* TolC contributes to resistance against antimicrobial peptides (AMPs) (Shen *et al.*, 2009). Under iron-depleted

conditions, *V. parahaemolyticus* has the ability to acquire iron through the action of the siderophore vibrioferrin and is able to utilize haem compounds as iron sources (Koga & Takumi, 1995; Yamamoto *et al.*, 1994, 1995a). PvuA1 (formerly named PsuA) and PvuA2 (formerly named PvuA) were identified as ferric vibrioferrin receptors in *V. parahaemolyticus* (Dai *et al.*, 1992; Funahashi *et al.*, 2002; Tanabe *et al.*, 2011; Yamamoto *et al.*, 1995a, b). These two vibrioferrin receptors require energy from different TonB systems. *V. parahaemolyticus* contains three TonB systems, TonB1, TonB2 and TonB3 (Kustusch *et al.*, 2011). PvuA1 obtains energy exclusively from TonB2 whereas PvuA2 obtains energy from both TonB1 and TonB2 (Tanabe *et al.*, 2011). PvuA1 and PvuA2 also have immunogenic properties and were suggested to be vaccine candidates against *V. parahaemolyticus* infection in the large yellow croaker, a native Asian fish (*Larimichthys crocea*) (Mao *et al.*, 2007b). Furthermore, a homologue of *Iut*, a gene encoding an OMP receptor for ferric aerobactin in *E. coli*, has been described in *V. parahaemolyticus* (Funahashi *et al.*, 2003). The role of other OMPs as potential vaccine candidates against pathogenic *V. parahaemolyticus* have also been described in previous studies (Li *et al.*, 2010b, c; Ningqiu *et al.*, 2008; Yuan *et al.*, 2011).

OMP-encoding genes can be predicted from the genome by using bioinformatic approaches (Gromiha & Suwa, 2006; Gromiha, 2005; Jackups *et al.*, 2006; Juncker *et al.*, 2003). OMP predictive tools are able to predict the OMPs encoded by the genome by determining subcellular localization, β-barrel conformation and lipoprotein composition from the amino acid sequences of total open reading frames in the genome (Bagos *et al.*, 2004; Berven *et al.*, 2004, 2006; Gardy *et al.*, 2005; Garrow *et al.*, 2005; Imai *et al.*, 2008; Juncker *et al.*, 2003; Ou *et al.*, 2008; Yu *et al.*, 2004). Integration of predicted proteins by these

software tools will generate a list of putative OMPs from a given bacterial genome (E-komon *et al.*, 2012). Furthermore, the outer membrane proteome of bacterial isolates can be analyzed using a combination of techniques including SDS-PAGE and mass-spectrometry.

The aim of the present study was to predict OMPs encoded by the genome of the clinical *V. parahaemolyticus* isolate RIMD2210633 using bioinformatic approaches, and then to apply comparative proteomics to identify the OMPs present in outer membrane fractions of eight representative *V. parahaemolyticus* isolates recovered from different sources including clinical samples, human carriers, seafood and water in Thailand. Comparative analysis of the outer membrane proteomes of these different strains of *V. parahaemolyticus* will contribute to our understanding of the molecular adaptation of this organism to different ecological niches. In particular, this knowledge will improve our understanding of the molecular basis of virulence in *V. parahaemolyticus.*

## 4.2 Materials and methods

### 4.2.1 Bioinformatic prediction of OMPs from the genome of *V. parahaemolyticus isolate* RIMD2210633

The publicly available genome of a clinical *V. parahaemolyticus* isolate (GenBank ID: BA00031.2 and BA00032.2) was used for the bioinformatic analysis. All *V. parahaemolyticus* protein sequences (4,832 open reading frames) were retrieved from NCBI. The genome was examined by bioinformatic approaches according to the workflow described previously to predict proteins which localize to the outer membrane (E-komon *et al.*, 2012). The OMPs of *V. parahaemolyticus* were

predicted by prediction software used in this study with the exception that Proteome Analyst (PA) was not used (since it is no longer available). The genome was analyzed by three categories of bioinformatic prediction software, using a total of nine prediction tools. Subcellular localization predictors included PSORTb (Gardy *et al.*, 2005), CELLO (Yu *et al.*, 2004) and SOSUI-GramN (Imai *et al.*, 2008); ß-barrel predictors included TMB-Hunt (Garrow *et al.*, 2005), TMBETADISC-RBF (Ou *et al.*, 2008), MCMBB (Bagos *et al.*, 2004) and BOMP (Berven *et al.*, 2004); and outer membrane lipoprotein predictors included LipoP (Juncker *et al.*, 2003) and LIPO (Berven *et al.*, 2006). A consensus prediction framework was developed according to the following scheme (Fig. 4.1).

Proteins that were predicted to be localized to the outer membrane by at least two subcellular localization predictors or to have a ß-barrel conformation by at least three ß-barrel predictors or to be outer membrane lipoproteins by at least one lipoprotein predictor, were considered to be putative OMPs. A list of putative OMPs was produced by integrating the results from each of the predictor categories. These OMPs were further examined using additional domain, homology and public database searches (textmining) to assign likely molecular functions and to predict their subcellular localizations with a higher degree of confidence. Based on this additional information, each putative OMP was assigned to one of three categories: (1) confidently predicted OMPs, (2) putative OMPs whose subcellular locations cannot confidently be assigned, or (3) false positives.
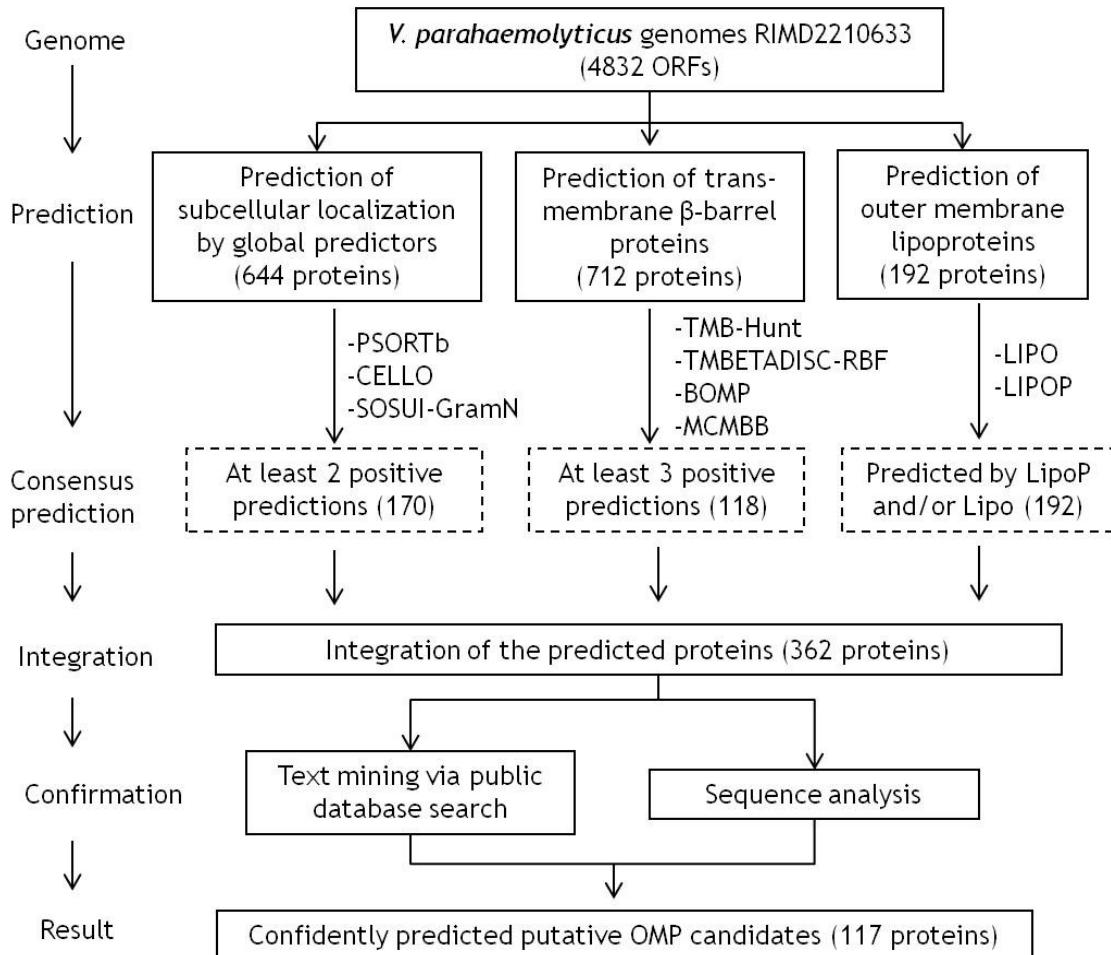
**Figure 4.1.** Diagram representing the workflow of bioinformatic prediction of putative OMPs from the genome of *V. parahaemolyticus*. Nine predictors were categorized into 3 groups: subcellular localization predictors, transmembrane ß-barrel protein predictors and outer membrane lipoprotein predictors. This diagram is adapted from the bioinformatic workflow developed by E. Komon *et al.* (2012)

## *4.2.2 Bacterial isolates and growth conditions*

Since clinical isolates from Thailand were resolved into two main genetic clusters characterized by MLST (Fig. 2.52; clusters 2 and 4), nine isolates from clinical clusters 2 (4 isolates) and 4 (5 isolates) were first selected to study variation of OMP profiles of isolates representing identical STs or the same genetic cluster. The properties of these strains are shown in Table 4.1. This comparison was used to select two clinical strains (one from each of clusters 2 and 4) for proteomic analysis (see below).

Eight *V. parahaemolyticus* isolates from different epidemiological sources were selected for comparative proteomic analysis. The isolates were selected to represent important lineages/clonal groups among the 101 Thai isolates and a Japanese type strain (ATCC 17802$^T$) previously characterized by MLST (Fig. 2.52). The eight representative isolates were recovered from clinical samples (2 isolates), human carriers (2 isolates), oyster (1 isolate), shrimp tissue (2 isolates) and water from a shrimp farm (1 isolate). The properties of each strain are shown in Table 4.2.

The isolates were stored at -80°C in 50% (v/v) glycerol in tryptone soy broth (TSB) containing 3% (w/v) NaCl and were subcultured on tryptone soy agar (TSA) containing 3% (w/v) NaCl overnight at 37°C. For preparation of outer membrane fractions, liquid starter cultures were prepared by inoculating a few colonies of overnight growth into 15 ml volumes of TSB containing 3% (w/v) NaCl and incubating overnight at 37°C with shaking at 120 rpm. Eight hundred microlitres of overnight culture were inoculated into a 2-litre Ehrlenmeyer flask containing 400 ml of TSB containing 3% (w/v) NaCl, which was incubated at 37°C with shaking at 120 rpm until an OD$_{600nm}$ of 0.8-0.9 was achieved.

**Table 4.1.** Properties of nine clinical isolates of *V. parahaemolyticus* selected for comparative OMP analysis

| Isolate | Source | MLST cluster | Year of Isolation | Serotype (O:K) | Sequence type (ST) | Haemolysin gene (*tdh/trh*) |
|---|---|---|---|---|---|---|
| VP2 | Food poisoning agent (type strain) | Cluster 2 | 1950 | O1:K1 | 1 | -/- |
| VP166 | Clinical sample | Cluster 2 | 1990 | O1:K1 | 83 | +/+ |
| VP172 | Clinical sample | Cluster 2 | 1990 | O1:K1 | 83 | +/+ |
| VP176 | Clinical sample | Cluster 2 | 1990 | O1:K1 | 264 | +/+ |
| VP178 | Clinical sample | Cluster 4 | 1991 | O1:K69 | 262 | +/+ |
| VP180 | Clinical sample | Cluster 4 | 1990 | O8:K22 | 262 | +/- |
| VP182 | Clinical sample | Cluster 4 | 1990 | O1:K69 | 262 | -/+ |
| VP184 | Clinical sample | Cluster 4 | 1990 | O4:K11 | 262 | +/- |
| VP188 | Clinical sample | Cluster 4 | 1991 | O1:K69 | 262 | +/+ |

**Table 4.2.** Properties of eight representative isolates of *V. parahaemolyticus* selected for comparative proteomic analysis

| Isolate | Source | Year of isolation | Serotype (O:K) | Sequence type (ST) | Haemolysin gene (*tdh/trh*) |
|---|---|---|---|---|---|
| VP166 | Clinical sample | 1990 | O1:K1 | 83 | +/+ |
| VP178 | Clinical sample | 1991 | O1:K69 | 262 | +/+ |
| VP132 | Human carriage | 2003 | O3:K46 | 3 | +/- |
| VP138 | Human carriage | 2003 | O11:K5 | 255 | +/+ |
| VP204 | Oyster | 2003 | O1:K64 | 267 | -/- |
| VP84 | Shrimp tissue | 2007 | O10:K71 | 251 | -/- |
| VP112 | Shrimp tissue | 2007 | O9:K44 | 246 | -/- |
| VP44 | Shrimp-farm water | 2008 | O9:K23 | 244 | -/- |

## *4.2.3 Preparation of OMPs*

Outer membrane proteins were prepared by Sarkosyl extraction as previously described (Davies, 2003; Davies *et al.*, 2003a, b, 2004). Bacterial growth was stopped by chilling the growth media in iced water for 5 min. The bacterial cells were harvested by centrifugation at 13,000 × *g* for 20 min at 4°C. The pellet was washed in 50 ml of 20 mM Tris/HCl (pH 7.2) and centrifuged at 12,000 × *g* for 20 min at 4°C. The cell pellet was resuspended in 7 ml of 20 mM Tris/HCl (pH 7.2) and sonicated on ice for 5 min using a Soniprep sonicator (12 microns amplitude). The sonicated samples were adjusted to a total volume of 10 ml with 20 mM Tris/HCl (pH 7.2) and centrifuged at 11,000 × *g* for 30 min at 4°C to remove unbroken cells. The supernatants were centrifuged at 84,000 × *g* for 1 h at 4°C in a Sorvall ultracentrifuge to pellet the cell envelopes. The gelatinous pellets were thoroughly resuspended in 0.5 % sodium *N*-lauroylsarcosine (Sarkosyl; Sigma) for 20 min at room temperature to solubilise the cytoplasmic membranes and centrifuged at 84,000 × *g* for 1 h at 4°C to pellet the outer membranes. The gelatinous outer membranes were resuspended in 20 mM Tris/HCl (pH 7.2) and centrifuged at 84,000 × *g* for 1 h at 4°C. The final pellets were resuspended in approximately 1 ml of 20 mM-Tris/HCl (pH 7.2). Fifty microlitre aliquots of these suspensions were transferred to separate tubes and the protein concentrations determined by the modified Lowry procedure (Markwell *et al.*, 1978). One hundred microlitre aliquots of the outer membrane suspensions were adjusted to 2 mg ml$^{-1}$ with 20 mM Tris/HCl (pH 7.2) and stored at -80°C.

## *4.2.4 Gel-based proteomic analysis*

Twenty micrograms of each OMP sample were separated by 1D SDS-PAGE in a 12% linear polyacrylamide gel using the SDS discontinuous system (Laemmli, 1970) and the Hoefer SE600 electrophoresis equipment as previously described (Davies, 2003; Davies *et al.*, 2003a, b, 2004). Proteins were visualised by staining the polyacrylamide gel with Coomassie brilliant blue. A total of 158 gel pieces were manually excised and individual gel pieces were stored in separate wells of 96-well plates to be subjected to in-gel digestion for protein extraction prior to identification via mass spectrometry analysis. The gel pieces included protein bands from all eight isolates and gel fractions without any visible proteins for VP132. The gel pieces were washed with 100 mM $NH_4HCO_3$ (Cat No. V5111, Promega, Madison, WI, USA) for 30 min and then for 1 h with 100 mM $NH_4HCO_3$ in 50% (v/v) acetonitrile. After each wash all solvent was discarded. The gel slices were then dehydrated with 100% (v/v) acetonitrile for 10 min prior to solvent being removed and the slices dried completely by vacuum centrifuge. The dry gel pieces were then rehydrated with 10 µl trypsin at a concentration of 20 ng $µl^{-1}$ in 25 mM $NH_4HCO_3$ and proteins allowed to digest overnight at 37˚C.

The liquid contents of each well were transferred to a fresh 96-well plate, and the gel pieces were washed for 10 min at room temperature with 10 µl of 50% (v/v) acetonitrile. This wash was pooled with the first extract and the tryptic peptides were dried by vacuum centrifugation. A sufficient amount of 1% (v/v) formic acid was added to cover the gel pieces and these were incubated for 10 min at room temperature. The liquid was pooled with the dried tryptic peptide from the previous extract. A sufficient amount of 50% (v/v) acetonitrile was added to cover the gel pieces and these were incubated for 10 min at room

temperature. The liquid was pooled with the dried tryptic peptide from the previous extract. The gel pieces were repeatedly treated one more time by formic acid extraction and acetronitrile washing, respectively, as described above. In each case, the liquid was pooled with the previous extract and finally dried down by vacuum centrifugation. The samples were stored at -20°C.

## 4.2.5 ESI-TRAP and data analysis

Tryptic peptides were solubilized in 0.5 % (v/v) formic acid and fractionated on a nanoflow UHPLC system (Thermo RSLCnano) before being analyzed by electrospray ionisation (ESI) mass spectrometry on an Amazon Ion Trap MS/MS (Bruker Daltonics). Peptide separation was performed on a Pepmap C18 reverse phase column (LC Packings), using a 5 - 85% (v/v) acetonitrile gradient (in 0.5% (v/v) formic acid) run over 45 min at a flow rate of 0.2 $\mu$l min$^{-1}$. Mass spectrometric (MS) analysis was performed using a continuous duty cycle of survey MS scan followed by up to five MS/MS analyses of the most abundant peptides, choosing the most intense multiply charged ions with dynamic exclusion for 120 s.

MS data were processed using Data Analysis software (Bruker) and the automated Matrix Science Mascot Daemon server (v2.1.06). Protein identifications were assigned using the *V. parahaemolyticus* RIMD2210633 protein database with methionine oxidation selected as a variable modification and carbamidomethylation as a fixed modification, allowing a mass tolerance of 0.4 Da for both MS and MS/MS analyses, and one possible missed cleavage per peptide. Only proteins identified with a significant MOWSE score ($p \leq 0.005$) were accepted.

# 4.3 Results

## *4.3.1 Bioinformatic prediction of OMPs in the V. parahaemolyticus genome*

Nine different bioinformatic tools were used to predict putative OMPs encoded within the genome sequence of clinical *V. parahaemolyticus* isolate RIMD2210633 following the bioinformatic workflow of E-Komon *et al*. (2012) (Fig. 4.1). The bioinformatic tools used in this study were categorized into three groups: subcellular localization predictors (PSORTb, CELLO, SOSUI-GramN), ß-barrel protein predictors (TMB-Hunt, TMBETADISC-RBP, BOMP, MCMBB), and outer membrane lipoprotein predictors (LIPO and LipoP). Six hundred and forty-four proteins were predicted by the subcellular localization predictors, 712 proteins were predicted by the transmembrane ß-barrel protein predictors, and 192 proteins were predicted by the outer membrane lipoprotein predictors (Fig. 4.2). Predicted OMPs from these three categories of bioinformatic tools were processed through a consensus prediction in a prediction framework (Fig. 4.1) and the consensus proteins from each category were subsequently integrated.

(a) Subcellular localization (644)

PSORTb (123)   CELLO (496)

18    36    334

61

8    65

122

SOSUI-GramN (256)

(b) Trans membrane beta barrel (712)

TMB-Hunt (200)    TMBETADISC-RBF (464)

91    69

16    23    14

111    43    2

24    24

241    0

48    1

MCMBB (489)    5    BOMP (118)

(c) Lipoprotein (192)

LIPO (166)    LIPOP (160)

32    152    8
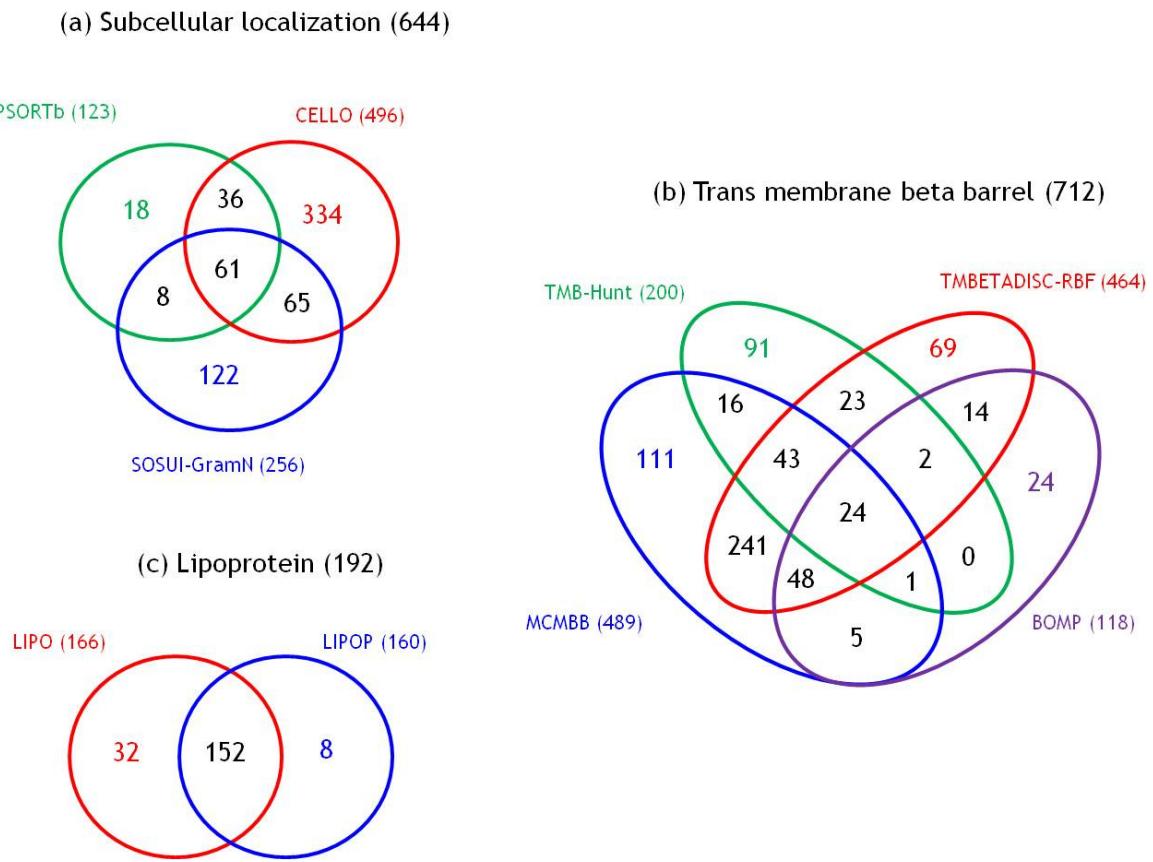
**Figure 4.2.** Within-group comparisons of numbers of predicted proteins by three groups of predictors: (a) subcellular localization, (b) transmembrane ß-barrel protein and (c) outer membrane lipoprotein predictors. The corresponding colour of each predictor and numbers represents the number of proteins predicted by that predictor. Black represents the number of proteins predicted by at least two predictors in that group.

One hundred and seventy consensus proteins were predicted by the subcellular localization predictors, 118 consensus proteins were predicted by the transmembrane ß-barrel protein predictors and 192 consensus proteins were predicted by the outer membrane lipoprotein predictors (Figs. 4.1 and 4.3). After integration of these predicted proteins, 362 annotated proteins were predicted to be putative OMPs in the *V. parahaemolyticus* genome (Figs. 4.1 and 4.3). Eight OMPs were predicted by predictive tools from all three groups, 102 proteins were predicted by predictive tools from two different groups, and 252 OMPs were predicted by predictive tools from only one group (Fig. 4.3).

The 362 predicted OMPs were evaluated by BLAST searching of public databases (http://www.uniprot.org), and by homology and literature searches, to confirm the sub-cellular localization of the predicted OMPs with a higher degree of confidence (Fig. 4.1). By this process, 117 (32.3%) proteins were identified as confidently predicted OMPs (Figs. 4.1 and 4.4). However, 229 (63.3%) of the predicted proteins could not be localized to any particular subcellular compartment and were considered to be non-OMPs (Fig. 4.4). Furthermore, 16 false positive predictions were identified and these included six (1.7%) proteins localizing in the periplasm, four (1.1%) inner membrane proteins, three (0.8%) extracellular proteins and three (0.8%) cytoplasmic proteins (Fig. 4.4). Although these proteins were considered to be false positives since they are not localized in the outer membrane but were predicted by OMP predictors, they likely include some true OMPs.
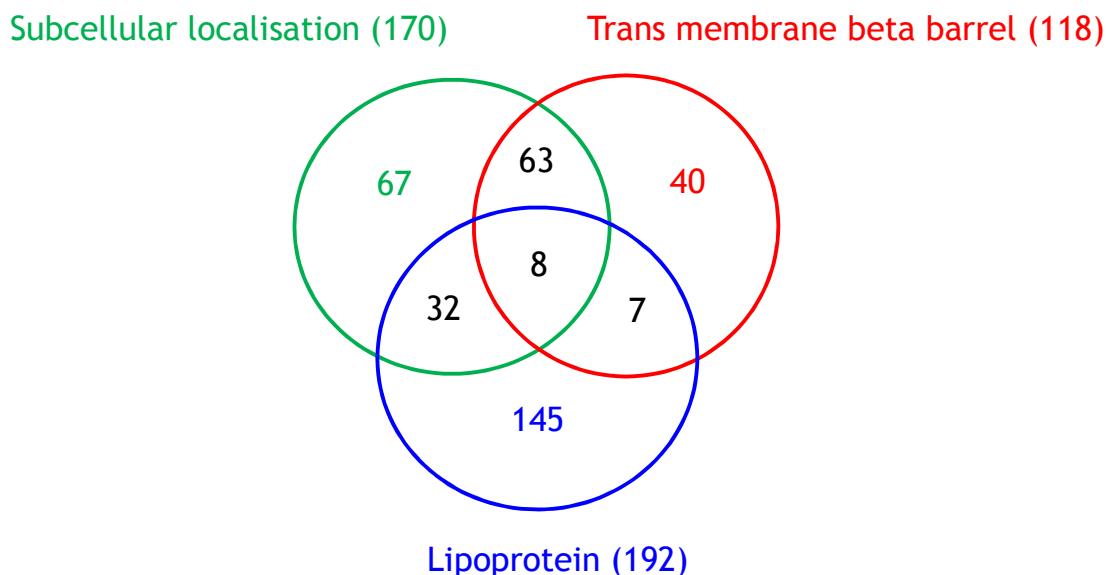
**Figure 4.3.** Between-group comparison of the numbers of proteins predicted by the three groups of predictors: subcellular location predictors, transmembrane ß-barrel protein predictors and outer membrane lipoprotein predictors. The corresponding colour of each group of predictors and numbers represent the number of proteins predicted by that group. Black represents the number of shared proteins predicted by at least two groups of predictors
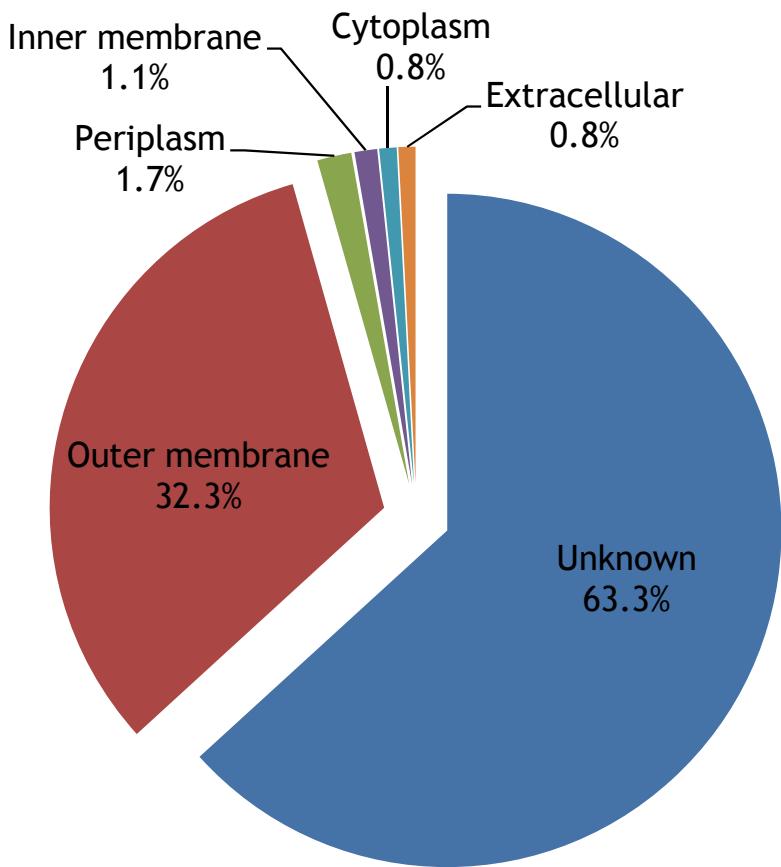
**Figure 4.4.** Subcellular locations of 362 putative OMPs predicted by 9 bioinformatic prediction tools of *V. parahaemolyticus* proteome after domain, homology and literature searches had been performed on each protein

## *4.3.2 Functional classifications of confidently predicted OMPs*

The functional classification of the 117 confidently predicted OMPs is detailed in Table 4.3 and summarized in Fig. 4.5. The distribution of the confidently predicted OMPs between chromosomes 1 and 2 is also shown in Fig. 4.6. Of 117 predicted OMPs, 64 are located in chromosome 1 and 53 are located in chromosome 2 (Fig. 4.6A). The predicted OMPs can be classified into seven different functional groups (Fig. 4.5). Thirty three (28.2%) proteins were predicted to be involved in outer membrane biogenesis and integrity (Fig. 4.5), of which 22 are located in chromosome 1 and 11 in chromosome 2 (Fig. 4.6B). Forty nine (41.9%) proteins were predicted to be involved in transport and receptor activity (excluding those involved in iron uptake) (Fig. 4.5), of which 25 are located in chromosome 1 and 24 in chromosome 2 (Fig. 4.6B). Nine (7.7%) proteins were predicted to be involved in iron binding and TonB receptor activity (Fig. 4.5), of which two are located in chromosome 1 and seven in chromosome 2 (Fig. 4.6B). Eight (6.8%) proteins were predicted to be involved in flagella and motor activity (Fig. 4.5), of which three are located in chromosome 1 and five in chromosome 2 (Fig. 4.6B). Eight (6.8%) proteins were predicted to be involved in enzyme activity (Fig. 4.5), of which six are located in chromosome 1 and two in chromosome 2 (Fig. 4.6B). Six (5.1%) proteins were predicted to be involved in adherence and colonization (Fig. 4.5), of which three are located in chromosome 1 and three in chromosome 2 (Fig. 4.6B). Four (3.4%) proteins were predicted to be involved in the other activities (Fig. 4.5), of which three are located in chromosome 1 and one in chromosome 2 (Fig. 4.6B).
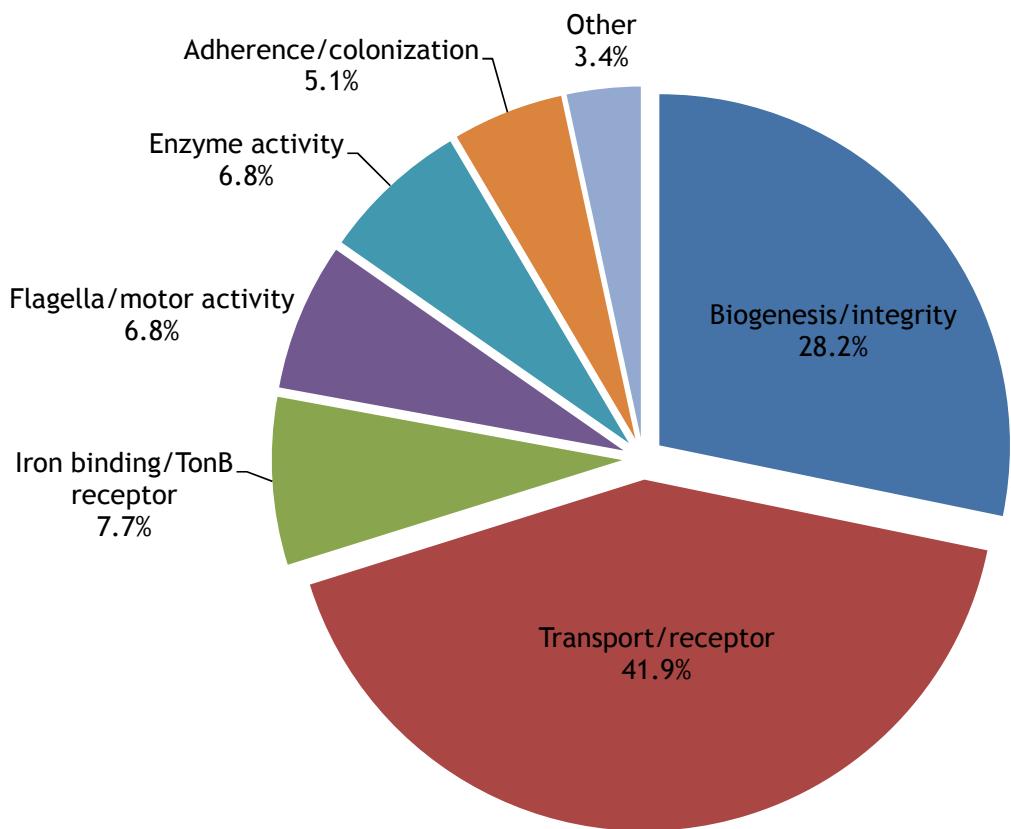
**Figure 4.5.** Functional classification of 117 confidently predicted OMPs present in the *V. parahaemolyticus* genome after the text mining process and further domain and homology searches.

(A)



(B)
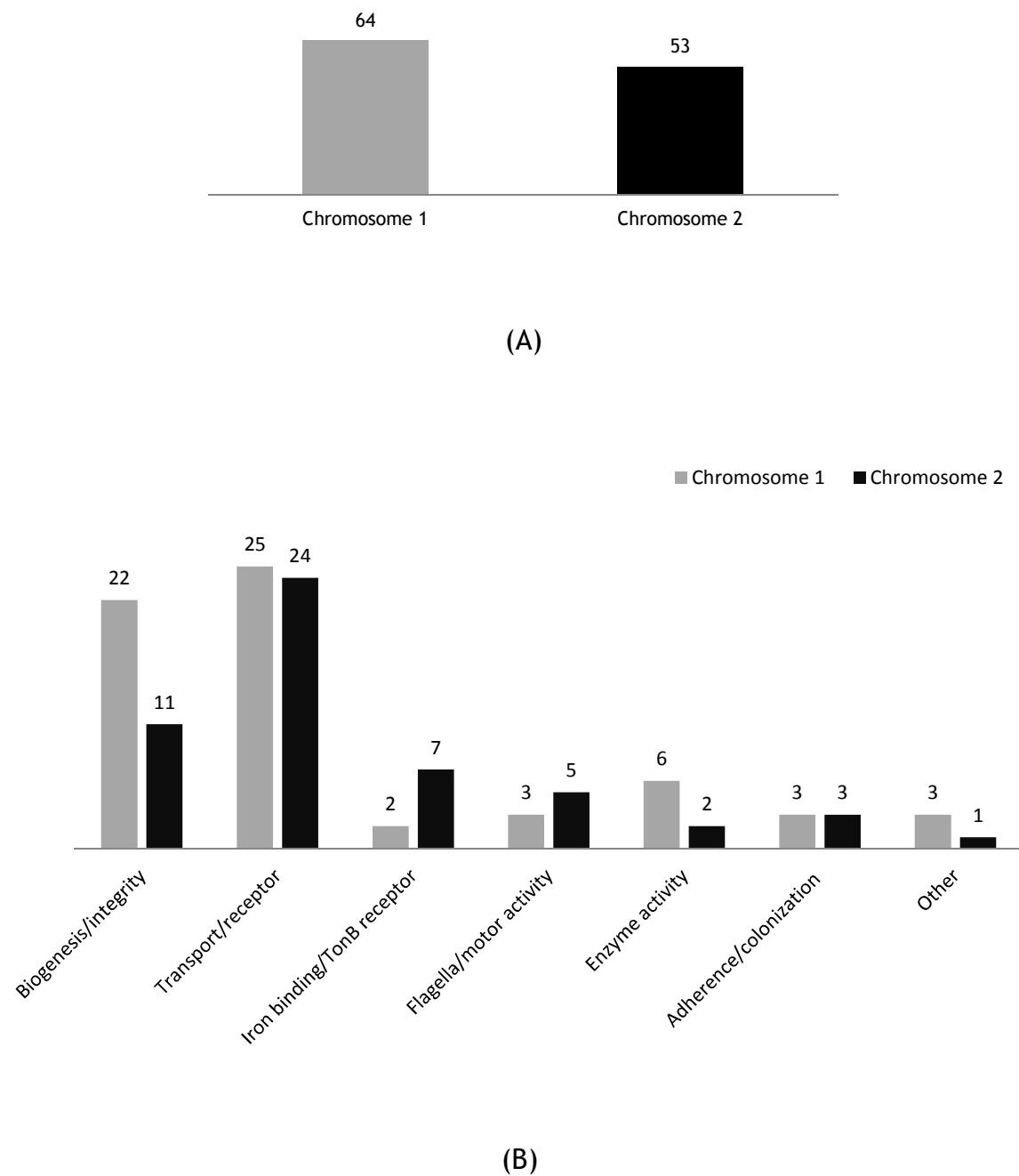
**Figure 4.6.** (A) Distribution of 117 confidently predicted proteins from the *V. parahaemolyticus* genome between chromosomes 1 and 2 and (B) distribution of 117 confidently predicted proteins categorized by functional class among chromosomes 1 and 2. Numbers above each bar represent the number of predicted OMPs identified in that group.

**Table 4.3.** Functional classifications of 117 confidently predicted OMPs encoded by *V. parahaemolyticus* strain RIMD2210633 genome

| No. | Protein ID | Gene name | Chromosome | Function and property |
|---|---|---|---|---|
| **1. Outer membrane biogenesis/integrity (33)** | | | | |
| 1 | VP0764 | OmpA1 | 1 | Cell integrity, surface antigen, flagellar motility |
| 2 | VP0741 | LolB | 1 | Sorting and outer membrane localization of lipoproteins, outer membrane protein receptor |
| 3 | VP2310 | YaeT | 1 | Bacterial surface antigen, outer membrane protein assembly |
| 4 | VP0636 | OmpA domain protein | 1 | Structural integrity |
| 5 | VP0339 | LptD/organic solvent tolerance protein | 1 | Outer membrane assembly, response to organic substance |
| 6 | VP0726 | Rare lipoprotein B | 1 | Outer membrane assembly |
| 7 | VP0611 | YfgL lipoprotein | 1 | Outer membrane assembly |
| 8 | VP0868 | Slp lipoprotein | 1 | Starvation inducable protein |
| 9 | VP0647 | SmpA/ small protein A | 1 | Structural integrity |
| 10 | VP1061 | Peptidoglycan-associated lipoprotein | 1 | Structural integrity |
| 11 | VP0944 | Putative outer membrane protein | 1 | Structural integrity |
| 12 | VP1632 | Putative outer membrane protein | 1 | Structural integrity |
| 13 | VP0967 | Hypothetical protein VP0967 | 1 | Curli production assembly/transport |
| 14 | VP0219 | Hypothetical protein VP0219 | 1 | OmpA family, transmembrane domain |
| 15 | VP0307 | Hypothetical protein VP0307 | 1 | Surface antigen |
| 16 | VP2733 | Hypothetical protein VP2733 | 1 | Surface antigen |
| 17 | VP1410 | Hypothetical protein VP1410 | 1 | OmpA-like, transmembrane domain |
| 18 | VP1475 | Hypothetical protein VP1475 | 1 | OmpA family |
| 19 | VP1455 | Hypothetical protein VP1455 | 1 | OmpA family |

Table 4.3. (continued)

| No. | Protein ID | Gene name | Chromosome | Function and property |
|---|---|---|---|---|
| 20 | **VP1356** | Hypothetical protein VP1356 | 1 | Bacterial surface antigen |
| 21 | **VP1390** | Hypothetical protein VP1390 | 1 | OmpA family |
| 22 | **VP0558** | Hypothetical protein VP0558 | 1 | Outer membrane assembly lipoprotein YfiO, structural molecule activity |
| 23 | **VPA0248** | OmpA2 | 2 | Structural integrity, surface antigen |
| 24 | **VPA0318** | OmpV | 2 | Structural integrity |
| 25 | **VPA1404** | CpsB | 2 | Regulation of capsular polysaccharide biosynthesis |
| 26 | **VPA1469** | Outer membrane lipoprotein | 2 | Murein-lipoprotein found in the enterobacterial outer membrane lipoprotein |
| 27 | **VPA0312** | Hypothetical protein VPA0312 | 2 | Surface antigen |
| 28 | **VPA1353** | Putative outer membrane protein | 2 | OmpA family |
| 29 | **VPA0810** | Hypothetical protein VPA0810 | 2 | OMP beta barel, OMP85_target, OmpA family |
| 30 | **VPA0242** | Hypothetical protein VPA0242 | 2 | OMP beta barel, OMP85_target, OmpA family |
| 31 | **VPA0548** | Hypothetical protein VPA0548 | 2 | OmpA family |
| 32 | **VPA0731** | Hypothetical protein VPA0731 | 2 | OmpA family |
| 33 | **VPA1440** | Hypothetical protein VPA1440 | 2 | OmpA family |

**2. Transport/receptor (49)**

| No. | Protein ID | Gene name | Chromosome | Function and property |
|---|---|---|---|---|
| 34 | **VP0425** | TolC | 1 | Outer membrane protein efflux involving in type I secretion |
| 35 | **VP2467** | OmpU | 1 | Ion transport, porin activity |
| 36 | **VP2362** | OmpK | 1 | Nucleoside transmembrane transporter activity |
| 37 | **VP1901** | OmpX | 1 | Transport activity |
| 38 | **VP2938** | BtuB | 1 | Active translocation of vitamin B12, vitamin B12 receptor |
| 39 | **VP1690** | VscJ | 1 | Lipoprotein involved in type III secretion system |

Table 4.3. (continued)

| No. | Protein ID | Gene name | Chromosome | Function and property |
|---|---|---|---|---|
| 40 | **VP1696** | YscC | 1 | Belongs to the GSP D family, Transport |
| 41 | **VP0132** | GspC | 1 | Transporter activity, involved in type II general secretion pathway C |
| 42 | **VP0133** | GspD | 1 | Transporter activity, involved in type II general secretion pathway D |
| 43 | **VP2746** | PilQ | 1 | Transporter activity, pilus assembly |
| 44 | **VP1700** | YscW | 1 | Lipoprotein involved in type III secretion system |
| 45 | **VP0802** | OprD family | 1 | Outer membrane porin |
| 46 | **VP1667** | Putative PopN | 1 | Type III secretion regulator |
| 47 | **VP1998** | Putative TolC | 1 | Transporter activity, TolC family |
| 48 | **VP0760** | Putative chitoporin  ChiP | 1 | Transport glycosidases for extracellular chitooligosaccharides |
| 49 | **VP1008** | Porin qsr prophage protein | 1 | Ion-selective channels for small hydrophilic molecules |
| 50 | **VP2212** | Long-chain fatty acid transport protein | 1 | Translocation of long-chain fatty acids, a receptor for the bacteriophage T2, FadL related protein |
| 51 | **VP2213** | Long-chain fatty acid transport protein | 1 | Translocation of long-chain fatty acids, a receptor for the bacteriophage T2, FadL related protein |
| 52 | **VP1631** | AggA | 1 | Outer membrane protein efflux |
| 53 | **VP1634** | AggA | 1 | Outer membrane protein efflux |
| 54 | **VP0168** | Hypothetical protein VP0168 | 1 | Transporter and receptor activity |
| 55 | **VP0756** | Hypothetical protein VP0756 | 1 | Specific porin, KdgM family |
| 56 | **VP1757** | Hypothetical protein VP1757 | 1 | Porin family |
| 57 | **VP1713** | Hypothetical protein VP1713 | 1 | Transporter activity |
| 58 | **VP1412** | Hypothetical protein VP1412 | 1 | Lipoproteins involved in type VI secretion |

Table 4.3. (continued)

| No. | Protein ID | Gene name | Chromosome | Function and property |
|---|---|---|---|---|
| 59 | **VPA0166** | OmpN | 2 | Non specific porin |
| 60 | **VPA0096** | OmpW | 2 | Transport of small hydrophobic molecules |
| 61 | **VPA1644** | Maltoporin LamB | 2 | Involved in the transport of maltose and maltodextrins |
| 62 | **VPA1339** | VscC2 | 2 | Transport activity , involved in type III secretion |
| 63 | **VPA1602** | Putative Wza | 2 | Capsular polysaccharide transmembrane transporter activity |
| 64 | **VPA0860** | Long-chain fatty acid transport protein | 2 | Translocation of long-chain fatty acids, a receptor for the bacteriophage T2, FadL related protein |
| 65 | **VPA0954** | AggA | 2 | Outer membrane protein efflux |
| 66 | **VPA1018** | Lipoprotein Blc | 2 | Transport acitvity, lipid binding |
| 67 | **VPA0526** | Putative OmpU | 2 | Transporter activity |
| 68 | **VPA0320** | Putative lipoprotein | 2 | Involved in signalling by the Cpx pathway |
| 69 | **VPA0807** | Putative multidrug resistance protein | 2 | Transmembrane transport, efflux pump |
| 70 | **VPA0364** | Putative efflux protein | 2 | Transmembrane transport, efflux pump |
| 71 | **VPA0362** | putative outer membrane protein | 2 | Transport acitvity, lipid binding |
| 72 | **VPA1466** | Putative TonB system receptor | 2 | Receptor activity, transporter activity, sequence similarity tonB-dependent receptor family |
| 73 | **VPA1579** | Putative outer membrane protein | 2 | Transporter activity, outer membrane efflux lipoprotein |
| 74 | **VPA0225** | Putative efflux pump channel protein | 2 | Transporter activity, lipid binding |
| 75 | **VPA1745** | Putative outer membrane protein | 2 | Porin domain |
| 76 | **VPA0482** | Putative outer membrane cation efflux protein | 2 | Transporter activity |

Table 4.3. (continued)

| No. | Protein ID | Gene name | Chromosome | Function and property |
|---|---|---|---|---|
| 77 | **VPA0472** | putative long-chain fatty acid transport protein | 2 | Translocation of long-chain fatty acids, a receptor for the bacteriophage T2, FadL related protein |
| 78 | **VPA0316** | Putative outer membrane protein | 2 | Transmembrane and porin domain |
| 79 | **VPA0211** | Hypothetical protein VPA0211 | 2 | Outer membrane receptor protein |
| 80 | **VPA0085** | Hypothetical protein VPA0085 | 2 | Specific porin, KdgM family |
| 81 | **VPA0018** | Hypothetical protein VPA0018 | 2 | Receptor activity |
| 82 | **VPA1042** | Hypothetical protein VPA1042 | 2 | Lipoprotein involved in type VI secretion |
| **3. Iron binding/TonB receptor (9)** | | | | |
| 83 | **VP2602** | IrgA | 1 | Transport and receptor activity |
| 84 | **VP1220** | Putative 83 kDa decaheme outer membrane cytochrome c | 1 | Heme binding |
| 85 | **VPA1657** | PvuA1 (PsuA) | 2 | Ferric siderophore receptor |
| 86 | **VPA1656** | PvuA2 (PvuA) | 2 | Ferric vibrioferrin receptor |
| 87 | **VPA0979** | LutA | 2 | Ferric aerobactin receptor |
| 88 | **VPA0150** | FhuE | 2 | Ferrichrome-iron receptor |
| 89 | **VPA1435** | FhuA | 2 | Iron(III) compound receptor, siderophore transport |
| 90 | **VPA0882** | HutA | 2 | Heme transport |
| 91 | **VPA0664** | Putative Fe-regulated protein B | 2 | Enterobactin receptor |
| **4. Flagella/motor activity (8)** | | | | |
| 92 | **VP1267** | Putative lipoprotein | 1 | Flagellar motility, motor activity |
| 93 | **VP0782** | FlgH1 | 1 | Flagellar motility, motor activity |
| 94 | **VP2111** | MotY | 1 | Flagellar motility, motor activity |
| 95 | **VPA0270** | FlgH2 | 2 | Flagellar motility, motor activity |
| 96 | **VPA0271** | FlgL1 | 2 | Flagellar motility, motor activity |

Table 4.3. (continued)

| No. | Protein ID | Gene name | Chromosome | Function and property |
|---|---|---|---|---|
| 97 | **VPA1539** | Putative sodium-type flagellar protein MotY | 2 | Flagellar motility, motor activity, ompA family |
| 98 | **VPA1186** | Outer membrane protein OmpA | 2 | Flagellar motility, motor activity, ompA family |
| 99 | **VPA1503** | CsuE | 2 | Spore coat protein U domain, motility, biofilm formation |
| **5. Enzyme activity (8)** | | | | |
| 100 | **VP0748** | NutA | 1 | Degradation of extracellular 5'-nucleotides for nutritional requirement |
| 101 | **VP2369** | MtlA | 1 | Murein degradation, peptidoglycan metabolic process |
| 102 | **VP2628** | MtlC | 1 | Murein degradation, peptidoglycan metabolic process |
| 103 | **VP0665** | MtlF | 1 | Murein degradation, peptidoglycan metabolic process |
| 104 | **VP1260** | Outer membrane phospholipase subunit A | 1 | Lipid metabolic process, phospholipase activity |
| 105 | **VP2496** | Hypothetical protein VP2496 | 1 | Lipid metabolic process |
| 106 | **VPA1615** | Putative outer membrane protein | 2 | Protein disulfide oxidoreductase activity |
| 107 | **VPA0514** | Putative transmembrane protein | 2 | Permease activity |
| **6. Adherence/colonization (6)** | | | | |
| 108 | **VP2704** | MshL | 1 | Pilus assembly, protein secretion |
| 109 | **VP1752** | PilF | 1 | Binding activity, involved in type IV pilus biogenesis |
| 110 | **VP1767** | Hypothetical protein VP1767 | 1 | Adhesion, homologous to the Invasins of pathogenic *Yersinia* and intimins of pathogenic *E. coli* |
| 111 | **VPA1442** | Putative hemagglutinin/hemolysin-like protein | 2 | Adhesion and binding activity |
| 112 | **VPA1376** | AcfD | 2 | Accessory colonizing factor |
| 113 | **VPA0695** | AcfA | 2 | Accessory colonizing factor |

Table 4.3. (continued)

| | No. | Protein ID | Gene name | Chromosome | Function and property |
|---|---|---|---|---|---|
| **7. Other (4)** | | | | | |
| | 114 | **VP1192** | Pcp | 1 | Unknown |
| | 115 | **VP2272** | Lipoprotein-34 NlpB | 1 | Composition of outer membrane vesicle, unknown function |
| | 116 | **VP2042** | Hypothetical protein VP2042 | 1 | Unknown |
| | 117 | **VPA0396** | Putative outer membrane lipoprotein | 2 | Unknown |

From these results, it can be seen that the location of predicted OMPs involved in outer membrane transport and receptor function, flagella and motor activity, and adherence and colonization, are similarly distributed between both chromosomes. In contrast, twice as many OMPs involved in outer membrane biogenesis and integrity are located in chromosome 1 (n=22) than in chromosome 2 (n=11). Furthermore, iron binding and TonB receptor proteins are located predominantly in chromosome 2, whereas proteins involved in enzyme activity are located predominantly in chromosome 1.

In the group involved in biogenesis and integrity activity (Table 4.3), two OmpA proteins, OmpA1 and OmpA2, were predicted in chromosome 1 and chromosome 2, respectively. Other predicted OMPs in this group included proteins involved in outer membrane assembly (LolB, YaeT, Rare lipoprotein B and YfgL lipoprotein), proteins involved in adaptation to stressful conditions such as organic substance tolerance and starvation (LptD and Slp lipoprotein), a protein involved in capsular polysaccharide biosynthesis (CpsB) and proteins involved in structural integrity (peptidoglycan-associated lipoprotein and OmpV). Sixteen hypothetical proteins were predicted in this group, and of these ten were predicted to be OmpA family proteins.

In the group of transport and receptor activity proteins (Table 4.3), several predicted proteins included OMPs involved in bacterial secretion systems; type I (TolC), type II (GspC and GspD), type III (YscW, VscJ, YscC, putative PopN, and VscC2), and type VI (hypothetical protein VP1412 and VPA1042). Three specific porins, BtuB, a putative chitoporin (ChiP), and maltoporin (LamB), and four non specific porins, OmpU, OprD family outer membrane protein, porin qsr prophage, and OmpN were predicted in this group. Predicted channel-forming

proteins included OmpK, OmpX and OmpW. Furthermore, three copies of AggA, of which two (VP1631 and VP1634) are located in chromosome 1 and the third (VPA0954) in chromosome 2, were also predicted in this group. Three copies of long-chain fatty acid transport proteins, of which two (VP2212 and VP2213) are located in chromosome 1 and the third (VPA0860) in chromosome 2, were also predicted in this group. Nine hypothetical proteins were also predicted in this group (Table 4.3).

In the group involved in iron binding and TonB receptor activity (Table 4.3), nine OMPs, IrgA, putative 83 kDa decaheme outer membrane cytochrome c, PvuA1, PvuA2, LutA, FhuE, FhuA, HutA, and putative Fe-regulated protein B, were predicted. It is well established that all of these OMPs have an important role in iron uptake in many bacteria including *V. parahaemolyticus* (http://www.uniprot.org/). No hypothetical proteins were predicted to be involved in this group. In the group related to flagella and motor activity (Table 4.3), eight proteins including putative lipoprotein VP1267, FlgH1, MotY, FlgH2, FlgL1, putative MotY VPA1539, the OmpA family-related OMP VPA1186 and CsuE, were predicted.

In the group of enzyme activity (Table 4.3), three (MtlA, MtlC, and MtlF) out of eight OMPs in this category are involved in murein degradation and peptidoglycan metabolism. Two predicted OMPs (VP1260 and VP2496) are involved in lipid metabolic processes and two other OMPs (VPA1615 and VPA0514) are involved in protein catabolism activity. One predicted OMP (NutA) in this group is involved in extracellular nucleotide degradation.

In the group with adherence and colonization activity (Table 4.3), two proteins predicted to be involved in pilus assembly and stability included MshL and PilF. A

putative haemaglutinin, VPA1442, was predicted to be involved in red blood cell binding activity. Two OMPs functioning as colonizing factors, AcfA and AcfD, were predicted in this category. One hypothetical protein (VP1767) was predicted in this group. This protein may be involved in pathogenicity in *V. parahaemolyticus* since its amino acid sequence is homologous to that of invasin, a protein which allows enteric bacteria, including pathogenic *Yersinia spp.* and *E. coli*, to penetrate mammalian cells (http://www.uniprot.org/).

Although four predicted OMPs (Pcp, NlpB, hypothetical protein VP2042, and putative outer membrane lipoprotein VPA0396) were of unknown functions, their subcellular locations nevertheless suggested that they are located in the outer membrane.

## *4.3.3 Variation of OMP profiles of clinical V. parahaemolyticus isolates from Thailand*

Variation of the OMP profiles of clinical *V. parahaemolyticus* isolates within the same clonal group was demonstrated by SDS-PAGE (Fig. 4.7). The OMP profiles of nine clinical *V. parahaemolyticus* isolates, including four isolates (VP2, VP166, VP172 and VP176) from clinical cluster 2 and five isolates from clinical cluster 4 (VP178, VP180, VP182, VP184 and VP188) were analyzed. The phylogenetic relationships of these isolates are shown in the Neighbour-Joining tree based on the MLST analysis of the Thai *V. parahaemolyticus* isolates (Fig. 2.52). Isolates VP166 and VP172 represent similar OMP profiles, which differ slightly from those of isolates VP2 and VP176 (Fig. 4.7). Isolates VP166 and VP172 represent ST83, whereas VP2 and VP176 represent ST1 and ST264, respectively.
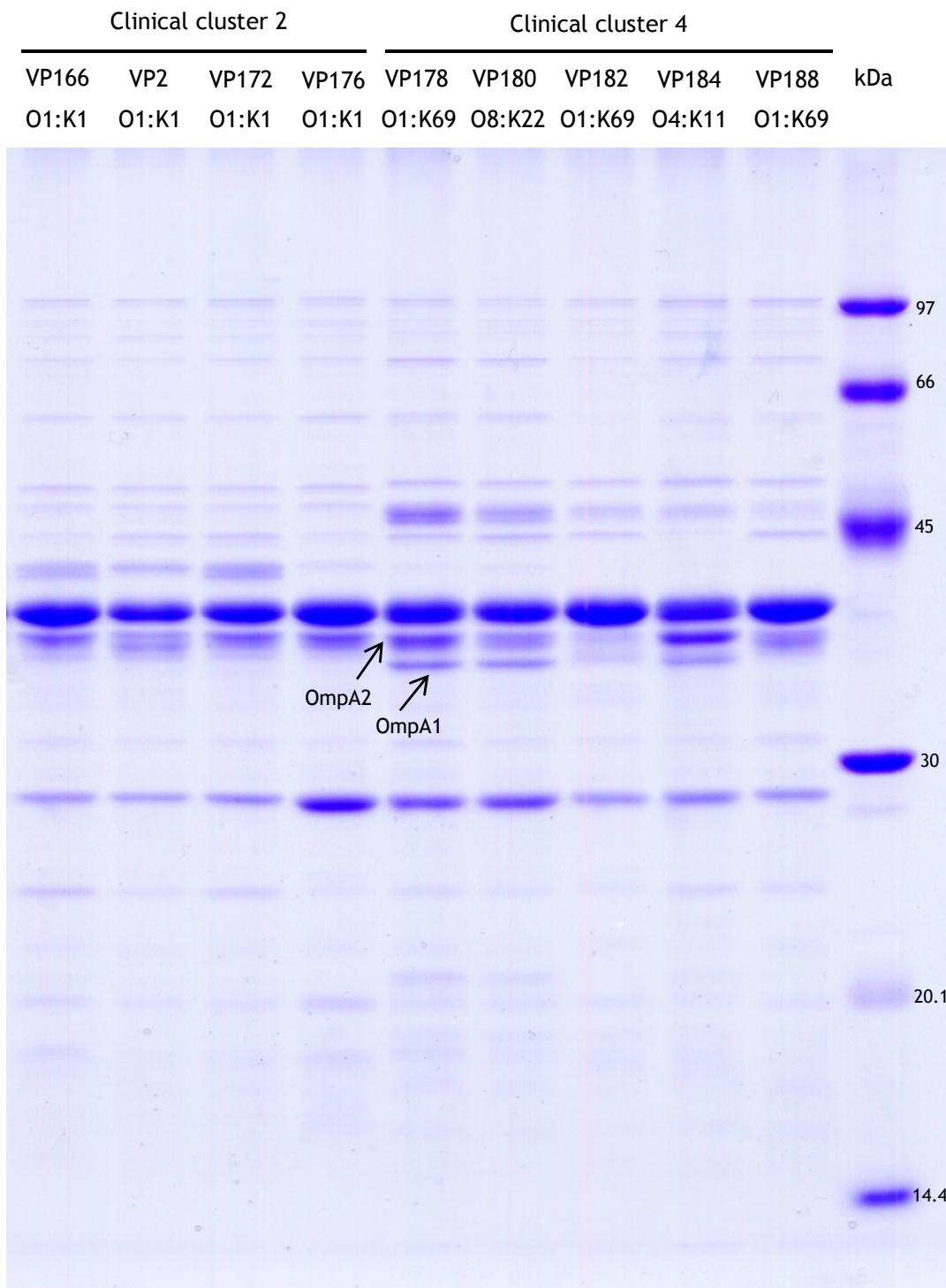
**Figure 4.7.** 1-D 12% SDS-polyacrylamine gel representing OMP profiles of nine clinical *V. parahaemolyticus* including isolates from clinical cluster 2 (VP2, VP166, VP172 and VP176) and clinical cluster 4 (VP178, 180, VP182, VP184, and VP188) represented in the MLST phylogenetic tree (Fig. 2.52). Serotypes of each isolate are indicated, together with the isolate designation, at the top of the gel.

Since isolates VP166, VP172, VP2 and VP176 are of the same O1:K1 serotype, and the OMP profiles of VP166 and VP172 differ from those of VP2 and VP176, the OMP profiles of these isolates appear to be linked to the ST rather than the serotype. Furthermore, variation of the OMPs profiles of isolates VP178, VP180, VP182, VP184 and VP188 from clinical cluster 4 was also demonstrated. In this case, isolates represent the same ST262 but have different O:K serotypes (Fig. 4.7). Isolates VP182 and VP188 share a similar OMP profile that differs from that of the other three isolates, VP178, VP180 and VP184. OmpA1 and OmpA2 of isolates VP178, VP180 and VP184 are more abundant and the protein bands more clearly separated than those from VP182 and VP188. The OMP profile of VP178 differs from those of VP182 and VP188, although they represent the same serotype O1:K69. Thus, no clear association of serotype, ST and OMP profile was observed among the clinical *V. parahaemolyticus* isolates of cluster 4. Based on these OMP profiles, isolates VP166 and VP178 were selected to represent clinical isolates for the proteomic analyses.

## 4.3.4 Identification of *V. parahaemolyticus* OMPs by gel-based proteomic approaches

The outer membrane fractions of eight *V. parahaemolyticus* isolates were prepared by Sarkosyl extraction and analyzed using the gel-based proteomic method. A 1-D SDS-polyacrylamide gel showing the OMP profiles of the eight representative isolates is shown in Fig. 4.8. Proteomic analyses identified several OMPs associated with the majority of protein bands and the protein identification numbers are shown. The numbers associated with each protein band correspond to the protein identification numbers provided in Table 4.4. To simplify protein identification in Fig. 4.8, OMPs with the most significant MOWSE score associated with each band are shown in Fig. 4.9.

261
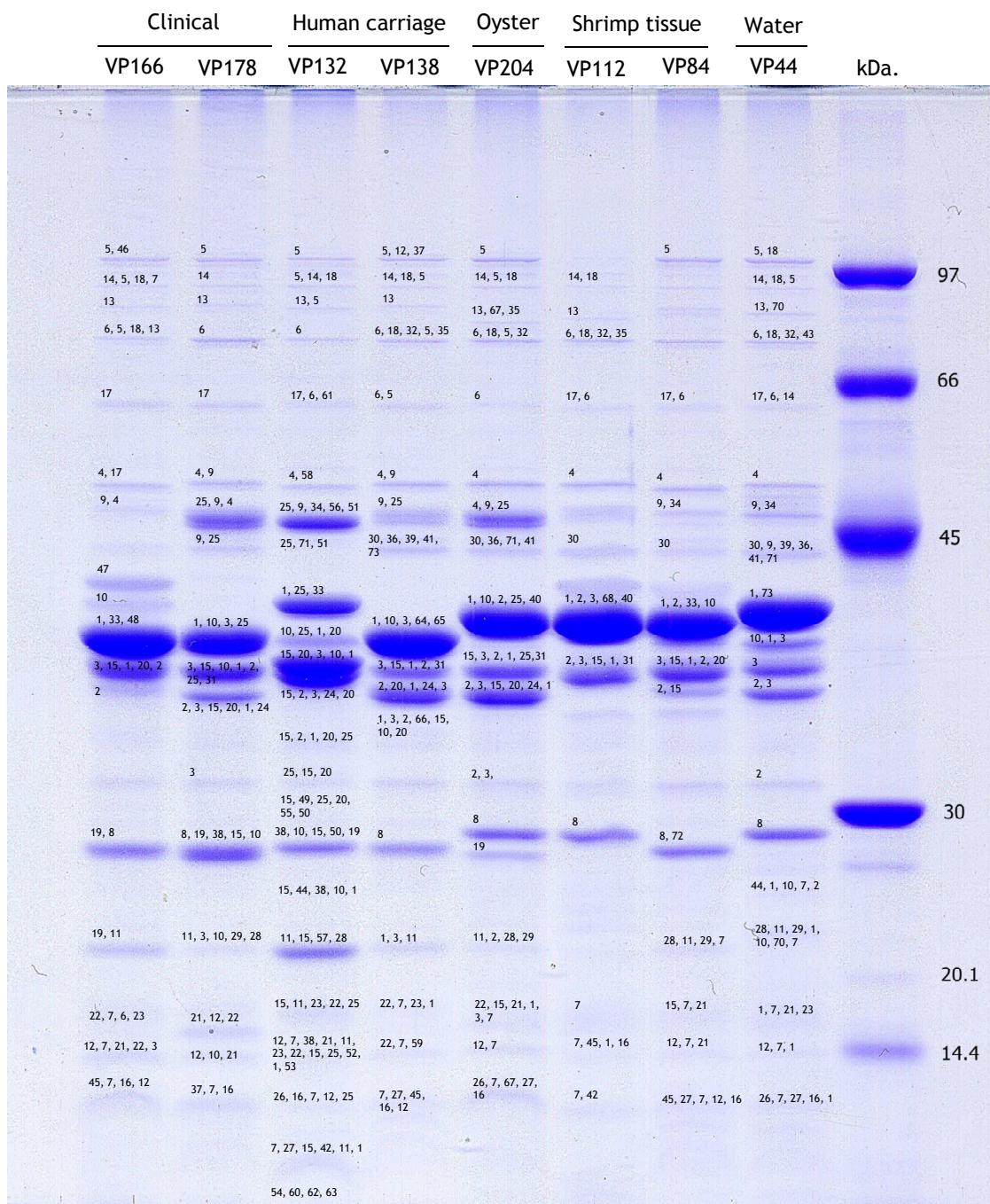
**Figure 4.8.** 1-D 12% SDS-polyacrylamide gel representing the gel-based proteomic identification of the OMPs from eight representative *V. parahaemolyticus* isolates recovered from different sources including clinical, human carrier, various seafood and water. Twenty micrograms of protein were loaded per lane and molecular mass markers (KDa) are shown on the right. Labelled numbers on the gel correspond to the identification numbers of the proteins provided in Table 4.4.
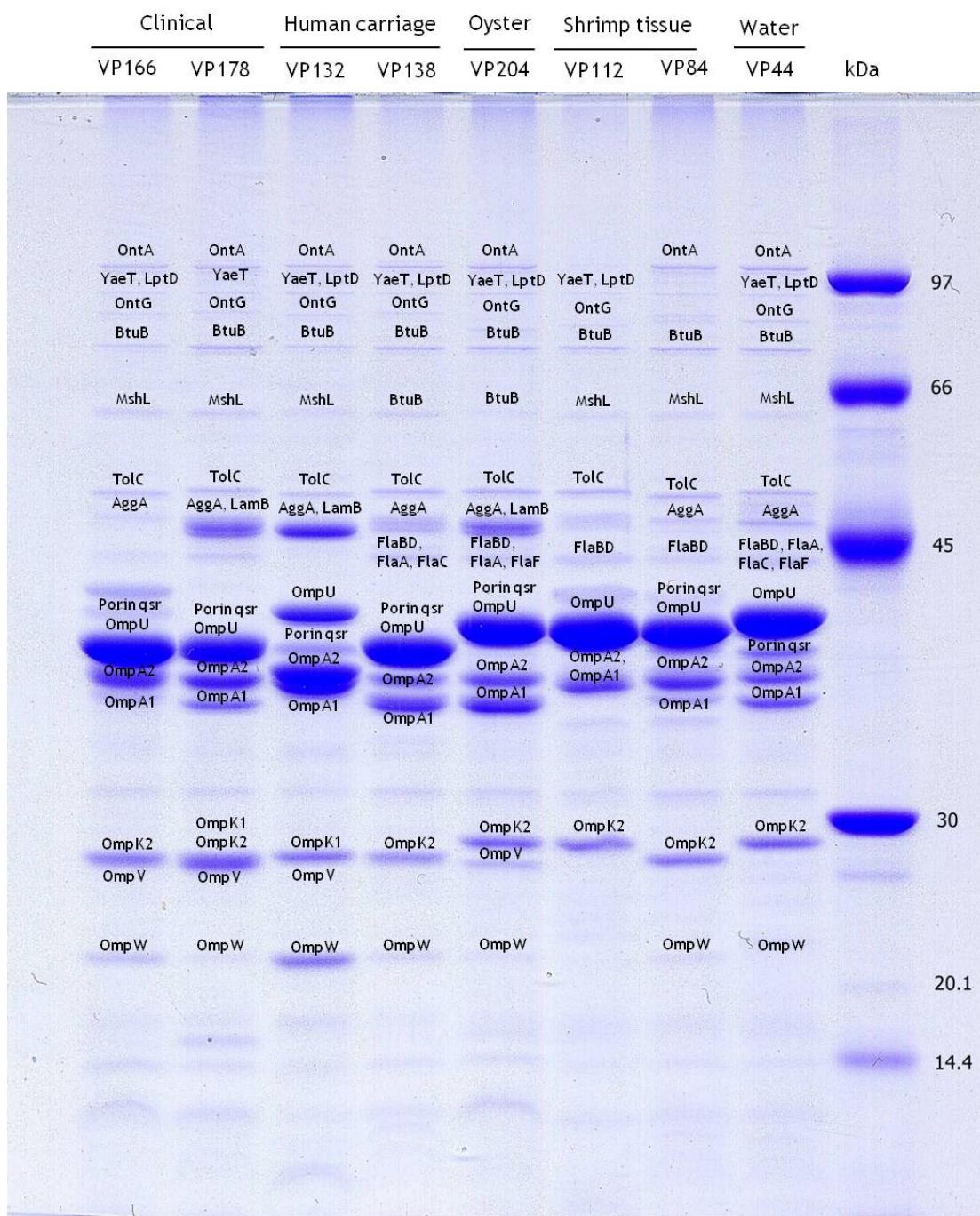
**Figure 4.9.** 1-D 12% SDS-polyacrylamine gel representing the gel-based proteomic identification of the OMPs from eight representative *V. parahaemolyticus* isolates recovered from different sources including clinical, human carrier, various seafood, and water. Twenty micrograms of protein were loaded per lane and molecular mass markers (KDa) are shown on the right. OMPs with highly significant prediction scores for individual bands are indicated.

A total of 73 different OMPs were identified in the eight isolates using the gel-based method and information about these proteins is summarized in Table 4.4. Seventy six identified proteins which were not annotated as OMPs but were localized to other subcellular locations such as the inner membrane, periplasm and cytoplasm are not included in Table 4.4. Details of these proteins are provided in Table A4, Appendix 3.

Of the 73 putative OMPs identified, 32 proteins were identified by both bioinformatic prediction and by the gel-based proteomic analysis (Fig. 4.10). These 32 OMPs are highlighted in grey shading in Table 4.4. Eighty five OMPs predicted from the *V. parahaemolyticus* RIMD2210633 genome by the bioinformatic approach were not identified by the gel-based proteomic analysis of eight representative isolates of *V. parahaemolyticus* (Fig. 4.10). Of the 73 OMPs identified by the gel-based proteomic analysis, 41 were not predicted by bioinformatic prediction and 24 of these were of unknown function.

The OMPs predicted from *V. parahaemolyticus* isolate RIMD2210633, which represents ST3, were compared with OMPs identified in isolate VP132 which also represents ST3 although it was recovered from human carriage. To compare the predicted OMPs of *V. parahaemolyticus* isolate RIMD2210633 with those of isolate VP132, gel pieces of the OMP profile covering the entire lane (i.e. gel pieces containing Coomasssie blue-stained bands and blank regions) of VP132 were excised for protein identification. A total of 46 OMPs were identified from isolate VP132. This represents 39.3% (46/117) of the total proteins predicted by bioinformatic prediction from *V. parahaemolyticus* RIMD2210633.

**Table 4.4.** Proteins identified in the outer membrane fractions of eight representative *V. parahaemolyticus* isolates

| Protein no. | Gene locus | Protein name | Function | Clinical | | Human carriage | | Oyster | Shrimp tissue | | Water |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | VP 166 | VP178 | VP132 | VP138 | VP204 | VP112 | VP84 | VP44 |
| 1 | VP2467 | OmpU | Transport activity | + | + | + | + | + | + | + | + |
| 2 | VP0764 | OmpA1 | Structural and integrity | + | + | + | + | + | + | + | + |
| 3 | VPA0248 | OmpA2 | Structural and integrity | + | + | + | + | + | + | + | + |
| 4 | VP0425 | TolC | Transport activity | + | + | + | + | + | + | + | + |
| 5 | VP0220 | OtnA | Transport activity | + | + | + | + | + | + | + | + |
| 6 | VP2938 | BtuB Vitamin B12 transporter | Transport activity | + | + | + | + | + | + | + | + |
| 7 | VPA1469 | Murein lipoprotein | Structural and integrity | + | + | + | + | + | + | + | + |
| 8 | OMPK2_VIBPA | OmpK2 | Phage receptor activity | + | + | - | + | + | + | + | + |
| 9 | VP1634 | AggA | Transport activity | + | + | + | + | + | - | + | + |
| 10 | VP1008 | Porin qsr prophage | Transport activity | + | + | + | + | + | - | + | + |
| 11 | VPA0096 | OmpW | Transport activity | + | + | + | + | + | - | + | + |
| 12 | VP0967 | Putative uncharacterized protein | Production of adhesive surface fibre (Curli) | + | + | + | + | + | - | + | + |
| 13 | VP0215 | OtnG | Unknown | + | + | + | + | + | + | - | + |
| 14 | VP2310 | YaeT | Surface antigen | + | + | + | + | + | + | - | + |
| 15 | VPA1186 | OmpA2 | Structural and integrity | + | + | + | + | + | + | + | - |
| 16 | VP0726 | Rare lipoprotein B | Outer membrane biogenesis | + | + | + | + | + | + | + | - |
| 17 | VP2704 | MSHA biogenesis protein MshL | Pillus assembly | + | + | + | - | - | + | + | + |
| 18 | VP0339 | LptD LPS-assembly protein | Outer membrane biogenesis | + | - | + | + | + | + | - | + |
| 19 | VPA0318 | OmpV | Structural and integrity | + | + | + | - | + | - | - | - |
| 20 | VPA0166 | Putative porin outer membrane protein | Unknown | + | + | + | + | + | - | + | - |
| 21 | VPA0810 | Putative uncharacterized protein | Structural and integrity | - | + | + | - | + | - | + | + |

*Grey-shaded proteins represent the OMPs that were also predicted by bioinformatic prediction tools from *V. parahaemolyticus* RIMD2210633 genome

# Table 4.4. (continued)

| Protein no. | Gene locus | Protein name | Function | Isolates of different sources | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Clinical | | Human carriage | | Oyster | Shrimp tissue | | Water |
| | | | | VP 166 | VP178 | VP132 | VP138 | VP204 | VP112 | VP84 | VP44 |
| 22 | VP2354 | Putative lipoprotein | Unknown | + | + | + | + | + | - | - | - |
| 23 | VP1061 | Peptidoglycan-associated lipoprotein | Structural and integrity | + | - | + | + | - | - | - | + |
| 24 | VPA0527 | OmpN | Structural and integrity | - | + | + | + | + | - | - | - |
| 25 | VPA1644 | Maltoporin LamB | Transport activity | - | + | + | + | + | - | - | - |
| 26 | VP1243 | Putative uncharacterized protein VP1243 | Unknown | - | - | + | + | + | - | + | - |
| 27 | VP2850 | Putative uncharacterized protein VP2850 | Unknown | - | - | + | + | + | - | + | - |
| 28 | VP2724 | Putative uncharacterized protein VP2724 | Unknown | - | - | + | - | + | - | + | + |
| 29 | VP0434 | Putative uncharacterized protein VP0434 | Unknown | - | + | - | - | + | - | + | + |
| 30 | VP2259, VP0790 | Polar flagellin B/D FlaBD | Flagella biogenesis | - | - | - | + | + | + | + | + |
| 31 | VP0712 | Putative lipoprotein protein VP0712 | Unknown | - | + | - | + | + | + | - | - |
| 32 | VP0449 | PBP activator LpoA | Outer membrane biogenesis, penicillin-binding protein | - | - | - | + | + | + | - | + |
| 33 | VPA1602 | Capsular polysaccharide transport protein | Transport activity | + | - | + | - | - | - | + | - |
| 34 | VP1631 | AggA | Transport activity | - | - | + | - | - | - | + | + |
| 35 | VPA1435 | Putative iron(III) compound receptor VPA1435 | Receptor activity | - | - | - | + | + | + | - | - |
| 36 | VP2258 | Polar flagellin A FlaA | Flagella biogenesis | - | - | - | + | + | - | - | + |
| 37 | VPA0242 | Putative uncharacterized protein VPA0242 | Unknown | - | + | - | + | + | - | - | - |
| 38 | VP2362 | OmpK1 | Phage receptor activity | - | + | + | - | - | - | - | - |
| 39 | VP0788 | Polar flagellin C FlaC | Flagella biogenesis | - | - | - | + | - | - | - | + |
| 40 | VP2272 | Lipoprotein-34 NlpB | Composition of outermembrane vesicle | - | - | - | - | + | + | - | - |
| 41 | VP2261 | Polar flagellin F FlaF | Flagella biogenesis | - | - | - | - | + | - | - | + |
| 42 | VP1454 | Putative uncharacterized protein VP1454 | Unknown | - | - | + | - | - | + | - | - |
| 43 | VP1594 | Putative uncharacterized protein VP1594 | Unknown | - | - | + | - | - | - | - | + |
| 44 | VP1267 | Putative lipoprotein | Flagellar motility | - | - | + | - | - | - | - | + |

## Table 4.4. (continued)

| Protein no. | Gene locus | Protein name | Function | Isolates of different sources | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | Clinical | | Human carriage | | Oyster | Shrimp tissue | | Water |
| | | | | VP 166 | VP178 | VP132 | VP138 | VP204 | VP112 | VP84 | VP44 |
| 45 | VP1243 | Putative uncharacterized protein VP1243 | Unknown | + | - | - | - | - | + | - | - |
| 46 | VP0168 | Putative uncharacterized protein VP0168 | Transporter activity, TonB-dependent receptor | + | - | - | - | - | - | - | - |
| 47 | VP1998 | Putative outer membrane protein TolC | Transport activity | + | - | - | - | - | - | - | - |
| 48 | VPA0111 | Putative uncharacterized protein VPA0111 | Unknown | + | - | - | - | - | - | - | - |
| 49 | VP0704 | Lipoprotein YaeC | Surface antigen | - | - | + | - | - | - | - | - |
| 50 | VP0374 | Putative outer membrane protein VP0374 | Unknown | - | - | + | - | - | - | - | - |
| 51 | VPA0363 | Acriflavin resistance protein | Transport activity | - | - | + | - | - | - | - | - |
| 52 | VP2042 | Putative uncharacterized protein VP2042 | Transport activity | - | - | + | - | - | - | - | - |
| 53 | VP2698 | Putative V10 pilin | Flagella biogenesis | - | - | + | - | - | - | - | - |
| 54 | VP1192 | Putative outer membrane lipoprotein Pcp | Surface antigen | - | - | + | - | - | - | - | - |
| 55 | VP0216 | OprD | Outer membrane porin | - | - | + | - | - | - | - | - |
| 56 | VP0039 | Membrane fusion protein | Transport activity | - | - | + | - | - | - | - | - |
| 57 | VP1901 | OmpX | Transport activity | - | - | + | - | - | - | - | - |
| 58 | VP0802 | Putative exported protein | Transport activity | - | - | + | - | - | - | - | - |
| 59 | VP0868 | Putative outer membrane lipoprotein Slp | Unknown | - | - | - | + | - | - | - | - |
| 60 | VP0966 | Putative uncharacterized protein VP0966 | Unknown | - | - | + | - | - | - | - | - |
| 61 | VPA1648 | Putative uncharacterized protein VPA1648 | Unknown | - | - | + | - | - | - | - | - |
| 62 | VP0898 | Putaive lipoprotein VP0898 | Unknown | - | - | + | - | - | - | - | - |
| 63 | VP0541 | Putative uncharacterized protein VP0541 | Unknown | - | - | + | - | - | - | - | - |
| 64 | VP1847 | Putative uncharacterized protein VP1847 | Unknown | - | - | - | + | - | - | - | - |
| 65 | VP1268 | Putative uncharacterized protein VP1268 | Unknown | - | - | - | + | - | - | - | - |
| 66 | VP0993 | Putative uncharacterized protein VP0993 | Unknown | - | - | - | + | - | - | - | - |
| 67 | VP0133 | Secretion pathway protein D | Protein secreted by the type II secretion system | - | - | - | - | + | - | - | - |
| 68 | VP2325 | Putative uncharacterized protein VP2325 | Unknown | - | - | - | - | - | + | - | - |
| 69 | VPA0448 | Putative uncharacterized protein VPA0448 | Unknown | - | - | - | - | - | - | - | + |
| 70 | VP0768 | Putative uncharacterized protein VP0768 | Unknown | - | - | - | - | - | - | - | + |
| 71 | VP2212 | long-chain fatty acid transport protein | Translocation of long-chain fatty acids across the outer membrane | - | - | + | - | + | - | - | + |
| 72 | VPA0291 | Putative uncharacterized protein VPA0291 | Unknown | - | - | - | - | - | - | + | - |
| 73 | VPA0749 | Putative uncharacterized protein VPA0749 | Unknown | - | - | - | - | - | - | - | + |
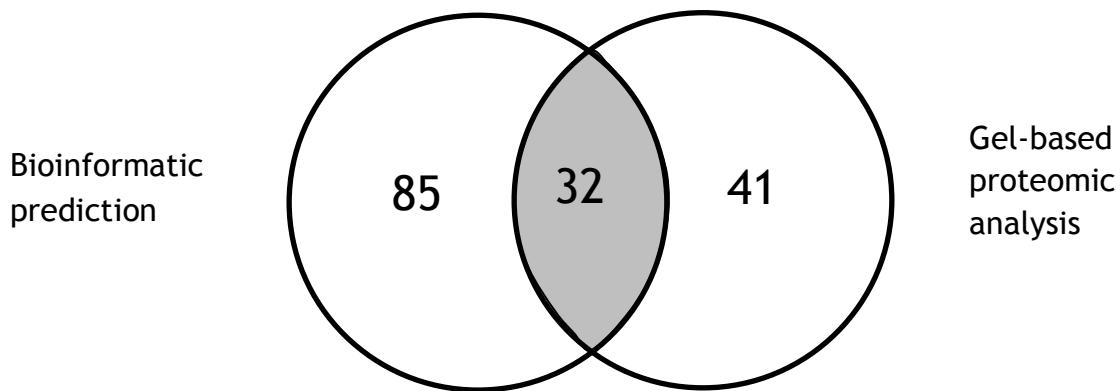
**Figure 4.10.** Comparison of OMPs predicted by bioinformatic prediction of *V. parahaemolyticus* RIMD2210633 and identified by gel-based proteomic analysis of eight representative *V. parahaemolyticus* isolates. The area shaded in grey represents the number of proteins predicted by bioinformatic approaches and identified by gel-based proteomic analysis.

Eleven OMPs recovered from VP132 were not predicted from the *V. parahaemolyticus* RIMD2210633 genome. Of these eleven proteins, six included YaeC, putative efflux protein_VPA0363, putative V10, OprD, membrane fusion protein_VP0039 and putative exported protein_VP0802, whereas the remaining five are of unknown function.

Proteins from gel pieces containing protein bands and blank regions of the gel representing isolate VP132 were analyzed. Seven OMPs including YaeC, putative protein_VP0374, OprD, putative lipoprotein_VP1192, putative proteins_VP0966, putative proteins_VP0898, and putative proteins_VP0541, were identified from the blank regions of the gel representing isolate VP132. Of these seven proteins, only the putative lipoprotein_VP1192 was also predicted from the *V. parahaemolyticus* RIMD2210633 genome. Since these seven OMPs were identified

in the blank regions of the gel only of isolate VP132, it was not possible to determine whether or not the other seven isolates also contain these proteins.

Seven OMPs were identified in all eight representative strains (Table 4.4). These OMPs included OmpU, OmpA1, OmpA2, TolC, OtnA, BtuB, and murein lipoprotein. OmpU was identified as a major OMP and has the most abundant band intensity in all eight strains (Fig. 4.9). However, the bands representing OmpU have different molecular masses among the eight strains. In particular, OmpU of isolate VP132 has a higher molecular mass (and is less abundant) than OmpU of the other isolates. Variation in protein molecular mass in VP132 was also demonstrated in OmpA1 and OmpA2. These two OMPs are encoded by genes located in different chromosomes; *ompA1* is located in chromosome 1 and *ompA2* is located in chromosome 2.

The OmpK2, AggA_VP1634, porin qsr prophage, OmpW, putative uncharacterized protein VP0967, OtnG, YaeT, OmpA2_VPA1186 and rare lipoprotein B were present in seven isolates. OmpK is recognized as a receptor for the broad-host-range vibriophage KVP40 and its amino acid sequence is closely related to that of a specific channel-forming OMP (Tsx) of enteric bacteria (Inoue *et al.*, 1995a, b). OmpK1 of clinical isolate VP178 and human carriage isolate VP132 were best matched with OmpK1, the OmpK protein from clinical *V. parahaemolyticus* isolate RIMD2210633 (http://blast.ncbi.nlm.nih.gov). However, the OmpK proteins of the other isolates, VP166, VP178, VP138, VP204, VP84, VP112 and VP44, were best matched with OmpK2, the OmpK protein from environmental *V. parahaemolyticus* isolate RIMD2210001 (http://blast. ncbi.nlm.nih.gov). OmpK1 and OmpK2 share 81.8% amino acid sequence similarity. A BLAST search against the protein public database (http://www.uniprot.org) showed that the amino

acid sequence of OmpK1 shares 100% similarity with OmpK from clinical *V. parahaemolyticus* isolate K5030, whereas the amino acid sequence of OmpK2 is best matched (89.0%) with OmpK of *Vibrio harveyi HY01*. OmpK exhibits molecular mass variation among the eight representative *V. parahaemolyticus* isolates (Fig. 4.9). Notably, OmpK from clinical (VP166 and VP178) and human carriage (VP132 and VP138) isolates were of lower molecular mass than OmpK of environmental isolates recovered from oyster (VP204), shrimp tissue (VP112) and water (VP44) (Fig. 4.9). However, an environmental isolate recovered from shrimp tissue, VP84, possessed an OmpK protein with a similar molecular mass to those of the clinical (VP166 and VP178) and human carriage (VP132 and VP138) isolates (Fig. 4.9).

The OtnG and YaeT proteins were identified in all isolates except the shrimp tissue isolate VP84, whereas AggA_VP1634 (labelled as Agg in Fig. 4.9), porin qsr prophage, OmpW and putative uncharacterized protein VP0967 were identified in all isolates except the shrimp tissue isolate VP112. Two copies of AggA, AggA_VP1634 and AggA_VP1631, were detected in three strains; human carriage strain VP132, shrimp tissue strain VP84 and water strain VP44. In contrast to OmpA, the genes-encoding AggA_VP1634 and AggA_VP1631 are both located in chromosome 1 of *V. parahaemolyticus* RIMD2210633 and the proteins have molecular masses of 48.6 and 50.7 kDa, respectively. Although AggA_VP1634 was present in all strains, except shrimp tissue strain VP112, its expression was more abundant in clinical strain VP178, human carriage strain VP132 and oyster strain VP204. AggA_VP1631 was detected in three strains; human carriage strain VP132, shrimp tissue strain VP84 and water strain VP44.

Proteins MshL, LptD, and putative porin outer membrane_VPA0166 were identified in six isolates. MshL, a protein involved in adherence, was present in all isolates except human carrier isolate VP138 and oyster isolate VP204. LptD, a protein involved in lipoprotein biogenesis, was present in all isolates except clinical isolate VP178 and shrimp tissue isolate VP84. Putative uncharacterized protein_VPA0810 was identified in five isolates, clinical isolate VP178, human carrier isolate VP132, oyster isolate VP204, shrimp tissue isolate VP84 and water isolate VP44. Putative lipoprotein_VP2354 was also identified in five isolates, clinical isolates VP166 and VP178, human carrier isolates VP132 and VP138 and oyster isolate VP204.

OmpV was identified in four isolates, including clinical isolates VP166 and VP178, human carrier isolate VP132 and oyster isolate VP204. The position of OmpV in the SDS-polyacrylamide gel is the same as that of OmpK1 and OmpK2, with the exception of isolate VP204 (Fig. 4.9). Noteworthy, four polar flagellins, FlaA, FlaBD, FlaC and FlaF, which are the OMPs involved in flagella biogenesis, were not identified in clinical isolates VP166 and VP178 and human carrier isolate VP132 (Fig. 4.9 and Table 4.4). The FlaA protein was identified in human carrier isolate VP138, oyster isolate VP204 and water isolate VP44. The FlaBD protein was identified in human carrier isolate VP138, oyster isolate VP204, shrimp tissue VP112 and VP84 and water isolate VP44. The FlaC protein was identified in human carrier isolate VP138 and water isolate VP44. The FlaF protein was identified in oyster isolate VP204 and water isolate VP44. However, these four flagellins (FlaA, FlaBD, FlaC and FlaF) were not predicted by bioinformatic analysis from the *V. parahaemolyticus* RIMD2210633 genome although these proteins-encoding genes are present in the genome. These results from the proteomic analyses suggest that flagellin proteins are likely to be present in

isolates recovered from the environment rather than from human-associated samples i.e. clinical and human carrier. Flagellin proteins are involved in bacterial flagellation. Certain flagellin proteins are also involved in bacterial virulence. For example, FlaA and FlaD are involved in *V. anguillarum* virulence in fish (McGee *et al.*, 1996; Milton *et al.*, 1996). Furthermore, FlaA is essential for symbiotic colonization of *V. fischeri* in squid (Millikan & Ruby, 2004). FlaC and FlaD are capable of inducing inflammation in colonic cells that were infected by *V. cholerae* O1 (Xicohtencatl-Cortés *et al.*, 2006).

A number of OMPs predicted by bioinformatic analysis to be involved in type III secretion and iron uptake systems of *V. parahaemolyticus* isolate RIMD2210633 were not identified by proteomic analyses of the eight representative isolates. Five OMPs involved in type III secretion systems, VscC2, YscC, PopN, YscW and VscJ were not recovered from any of the eight representative isolates, although they were predicted from the *V. parahaemolyticus* RIMD2210633 isolate (Table 4.3). Furthermore, a total of nine iron binding OMPs, including the IrgA, PvuA1, PvuA2, LutA, FhuE, FhuA, HutA, the putative 83 kDa decaheme outer membrane cytochrome c_VP1220 and the putative Fe-regulated protein B were predicted from *V. parahaemolyticus* isolate RIMD2210633 by the bioinformatic approach (Table 4.3). Only FhuA (VPA1435) was identified by proteomic analyses in human carriage isolate VP138, oyster isolate VP204 and shrimp tissue isolate VP112 (Table 4.4). The other eight predicted iron-binding OMPs were not identified in any of the eight representative isolates.

## 4.4 Discussion

Three hundred and sixty two genome-encoded OMPs of pandemic *V. parahaemolyticus* serotype O3:K6 isolate RIMD2210633 were predicted by bioinformatic approaches involving consensus prediction (Fig. 4.1). Although these proteins were predicted by bioinformatic tools that are supposed to predict OMPs based on three different characteristics, namely subcellular location, transmembrane ß-barrel protein prediction and outer membrane lipoprotein prediction, some proteins that are not likely to be OMPs (i.e. proteins localizing in the cytoplasm, inner membrane, periplasm, and extracellular compartment) were also predicted (Fig. 4.4). Of the 362 predicted genome-encoded OMPs, 117 OMPs were confidently predicted after integration and text mining processes (Fig. 4.1). The 229 predicted proteins with unidentified locations and 16 falsely-predicted proteins (i.e. proteins in other subcellular compartments) were not included among the confidently predicted OMPs. The falsely-predicted proteins that are unlikely to be OMPs, six were periplasmic localizing proteins, four were inner membrane proteins, three were extracellular proteins and three were cytoplasmic proteins. Since the outer membrane is adjacent to the periplasm, and some proteins are associated with both compartments (Costerton *et al.*, 1974), there is a possibility that the prediction tools misidentified proteins in the periplasmic space as OMPs. Of the six periplasmic proteins predicted by the bioinformatic tools, three were also identified by the gel-based proteomic analysis (Table A4 in Appendix 3).

The 117 confidently predicted proteins were classified into six functional groups, biogenesis and integrity, transportation and receptor activity, iron binding activity, flagella and motor activity, enzyme activity, adherence and others. The

33 predicted proteins are involved in OMP biogenesis and integrity. The majority (10/16) of the hypothetical proteins predicted in this group represent the OmpA family. This finding is consistent with the fact that OmpA is highly abundant and is a predominant antigen in the enterobacterial outer membrane (Koebnik *et al.*, 2000). OmpA not only functions in outer membrane biogenesis and integrity but is also involved in other activities such as adhesion, immune invasion and biofilm formation (Smith *et al.*, 2007).

The largest proportion (41.9%) of the 117 predicted OMPs are involved in transportation and receptor activity. This large proportion of predicted OMPs involved in transportation and receptor activity highlights an important role of these proteins in respect to bacterial adaptation by controlling nutrient uptake under different conditions encountered in diverse habitats. Seven porin proteins were predicted, of which three are specific porins (BtuB, chitoporin [Chip], and maltoporin [LamB]) and four are non-specific porins (OmpU, OmpN, OprD, and porin qsr prophage). With the exception of the chitoporin, these predicted porins were all identified by proteomic analyses of eight representative *V. parahaemolyticus* isolates. The chitoporin (ChiP) is involved in the catabolic breakdown of chitin; it mediates the uptake of glucose derivatives following the breakdown of chitin, the main constituent of the crustacean exoskeleton (Park *et al.*, 2000). The mechanism of the chitin catabolic pathway was proposed in *V. cholerae* (Hunt *et al.*, 2008). Chitoporin functions as a specific channel that allows extracellular chito-oligosaccharides [$N$-acetylglucosamine (GlcNAc)$_n$], the digested products of chitin produced by chitinase, to be accessible through the outer membrane (Hunt *et al.*, 2008). These chito-oligosaccharides are subsequently processed by catabolic enzymes and certain proteins localized in the periplasmic space, the inner membrane and the cytoplasm (Hunt *et al.*,

2008; Keyhani & Roseman, 1999). The different chitin degradation pathways of 19 individual *Vibrio* species were reported in a previous study suggesting a dynamic evolution of chitin metabolism in this species (Hunt *et al.*, 2008). Since the expression of chitoporin is induced by chito-oligosaccharides, failure to detect chitoporin in the eight representative *V. parahaemolyticus* isolates may suggest a substrate-regulated role of chitoporin. It would be interesting to investigate whether chitoporin expression varies among clinical and environmental *V. parahaemolyticus* isolates grown in medium supplemented with chitin or chito-oligosaccharides.

The distribution of the predicted OMPs between chromosomes 1 and 2 in the present study was consistent with the *V. parahaemolyticus* genomic analysis of Makino *et al*. (2003). Predicted OMPs involved in OMP biogenesis were located predominantly in chromosome 1 (22/33) whereas predicted OMPs involved in the iron binding/TonB receptor were located predominantly in chromosome 2 (7/9) (Fig 4.6b). These results are in agreement with the previous genomic study which demonstrated that chromosome 1 tends to contain the genes required for growth, biogenesis and viability, whereas chromosome 2 contains the genes for bacterial adaptation which includes the gene encoding proteins involved in iron uptake, transport of various substrates and transcriptional regulation (Makino *et al.*, 2003).

The OMP profile variation among clinical *V. parahaemolyticus* isolates, within two distinct MLST clusters (i.e. clusters 2 and 4), was observed by SDS-PAGE (Fig. 4.7). No associations between OMP type, ST and serotypes were noted among these isolates. Isolates sharing the same ST may have different OMP profiles as well as different serotypes (Fig. 4.7). These results indicate that MLST

has less resolution than serotyping and OMP typing. This is because MLST analysis is based on housekeeping enzyme gene sequences (Maiden, 2006). MLST is a reliable typing method although housekeeping enzyme genes are more conserved than OMP- and serotype-encoding genes which are recognized as accessory genes in the genome. Thus, the genetic classification based on housekeeping enzyme genes will reflect the under-lying evolution of the organism and is unlikely to be biased by significant horizontal gene transfer, except in the case of highly recombining organisms. In a previous study, an advantage of using molecular typing by a combination of OMP type, serotype and lipopolysaccharide type for subgroup classification was demonstrated in *Pasteurella trehalosi* isolates from different origins (Davies & Quirie, 1996). However, a much greater number of isolates (n=60) was used in that study, whereas only eight selected isolates were used in the present study. Comparative OMP profiles of isolates from wider origins including pandemic, non-pandemic and environmental *V. parahaemolyticus* isolates will contribute to a better understanding of the role of OMPs in the virulence and evolution of this organism.

Variation in the OMP profiles of eight representative *V. parahaemolyticus* isolates recovered from different sources was demonstrated by SDS-PAGE analysis (Fig. 4.8-4.9). The presence of OmpU, OmpA1, OmpA2, TolC, OtnA, BtuB and murein lipoprotein in all eight isolates indicates an important role for these OMPs in this organism. **OmpU** participates in bacterial virulence in several ways. It plays a vital role in bacterial survival in bile-containing environments and is involved in cell adherence in pathogenic *Vibrio* species including *V. parahaemolyticus*, *V. cholerae* and *V. fischeri* (Aeckersberg *et al.*, 2001; Sperandio *et al.*, 1995; Whitaker *et al.*, 2012; Wibbenmeyer *et al.*, 2002). In particular, Whitaker *et al*. (2012) showed that *V. parahaemolyticus* OmpU is

necessary for stress tolerance in the presence of bile salts, acetic acid, and sodium dodecyl sulphate (SDS) as well as colonization of the mouse intestine. These authors also demonstrated that expression of OmpU is regulated by ToxRS, a virulence factor that also regulates the transcription of the cholera toxin-encoding gene (*ctx*) in *V. cholerae* and the thermostable direct haemolysin gene (*tdh*) in *V. parahaemolyticus* (Lin *et al.*, 1993). OmpU could be an important virulence factor in *V. parahaemolyticus*, prolonging bacterial infection under stressful environmental conditions, and could also facilitate the colonization of the human gut. The presence of OmpU in all eight isolates suggests that this protein has an important function in *V. parahaemolyticus* regardless of the epidemiological source of the strains. Although the molecular structure of OmpU has not been established, OmpU has equivalent functions to OmpF in *E. coli* (Chakrabarti *et al.*, 1996). OmpF is a porin that consists of three identical subunits and each subunit contains a 16-stranded anti-parallel ß-barrel (Cowan *et al.*, 1992). However, different features between OmpU and OmpF were reported in a previous study (Chakrabarti *et al.*, 1996). This study demonstrated that OmpU has a slightly bigger pore size (1.6 nm) than OmpF (1.2 nm). There is also a lack of both nucleotide sequence homology and immunological relatedness between OmpF and OmpU. In *V. cholerae,* OmpU constituted 30% of the outer membrane proteome when bacteria were grown in standard growth medium (nutrient broth supplemented with 2% NaCl), but almost 60% of the outer membrane proteome when bacteria were grown in the same medium without NaCl (Chakrabarti *et al.*, 1996). It should be noted that the *V. parahaemolyticus* cultures in the present study were grown in TSB supplemented with 3% NaCl, thus the abundance of OmpU may change under growth conditions without NaCl.

**OmpA** is highly conserved and is present in all Gram-negative bacteria (Beher *et al.*, 1980; Smith *et al.*, 2007). It is involved in host invasion and biofilm formation, it acts as a bacteriophage receptor and it interferes with host defence mechanisms (Morona *et al.*, 1985; Smith *et al.*, 2007). In the present study, three OmpA gene copies were identified; one was located in chromosome 1 (OmpA1) and two were located in chromosome 2 (OmpA2_VPA0248 and OmpA2_VPA1186). From a previous study, two types of OmpA with different alleles were identified in *E. coli* (Power *et al.*, 2006). These two OmpAs (OmpA1 and OmpA2) in *E. coli* differ in their amino acid sequence at the region encoding the surface-exposed loops; these function as bacteriophage receptors (Koebnik, 1999; Power *et al.*, 2006). These authors suggested that OmpA2 has a selective advantage in human isolates since it occurs with a greater frequency in isolates recovered from humans compared to other vertebrates, including the Tasmanian devil, the mountain possum, the brushtail possum, the eastern grey kangaroo and the varied honeyeater (Power *et al.*, 2006). However, the OmpA1 and OmpA2_VPA0248 of *V. parahaemolyticus* are unlikely to have the same properties as OmpA1 and OmpA2 in *E. coli*. The amino acid sequences of *V. parahaemolyticus* OmpA1 and OmpA2_VPA0248 share only 40.1% similarity, whereas OmpA2_VPA0248 and OmpA2_VPA1186 share 60.0% similarity. The amino acid sequence polymorphisms among these OMPs are present intermittently throughout the entire protein, indicating that the variable sites are not associated with particular regions such as the surface-exposed loops. In a previous study, OmpA1 and OmpA2 from *V. parahaemolyticus* were demonstrated to respond to salt concentration (Yang *et al.*, 2010). Expression of OmpA2 was induced in culture media supplemented with 2% NaCl compared to 0.66% NaCl, whereas there was less expression of OmpA1 under 2% NaCl compared to 0.66% NaCl (Yang *et al.*, 2010). An association between

hypervariable domains of the OmpA protein and host species (bovines and ovines), was also demonstrated in *M. haemolytica* (Davies & Lee, 2004).

**TolC** functions as a multidrug efflux pump and is essential for the transportation of diverse molecules across the cell membranes of Gram-negative bacteria (Koronakis *et al.*, 2004). Indeed, it is necessary for bacterial adaptation particularly in organisms inhabiting a wide range of ecological niches such as *V. parahaemolyticus*. **OtnA** is involved in capsular and O antigen synthesis of *V. cholerae* (Bik *et al.*, 1996) and is a nutrient-regulated protein in *V. alginolyticus* and *V. parahaemolyticus* (Abdallah *et al.*, 2010). It is logical that all eight *V. parahaemolyticus* isolates contain OtnA because it plays an important role in capsular polysaccharide transport, which contributes to the biogenesis of O and K antigens. Although the molecular function of OtnA in *V. parahaemolyticus* has not been studied, OtnA may be involved in the serotypic variation of isolates from diverse sources since expression of OtnA tends to vary with different growth conditions (Abdallah *et al.*, 2010). **BtuB** is involved in vitamin B12 uptake (Aufrere *et al.*, 1986), and the occurrence of BtuB in all eight representative isolates represents the necessity of vitamin B12 for *V. parahaemolyticus* growth. The ability of *V. parahaemolyticus* to take up vitamin B12 may have implications to the nutrient cycle in marine ecosystems since vitamin B12 is limited in the marine environment and is also required by other marine organisms including phytoplankton, algal flagellates and diatoms (Droop, 1957). Thus, BtuB is likely to be an important OMP for *V. parahaemolyticus* to obtain vitamin B12 from the environment. In fact, certain bacterial species including *Pseudomonas, Bacillus, Acetobacterium* and *Mycobacterium* are capable of vitamin B12 synthesis and these bacteria have an important role in the carbon cycle by being a primary source of vitamin B12 (Bertrand *et al.*, 2011; Rodionov *et al.*, 2003).

Nine OMPs, OmpK2, AggA_VP1634, Porin qsr prophage, OmpW, putative OtnG, YaeT, OmpA2_VPA1186, rare lipoprotein B and putative uncharacterized protein_VP0967 were identified in seven isolates, also indicating an important role for these OMPs in *V. parahaemolyticus*. **OmpK** acts as a receptor of vibriophage KVP40 in *V. parahaemolyticus* (Inoue *et al.*, 1995a), which was initially recovered from *V. parahaemolyticus* in sea water (Matsuzaki *et al.*, 1992). It has a broad host range among *Vibrio* species and its genome contains the components required for horizontal gene transfer, such as the recombinase A-like enzyme and endonuclease which are involved in genetic recombination (Miller *et al.*, 2003). Thus, the vibriophage KVP40 might play an important role in horizontal gene transfer though transduction among *Vibrio* species. The OmpK protein in isolates VP166, VP138, VP204, VP112, VP84 and VP44 was identified as OmpK2, the same OmpK protein as in environmental *V. parahaemolyticus* isolate RIMD2210001. In contrast, OmpK from VP132 and VP178 was identified as OmpK1, the same OmpK protein as in pandemic *V. parahaemolyticus* isolate RIMD2210633. Both OmpK1 and OmpK2 were identified in clinical strain VP178. However, it was not possible to determine whether VP178 contains either OmpK1 or OmpK2 or both OmpK1 and OmpK2 because the peptide search will identify all possible OMPs from amino acid peptide hits of the OMP fractions being analyzed against the protein database. Thus, if the amino acid sequence of OmpK from VP178 is similar to that of OmpK1 as well as OmpK2, both OmpK1 and OmpK2 can be identified. Since a BLAST search of the amino acid sequence of *V. parahaemolyticus* OmpK2 shares 89% similarity with OmpK from another species, namely *V. harveyi* HY01, and has a lower similarity (81%) with OmpK from *V. parahaemolyticus* RIMD2210633, it is feasible that strains VP166, VP138, VP204, VP112, VP84 and VP44 may have acquired *ompK2* from *V. harveyi* HY01 by transduction.

**AggA** was proposed to be a TolC-like protein, having a channel-forming structure that connects the inner and outer membranes of *Shewanella oneidensis,* an anaerobic deep-sea bacterium (Theunissen *et al.,* 2009). Previous studies demonstrated that AggA is upregulated in the biofilm-forming strain of *S. oneidensis* and is also upregulated when the organism is grown under aerobic conditions rather than anaerobic conditions (Beliaev *et al.,* 2002; De Vriendt *et al.,* 2005). Furthermore, AggA has been identified in *Pseudomonas putida,* a plant pathogenic bacterium, and shown to have a role in adherence to plant roots (Buell & Anderson, 1992). The role of AggA in *V. parahaemolyticus* has not been studied to date. However, from the studies cited above, *V. parahaemolyticus* AggA may be involved in biofilm formation and bacteria adherence of this organism.

Putative outer membrane **porin qsr protein_VP1008**, encoded by a gene at the locus of qsr prophage insertion, was found in seven representative *V. parahaemolyticus* isolates. In *E. coli*, a porin-encoding gene, *nmpC,* is located at the locus of qsr prophage insertion on the bacterial chromosome (Highton *et al.,* 1985; Hindahl *et al.,* 1984). Lack of *nmpC* in *E. coli* contributes to the production of a substitute protein, namely Lc, which has an equivalent function to NmpC suggesting an important role of the porin protein encoded by the gene that is acquired from the qsr phophage (Blasband *et al.,* 1986; Highton *et al.,* 1985). The extensive presence of the porin encoded by the locus of qsr prophage in *V. parahaemolyticus* isolates in this study also suggests that the bacteriophage plays an important role in the molecular evolution of this organism.

**OmpW** is an osmotic stress responsive protein that is generally found in many *Vibrio* species including *V. parahaemolyticus* (Yang *et al.,* 2010)*, V. alginolyticus*

(Xu *et al.*, 2005)*, and V. cholerae* (Jalajakumari & Manning, 1990). Since the nucleotide sequence of OmpW is conserved among different biotypes and serogroups of *V. cholerae*, it was suggested to be a species-specific marker for *V. cholerae* (Nandi *et al.*, 2000). OmpW of *V. parahaemolyticus* and *V. alginolyticus* is upregulated under high NaCl concentrations (Xu *et al.*, 2005; Yang *et al.*, 2010). In the present study, expression of OmpW was relatively low in all of the isolates examined, except in human carrier isolate VP132, although the isolates were cultured in growth media (TSB) supplemented with 3% NaCl (Fig. 4.9).

Functional characterization of putative **OtnG** has not been studied in V. *parahaemolyticus*. OtnG of *V. parahaemolyticus* is encoded within the same chromosome (chromosome 1) as OtnA. Both proteins are involved in capsular polysaccharide transport in *V. parahaemolyticus* (Makino *et al.*, 2003) and *V. cholerae* (Bik *et al.*, 1995). Since the OtnA- and OtnG-encoding genes are located close to each other (2.9 Kb) in chromosome 1 in *V. parahaemolyticus*, it is possible that *otnA* and *otnG* are in the same operon and has an important role in capsular polysaccharide production in this bacterium.

**YaeT** is a member of the Omp85 family which play an important role in outer membrane transport and assembly (Jain & Goldberg, 2007). Loss of YaeT contributes to defective outer membrane synthesis (protein organization, folding, insertion, etc.) in *E. coli* (Doerrler & Raetz, 2005; Jain & Goldberg, 2007; Werner & Misra, 2005). Extensive expressions of the YaeT in all isolates (except VP84) in the present study confirm a significant role of this protein for bacterial survival in *V. parahaemolyticus*.

A protein domain homology search from the database (http://www.uniprot.org) showed that the **rare lipoprotein B** and **putative uncharacterized outer membrane protein_VP0967** share similar domain homology with proteins involved in lipopolysaccharide (LPS) transport machinery and with the penicillin-binding protein activator (LpoB), respectively. Since these OMPs were extensively identified in seven representative *V. parahaemolyticus* isolates from diverse sources, they are likely to have significant roles in the survival of *V. parahaemolyticus*.

Comparative OMPs analysis of eight representative strains revealed that four polar flagellins (FlaA, FlaBD, FlaC and FlaF) are likely present in environmental isolates (oyster, shrimp tissue and water isolates) rather than in human-associated isolates (clinical and human carrier isolates) (Fig. 4.9). It is known that *V. parahaemolyticus* possesses both polar and lateral flagellar gene systems (McCarter, 1995, 2001). Polar flagella, which propels the swimmer cells in liquid environments, are continuously produced whereas lateral flagella are produced only when the organism is grown on surface or solid environments (Kim *et al.*, 2000; Stewart & Mccarter, 2003). Since growth conditions in the environment are more varied than in the human body, it is possible that environmental strains express more abundant flagellin proteins than do the human-associated strains to benefit bacteria in diverse environmental conditions.

Although no correlation between common OMP pattern and epidemiological source of *V. parahaemolyticus* was identified, isolate VP132 contained a distinctive molecular mass of OmpU, OmpA1 and OmpA2 proteins compared to those of the seven other strains. Sixteen OMPs were identified only in VP132. These OMPs included four annotated OMPs (YaeC, OprD, OmpX and a membrane

fusion protein) and 12 hypothetical OMPs (putative uncharacterized proteins VP0374, VP2042, VP0039, VP0966, VP1648 and VP0541, putative lipoproteins VP1267, VP1192 and VP0948, putative efflux protein VPA0363, putative V10 pilin and putative exported protein VP0802). **YaeC** is involved in D-methionine transport in *E. coli* (Gál *et al.*, 2002). The function of YaeC in *Vibrio* species has not been characterized to date. **OprD** is a specific porin that is responsible for the uptake of basic amino acids and related metabolites in *Pseudomonas aeruginosa* (Tamber *et al.*, 2006). OprD is also permeable to imipenem and cabapenem, ß-lactem antibiotics that are active against *P. aeruginosa* (Chen *et al.*, 1995; Trias & Nikaido, 1990). Previous studies demonstrated that loss of OprD contributes to the resistance of these antibiotics in *P. aeruginosa* (Köhler *et al.*, 1999; Naenna *et al.*, 2010; Ochs *et al.*, 1999). **OmpX** is involved in adhesion and also promotes bacterial resistance against the bactericidal effects of complement (Mecsas *et al.*, 1995). OmpX consists of an eight-stranded antiparallel ß-barrel that contains exposed ß-sheets (Vogt & Schulz, 1999). This ß-sheet topology of OmpX is similar to that of OmpA, although they differ at the level of internal hydrogen bonding (Vogt & Schulz, 1999). The extracellular ß-sheet edge of OmpX was suggested to have a binding affinity that is associated with adhesion and also promotes resistance against human complement defence mechanisms (Vogt & Schulz, 1999). In this study, OmpX was predicted by bioinformatic methods from the clinical *V. parahaemolyticus* isolate RIMD2210633, and was also identified in isolate VP132 recovered from human carriage. However, in a previous study, OmpX was unable to be identified from the outer membrane proteome of a fish pathogenic *V. parahaemolyticus* isolate by 2-DE gel analysis (Li *et al.*, 2010a). These results suggest that the expression of OmpX may vary between *V. parahaemolyticus* isolates recovered from human and other animal hosts. The ability to resist the bactericidal activity of the

human immune system might be beneficial to this strain in allowing it to colonize the human intestinal tract and survive in the carrier state. Since OmpX is involved in adhesion and is capable of resisting the immune response, studying the presence and sequence variation of OmpX in *V. parahaemolyticus* isolates from multiple hosts will contribute to a better understanding of OmpX and host-specific interactions.

Isolate VP132 represents ST3, which also includes pandemic O3:K6 *V. parahaemolyticus* strains (http://pubmlst.org/vparahaemolyticus/). A previous study showed that the pandemic O3:K6 strain evolved as a result of acquisition of *tdh*, genomic island VPaI-5, and other unidentified genes (Han *et al.*, 2008). However, these authors did not consider how proteins have evolved in this strain. There are no reports to date about the evolution of outer membrane proteins of pandemic and non-pandemic strains of *V. parahaemolyticus*. Although VP132 represents the pandemic ST3, it was recovered from a healthy carrier and represents serotype O3:K46. Thus, it was not possible to determine from this study whether the OMP profile of VP132 represents the OMP profile of the pandemic O3:K6 strains, although they do share the common ST3.

Comparison of the OMPs predicted by the bioinformatic approach with those identified by proteomic analysis of the eight selected strains of *V. parahaemolyticus* showed that OMPs involved in Type III secretion (T3SS) were predicted by the bioinformatic process but were not identified in any of the eight selected isolates by proteomics. However, these isolates were cultured under *in vitro* conditions that are very different from the conditions found in the human intestine. Variation of OMP profiles under *in vitro* and *in vivo* culture conditions has been demonstrated in *Vibrio salmonicida* (Colquhoun & Sorum,

1998) and in other Gram-negative species including *M. haemolytica* (Davies *et al.*, 1994). In a previous study, a much higher abundance of T3SS2-related proteins including VscC2 was detected in *V. parahaemolyticus* when the bacteria were cultured at 37 and 42°C, which corresponds to the temperature in the intestine, compared to lower temperatures of 20, 25 and 30°C (Gotoh *et al.*, 2010). This study also demonstrated that host–derived inducers of virulence phenotypes such as bile acid can enhance the expression of T3SS2-related proteins.

Nine predicted iron uptake OMPs, IrgA, putative 83 kDa decaheme outer membrane cytochrome c, PvuA1, PvuA2, LutA, FhuE, FhuA, HutA and putative Fe-regulated protein B, were predicted from *V. parahaemolyticus* isolate RIMD2210633 by bioinformatic analyses (Table 4.3). With the exception of FhuA (VPA1435), which was found in human carrier isolate VP138, oyster isolate VP204 and shrimp tissue isolate VP112, no OMPs involved in iron uptake were identified by proteomic analyses among eight representative isolates. **FhuA** is an integral OMP that is essential for siderophore transport in Gram-negative bacteria (Braun, 2001). The FhuA structure contains binding sites for ferrichrome, phages T1, T5, Φ80, colicin M, bacteriocin produced by *E. coli* and albomycin, a broad-host range anitibiotic, at the exposed-surface (Ferguson *et al.*, 2000; Killmann *et al.*, 1995). The N-terminus of FhuA interacts with the TonB system whereby the energy from the cytoplasmic membrane is transferred to the outer membrane for iron transport activity (Ferguson *et al.*, 2000). The present study was unable to demonstrate expression of iron uptake OMPs among the Thai clinical and environmental isolates under standard iron-replete growth conditions. Although more iron uptake OMPs including IrgA, putative 83 kDa decaheme outer membrane cytochrome c, PvuA1, PvuA2, LutA, FhuE, HutA and putative Fe-

regulated protein B, were expected to be identified using the gel-based approach, it is likely that these iron-uptake OMPs are not expressed under normal bacterial culture conditions (i.e. TSB and 3% NaCl). The expression of iron-uptake OMPs is likely to increase when bacteria are grown under iron-depleted conditions. The absence of T3SS and iron-uptake-related OMPs in the eight representative strains may suggest an important requirement for *in vivo* growth conditions to induce expression of important virulence factor-related OMPs in *V. parahaemolyticus*.

The OMP profile of the shrimp tissue isolate VP84 (ST251, cluster 5), which represents a distinct lineage in the Neighbour-Joining tree based on the MLST data (Fig. 2.52), is similar to those of the other seven strains. Although VP84 represents a very distinct lineage in the phylogenetic tree as a consequence of horizontal transfer of the *recA* gene, this result suggests that the evolution of housekeeping genes and OMP-encoding genes may not occur in parallel. The phylogenetic relationships of housekeeping genes and OMP-encoding genes of *V. vulnificus* biotype 3 were shown to be incongruent (Bisharat *et al.*, 2007). In this study, a Neighbour-Joining tree based on MLST analysis showed that *V. vulnificus* biotype 3 strains were present in an intermediate position between biotype 1 and 2 populations, whereas the analysis of OMP-encoding genes grouped biotype 3 with one of the two main clusters.

In conclusion, 117 OMPs were predicted from the genome of *V. parahaemolyticus* isolate RIMD2210633. These predicted OMPs comprise proteins involved in biogenesis and integrity, transportation and receptor activity, iron binding and TonB receptor, flagella and motor activity, enzyme activity, adherence and colonization and other activities. Seventy three proteins were identified from

eight *V. parahaemolyticus* isolates recovered from clinical samples, human carriers, seafood and water by gel-based proteomic analysis. Thirty two OMPs were detected by both bioinformatic and proteomic analyses, whereas 85 OMPs were predicted only by bioinformatic prediction and 41 were identified only by the proteomic approach. OMPs involved in TTSSs (YscW, YscJ, YscC, PopN and VscC2) and iron uptake (IrgA, putative 83 Da decaheme outer membrane cytochrome C, PvuA1, PvuA2, LutA, FhuE, HutA and putative-regulated protein B) were predicted from the genome of *V. parahaemolyticus* isolate RIMD2210633, but were not recovered from any of the eight Thai isolates. With the exception of the shrimp tissue isolate VP112, proteins involved in bacteriophage-related activity (i.e. OmpK and porin qsr prophage) were extensively present in all representative isolates, and also were predicted in the *V. parahaemolyticus* isolate RIMD2210633. The gel-based analysis showed that OmpU is a major porin protein which represents the most abundantly expressed protein in all eight *V. parahaemolyticus* isolates grown under *in vitro* conditions. In human carrier isolate VP132, OmpU, OmpA1 and OmpA2 differed in protein abundance and molecular mass compared to other strains. Although no clear association between the OMP profile and the source of isolation, the ST or the serotype was observed, there was nevertheless a high degree of variation of OMP profiles in strains isolated from different origins. OMP profile variation was also observed among the clinical isolates representing identical STs. This study therefore contributes to a better understanding of the OMPs in *V. parahaemolyticus* isolates from different epidemiological sources, and also provides a guideline for further studies that focus on the evolution of virulence-related OMPs in this organism.

# 5. FINAL DISCUSSION AND CONCLUSION

The application of a combined research approach involving MLST, DNA sequence analyses of virulence genes and proteomics in the present study has provided a more thorough understanding of the molecular evolutionary relationships and epidemiology of *V. parahaemolyticus* in Thailand. Although analyses in this study have elucidated the genetic relationships of *V. parahaemolyticus* isolates on a relatively local scale, the outcomes of this research are applicable to isolates from other parts of the world, particularly in the countries where seafood is widely consumed.

The first objective was to study genetic relationships and the population structure of *V. parahaemolyticus* in Thailand using MLST. The difficulty in obtaining PCR products of housekeeping gene fragments from each of the 102 individual isolates using the same primers for each gene has highlighted the high level of nucleotide variation in this organism. Phylogenetic analysis of seven concatenated housekeeping gene sequences of Thai *V. parahaemolyticus* revealed that isolates from clinical samples, shrimp tissue and water were resolved into five distinct clusters (cluster 1 = shrimp tissue isolates from farm 1; cluster 2 = clinical isolates; cluster 3 = water isolates from farm 1; cluster 4 = clinical isolates; cluster 5 = shrimp tissue isolates from farm 2) (Fig. 2.52). In contrast, isolates from human carriers, frozen shrimps from a processing plant and various fresh seafood products were genetically unrelated. STs representing clinical isolates did not include any other isolates from human carriers, frozen shrimps, shrimp tissue, seafood or water. Thus, the MLST analysis was unable to determine the likely sources of the *V. parahaemolyticus* infections. Genetic association of environmental isolates from the same source at the same time

point of isolation was evident in shrimp tissue at farm 1 (cluster 1; Fig. 2.52) and farm 2 (cluster 5; Fig. 2.52) and water isolates at farm 1 (cluster 3; Fig. 2.52). In contrast, genetic relatedness of isolates from frozen shrimps was not observed, although these were obtained from the same processing factory. However, these frozen shrimp probably originated from various locations, since they were sourced by the processing factory from different farming areas. This could be one explanation for the genetic diversity of isolates from the frozen shrimp observed in the present study.

The presence of two clonal groups of clinical isolates responsible for gastroenteritis cases in Thailand during 1990-1991 indicates that these two clones may have greater fitness to survive in the regional seafood or in human hosts, enabling them to persist and be capable of causing disease in consumers who ingest contaminated seafood during that period. Emergence of pathogenic strains that cause outbreaks of gastroenteritis is possibly due to the horizontal transmission of virulence factor-encoding genes from pathogenic strains to non-pathogenic strains (Smith, 2001). The introduction of pathogenic *V. parahaemolyticus* strains in the faeces of gastroenteritis patients into the environment allows pathogenic and environmental strains to exchange genetic material including virulence factor-encoding genes. By this process, environmental strains that become pathogenic by acquiring virulent factor-encoding genes are increased in a certain area. Consequently, there is a greater chance that humans who consume seafood in the same area also ingest these pathogenic strains that are capable of causing gastroenteritis.

Serotypic variation within isolates from the same genetic cluster was present in clinical cluster 4 but not in clinical cluster 2. All isolates in clinical cluster 2,

with the exception of VP194 (OUT:KUT), contained the identical serotype O1:K1, whereas isolates in clinical cluster 4 contained four different serotypes. However, serotypic conversion in clinical isolates in the present study was not as extensive as it was in environmental isolates, including those from shrimp tissue clusters 1 (seven serotypes) and 5 (eight serotypes) and water cluster 3 (six serotypes). These results suggest that rapid change in O and K antigens of *V. parahaemolyticus* occurred not only in pandemic strains (Chen *et al.*, 2010, 2011) but also extensively in environmental isolates.

MLST analyses showed that the majority of isolates from human carriers were genetically different from the isolates that were commonly found in clinical samples and seafood. Evidence of a high rate of recombination, particularly at the *recA* locus in human carrier isolates (Fig. 2.54) suggests that the human intestinal tract may serve as a reservoir providing an environment for the emergence of new strains. However, three human carrier isolates (VP138, VP156 and VP162) were closely related to isolates of clinical clusters 2 or 4, and five human carrier isolates (VP132, VP140, VP154, VP158 and VP199) represented STs identical with clinical isolates from worldwide distributions. In particular, VP132 and VP158 represented ST3 which includes the worldwide pandemic serotype O3:K6 strains and its serovarients (http://pubmlst.org/vparahaemolyticus/). These results suggest close genetic relationships between the isolates inhabiting human hosts and clinical isolates, and also the possibility that healthy workers in seafood factories can be carriers of pathogenic *V. parahaemolyticus*. According to the latest risk assessment report of *V. parahaemolyticus* in seafood established by the World Health Organization (WHO) and the Food and Agriculture Organization of the United Nations (FAO), factors involved in seafood production such as temperature control and hygiene practices after harvesting

have critical roles in minimizing the level of *V. parahaemolyticus* in seafood (Boonyawantang *et al.*, 2012; WHO & FAO, 2011; Xie *et al.*, 2012). This report provides information about the characteristics of the human hosts (e.g. sex, ages and ethnic group) as consumers, but does not include information about the human host as carriers of potentially pathogenic *V. parahaemolyticus*. Surveillance of *V. parahaemolyticus* in seafood production in Japan revealed that the decrease of *V. parahaemolyticus* infection was a consequence of microbial controlling schemes by the government rather than the reduction of *V. parahaemolyticus* contamination in retail seafood (Hara-Kudo *et al.*, 2012). Together with results from the present study, this suggests that the risk of contamination from workers involved in seafood manufacture should be included in the risk assessment scheme of *V. parahaemolyticus* in seafood.

Knowledge of the genetic relationships of *V. parahaemolyticus* isolates based on housekeeping genes facilitated the selection of representative isolates from different clonal groups for the further study of the molecular evolution of virulence-determining encoding genes. This concept led to the second objective of this research, which was a study of the distribution and DNA sequence variation of haemolysin (*tdhA, tdhS, trh1* and *trh2*) and TTSS1-related (*vcrD1, vscC2* and *VP1680*) genes. Nine out of 18 (50%) human carrier isolates contained *tdhA* and the majority of these isolates also contained *tdhS*. The presence of these haemolysin genes in human carrier isolates indicates that healthy human carriers are able to harbour *V. parahaemolyticus* possessing virulence factor-encoding genes. Nucleotide sequence analyses of selected virulence genes revealed that human carrier isolates VP132 share identical *tdhA, tdhS, vcrD1, vscC2* and *VP1680* genes with clinical isolate RIMD2210633 and the human carrier isolates VP138 share identical *tdhA, tdhS, vcrD1, vscC2* and *VP1680* genes with

clinical isolate VP178 (Fig. 3.10 and Figs. 3.12-3.14). The human carrier isolate VP132 also shared identical TTSS1 genes *vcrD1, vscC2* and *VP1680* genes with other pathogenic strains, mainly serotype O3:K6, from worldwide sources (Fig. 3.12-3.14). Furthermore, the clinical isolate VP178 also shared very similar *trh1* fragment sequences with the human carrier isolate VP138 (Fig. 3.11). These results highlight that *V. parahaemolyticus* isolates inhabiting healthy human carriers possess identical virulence genes to those of pathogenic isolates. However, a comparative phenotypic test of haemolysin and TTSS1 genes from clinical and human carrier isolates is required to determine whether these virulence genes are equally functional in both clinical and human carrier isolates.

The finding of very similar *trh2* fragment sequences in clinical (VP166), human carrier (VP132) and seafood (VP216) isolates (Fig. 3.11) suggests that *trh2* has been transferred across genetically unrelated isolates (Fig. 2.52) that were recovered from different sources. Evidence of horizontal gene transfer of haemolysin genes was also provided by the different *tdh*/*trh* gene profiles in clinical isolates of the same STs (ST83 and ST262). These results suggest that horizontal gene transfer may play an important role in the evolution of virulence genes among human-associated and environmental *V. parahaemolyticus* isolates (Han *et al.*, 2008; Harth *et al.*, 2009). It has been shown that pandemic *V. parahaemolyticus* strains of serotype O3:K6 contain distinct genetic elements including ORF8 of phage f237 (Nasu *et al.*, 2000) and unique *toxRS* gene sequences (Matsumoto *et al.*, 2000), and also have unique genetic profiles generated by AP-PCR (Matsumoto *et al.*, 2000; Okuda *et al.*, 1997) and PFGE (Wong *et al.*, 2000). Several studies have demonstrated that the emergence of pandemic *V. parahaemolyticus* strains of serotype O3:K6 is a consequence of

substantial genomic flux, particularly in pathogenicity islands (VPaIs) which are located in both chromosome 1 (VPaI 1-5) and chromosome 2 (VPaI 6-7) (Boyd *et al.*, 2008; Chao *et al.*, 2009a, 2010; Chen *et al.*, 2011; Han *et al.*, 2008; Harth *et al.*, 2009). However, genetic variation in TTSS1 genes among pandemic *V. parahaemolyticus* serotype O3:K6 and related strains, as well as isolates from other sources, has not been studied. DNA sequence analyses in the present study indicate that the TTSS1 *vcrD1, vscC1* and *VP1680* genes from pandemic *V. parahaemolyticus* serotype O3:K6 (RIMD2210633) and related strains (AQ3810, AQ4037, Peru466, AN5034 and K5030) differ from that of Thai isolates from clinical samples (VP178), human carriers (VP138), seafood (VP216), frozen shrimp (VP36) and the Japanese type strain (VP2). However, an ST3 isolate (VP132) recovered from a human carrier possessed the same TTSS1 genes as the pandemic *V. parahaemolyticus* serotype O3:K6 (RIMD2210633) and related strains. Thus, this study suggests that healthy human carriers harbour *V. parahaemolyticus* strains possessing virulence genes identical with pandemic strains of serotype O3:K6.

The third objective of this research was to study the outer membrane proteome of *V. parahaemolyticus* isolates from various sources. One hundred and seventeen OMPs were predicted from the genome of *V. parahaemolyticus* isolate RIMD2210633. These predicted OMPs were classified into seven functional groups (i.e. biogenesis and integrity, transportation activity, iron-binding activity, flagellar and motility activity, enzyme activity, adherence activity and other activity). Among these seven functional groups, proteins involved in transportation represented the largest group suggesting an important role for OMPs involved in nutrient uptake in this organism. Proteins involved in biogenesis and integrity represented the second largest group followed by iron-

binding activity, flagellar and motility activity, enzyme activity, adherence and other activity, respectively (Table 4.3).

Seventy-three unique OMPs were identified from eight isolates by proteomic analyses. Predicted OMPs of isolate RIMD2210633 and OMPs identified in the human carrier isolate VP132 were compared, since these two isolates share the same sequence type (ST3). Seven porin proteins, BtuB, ChiP, LamB, OmpU, OmpN, OprD and qsr prophage were predicted in isolate RIMD2210633. With the exception of ChiP, a specific porin protein involved in chitooligosacharide uptake, all of these predicted porin proteins were also identified in the human carrier isolate VP132. Expression of ChiP was also not detected in the other seven selected isolates from Thailand, although isolates from shrimp and oyster tend to be capable of utilizing chitin from the shells of host species. Lack of expression of ChiP may be due to insufficient chitin compound in the culture media. It may be possible to induce the expression of ChiP using growth media with added chitin supplement.

OMPs involved in TTSSs (YscW, YscJ, YscC, PopN and VscC2) and iron uptake (IrgA, putative 83 Da decaheme outer membrane cytochrome C, PvuA1, PvuA2, LutA, FhuE, HutA and putative-regulated protein B) were predicted from the *V. parahaemolyticus* isolate RIMD2210633 genome, but were not identified in any of the eight Thai isolates. Only one predicted iron uptake OMP (FhuA) from isolate RIMD2210633 genome was identified in isolates VP138, VP204 and VP112. Since the TTSS-related and iron uptake proteins are involved in bacterial virulence, expression of these proteins may be induced by the host environment (i.e. human intestinal tract) (Kuntumalla *et al.*, 2011). OMPs involved in iron-binding activity including FhuA are probably induced when the isolates are grown under

iron-depleted condition (Funahashi *et al.*, 2002, 2009; Miyamoto *et al.*, 2009). A previous study demonstrated that TTSS2 was involved in *V. parahaemolyticus* colonization which subsequently caused pathogenicity in the infant rabbit intestine (Ritchie *et al.*, 2012). Since the isolates used for the proteomic analyses were cultured under *in vitro* conditions, which are very different from the environment in the human host, it is possible that expression of TTSS2-related OMPs in these isolates would be induced when the isolates are grown under *in vivo* condition.

Interestingly, OmpX, a protein that was predicted from the *V. parahaemolyticus* isolate RIMD2210633 genome, was identified in the human carrier isolate VP132 but not in the other seven isolates. Expression of OmpX in isolate VP132 may suggest an important role for this protein in the isolate representing pandemic ST3. OmpX is involved in adhesion and in resistance to the human immune response in *E. coli* (Vogt & Schulz, 1999) and is an important virulence factor for *Yersinia pestis* (also known as Ail protein in *Y. pestis*) (Kolodziejek *et al.*, 2012). The roles of OmpX in virulence of *Y. pestis* include serum resistance (Bartra *et al.*, 2008; Kolodziejek *et al.*, 2007, 2010), adhesion and internalization to host cells (Felek & Krukonis, 2009; Yamashita *et al.*, 2011), delivery TTSS effector protein (Yop) (Felek & Krukonis, 2009; Tsang *et al.*, 2010), biofilm formation (Kolodziejek *et al.*, 2007) and inhibition of inflammatory response (Felek & Krukonis, 2009; Hinnebusch *et al.*, 2011). It may be one of the factors that facilitates the colonization of *V. parahaemolyticus* RIMD2210633 and VP132 in the human intestinal tract, and could potentially serve as a virulence factor for pathogenic *V. parahaemolyticus.*

The OMP profile of isolate VP132 represented different molecular masses of the major OMPs OmpU, OmpA1 and OmpA2_VPA2048 compared to those of the other seven isolates (Fig. 4.9). Further study of the OMP profiles of clinical isolates of the pandemic ST3 is required to confirm that the unique OMP profile of the human carrier VP132 (ST3) represents a common characteristic of *V. parahaemolyticus* ST3. With the exception of VP132, the OmpU protein in clinical (VP166 and VP178) and human carrier (VP138) isolates had a lower molecular mass than that of environmental isolates (VP204, VP122, VP84 and VP44). Since the OmpU is responsible for stress tolerance (e.g. to bile salt and acid) and intestinal colonization in mice (Whitaker *et al.*, 2012), the OmpU protein of human-associated isolates may differ from that of environmental isolates.

Two different vibriophage (KVP40) receptor OmpK proteins (OmpK1 and OmpK2) were identified in eight isolates. The clinical isolate VP178 and human carrier isolate VP132 contained OmpK1, the OmpK protein from pandemic *V. parahaemolyticus* isolate RIMD2210633, whereas the other isolates, VP166, VP138, VP204, VP84, VP112 and VP44, contained OmpK2, the OmpK protein from environmental *V. parahaemolyticus* isolate RIMD2210001. The presence of the OmpK2 protein in clinical (VP166) and human carrier (VP138) isolates, as well as environmental isolates, suggests that horizontal gene transfer of OmpK-encoding genes has occurred among human-associated and environmental isolates. Furthermore, amino acid sequence analysis of OmpK2 showed a best match with the OmpK of other *Vibrio* species such as *V. harveyi*. These results suggest mobility of OmpK-encoding genes across *Vibrio* species and this may have implications for the evolution of *V. parahaemolyticus*. Furthermore, extensive expression of the phage-derived VP1008 porin protein encoded by a gene at the

locus of qsr prophage insertion in all isolates except VP112 suggests an important role of functional proteins introduced by bacteriophages in this organism.

Acquisition of bacteriophage-encoded virulence factors has been reported in many Gram-negative pathogenic bacteria including *E. coli, Shigella spp., P. aeruginosa, V. cholerae* (Boyd & Brüssow, 2002). It is also known that the pandemic *V. parahaemolyticus* serotype O3:K6 is associated with a filamentous phage (f237) (Nasu *et al.*, 2000). A previous study demonstrated that bacteriophages isolated from both shellfish and finfish were capable of infecting pandemic *V. parahaemolyticus* serotype O3:K6 (Bastías *et al.*, 2010). These bacteriophages were able to multiply in a *V. parahaemolyticus* pandemic strain, and this process allowed bacteriophages to exchange genetic material between the phage DNA and host chromosome. After lysis stage of bacteriophage infection, released bacteriophages possessing a DNA fragment from the host cell were able to further infect other strains located nearby. Previous study demonstrated that bacteriophage Vp1, which was isolated from shrimp pond water, seawater, estuarine water, shrimp surface and tissue, has specific infectivity to *V. parahaemolyticus*, but not to other *Vibrio* species including *V. alginolyticus* and *V. harveyi* (Alagappan *et al.*, 2010). These findings suggest that transduction-mediated horizontal gene transfer may be one of the key factors in the evolutionary change of *V. parahaemolyticus*.

As whole genome sequencing becomes more available as a routine practice, genomic analyses of *V. parahaemolyticus* isolates from various sources will provide a more comprehensive understanding of the fine scale molecular evolution of the examined strains. Complete genome sequencing makes it possible to generate other typing analyses, including single nucleotide

polymorphism (SNP) typing (Hendriksen *et al.*, 2011) and pangenome family tree construction (Snipen & Ussery, 2010). Whole genome sequencing yields high resolution for evolutionary analyses of recent epidemic clones whereas MLST has limited resolution to distinguish between closely related isolates (Achtman, 2008). Thus, microevolution within emerging clones may not be resolved by MLST. To date, MLST data can be interrogated from whole genome sequencing data (Larsen *et al.*, 2012). Furthermore, genome sequence analyses have been widely used to study the epidemiology and evolution of bacterial pathogens including *Y. pestis, V. cholerae, C. difficile,* MRSA, *S. pneumoniae* and Group A *Streptococcus* (Parkhill & Wren, 2011). However, MLST is still considered as a 'gold standard' typing method because it has more standardized implementation compared with other genomic sequence-based typing methods (e.g. SNP typing and pangenome family tree) (Larsen *et al.*, 2012).

In conclusion, MLST analyses reveal high genetic diversity of *V. parahaemolyticus* isolated from Thailand. Extensive recombination involving the *recA* locus influences the topology of the phylogenetic tree constructed from seven housekeeping genes. Genetic association between clinical isolates and other isolates from human carriers, seafood, shrimp tissue, frozen shrimp and water was not resolved by MLST. Healthy human carriers potentially act as reservoirs of pathogenic *V. parahaemolyticus,* and may also provide an environment for the emergence of new strains. Remarkably high serotypic conversion among isolates of the same genetic clusters was observed in both clinical and environmental (shrimp tissue and water) isolates, although more serotypic variation was present in the environmental clusters. Comparative nucleotide sequence analysis of the virulence genes *tdhA, tdhS, trh1, trh2, vcrD1, vscC1* and *VP1680* demonstrated that human carriers can harbour *V. parahaemolyticus* isolates with

identical virulence genes to those of pathogenic strains. Human carrier isolate VP132 (ST3) shares identical TTSS1 genes *vcrD1, vscC1* and *VP1680* with pandemic *V. parahaemolyticus* serotype O3:K6 isolate RIMD2210633 (ST3) and related strains, and these sequences are distinct from those of other Thai isolates including clinical, human carrier, frozen shrimp, and seafood isolates and the Japanese type strain. Comparison of the OMP profiles of eight *V. parahaemolyticus* isolates revealed a highly abundant OmpU protein in all isolates examined. The OMP profiles of the eight isolates were not associated with epidemiological source, ST or serotype. However, the human isolate VP132 had a distinct OMP profile with OmpU, OmpA1 and OmpA2_VPA2048 proteins of different molecular masses compared to those of the other seven isolates. OMPs involved in TTSS, iron uptake activity and chitin transportation could not be identified in any of the eight isolates, although they were predicted from the genome of *V. parahaemolyticus* isolate RIMD2210633. Furthermore, the common presence of a porin protein encoded by a gene located at the qsr prophage insertion site in most isolates suggests a contribution of bacteriophages in *V. parahaemolyticus* evolution.

The present study has demonstrated the evolutionary relationships of *V. parahaemolyticus* with respect to analyses of seven housekeeping genes (MLST), virulence gene sequences and OMPs. The bacterial collection used in this study represents isolates recovered from various sources throughout the seafood production chain, sources which were not represented for isolates examined in previous MLST studies of *V. parahaemolyticus* (Chao *et al.*, 2011; Chowdhury *et al.*, 2004; Ellis *et al.*, 2012; González-Escalona *et al.*, 2008; Yan *et al.*, 2011; Yu *et al.*, 2011). Although the likely source of *V. parahaemolyticus* infection in seafood production could not be determined, the major outcomes of this

research raises awareness of the contamination of workers involved in the seafood industry with pathogenic *V. parahaemolyticus*. This concern should be included in the risk assessment of *V. parahaemolyticus* in seafood production in Thailand, as well as other seafood exporting countries, in order to elevate the standard of seafood safety in both domestic and global supplies.

### *Future research*

The knowledge from the present study points to the requirement for a far higher genetic resolution [such as provided by next-generation genome sequencing (NGS)], combined with a focus on specific clones, in order to robustly pin-point the most significant sources of contamination. NGS is a cost-effective high-throughput sequencing technology that generates genomic sequence data from organisms in much less time (Mardis, 2008; Metzker, 2010). Sequences data obtained by this approach allow us to study comparative genomics of various bacterial isolates and enables the identification of virulent determinants of pathogenic strains. In particular, comparative genomic analyses of pandemic *V. parahaemolyticus* strain RIMD2210633 (ST3) and the human carrier isolate VP132 (ST3) will provide a better understanding of the genetic evolution of potentially pathogenic *V. parahaemolyticus* isolates inhabiting healthy human carriers involved in seafood production. Further advantages of using NGS include bacterial species identification, antibiotic resistance profile analyses and pathogenic strains outbreak detection (Didelot *et al.*, 2012; MacLean *et al.*, 2009). For example, application of NGS to study the comparative genomics of four *V. vulnificus* strains enabled possible virulence factors including Flp pili, GGDEF proteins and genomic island XII to be identified (Gulig *et al.*, 2010). Genomic data analyses by NGS of five clinical *V. cholerae* strains from worldwide

distributions revealed that the recent Haitian cholera outbreak in 2010 was caused by the introduction of El Tor O1 strains from South Asia, a distant geographic region, rather than from South America (Chin *et al.*, 2011; Dasgupta *et al.*, 2012).

In the present study, *V. parahaemolyticus* strains isolated from two different farms represented two distinct clusters (Fig. 2.52). It would be interesting to further investigate geographical variation of environmental *V. parahaemolyticus* by MLST analysis using isolates collected from various shrimp farms at different locations and time points. Furthermore, study of OMP profile of isolates that are grown under conditions imitating human host environment such as iron-depleted and bile-containing conditions will enable us to understand the role of OMPs in host adaptation of this organism. Lastly, comparative nucleotide sequence analysis of genes encoding OMPs such as OmpU and OmpK, OMPs that exhibited variation of molecular mass among isolates from different sources (Fig 4.9), will contribute to a better understanding of molecular evolution of OMP in pathogenic and non-pathogenic strains of *V. parahaemolyticus.*

# 6. REFERENCES

**Abdallah, F., Ellafi, A., Lagha, R., Bakhrouf, A., Namane, A., Rousselle, J. C., Lenormand, P. & Kallel, H. (2010).** Identification of outer membrane proteins of *Vibrio parahaemolyticus* and *Vibrio alginolyticus* altered in response to γ-irradiation or long-term starvation. *Res Microbiol* **161**, 869–875.

**Abdallah, F. Ben, Bakhrouf, A., Ayed, A. & Kallel, H. (2009a).** Alterations of outer membrane proteins and virulence genes expression in gamma-irradiated *Vibrio parahaemolyticus* and *Vibrio alginolyticus*. *Foodborne Pathog Dis* **6**, 1171–1176.

**Abdallah, F. Ben, Kallel, H. & Bakhrouf, A. (2009b).** Enzymatic, outer membrane proteins and plasmid alterations of starved *Vibrio parahaemolyticus* and *Vibrio alginolyticus* cells in seawater. *Mol Microbiol* **191**, 493–500.

**Achtman, M. (2008).** Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* **62**, 53–70.

**Achtman, M., Wain, J., Weill, F. X., Nair, S., Zhou, Z., Sangal, V., Krauland, M. G., Hale, J. L., Harbottle, H., & other authors. (2012).** Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS pathog* **8**, e1002776.

**Aeckersberg, F., Lupp, C., Feliciano, B. & Ruby, E. G. (2001).** *Vibrio fischeri* outer membrane protein OmpU plays a role in normal symbiotic colonization. *J of bacteriol* **183**, 6590–6597.

**Akeda, Y., Okayama, K., Kimura, T., Dryselius, R. & Kodama, T. (2009).** Identification and characterization of a type III secretion-associated chaperone in the type III secretion system 1 of *Vibrio parahaemolyticus*. *FEMS Microbiol Lett* **296**, 18–25.

**Alagappan, K. M., Deivasigamani, B., Somasundaram, S. T. & Kumaran, S. (2010).** Occurrence of *Vibrio parahaemolyticus* and its specific phages from shrimp ponds in east coast of India. *Curr Microbiol* **61**, 235-240.

**Alam, M., Chowdhury, W. B., Bhuiyan, N. A., Hasan, N. A., Nair, G. B., Watanabe, H., Huq, A., Sack, R. B., Akhter, M. Z., & other authors. (2009).** Serogroup, virulence and genetic traits of *Vibrio parahaemolyticus* in the estuarine ecosystem of Bangladesh. *Appl Environ Microbiol* **75**, 6268–6274.

**Alm, R. (1986).** Detection of an OmpA-like protein in *Vibrio cholerae*. *FEMS Microbiol Lett* **37**, 99–104.

**Amin, S. a, Green, D. H., Küpper, F. C. & Carrano, C. J. (2009).** Vibrioferrin, an unusual marine siderophore: iron binding, photochemistry, and biological implications. *Inorg Chem* **48**, 11451-11458.

Andersen, C., Hughes, C. & Koronakis, V. (2000). Chunnel vision: Export and efflux through bacterial channel-tunnels. *EMBO Rep* **1**, 313-318.

Anonymous. (1997). Outbreak of *Vibrio parahaemolyticus* related to raw oyster in British columbia. *Can Commun Dis* **23-19**, 145–148.

Anonymous. (2007). Quality reference criteria, Division of fish inspection and quality control, Department of Fisheries, Thailand.

Anonymous. (2009). Quality reference criteria, Division of fish inspection and quality control, Department of Fisheries, Thailand. Available online http://www.fisheries.go.th/quality/std%20micro.html.

Ansede-Bermejo, J., Gavilan, R. G., Triñanes, J., Espejo, R. T. & Martinez-Urtaza, J. (2010). Origins and colonization history of pandemic *Vibrio parahaemolyticus* in South America. *Mol Ecol* **19**, 3924–3937.

Assavanig, A., Suthienkul, O., Lertsiri, S., A, B., Utrarachkij, F. & Komolprasert, V. (2008). Report of diagnosis and surveillance of enteropathogens and toxic chemicals in white shrimp (*Penaeus vannamei*) in north Andaman provinces (DSECS) (Phase I - Phang-nga). Faculty of Public Health, Mahidol University, Thailand.

Athajariya, T. (2004). The study of serotypes and virulence genes of *Vibrio parahaemolyticus* in healthy carriers at frozen seafood plants. *A thesis submitted in partial fulfillment of the requirement for the degree of Master of Science, Faculty of Public Health, Mahidol University, Thailand.*

Atthasampunnpa, P. (1974). *Vibrio parahaemolyticus* food poisoning in Thailand. In *International Symposium on Vibrio parahaemolyticus*, pp. 21–26. Edited by T. Fujino, G. Sakaguchi, R.Sakazaki & T. Y. Tokyo: Saigon publishing.

Aufrere, R., Temp, M. & Bohin, J. (1986). Regulation of expression of the gene for vitamin B12 receptor cloned on a multicopy plasmid in *Escherichia coli*. *Mol Gen Genet* **467**, 358–365.

Austin, B. & Zhang, X.-H. (2006). *Vibrio harveyi*:a significant pathogen of marine vertebrates and invertebrates. *Lett Appl Microbiol* **43**, 119-124.

Bagos, P. G., Liakopoulos, T. D., Spyropoulos, I. C. & Hamodrakas, S. J. (2004). A Hidden Markov Model method, capable of predicting and discriminating beta-barrel outer membrane proteins. *BMC bioinformatics* **5**, 29.

den Bakker, H. C., Didelot, X., Fortes, E. D., Nightingale, K. K. & Wiedmann, M. (2008). Lineage specific recombination rates and microevolution in *Listeria monocytogenes*. *BMC Evol Biol* **8**, 277.

Barker, W. H., Mackowiak, P. a, Fishbein, M., Morris, G. K., D'Alfonso, J. a, Hauser, G. H. & Felsenfeld, O. (1974). *Vibrio parahaemolyticus* gastroenteritis outbreak in Covington, Louisiana, in August 1972. *Am J Epidemiol* **100**, 316–323.

Bartra, S. S., Styer, K. L., O'Bryant, D. M., Nilles, M. L., Hinnebusch, B. J., Aballay, A. & Plano, G. V. (2008). Resistance of *Yersinia pestis* to

complement-dependent killing is mediated by the Ail outer membrane protein. *Infect Immun* **76**, 612–622.

**Baslé, A., Rummel, G., Storici, P., Rosenbusch, J. P. & Schirmer, T. (2006).** Crystal structure of osmoporin OmpC from *E. coli* at 2.0 A. *J Mol Biol* **362**, 933–942.

**Bastías, R., Higuera, G., Sierralta, W. & Espejo, R. T. (2010).** A new group of cosmopolitan bacteriophages induce a carrier state in the pandemic strain of *Vibrio parahaemolyticus*. *Environ Microbiol* **12**, 990–1000.

**Beher, M. G., Schnaitman, C. a & Pugsley, a P. (1980).** Major heat-modifiable outer membrane protein in Gram-negative bacteria: comparison with the OmpA protein of *Escherichia coli*. *J Bacteriol* **143**, 906–913.

**Beliaev, A. S., Thompson, D. K., Khare, T., Lim, H., Brandt, C. C., Li, G., Murray, A. E., Heidelberg, J. F., Giometti, C. S., & other authors. (2002).** Gene and protein expression profiles of *Shewanella oneidensis* during anaerobic growth with different electron acceptors. *OMICS J Integr Biol* **6**, 39–60.

**Bentley, M. G. (2011).** The global spread of the Chinese mitten crab *Eriocheir sinensis*. In *the Wrong Place - Alien Marine Crustaceans: Distribution, Biology and Impacts, Invading*, pp. 107–127. Edited by B. S. Galil. Dordrecht: Springer Netherlands.

**Bertrand, E. M., Saito, M. a., Jeon, Y. J. & Neilan, B. a. (2011).** Vitamin B12 biosynthesis gene diversity in the Ross Sea: the identification of a new group of putative polar B12 biosynthesizers. *Environ Microbiol* **13**, 1285–1298.

**Berven, F. S., Flikka, K., Jensen, H. B. & Eidhammer, I. (2004).** BOMP: a program to predict integral beta-barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. *Nucleic Acids Res* **32**, W394–W399.

**Berven, F. S., Karlsen, O. A., Straume, A. H., Flikka, K., Murrell, J. C., Fjellbirkeland, A., Lillehaug, J. R., Eidhammer, I. & Jensen, H. B. (2006).** Analysing the outer membrane subproteome of *Methylococcus capsulatus* (Bath) using proteomics and novel biocomputing tools. *Arch Microbiol* **184**, 362–377.

**Bhattacharjee, R. N., Park, K. S., Okada, K., Kumagai, Y., Uematsu, S., Takeuchi, O., Akira, S., Iida, T. & Honda, T. (2005).** Microarray analysis identifies apoptosis regulatory gene expression in HCT116 cells infected with thermostable direct hemolysin-deletion mutant of *Vibrio parahaemolyticus*. *Biochem Biophys Res Commun* **335**, 328–334.

**Bhattacharjee, R. N., Park, K.-S., Kumagai, Y., Okada, K., Yamamoto, M., Uematsu, S., Matsui, K., Kumar, H., Kawai, T., & other authors. (2006).** VP1686, a *Vibrio* type III secretion protein, induces toll-like receptor-independent apoptosis in macrophage through NF-kappaB inhibition. *J Biol Chem* **281**, 36897–36904.

**Bhattacharya, M., Roy, S. S., Biswas, D. & Ranajit, K. (2000).** Effect of $Mg^{2+}$ ion in protein secretion by magnesium-resistant strains of *Pseudomonas*

*aeruginosa* and *Vibrio parahaemolyticus* isolated from the coastal water of Haldia port. *Microbiology* **185**, 151–156.

**Bhoopong, P., Palittapongarnpim, P., Pomwised, R., Kiatkittipong, A., Kamruzzaman, M., Nakaguchi, Y., Nishibuchi, M., Ishibashi, M. & Vuddhakul, V. (2007).** Variability of properties of *Vibrio parahaemolyticus* strains isolated from individual patients. *J Clin Microbiol* **45**, 1544–1550.

**Bik, E., AE, B., RJ, W., AC, C. & FR., M. (1996).** Genetic organization and functional analysis of the *otn* DNA essential for cell-wall polysaccharide synthesis in *Vibrio cholerae* O139. *Mol Microbiol* **20**, 799–811.

**Bik, E. M., Bunschoten, A. E., Gouw, R. D. & Mooi, F. R. (1995).** Genesis of the novel epidemic *Vibrio cholerae* 0139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. *EMBO J* **14**, 209–216.

**Bilung, L. M., Radu, S., Bahaman, A. R., Rahim, R. A., Napis, S., Ling, M. W. C. V., Tanil, G. B. & Nishibuchi, M. (2005).** Detection of *Vibrio parahaemolyticus* in cockle (*Anadara granosa*) by PCR. *FEMS Microbiol Lett* **252**, 85–88.

**Binet, R., Létoffé, S., Ghigo, J. M., Delepelaire, P. & Wandersman, C. (1997).** Protein secretion by Gram-negative bacterial ABC exporters – a review. *Gene* **192**, 7–11.

**Bisharat, N., Cohen, D. I., Harding, R. M., Falush, D., Crook, D. W., Peto, T. & Maiden, M. C. (2005).** Hybrid *Vibrio vulnificus*. *Emerg Infect Dis* **11**, 30–35.

**Bisharat, N., Llorens, A., Amaro, C. & Cohen, D. I. (2007).** Serological and molecular characteristics of *Vibrio vulnificus* biotype 3: evidence for high clonality Printed in Great Britain. *Microbiology* **153**, 847–856.

**Blasband, A. J., Marcotte, W. R. & Schnaitman, A. (1986).** Structure of the *IC* and *nmpC* outer membrane porin protein genes of lambdoid bacteriophage. *J Biol Chem* **261**, 12723–12732.

**Bondad-Reantaso, M. G., Subasinghe, R. P., Josupeit, H., Junning, C. & Xiaowei, Z. (2012).** The role of crustacean fisheries and aquaculture in global food security: past, present and future. *J Invertebr Pathol* **110**, 158–165.

**Boonyawantang, A., Mahakarnchanakul, W., Rachtanapun, C. & Boonsupthip, W. (2012).** Behavior of pathogenic *Vibrio parahaemolyticus* in prawn in response to temperature in laboratory and factory. *Food Control* **26**, 479–485.

**Bos, M. P., Robert, V. & Tommassen, J. (2007).** Biogenesis of the Gram-negative bacterial outer membrane. *Annu Rev Microbiol* **61**, 191–214.

**Boyd, E. F. & Brüssow, H. (2002).** Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol* **10**, 521–529.

**Boyd, E. F., Cohen, A. L. V, Naughton, L. M., Ussery, D. W., Binnewies, T. T., Stine, O. C. & Parent, M. A. (2008).** Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*. *BMC Microbiol* **14**, 1–14.

**Braun, V. (2001).** Iron uptake mechanisms and their regulation in pathogenic bacteria. *Int J Med Microbiol* **291**, 67-79.

**Brayton, P. R., Bode, R. B., Colwell, R. R., Macdonell, M. T., Hall, H. L. & Grimes, D. J. (1986).** *Vibrio cincinnatiensis* sp. nov., a new human pathogen. *Microbiology* **23**, 104-108.

**Brock, J.A. and Lightner, D.V. (1990).** Chapter 3: Diseases of crustacea. *In: O. Kinne (ed.) Diseases of Marine Animals* **3**, Biologische Anstalt Helgoland, Hamburg, 245-424.

**Broz, P., Ohlson, M. B., Monack, D. M. & Immunity, I. (2012).** Innate immune response to *Salmonella typhimurium*, a model enteric pathogen. *Microbiology* **3**, 62–70.

**Buchanan, K. (1999).** ß-Barrel proteins from bacterial function and refolding outer membranes: structure, function and refolding. *Curr Opin Struc Biol* **9**, 455–461.

**Buell, C. R. & Anderson, A. J. (1992).** Genetic analysis of the *aggA* locus involved in agglutination and adherence of *Pseudomonas putida*, beneficial fluorescent Pseudomonad. *Mol Plant Microbe In* **5**, 154-162.

**Burns, D. L. (1999).** Biochemistry of type IV secretion. *Curr Opin Microbiol* **2**, 25-29.

**Byun, R., Elbourne, L. D., Lan, R. & Reeves, P. R. (1999).** Evolutionary relationships of pathogenic clones of *Vibrio cholerae* by sequence analysis of four housekeeping genes. *Infect Immun* **67**, 1116–1124.

**Büttner, D. & Bonas, U. (2002).** Port of entry-the type III secretion translocon. *Trends Microbiol* **10**, 186–192.

**Caburlotto, G., Gennari, M., Ghidini, V., Tafi, M. & Lleo, M. M. (2009).** Presence of T3SS2 and other virulence-related genes in *tdh*-negative *Vibrio parahaemolyticus* environmental strains isolated from marine samples in the area of the Venetian Lagoon, Italy. *FEMS Microbiol Ecol* **70**, 506-514.

**Caburlotto, G., Lleò, M. M., Hilton, T., Huq, A., Colwell, R. R. & Kaper, J. B. (2010).** Effect on human cells of environmental *Vibrio parahaemolyticus* strains carrying type III secretion system 2. *Infect Immun* **78**, 3280–3287.

**Cai, H., Rodríguez, B. T., Zhang, W., Broadbent, J. R. & Steele, J. L. (2007).** Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity. *Microbiology* **153**, 2655-2665.

**Castillo-Ramírez, S., Harris, S. R., Holden, M. T. G., He, M., Parkhill, J., Bentley, S. D. & Feil, E. J. (2011).** The impact of recombination on *dN/dS* within recently emerged bacterial clones. *PLoS Pathog* **7**, e1002129 **Chakrabarti, M. K., Sinha, A. K. & Biswas, T. (1991).** Adherence of *Vibrio parahaemolyticus* to rabbit intestinal epithelial cells *in vitro*. *FEMS Microbiol Lett* **84**, 113-118.

**Chakrabarti, S. R., Chaudhuri, K., Sen, K. & Das, J. (1996).** Porins of *Vibrio cholerae*: purification and characterization of OmpU. *J Bacteriol* **178**, 524-530.

Chao, G., Jiao, X., Zhou, X., Yang, Z., Pan, Z., Huang, J., Zhou, L. & Qian, X. (2009a). Systematic functional pandemic strain-specific genes, three genomic islands, two T3SSs in foodborne, and clinical *Vibrio parahaemolyticus* isolates in China. *Foodborne Pathog Dis* **6**, 689–698.

Chao, G., Jiao, X., Zhou, X., Yang, Z., Huang, J., Zhou, L. & Qian, X. (2009b). Distribution, prevalence, molecular typing, and virulence of Vibrio parahaemolyticus isolated from different sources in coastal province Jiangsu, China. *Food Control* **20**, 907–912.

Chao, G., Jiao, X., Zhou, X., Wang, F., Yang, Z., Huang, J. & Zhiming Pan, 1 Liping Zhou, 2 and X. Q. (2010). Distribution of genes encoding four pathogenicity islands (VPaIs), T6SS, biofilm, and type I pilus in food and clinical strains of *Vibrio parahaemolyticus* in China. *Foodborne Pathog Dis* **7**, 649–658.

Chao, G., Wang, F., Zhou, X., Jiao, X., Huang, J., Pan, Z., Zhou, L. & Qian, X. (2011). Origin of *Vibrio parahaemolyticus* O3:K6 pandemic clone. *Int J Food Microbiol* **145**, 459–463.

Chen, H. Y., Yuan, M. & Livermore, D. M. (1995). Mechanisms of resistance to beta-lactam antibiotics amongst *Pseudomonas aeruginosa* isolates collected in the UK in 1993. *J Med Microbiol* **43**, 300–309.

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

Chen, Y., Stine, O. C., Badger, J. H., Gil, A. I., Nair, G. B., Nishibuchi, M. & Fouts, D. E. (2011). Comparative genomic analysis of *Vibrio parahaemolyticus*: serotype conversion and virulence. *BMC genomics* **12**, 294.

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., & other authors. (2011). The origin of the Haitian cholera outbreak strain. *N Engl J Med* **364**, 33–42.

Chiou, S. H. & Wu, S. H. (1999). Evaluation of commonly used electrophoretic methods for the analysis of proteins and peptides and their application to biotechnology. *Anal Chim Acta* **383**, 47–60.

Chowdhury, N. R., Chakraborty, S. & Ramamurthy, T. (2000). Molecular evidence of clonal *Vibrio parahaemolyticus* pandemic strains. *Emerg Infect Dis* **6**, 631–636.

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

Chun, D., Chung, J. K., Tak, R. & Seol, S. Y. (1975). Nature of Kanagawa phenomenon of *Vibrio parahaemolyticus*. *Infect Immun* **12**, 81–87.

Ch'ng, S. L., Octavia, S., Xia, Q., Duong, A., Tanaka, M. M., Fukushima, H. & Lan, R. (2011). Population structure and evolution of pathogenicity of *Yersinia pseudotuberculosis*. *Appl Environ Microbiol* **77**, 768–775.

Clarke, T. E., Tari, L. W. & Vogel, H. J. (2001). Structural biology of bacterial iron uptake systems. *Curr Top Med Chem* **1**, 7–30.

Codex. (2003). Codex alimentarius commission discussion paper on risk management strategies for *Vibrio* spp. in seafood. Orlando.

Colquhoun, D. J. & Sorum, H. (1998). Outer membrane protein expression during *in vivo* cultivation of *Vibrio salmonicida*. *Fish Shellfish Immun* **8**, 367–377.

Cooper, J. E. & Feil, E. J. (2004). Multilocus sequence typing-what is resolved? *Trends Microbiol* **12**, 373–377.

Cooper, V. S., Vohr, S. H., Wrocklage, S. C. & Hatcher, P. J. (2010). Why genes evolve faster on secondary chromosomes in bacteria. *PLoS Comput Biol* **6**, 1–11.

Corander, J. & Marttinen, P. (2006). Bayesian identification of admixture events using multilocus molecular markers. *Mol Ecol* **15**, 2833–43.

Corander, J. & Tang, J. (2007). Bayesian analysis of population structure based on linked molecular information. *Math Biosci* **205**, 19–31.

Cordwell, S. J. (2006). Technologies for bacterial surface proteomics. *Curr Opin Microbiol* **9**, 320–329.

Cornelis, G. R. (2006). The type III secretion injectisome. *Nat Rev Microbiol* **4**, 811–825.

Costerton, J. W., Ingram, J. M. & Cheng, K. J. (1974). Structure and function of the cell envelope of Gram-negative bacteria. *Bacteriol Rev* **38**, 87–110.

Cowan, S. W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, A., Jansonius, J. N. & Rosenbusch, J. P. (1992). Crystal structures explain functional properties of two *E. coli* porins. *Nature* **358**, 727–733.

Dai, J. H., Lee, Y. S. & Wong, H. C. (1992). Effects of iron limitation on production of a siderophore, outer membrane proteins, and hemolysin and on hydrophobicity, cell adherence, and lethality in mice of *Vibrio parahaemolyticus*. *Infect Immun* **60**, 2952–2956.

Dalmin, G., Kathiresan, K. & Purushothaman, A. (2001). Effect of probiotics on bacterial population and health status of shrimp in culture pond ecosystem. *Indian J Exp Biol* **39**, 939–942.

Daniels, N. A., Mackinnon, L., Bishop, R., Altekruse, S., Ray, B., Hammond, R. M., Thompson, S., Wilson, S., Bean, N. H. & Griffin, P. M. (2000). *Vibrio parahaemolyticus* Infections in the United States, 1973 – 1998. *J Infect Dis* **181**, 1661–1666.

Dasgupta, A., Banerjee, R., Das, S. & Basak, S. (2012). Evolutionary perspective on the origin of Haitian cholera outbreak strain. *J Biomol Struct Dyn* **30**, 338–346.

Davies, R. L. & Quirie, M. (1996). Intra-specific diversity within *Pasteurella trehalosi* based on variation of capsular polysaccharide, lipopolysaccharide and outer-membrane proteins. *Microbiology* **142**, 551–560.

**Davies, R. L., McCluskey, J., Gibbs, H. a, Coote, J. G., Freer, J. H. & Parton, R. (1994).** Comparison of outer-membrane proteins of *Pasteurella haemolytica* expressed *in vitro* and *in vivo* in cattle. *Microbiology* **140**, 3293–3300.

**Davies, R. L., Watson, J. & Caffrey, B. (2003a).** Comparative analyses of *Pasteurella multocida* strains associated with the ovine respiratory and vaginal tracts. *Vet Rec* **152**, 7–10.

**Davies, R. L. (2003).** Characterization and comparison of *Pasteurella multocida* strains associated with porcine pneumonia and atrophic rhinitis. *J Med Microbiol* **52**, 59–67.

**Davies, R. L. & Lee, I. (2004).** Sequence diversity and molecular evolution of the heat-modifiable outer membrane protein gene (*ompA*) of *Mannheimia* (*Pasteurella*) *haemolytica, Mannheimia glucosida*, and *Pasteurella trehalosi*. *J Bacteriol* **186**, 5741–5752.

**Davies, R. L., MacCorquodale, R. & Caffrey, B. (2003b).** Diversity of avian *Pasteurella multocida* strains based on capsular PCR typing and variation of the OmpA and OmpH outer membrane proteins. *Vet Microbiol* **91**, 169–182.

**Davies, R. L., MacCorquodale, R. & Reilly, S. (2004).** Characterisation of bovine strains of *Pasteurella multocida* and comparison with isolates of avian, ovine and porcine origin. *Vet Microbiol* **99**, 145–158.

**Deepanjali, A., Kumar, H. S. & Karunasagar, I. (2005).** Seasonal variation in abundance of total and pathogenic *Vibrio parahaemolyticus* bacteria in oysters along the southwest coast of India. *Appl Environ Microbiol* **71**, 3575–3580.

**Delcour, A. H. (2002).** Structure and function of pore-forming beta-barrels from bacteria. *J Mol Microb Biotech* **4**, 1–10.

**Dennis, W. & Bob, R. (1992).** World shrimp farming review. In *Proceedings of the special session on shrimp farming*, pp. 1–21. Edited by J. Wayban. LA.

**DePaola, A., Kaysner, C. A., Bowers, J., Cook, D. W., Ray, B., Wiles, K., Depaola, A., Cook, D., Kaysner, C. & Puhr, N. (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.

**Didelot, X., Urwin, R., Maiden, M. C. J. & Falush, D. (2009).** Genealogical typing of *Neisseria meningitidis*. *Microbiology* **155**, 3176–3186.

**Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. a. & Crook, D. W. (2012).** Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* **13**, 601–612.

**Dingle, K. E., Colles, F. M., Falush, D. & Maiden, M. C. (2005).** Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*. *J Clin Microbiol* **43**, 340–347.

**Do, T., Jolley, K. A., Maiden, M. C. J., Gilbert, S. C., Clark, D., Wade, W. G. & Beighton, D. (2009).** Population structure of *Streptococcus oralis*. *Microbiology* **155**, 2593–2602.

**Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. (2004).** Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**, 414–424.

**Doerrler, W. T. & Raetz, C. R. H. (2005).** Loss of outer membrane proteins without inhibition of lipid export in an *Escherichia coli* YaeT mutant. *J Biol Chem* **280**, 27679–27687.

**Droop, M. R. (1957).** Vitamin B12 in marine ecology. *Nature* **180**, 1041–1042.

**E-komon, T., Burchmore, R. J., Herzyk, P. & Davies, R. L. (2012).** Predicting the outer membrane proteome of *Pasteurella multocida* based on consensus prediction enhanced by results integration and manual confirmation. *BMC Bioinformatics* **13**, 63.

**Eckert, K. a & Kunkel, T. a. (1991).** DNA polymerase fidelity and the polymerase chain reaction. *Genome Res* **1**, 17–24.

**Economou, A. (1999).** Following the leader: bacterial protein export through the Sec pathway. *Trends Microbiol* **7**, 315–320.

**Ellis, C. N., Schuster, B. M., Striplin, M. J., Jones, S. H., Whistler, C. a & Cooper, V. S. (2012).** Influence of seasonality on the genetic diversity of *Vibrio parahaemolyticus* in New Hampshire shellfish waters as determined by multilocus sequence analysis. *Appl Environ Microbiol* **78**, 3778–3782.

**Enos-berlage, J. L., Guvener, Z. T., Keenan, C. E. & Mccarter, L. L. (2005).** Genetic determinants of biofilm development of opaque and translucent *Vibrio parahaemolyticus*. *Mol Microbiol* **55**, 1160–1182.

**Enright, M. C., Day, N. P. J., Davies, C. E. & Peacock, S. J. (2000).** Multilocus sequence typing for characterization of methicillin- resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol* **38**, 1008–1015.

**Eppens, E. F., Nouwen, N. & Tommassen, J. (1997).** Folding of a bacterial outer membrane protein during passage through the periplasm. *EMBO J* **16**, 4295–4301.

**European commission. (2007).** *EU import conditions for seafood and other fishery products*. Available online http://ec.europa.eu/food/international/trade/im_cond_fish_en.pdf.

**FAO. (2006).** State of world aquaculture. *Food and Agriculture Organization fisheries technical paper* 128pp. Rome.

**FAO. (2010).** Shrimp. *Food and Agriculture Organization globefish commodity update* 72pp. Rome.

**Farmer, J. J. & Janda, M. (2004).** Family I. *Vibrionaceae*. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn., pp. 491–546. Edited by M. Garrity, G. New York: Springer.

**FDA. (2005).** Quantitative risk assessment on the public health impact of pathogenic *Vibrio parahaemolyticus* in raw oysters. *Food and Drug Administration, US Department of Health and Human Services* 142pp.

**Feil, E. J. (2004).** Small change: keeping pace with microevolution. *Nat Rev Microbiol* **2**, 483–495.

**Feil, E. J. & Enright, M. C. (2004).** Analyses of clonality and the evolution of bacterial pathogens. *Curr Opin Microbiol* **7**, 308–313.

**Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P. & 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.

**Felek, S. & Krukonis, E. S. (2009).** The Yersinia pestis Ail protein mediates binding and Yop delivery to host cells required for plague virulence. *Infect Immun* **77**, 825–836.

**Ferguson, a D., Braun, V., Fiedler, H. P., Coulton, J. W., Diederichs, K. & Welte, W. (2000).** Crystal structure of the antibiotic albomycin in complex with the outer membrane transporter FhuA. *Protein Sci* **9**, 956–963.

**Ferguson, a. D. (1998).** Siderophore-Mediated Iron Transport: Crystal Structure of FhuA with Bound Lipopolysaccharide. *Science* **282**, 2215–2220.

**Francisco, A. P., Bugalho, M., Ramirez, M. & Carriço, J. a. (2009).** Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* **10**, 152.

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

**Fujino, T., Sakaguchi, G., Sakazaki, R. & Takeda, Y. (2002).** Discovery of *Vibrio parahaemolyticus*. *J Jpn Soc Intern Med* **91**, 2903–2906.

**Funahashi, T., Moriya, K., Uemura, S., Shinoda, S., Narimatsu, S. & Miyoshi, S. (2002).** Identification and characterization of *pvuA* , a gene encoding the ferric vibrioferrin receptor protein in *Vibrio parahaemolyticus*. *J Bacteriol* **184**, 936–946.

**Funahashi, T., Tanabe, T., Aso, H., Nakao, H., Fujii, Y., Okamoto, K., Shizuo, N. & Yamamoto, S. (2003).** An iron-regulated gene required for utilization of aerobactin as an exogenous siderophore in *Vibrio parahaemolyticus*. *Microbiology* **149**, 1217–1225.

**Funahashi, T., Tanabe, T., Shiuchi, K., Nakao, H. & Yamamoto, S. (2009).** Identification and characterization of genes required for utilization of desferri-ferrichrome and aerobactin in *Vibrio parahaemolyticus*. *Biol Pharm Bull* **32**, 359–365.

**Gardy, J. L., Laird, M. R., Chen, F., Rey, S., Walsh, C. J., Ester, M. & Brinkman, F. S. L. (2005).** PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* **21**, 617–623.

**Garfin, D. E. (2003).** Gel Electrophoresis of proteins. In *Essential Cell Biology*, pp. 197–268. Oxford University Press, Oxford UK.

**Garrido, A., Chapela, M.-J., Ferreira, M., Atanassova, M., Fajardo, P., Lago, J., Vieites, J. M. & Cabado, A. G. (2012).** Development of a multiplex real-time PCR method for pathogenic *Vibrio parahaemolyticus* detection (*tdh+* and *trh+*). *Food Control* **24**, 128–135.

**Garrow, A. G., Agnew, A. & Westhead, D. R. (2005).** TMB-Hunt: a web server to screen sequence sets for transmembrane beta-barrel proteins. *Nucleic Acids Res* **33**, 188–192.

**Gingras, S. P. & Howard, L. V. (1980).** Adherence of *Vibrio parahaemolyticus* to human epithelial cell lines. *Appl Environ Microbiol* **39**, 369-371.

**Gonzalez-Escalona, N., Strain, E. A., De Jesús, A. J., Jones, J. L. & Depaola, A. (2011).** Genome sequence of the clinical O4:K12 serotype *Vibrio parahaemolyticus* strain 10329. *J Bacteriol* **193**, 3405–3406.

**González-Escalona, N., Cachicas, V., Acevedo, C., Rioseco, M. L., Vergara, J. A., Cabello, F., Romero, J. & Espejo, R. T. (2005).** *Vibrio parahaemolyticus* diarrhea, Chile, 1998 and 2004. *Emerg Infect Dis* **11**, 2004–2006.

**González-Escalona, N., Blackstone, G. M. & DePaola, A. (2006).** Characterization of a *Vibrio alginolyticus* strain, isolated from Alaskan oysters, carrying a hemolysin gene similar to the thermostable direct hemolysin-related hemolysin gene (*trh*) of *Vibrio parahaemolyticus*. *Appl Environ Microbiol* **72**, 7925-7929.

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

**Gotoh, K., Kodama, T., Hiyoshi, H., Izutsu, K., Park, K.-S., Dryselius, R., Akeda, Y., Honda, T. & Iida, T. (2010).** Bile acid-induced virulence gene expression of *Vibrio parahaemolyticus* reveals a novel therapeutic potential for bile acid sequestrants. *PLoS One* **5**, e13365.

**Gromiha, M. M. (2005).** Motifs in outer membrane protein sequences: applications for discrimination. *Biophys Chem* **117**, 65–71.

**Gromiha, M. M. & Suwa, M. (2006).** Influence of amino acid properties for discriminating outer membrane proteins at better accuracy. *Biochim Biophys Acta* **1764**, 1493-1497.

**Gulig, P. a, de Crécy-Lagard, V., Wright, A. C., Walts, B., Telonis-Scott, M. & 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.

**Gundry, C. N. & Poulson, M. D. (2011).** Obtaining maximum PCR sensitivity and specificity. In *PCR troubles shooting and optimization,* pp. 79-96. Edited by S. Kennedy & O. Nick. Caister Academic Press, Norfolk, UK.

**Gál, J., Szvetnik, A. & Schnell, R. (2002).** The metD d-methionine transporter locus of *Escherichia coli* is an ABC transporter gene cluster. *J Bacteriol* **184**, 4930–4932.

**de Haan, C. P., Kivistö, R. I., Hakkinen, M., Corander, J. & Hänninen, M. L. (2010).** Multilocus sequence types of Finnish bovine *Campylobacter jejuni* isolates and their attribution to human infections. *BMC Microbiology* **10**, 200.

**Hackney, C. R., Kleeman, E. G., Ray, B. & Speck, M. L. (1980).** Adherence as a method for differentiating virulent and avirulent strains of *Vibrio parahaemolyticus*. *Appl Environ Microbiol* **40**, 652–658.

**Haddix, P. L. & Straley, S. C. (1992).** Structure and regulation of the *Yersinia pestis* yscBCDEF operon. *J Bacteriol* **174**, 4820–4828.

**Hagan, C. L., Silhavy, T. J. & Kahne, D. (2011).** β-Barrel membrane protein assembly by the Bam complex. *Annu Rev Biochem* **80**, 189–210.

**Haley, B. J., Grim, C. J., Hasan, N. a, Choi, S. Y., Chun, J., Brettin, T. S., Bruce, D. C., Challacombe, J. F., Detter, J. C., & other authors. (2010).** Comparative genomic analysis reveals evidence of two novel *Vibrio* species closely related to *V. cholerae*. *BMC Microbiol* **10**, 154.

**Han, H., Wong, H. C., Kan, B., Guo, Z., Zeng, X., Yin, S., Liu, X., Yang, R. & Zhou, D. (2008).** Genome plasticity of *Vibrio parahaemolyticus*: microevolution of the pandemic group. *BMC Genomics* **9**, 570.

**Hara-kudo, Y., Sugiyama, K., Nishibuchi, M., Chowdhury, A., Yatsuyanagi, J., Ohtomo, Y., Saito, A., Nagano, H., Nishina, T., & other authors. (2003).** Prevalence of pandemic thermostable direct hemolysin-producing *Vibrio parahaemolyticus* O3:K6 in seafood and the coastal environment in Japan. *Appl Environ Microbiol* **69**, 3883–3891.

**Hara-Kudo, Y., Saito, S., Ohtsuka, K., Yamasaki, S., Yahiro, S., Nishio, T., Iwade, Y., Otomo, Y., Konuma, H., & other authors. (2012).** Characteristics of a sharp decrease in *Vibrio parahaemolyticus* infections and seafood contamination in Japan. *Int J Food Microbiol* **157**, 95–101.

**Harth, E., Matsuda, L., Hernández, C., Rioseco, M. L., Romero, J., González-escalona, N., Martínez-urtaza, J. & Espejo, R. T. (2009).** Epidemiology of *Vibrio parahaemolyticus* outbreak, Southern Chile. *Emerg Infect Dis* **15**, 163–168.

**Haubold, B. & Hudson, R. R. (2000).** LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Bioinformatics* **16**, 847–848.

**Haug, M. C., Tanner, S. a, Lacroix, C., Stevens, M. J. a & Meile, L. (2011).** Monitoring horizontal antibiotic resistance gene transfer in a colonic fermentation model. *FEMS Microbiol Ecol* **78**, 210–219.

**Hayashi, S. & Wu, H. C. (1990).** Lipoproteins in bacteria. *J Bioenerg Biomembr* **22**, 451–471.

**Hendriksen, R. S., Price, L. B. & Schupp, J. M. (2011).** Origin of the Haitian Outbreak. *mBio* **2**.

**Highton, P. J., Chang, Y., Marcotte, W. R. & Schnaitman, C. A. (1985).** Evidence that the outer membrane protein gene nmpC of Escherichia coli K-12 lies within the defective *qsr'* prophage. *J Bacteriol* **162**, 256-262.

**Hindahl, M. S., Crockford, G. W. K. & Fiancockl, R. E. W. (1984).** Outer membrane protein NmpC of *Escherichia coli*: pore-forming properties in black lipid. *J Bacteriol* **159**, 3-6.

**Hinnebusch, B. J., Jarrett, C. O., Callison, J. a, Gardner, D., Buchanan, S. K. & Plano, G. V. (2011).** Role of the *Yersinia pestis* Ail protein in preventing a protective polymorphonuclear leukocyte response during bubonic plague. *Infect Immun* **79**, 4984–4989.

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

**Honda, S., Goto, I., Minematsu, N., Ikeda, N., Asano, N., Ishibashi, M., Kinoshita, Y., Nishibushi, M., Honda, T. & Miwatani, T. (1987).** Gastroenteritis due to Kanagawa negative *Vibrio parahaemolyticus*. *Lencet* **329**, 331–332.

**Honda, T., Takeda, Y., Miwatani, T., Kato, K. & Nimura, Y. (1976).** Clinical features of patients suffering from food poisoning due to *Vibrio parahaemolyticus* with special reference to changes in electrocardiograms. *Jpn J Assoc Infect Dis* **50**, 216–223.

**Honda, T., Ni, Y. X. & 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. *Infect Immun* **56**, 961–965.

**Honda, T., Abad-lapuebla, M. A., Ni, Y., Yamamoto, K. & 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.

**Hsieh, Y., Liang, S., Tsai, W., Chen, Y., Liu, T. & Liang, C. (2003).** Study of capsular polysaccharide from *Vibrio parahaemolyticus*. *Infect Immun* **71**, 3329–3336.

**Hueck, C. J. (1998).** Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol R* **62**, 379–433.

**Hunt, D. E., Gevers, D., Vahora, N. M. & Polz, M. F. (2008).** Conservation of the chitin utilization pathway in the *Vibrionaceae*. *Appl Environ Microbiol* **74**, 44-51.

**Hurley, C. C., Quirke, A., Reen, F. J. & Boyd, E. F. (2006).** Four genomic islands that mark post-1995 pandemic *Vibrio parahaemolyticus* isolates. *BMC genomics* **7**, 104.

**Iida, T. & 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.

**Iijima, Y., Yamada, H. & Shinoda, S. (1981).** Adherence of *Vibrio parahaemolyticus* and its relation to pathogenicity. *Can J Microbiol* **27**, 1252–1259.

**Imai, K., Asakawa, N., Tsuji, T., Akazawa, F., Ino, A. & Sonoyama, M. (2008).** SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in Gram-negative bacteria. *Bioinformatics* **2**, 417–421.

**Inoue, T., Matsuzaki, S. & Tanaka, S. (1995a).** A 26-kDa outer membrane protein, OmpK, common to *Vibrio* species is the receptor for a broad-host-range vibriophage, KVP40. *FEMS Microbiol Lett* **125**, 101–105.

**Inoue, T., Matsuzaki, S. & Tanaka, S. (1995b).** Cloning and sequence analysis of *Vibrio parahaemolyticus* ompK gene encoding a 26-kDa outer membrane protein, OmpK, that serves as receptor for a broad-host-range vibriophage, KVP40. *FEMS Microbiol Lett* **34**, 245–249.

**Izutsu, K., Kurokawa, K., Tashiro, K., Kuhara, S., Hayashi, T., Honda, T. & Iida, T. (2008).** Comparative genomic analysis using microarray demonstrates a strong correlation between the presence of the 80-kilobase pathogenicity island and pathogenicity in kanagawa phenomenon-positive *Vibrio parahaemolyticus* strains. *Infect Immun* **76**, 1016–1023.

**Jackups, R., Cheng, S. & Jie, L. (2006).** Sequence motifs and antimotifs in ß-barrel membrane proteins from a genome-wide analysis: the Ala-Tyr dichotomy and chaperone binding motifs. *J Mol Biol* **363**, 611–623.

**Jacobson, M. J., Lin, G., Whittam, T. S. & Johnson, E. A. (2008).** Phylogenetic analysis of *Clostridium botulinum* type A by multi-locus sequence typing. *Microbiology* **154**, 2408–2415.

**Jain, S. & Goldberg, M. B. (2007).** Requirement for YaeT in the outer membrane assembly of autotransporter proteins. *J Bacteriol* **189**, 5393–5398.

**Jalajakumari, M. B. & Manning, P. A. (1990).** Nucleotide sequence of the gene, *ompW*, encoding a 22kDa immunogenic outer membrane protein of *Vibrio cholerae*. *Nucleic Acids Res* **18**, 2180.

**Janda, J. M., Powers, C., Bryant, R. G. & Abbott, S. L. (1988).** Current perspectives on the epidemiology and pathogenesis of clinically significant *Vibrio* spp. *Clin Microbiol Rev* **1**, 245–267.

**Jeannin, P., Magistrelli, G., Goetsch, L., Haeuw, J.F., Thieblemont, N., Bonnefoy, J. Y. & Delneste, Y. (2002).** Outer membrane protein A (OmpA): a new pathogen-associated molecular pattern that interacts with antigen presenting cells-impact on vaccine strategies. *Vaccine* **20**, A23–A27.

**Jayasree, L., Janakiram, P and Madhavi, R. (2006).** Characterization of *Vibrio spp.* associated with diseased shrimp from culture ponds of Andhra Pradesh (India). *J World Aquacult Soc* **37**, 523.

**Jolley, K. A., Kalmusova, J., Feil, E. J., Gupta, S., Musilek, M., Kriz, P. & Maiden, M. C. J. (2000).** Carried Meningococci in the Czech Republic:a diverse recombining population. *J Clin Microbiol* **38**, 4492–4498.

Jones, J. L., Hara-Kudo, Y., Krantz, J. a, Benner, R. a, Smith, A. B., Dambaugh, T. R., Bowers, J. C. & Depaola, A. (2012a). Comparison of molecular detection methods for *Vibrio parahaemolyticus* and *Vibrio vulnificus*. *Food Microbiol* **30**, 105–111.

Jones, J. L., Lüdeke, C. H. M., Bowers, J. C., Garrett, N., Fischer, M., Parsons, M. B., Bopp, C. a & Depaola, A. (2012b). Biochemical, serological, and virulence characterization of clinical and oyster *Vibrio parahaemolyticus* isolates. *J Clin Microbiol* **50**, 2343–2352.

Joseph, S. W., Colwell, R. R. & Kaper, J. B. (1982). *Vibrio parahaemolyticus* and related halophilic *Vibrios*. *Crit Rev Microbiol* **10**, 77–124.

Josupeit, H. (2008). Towards sustainably in shrimp production, processing and trade. *Report of technical and trade conference on shrimp, Guangzhou, china* Available at http://www.fao.org/DOCREP/006/AC446E/.

Juncker, A. S., Willenbrock, H., Heijne, G. V. O. N., Brunak, S., Nielsen, H. & Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* **12**, 1652–1662.

Kamruzzaman, M., Bhoopong, P., Vuddhakul, V. & Nishibuchi, M. (2008). Detection of a functional insertion sequence responsible for deletion of the thermostable direct hemolysin gene (*tdh*) in *Vibrio parahaemolyticus*. *Gene* **421**, 67–73.

Keyhani, N. O. & Roseman, S. (1999). Physiological aspects of chitin catabolism in marine bacteria. *Biochim Biophys Acta* **1473**, 108–122.

Keymer, D. P. & Boehm, A. B. (2011). Recombination shapes the structure of an environmental *Vibrio cholerae* population. *Appl Environ Microbiol* **77**, 537–544.

De Keyzer, J., Van Der Does, C. & Driessen, A. J. M. (2003). The bacterial translocase: a dynamic protein channel complex. *Cell Mol Life Sci* **60**, 2034–2052.

Killmann, H., Videnov, G., Jung, G., Schwarz, H., Braun, V. & Jung, N. (1995). Identification of receptor binding sites by competitive peptide mapping: phages T1, T5, and phi 80 and colicin M bind to the gating loop of FhuA. *J Bacteriol* **177**, 694-698.

Kim, Y., Mccarter, L. L. & Carter, L. L. M. C. (2000). Analysis of the polar flagellar gene system of *Vibrio parahaemolyticus*. *J Bacteriol* **182**, 3693–3704.

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

Klebba, P. E. & Newton, S. M. (1998). Mechanisms of solute transport through outer membrane porins: burning down the house. *Curr Opin Microbiol* **1**, 238-248.

**Kodama, T., Rokuda, M., Park, K.-S., Cantarelli, V. V, Matsuda, S., Iida, T. & Honda, T. (2007).** Identification and characterization of VopT, a novel ADP-ribosyltransferase effector protein secreted via the *Vibrio parahaemolyticus* type III secretion system 2. *Cell Microbiol* **9**, 2598–2609.

**Kodama, T., Gotoh, K., Hiyoshi, H., Morita, M., Izutsu, K., Akeda, Y., Park, K.-S., Cantarelli, V. V, Dryselius, R., & other authors. (2010).** Two regulators of *Vibrio parahaemolyticus* play important roles in enterotoxicity by controlling the expression of genes in the Vp-PAI region. *PloS One* **5**, e8678.

**Koebnik, R. (1999).** Structural and functional roles of the surface-exposed loops of the beta-barrel membrane protein OmpA from *Escherichia coli*. *J Bacteriol* **181**, 3688–3694.

**Koebnik, R., Locher, K. P. & Van Gelder, P. (2000).** Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol Microbiol* **37**, 239-253.

**Koga, B. & Kawata, T. O. M. (1983).** Isolation and characterization of the outer membrane from *Vibrio parahaemolyticus*. *J Gen Microbiol* **129**, 3185-3196.

**Koga, T. & Takumi, K. (1995).** Siderophore production and outer membrane protein synthesis of *Vibrio parahaemolyticus* strain with different serotype under iron limited condition. *J Gen Appl Microbiol* **41**, 221-228.

**Kolodziejek, A. M., Sinclair, D. J., Seo, K. S., Schnider, D. R., Deobald, C. F., Rohde, H. N., Viall, A. K., Minnich, S. S., Hovde, C. J., & other authors. (2007).** Phenotypic characterization of OmpX, an Ail homologue of *Yersinia pestis* KIM. *Microbiology* **153**, 2941–2951.

**Kolodziejek, A. M., Schnider, D. R., Rohde, H. N., Wojtowicz, A. J., Bohach, G. a, Minnich, S. a & Hovde, C. J. (2010).** Outer membrane protein X (Ail) contributes to *Yersinia pestis* virulence in pneumonic plague and its activity is dependent on the lipopolysaccharide core length. *Infect Immun* **78**, 5233-5243.

**Kolodziejek, A. M., Hovde, C. J. & Minnich, S. a. (2012).** *Yersinia pestis* Ail: multiple roles of a single protein. *Front Cell Infect Microbiol* **2**, 1–10.

**Koronakis, V., Eswaran, J. & Hughes, C. (2004).** Structure and function of TolC: the bacterial exit duct for proteins and drugs. *Annu Rev Biochem* **73**, 467-89.

**Korteland, J., Tommassen, J. & Lugtenberg, B. (1982).** PhoE protein pore of the outer membrane of *Escherichia coli* K12 is a particularly efficient channel for organic and inorganic phosphate. *Biochim Biophys Acta* **690**, 282-289.

**Kuehl, C. J. & Crosa, J. H. (2011).** The TonB energy transduction systems in *Vibrio* species. *Future Microbiol* **5**, 1403–1412.

**Kumazawa, N. H., Hori, K., Fujimori, K., Iwade, Y. & Sugiyama, A. (1999).** Geographical features of estuaries for neritid gastropods including clithon retropictus to preserve thermostable direct hemolysin-producing *Vibrio parahaemolyticus*. *J Vet Med Sci* **61**, 721-724.

Kuntumalla, S., Zhang, Q., Braisted, J. C., Fleischmann, R. D., Peterson, S. N., Donohue-Rolfe, A., Tzipori, S. & Pieper, R. (2011). *In vivo* versus *in vitro* protein abundance analysis of *Shigella dysenteriae* type 1 reveals changes in the expression of proteins involved in virulence, stress and energy metabolism. *BMC microbiol* **11**, 147.

Kustusch, R. J., Kuehl, C. J. & Crosa, J. H. (2011). Power plays: iron transport and energy transduction in pathogenic vibrios. *Biometals* **24**, 559–566.

Köhler, T., Michea-Hamzehpour, M., Epp, S. F. & Pechere, J. C. (1999). Carbapenem activities against *Pseudomonas aeruginosa*: respective contributions of OprD and efflux systems. *Antimicrob Agents Ch* **43**, 424–427.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

Lang, P., Lefebure, T., Wang, W., Pavinski Bitar, P., Meinersmann, R. J., Kaya, K. & Stanhope, M. J. (2010). Expanded multilocus sequence typing and comparative genomic hybridization of *Campylobacter coli* isolates from multiple hosts. *Appl Environ Microbiol* **76**, 1913–1925.

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

Larsen, M. V, Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., Jelsbak, L., Sicheritz-Pontén, T., Ussery, D. W., & other authors. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* **50**, 1355–1361.

Lawrence, D. N., Blake, P. a, Yashuk, J. C., Wells, J. G., Creech, W. B. & Hughes, J. H. (1979). *Vibrio parahaemolyticus* gastroenteritis outbreaks aboard two cruise ships. *Am J Epidemiol* **109**, 71–80.

Lee, J. S., Jung, I. D., Lee, C. M., Park, J. W., Chun, S. H., Jeong, S. K., Ha, T. K., Shin, Y. K., Kim, D. J. & Park, Y. M. (2010). Outer membrane protein a of *Salmonella enterica* serovar Typhimurium activates dendritic cells and enhances Th1 polarization. *BMC microbiol* **10**, 263.

Lemee, L., Bourgeois, I., Ruffin, E., Collignon, A., Lemeland, J.-F. & Jean-Louis, P. (2005). Multilocus sequence analysis and comparative evolution of virulence-associated genes and housekeeping genes of *Clostridium difficile*. *Microbiology* **151**, 3171–3180.

Li, C. C., Crawford, J. a, DiRita, V. J. & Kaper, J. B. (2000). Molecular cloning and transcriptional regulation of *ompT*, a ToxR-repressed gene in *Vibrio cholerae*. *Mol Microbiol* **35**, 189–203.

Li, H., Ye, M., Peng, B., Wu, H., Xu, C., Xiong, X., Wang, C., Wang, S. & Peng, X. (2010a). Immunoproteomic identification of polyvalent vaccine candidates from *Vibrio parahaemolyticus* outer membrane proteins. *J Proteome Res* **9**, 2573–2583.

**Li, N., Yang, Z., Bai, J., Fu, X., Liu, L., Shi, C. & Wu, S. (2010b).** A shared antigen among *Vibrio species*: outer membrane protein-OmpK as a versatile Vibriosis vaccine candidate in Orange-spotted grouper (*Epinephelus coioides*). *Fish Shellfish Immun* **28**, 952–956.

**Li, Y. D., Ren, H. L., Lu, S. Y., Zhou, Y., Han, X., Gong, B. Bin, Zhang, Y. Y. & Liu, Z. S. (2010c).** Cloning, expression, and genus-specificity analysis of 28-kDa OmpK from *Vibrio alginolyticus*. *J Food Sci* **75**, M198–M203.

**Lightner, D. V. & Redman, R. M. (2010).** The global status of significant infectious diseases of farmed shrimp. *Asian Fish Sci* **23**, 383–426.

**Lin, J., Huang, S. & Zhang, Q. (2002).** Outer membrane proteins: key players for bacterial adaptation in host niches. *Microbes Infect* **4**, 325–331.

**Lin, Z., Kumagai, K., Baba, K., Mekalanos, J. J. & Nishibuchi, M. (1993).** *Vibrio parahaemolyticus* has a homolog of the *Vibrio cholerae toxRS* operon that mediates environmentally induced regulation of the thermostable direct hemolysin gene. *J Bacteriol* **175**, 3844–3855.

**Lozano-Leon, A., Torres, J., Osorio, C. R. & Martinez-Urtaza, J. (2003).** Identification of *tdh*-positive *Vibrio parahaemolyticus* from an outbreak associated with raw oyster consumption in Spain. *FEMS Microbiol Lett* **226**, 281–284.

**Lång, H. & Ferenci, T. (1995).** Sequence alignment and structural modelling of the LamB glycoporin family. *Biochem Bioph Res Co* **208**, 927–934.

**MacLean, D., Jones, J. D. G. & Studholme, D. J. (2009).** Application of "next-generation" sequencing technologies to microbial genetics. *Nat Rev Microbiol* **7**, 287–296.

**Mahmud, Z. H., Kassu, A., Mohammad, A., Yamato, M., Bhuiyan, N. A., Balakrish Nair, G. & Ota, F. (2006).** Isolation and molecular characterization of toxigenic *Vibrio parahaemolyticus* from the Kii Channel, Japan. *Microbiol Res* **161**, 25–37.

**Mahoney, J. C., Gerding, M. J., Jones, S. H. & Whistler, C. A. (2010).** Comparison of the pathogenic potentials of environmental and clinical *Vibrio parahaemolyticus* strains indicates a role for temperature regulation in virulence. *Appl Environ Microbiol* **76**, 7459–7465.

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

**Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., & other authors. (1998).** Multilocus sequence typing- A portable approach to the identification of clones within populations of clones within populations of pathogenic microorganisms. *Microbiology* **95**, 3140–3145.

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

**Mao, Z., Yu, L., You, Z., Wei, Y. & Liu, Y. (2007a).** Cloning, expression and immunogenicty analysis of five outer membrane proteins of *Vibrio parahaemolyticus* zj2003. *Fish Shellfish Immun* **23**, 567–575.

**Mao, Z., Yu, L., You, Z., Wei, Y. & Liu, Y. (2007b).** Expression and immunogenicity analysis of two iron-regulated outer membrane proteins of *Vibrio parahaemolyticus. Biochim Biophys Acta* **39**, 763–769.

**Mardis, E. R. (2008).** Next-generation DNA sequencing methods. *Annu Rev Genom Hum G* **9**, 387–402.

**Markwell, M. A. N. N. K., Haas, S. M. & Tolbert, N. E. (1978).** Determination of the Lowry procedure to simplify protein in membrane and lipoprotein. *Anal Biochem* **210**, 206–210.

**Marlina, Radu, S., Kqueen, C. Y., Napis, S. & Zakaria, Z. (2007).** Detection of *tdh* and *trh* genes in *Vibrio parahaemolyticus* isolated from Corbicula Moltkiana prime in West Sumatera, Indonesia. *Se Asian J Trop Med* **38**, 349–355.

**Marlovits, T. C. & Stebbins, C. E. (2010).** Type III secretion systems shape up as they ship out. *Curr Opin Microbiol* **13**, 47-52.

**Martin, D. P., Williamson, C. & Posada, D. (2005).** RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260–262.

**Matlawska-Wasowska, K., Finn, R., Mustel, A., O'Byrne, C. P., Baird, A. W., Coffey, E. T. & Boyd, A. (2010).** The *Vibrio parahaemolyticus* type III secretion systems manipulate host cell MAPK for critical steps in pathogenesis. *BMC Microbiol* **10**, 329.

**Matsumoto, C., Okuda, J., Ishibashi, M., Iwanaga, M., Garg, P., Rammamurthy, T., Wong, H. C., Depaola, A., Kim, Y. B., & other authors. (2000).** Pandemic spread of an O3: K6 clone of *Vibrio parahaemolyticus* and emergence of related strains evidenced by arbitrarily primed PCR and *toxRS* sequence analyses. *J Clin Microbiol* **38**, 578-585.

**Matsuzaki, S., Tanaka, S., Koga, T. & Kawata, T. (1992).** A broad-host-range vibriophage, KVP40, isolated from sea water. *Microbiol Immunol* **36**, 93–97.

**McCarter, L. L. (1995).** Genetic and molecular characterization of the polar flagellum of *Vibrio parahaemolyticus. J Bacteriol* **177**, 1595–1609.

**McCarter, L. L. (1999).** The multiple identities of *Vibrio parahaemolyticus. J Mol Microbiol Biotechnol* **1**, 1–7.

**McCarter, L. L. (2001).** Polar flagellar motility of the *Vibrionaceae. Microbiol Mol Biol R* **65**, 445–462.

**McGee, K., Hörstedt, P. & Milton, D. L. (1996).** Identification and characterization of additional flagellin genes from *Vibrio anguillarum. J Bacteriol* **178**, 5188–5198.

**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. & Middaugh, J.**

P. (2005). Outbreak of *Vibrio parahaemolyticus* gastroenteritis associated with Alaskan oysters. *New Engl J Med* **353**, 1463–1470.

Mecsas, J., Welch, R., Erickson, J. W. & Gross, C. a. (1995). Identification and characterization of an outer membrane protein, OmpX, in *Escherichia coli* that is homologous to a family of outer membrane proteins including Ail of *Yersinia enterocolitica*. *J Bacteriol* **177**, 799–804.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31–46.

Miller, E. S., Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Durkin, A. S., Ciecko, A., Feldblyum, T. V, White, O., Paulsen, I. T., & other authors. (2003). Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J Bacteriol* **185**, 5220–5233.

Miller, W. G., Englen, M. D., Kathariou, S., Wesley, I. V, Wang, G., Pittenger-alley, L., Siletz, R. M., Muraoka, W., Fedorka-cray, P. J. & Mandrell, R. E. (2006). Identification of host-associated alleles by multilocus sequence typing of *Campylobacter coli* strains from food animals. *Microbiology* **152**, 245-255.

Millikan, D. S. & Ruby, E. G. (2004). *Vibrio fischeri* flagellin A is essential for normal motility and for symbiotic competence during initial squid light organ colonization. *J Bacteriol* **186**, 4315–4325.

Milton, D. L., O'Toole, R., Horstedt, P. & Wolf-Watz, H. (1996). Flagellin A is essential for the virulence of *Vibrio anguillarum*. *J Bacteriol* **178**, 1310–1319.

Miragaia, M., Thomas, J. C., Couto, I., Enright, M. C. & de Lencastre, H. (2007). Inferring a population structure for *Staphylococcus epidermidis* from multilocus sequence typing data. *J Bacteriol* **189**, 2540-2552.

Miyamoto, K., Kosakai, K., Ikebayashi, S., Tsuchiya, T., Yamamoto, S. & Tsujibo, H. (2009). Proteomic analysis of *Vibrio vulnificus* M2799 grown under iron-repleted and iron-depleted conditions. *Microb Pathog* **46**, 171–177.

Miyamoto, Y., Kato, T., Obara, Y., Akiyama, S., Takizawa, K. & Yamai, S. (1969). *In vitro* hemolytic characteristic of *Vibrio parahaemolyticus*: Its close correlation with human pathogenicity. *J Bacteriol* **100**, 1147–1149.

Moeck, G. S. & Coulton, J. W. (1998). TonB-dependent iron acquisition: mechanisms of siderophore-mediated active transport. *Mol Microbiol* **28**, 675–681.

Molenda, J. R., Johnson, W. G., Fishbein, M., Wentz, B., Mehlman, I. J., Jr, D. & T.a. (1972). *Vibrio parahaemolyticus* gastroenteritis in Maryland: laboratory aspects. *Appl Microbiol* **24**, 444–448.

Morakot, T., Timothy W., F., Meerod, W., Grudloyma, U. & Pisamai, N. (2008). Aquacultural biotechnology in Thailand: the case of the shrimp industry. *Int J Biotechnology* **10**, 588–603.

**Morona, R., Kramer, C. & Henning, U. L. F. (1985).** Bacteriophage receptor area of outer membrane protein OmpA of *Escherichia coli* K-12. *J Bacteriol* **164**, 539–543.

**Naenna, P., Noisumdaeng, P. & Pongpech, P. (2010).** Detection of outer membrane porin protein, an imipenem influx channel, in *Pseudomonas Aeruginosa* clinical isolates. *Se Asian J Trop Med* **41**, 614-624.

**Nagayama, K., Yamamoto, K., Mitawani, T. & Honda, T. (1995).** Characterisation of a haemolysin related to Vp-TDH produced by a Kanagawa phenomenon-negative clinical isolate of *Vibrio parahaemolyticus*. *J Med Microbiol* **42**, 83–90.

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

**Nakaguchi, Y., Okuda, J., Iida, T. & Nishibuchi, M. (2003).** The urease gene cluster of *Vibrio parahaemolyticus* does not influence the expression of the thermostable direct hemolysin (TDH) gene or the TDH-related hemolysin gene. *Microbiol Immunol* **47**, 233–239.

**Nakasone, N. & Iwanaga, M. (1990).** Pili of a *Vibrio parahaemolyticus* strain as a possible colonization factor. *Infect Immun* **58**, 61–69.

**Nakasone, N., Insisengmay, S. & Iwanaga, M. (2000).** Characterization of the pili isolated from *Vibrio parahaemolyticus* O3:K6. *Se Asian J Trop Med* **31**, 5–10.

**Nandi, B., Nandy, R. K., Nair, G. B., Shimada, T. & Ghose, A. C. (2000).** Rapid method for species-specific identification of *Vibrio cholerae* using primers targeted to the gene of outer membrane protein OmpW. *J Clin Microbiol* **38**, 4145-4151.

**Nash, G. Nithimathachoke, C., Tungmandi, C., Arkarjamorn, A., Prathanpipat, P. and Ruamthaveesub, P. (1992).** Vibriosis and its control in pond-reared *Penaeus monodon* in Thailand. *In: M. Shariff, R.P. Subasinghe and J.R. Authur (eds.) Diseases in Asian Aquaculture* **1**. Fish Health Section, Asian Fisheries Society, Manila, Philippines, pp143-155.

**Nasu, H., Iida, T., Sugahara, T., Yamaichi, Y., Park, K. S., Yokoyama, K., Makino, K., Shinagawa, H. & Honda, T. (2000).** A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains. *J Clin Microbiol* **38**, 2156–2161.

**Neilands, J. B. (1995).** Siderophores: structure and function of microbial iron transport compounds. *J Biol Chem* **270**, 26723-26726.

**Ningqiu, L., Junjie, B., Shuqin, W., Xiaozhe, F., Haihua, L., Xing, Y. & Cunbin, S. (2008).** An outer membrane protein, OmpK, is an effective vaccine candidate for *Vibrio harveyi* in Orange-spotted grouper (*Epinephelus coioides*). *Fish Shellfish Immun* **25**, 829-833.

**Nishibuchi, M. & Kaper, J. B. (1990).** Duplication and variation of the thermostable direct haemolysin (*tdh*) gene in *Vibrio parahaemolyticus*. *Mol Microbiol* **4**, 87-99.

**Nishibuchi, M. & Kaper, J. B. (1995).** Thermostable direct hemolysin gene of *Vibrio parahaemolyticus*: a virulence gene acquired by a marine bacterium. *Infect Immun* **63**, 2093–2099.

**Nishibuchi, M., Kumagai, K. & Kaper, J. B. (1991).** Contribution of the *tdh* 1 gene of Kanagawa phenomenon-positive *Vibrio parahaemolyticus* to production of extracellular thermostable direct hemolysin. *Microb Pathog* **11**, 453–460.

**Nolan, C. M., Ballard, J., Kaysner, C. A., Lilja, J. L., Williams, L. P., Jr & Tenover, F. C. (1984).** *Vibrio parahaemolyticus* gastroenteritis: an outbreak associated with raw oysters in the Pacific Northwest. *Diagn Microbiol Infect Dis* **2**, 119–128.

**Ochman, H., Soncinit, F. C., Solomont, F. & Groismantt, E. A. (1996).** Identification of a pathogenicity island required for *Salmonella survival* in host cells. *Proc Natl Acad Sci USA* **93**, 7800–7804.

**Ochs, M. M., Mccusker, M. P., Bains, M. & Hancock, R. E. W. (1999).** Negative regulation of the *Pseudomonas aeruginosa* outer membrane porin OprD selective for imipenem and basic amino acids. *Antimicrob Agents Ch* **43**, 1085–1090.

**Octavia, S. & Lan, R. (2006).** Frequent recombination and low level of clonality within *Salmonella enterica* subspecies I. *Microbiology* **152**, 1099–1108.

**Oelschlaeger, T. a & Hacker, J. (2004).** Impact of pathogenicity islands in bacterial diagnostics. *APMIS* **112**, 930–936.

**Okada, N., Iida, T., Park, K., Goto, N., Yasunaga, T., Hiyoshi, H., Matsuda, S., Kodama, T. & 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. *Infect Immun* **77**, 904–913.

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

**Okuda, J., Ishibashi, M., Hayakawa, E., Nishino, T., Takeda, Y., Mukhopadhyay, A. K., Garg, S., Bhattacharya, S. K., Nair, G. B. & Nishibuchi, M. (1997).** Emergence of a unique O3:K6 clone of *Vibrio parahaemolyticus* in Calcutta, India, and isolation of strains from the same clonal group from Southeast Asian travelers arriving in Japan. *J Clin Microbiol* **35**, 3150–3155.

**Okura, M., Osawa, R., Tokunaga, A., Morita, M., Arakawa, E. & Watanabe, H. (2008).** Genetic analyses of the putative O and K antigen gene clusters of pandemic *Vibrio parahaemolyticus*. *Microbiol Immunol* **52**, 251–64.

**Olvera, A., Aragon, V. & Cerda, M. (2006).** Study of the population structure of *Haemophilus parasuis* by multilocus sequence typing. *Microbiology* **152**, 3683–3690.

**Ono, T., Park, K. S., Ueta, M., Tetsuya, I. & Honda, T. (2006).** Identification of proteins secreted via *Vibrio parahaemolyticus* type III secretion system 1. *Infect Immun* **74**, 1032–1042.

**Osawa, R. & Yamai, S. (1996).** Production of thermostable direct hemolysin by *Vibrio parahaemolyticus* enhanced by conjugated bile acids. *Appl Environ Microb* **62**, 3023–3025.

**Ottaviani, D., Leoni, F., Rocchegiani, E., Santarelli, S., Canonico, C., Masini, L., Ditrani, V. & Carraturo, A. (2008).** First clinical report of pandemic *Vibrio parahaemolyticus* O3:K6 infection in Italy. *J Clin Microbiol* **46**, 2144–2145.

**Ou, Y. Y., Gromiha, M. M., Chen, S. A. & Suwa, M. (2008).** TMBETADISC-RBF: Discrimination of beta-barrel membrane proteins using RBF networks and PSSM profiles. *Comput Biol Chem* **32**, 227–231.

**Pace, J. L., Chai, T. J., Rossi, H. A. & Jiang, X. (1997).** Effect of bile on *Vibrio parahaemolyticus*. *Appl Environ Microbiol* **63**, 2372–2377.

**Pal, D. & Das, N. (2010).** Isolation, identification and molecular characterization of *Vibrio parahaemolyticus* from fish samples in Kolkata. *Eur Rev Med Pharmaco* **14**, 545–549.

**Park, J. K., Keyhani, N. O. & Roseman, S. (2000).** Chitin catabolism in the marine bacterium *Vibrio furnissii*. *J Biol Chem* **275**, 33077–33083.

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

**Parkhill, J. & Wren, B. W. (2011).** Bacterial epidemiology and biology--lessons from genome sequencing. *Genome Biol* **12**, 230.

**Parveen, S., Hettiarachchi, K. A., Bowers, J. C., Jones, J. L., Tamplin, M. L., Mckay, R., Beatty, W., Brohawn, K., Dasilva, L. V & Depaola, A. (2008).** Seasonal distribution of total and pathogenic *Vibrio parahaemolyticus* in Chesapeake Bay oysters and waters. *Int J Food Microbiol* **128**, 354–361.

**Paulsen, I. T., Park, J. H., Choi, P. S. & Jr., M. H. S. (1997).** A family of Gram-negative bacterial outer membrne factors that function in the export of protiens, carbohydrates, drugs and heavy metals from Gram-negative bacteria. *FEMS Microbiol Lett* **156**, 1–8.

**Pembroke, J. T., Macmahon, C. & Mcgrath, B. (2002).** Cellular and molecular life sciences the role of conjugative transposons in the *Enterobacteriaceae*. *Cell Mol Life Sci* **59**, 2055–2064.

**Pinto, A. Di, Ciccarese, G., Corato, R. De, Novello, L. & Terio, V. (2008).** Detection of pathogenic *Vibrio parahaemolyticus* in southern Italian shellfish. *Food Control* **19**, 1037–1041.

**Plano, G. V & Straley, S. C. (1993).** Multiple effects of *lcrD* mutations in *Yersinia pestis*. *J Bacteriol* **175**, 3536–3545.

**Plano, G. V. & Straley, S. C. (1995).** Mutations in *yscC*, *yscD*, and *yscG* prevent high-level expression and secretion of V antigen and Yops in *Yersinia pestis*. *J Bacteriol* **177**, 3843–3854.

**Power, M. L., Ferrari, B. C., Littlefield-Wyer, J., Gordon, D. M., Slade, M. B. & Veal, D. a. (2006).** A naturally occurring novel allele of *Escherichia coli* outer membrane protein A reduces sensitivity to bacteriophage. *Appl Environ Microbiol* **72**, 7930-2.

**Priest, F. G., Barker, M., Baillie, L. W. J., Holmes, E. C. & Maiden, M. C. J. (2004).** Population structure and evolution of the *Bacillus cereus* group. *J Bacteriol* **186**, 7959-7970.

**Puohiniemi, R., Karvonen, M., Vuopio-Varkila, J., Muotiala, A., Helander, I. M. & Sarvas, M. (1990).** A strong antibody response to the periplasmic C-terminal domain of the OmpA protein of *Escherichia coli* is produced by immunization with purified OmpA or with whole *E. coli* or *Salmonella typhimurium* bacteria. *Infect Immun* **58**, 1691-1696.

**Qian, R.-H., Xiao, Z.-H., Zhang, C.-W., Chu, W.-Y., Wang, L.-S., Zhou, H.-H., Wei, Y. & Yu, L. (2008).** A conserved outer membrane protein as an effective vaccine candidate from *Vibrio alginolyticus*. *Aquaculture* **278**, 5-9.

**Qian, R., Chu, W., Mao, Z., Zhang, C., Wei, Y. & Yu, L. (2007).** Expression, characterization and immunogenicity of a major outer membrane protein from *Vibrio alginolyticus*. *Acta Bioch Bioph Sin* **39**, 194-200.

**Quilici, M.-L., Robert-Pillot, A., Picart, J. & Fournier, J.-M. (2005).** Pandemic *Vibrio parahaemolyticus* O3:K6 spread, France. *Emerg Infect Dis* **11**, 1148–1149.

**Raghunath, P., Acharya, S., Bhanumathi, A. & Karunasagar, I. (2008).** Detection and molecular characterization of *Vibrio parahaemolyticus* isolated from seafood harvested along the southwest coast of India. *Food Microbiol* **25**, 824–830.

**Reyes, A. L., Crawford, R. G., Spaulding, P. L., Peeler, J. T. & Twedt, R. M. (1983).** Hemagglutination and adhesiveness of epidemiologically distinct strains of *Vibrio parahaemolyticus*. *Microbiology* **39**, 721-725.

**Ritchie, J. M., Rui, H., Zhou, X., Iida, T., Kodoma, T., Ito, S., Davis, B. M., Bronson, R. T. & Waldor, M. K. (2012).** Inflammation and disintegration of intestinal villi in an experimental model for *Vibrio parahaemolyticus* induced diarrhea. *PLoS Pathog* **8**, e1002593.

**Rizvi, A. V. & Bej, A. K. (2010).** Multiplexed real-time PCR amplification of *tlh*, *tdh* and *trh* genes in *Vibrio parahaemolyticus* and its rapid detection in shellfish and Gulf of Mexico water. *Antonie Van Leeuwenhoek* **98**, 279-290.

**Rodionov, D. a, Vitreschak, A. G., Mironov, A. a & Gelfand, M. S. (2003).** Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *J Biol Chem* **278**, 41148-41159.

**Roque, A., Lopez-Joven, C., Lacuesta, B., Elandaloussi, L., Wagley, S., Furones, M. D., Ruiz-Zarzuela, I., de Blas, I., Rangdale, R. & Gomez-Gil, B. (2009).** Detection and identification of *tdh*- and *trh*-positive *Vibrio*

*parahaemolyticus* strains from four species of cultured bivalve molluscs on the Spanish Mediterranean Coast. *Appl Environ Microbiol* **75**, 7574–7577.

**Ruiz, N., Kahne, D. & Silhavy, T. J. (2006).** Advances in understanding bacterial outer-membrane biogenesis. *Nat Rev Microbiol* **4**, 57–66.

**Ruwandeepika, H. a D., Defoirdt, T., Bhowmick, P. P., Shekar, M., Bossier, P. & Karunasagar, I. (2010).** Presence of typical and atypical virulence genes in vibrio isolates belonging to the *Harveyi* clade. *J Appl Microbiol* **109**, 888–899.

**Sakazaki, R. (1992).** Bacteriology of *Vibrio* and related organisms. In *Cholera*, pp. 37–51. Edited by D. Barua & W. B. Greenough. New York: Plenum publishing.

**Sangal, V., Harbottle, H., Mazzoni, C. J., Helmuth, R., Guerra, B., Didelot, X., Paglietti, B., Rabsch, W., Brisse, S., & other authors. (2010).** Evolution and population structure of *Salmonella enterica* serovar Newport. *J Bacteriol* **192**, 6465–6476.

**Sarkar, B. L., Nair, G. B., Banerjee, A. K. & Pal, S. C. (1985).** Seasonal distribution of *Vibrio parahaemolyticus* in freshwater environs and in association with freshwater fishes in Calcutta. *Appl Environ Microbiol* **49**, 132–136.

**Sawabe, T., Ohtsuka, M. & Nakano, K. (1998).** *Vibrio halioticoli* sp. nov., a non-motile alginolytic marine bacterium isolated from the gut of the abalone *Haliotis discus hannai*. *Int J Syst Bacteriol* **48**, 573–580.

**Schulz, G. E. (2002).** The structure of bacterial outer membrane proteins. *Biochim Biophys Acta* **1565**, 308–317.

**Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. & Whittam3, T. S. (1986).** Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* **51**, 873–884.

**Serichantalergs, O., Bhuiyan, N. A., Nair, G. B., Chivaratanond, O., Srijan, A., Bodhidatta, L., Anuras, S. & Mason, C. J. (2007).** The dominance of pandemic serovars of *Vibrio parahaemolyticus* in expatriates and sporadic cases of diarrhoea in Thailand, and a new emergent serovar (O3:K46) with pandemic traits. *J Med Microbiol* **56**, 608-613.

**Seydel, a, Gounon, P. & Pugsley, a P. (1999).** Testing the "+2 rule" for lipoprotein sorting in the *Escherichia coli* cell envelope with a new genetic selection. *Mol Microbiol* **34**, 810-821.

**Shen, C., Kuo, T., Lin, C., Chow, L. & Chen, W. (2009).** Proteomic identification of membrane proteins regulating antimicrobial peptide resistance in *Vibrio parahaemolyticus*. *J Appl Microbiol* **108**, 1398–1407.

**Shime-Hattori, A., Iida, T., Arita, M., Park, K. S., Kodama, T. & Honda, T. (2006).** Two type IV pili of *Vibrio parahaemolyticus* play different roles in biofilm formation. *FEMS Microbiol Lett* **264**, 89-97.

Shimohata, T., Nakano, M., Lian, X., Shigeyama, T., Iba, H., Hamamoto, A., Yoshida, M., Harada, N., Yamamoto, H., & other authors. (2011). *Vibrio parahaemolyticus* infection induces modulation of IL-8 secretion through dual pathway via VP1680 in Caco-2 cells. *J Infect Dis* **203**, 537–544.

Simonet, V. C., Baslé, A., Klose, K. E. & Delcour, A. H. (2003). The *Vibrio cholerae* porins OmpU and OmpT have distinct channel properties. *J Biol Chem* **278**, 17539–17245.

Smith, J. (2001). The social evolution of bacterial pathogenesis. *Proc Biol Sci* **268**, 61–69.

Smith, J. M., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci USA* **90**, 4384–4388.

Smith, J. M., Feil, E. J. & Smith, N. H. (2000). Population structure and evolutionary dynamics of pathogenic bacteria. *BioEssays* **22**, 1115–1122.

Smith, S. G. J., Mahon, V., Lambert, M. a & Fagan, R. P. (2007). A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiol Lett* **273**, 1–11.

Smyth, D. S., Feil, E. J., Meaney, W. J., Hartigan, P. J., Tollersrud, T., Fitzgerald, J. R., Enright, M. C. & Smyth, C. J. (2009). Molecular genetic typing reveals further insights into the diversity of animal-associated *Staphylococcus aureus*. *J Med Microbiol* **58**, 1343–1353.

Snipen, L. & Ussery, D. W. (2010). Standard operating procedure for computing pangenome trees. *Stand Genomic Sci* **2**, 135–141.

Sonntag, I., Schwarz, H., Hirota, Y. & Henning, U. L. F. (1978). Cell Envelope and Shape of *Escherichia coli*: Multiple Mutants Missing the Outer Membrane Lipoprotein and Other Major Outer Membrane Proteins. *Proteins* **136**, 280–285.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**, 503–517.

Sparo, M., Urbizu, L., Solana, M. V, Pourcel, G., Delpech, G., Confalonieri, A., Ceci, M. & Sánchez Bruni, S. F. (2012). High-level resistance to gentamicin: genetic transfer between *Enterococcus faecalis* isolated from food of animal origin and human microbiota. *Lett Appl Microbiol* **54**, 119–125.

Sperandio, V., Girón, J. A., Silveira, W. D., Kaper, J. B., Silveira, W. D., Sperandio, V. & Giro, J. A. (1995). The OmpU outer membrane protein, a potential adherence factor of *Vibrio cholerae*. *Infect Immun* **63**, 4433–4438.

Sriratanaban, A. & Reinprayoon, S. (1982). *Vibrio patahaemolyticus*: a major cause of travellers' diarrhea in Bangkok. *ASTMH* **34**, 128 – 130.

Stathopoulos, C., Hendrixson, D. R., Thanassi, D. G., Hultgren, S. J., St Geme, J. W. & Curtiss, R. (2000). Secretion of virulence determinants by the general secretory pathway in Gram-negative pathogens: an evolving story. *Microbes Infect* **2**, 1061–1072.

Stecher, B., Denzler, R., Maier, L., Bernet, F., Sanders, M. J., Pickard, D. J., Barthel, M., Westendorf, A. M., Krogfelt, K. a, & other authors. (2012). Gut inflammation can boost horizontal gene transfer between pathogenic and commensal *Enterobacteriaceae*. *Proc Natl Acad Sci USA* **109**, 1269-1274.

Stewart, B. J. & Mccarter, L. L. (2003). Lateral flagellar gene system of *Vibrio parahaemolyticus*. *J Bacteriol* **185**, 4508-4518.

Stine, O. C., Sozhamannan, S., Gou, Q., Zheng, S., Morris, J. G. & Johnson, J. a. (2000). Phylogeny of *Vibrio cholerae* based on *recA* sequence. *Infect Immun* **68**, 7180-7185.

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

Sujeewa, A. K. W., Norrakiah, A. S. & Laina, M. (2009). Prevalence of toxic genes of *Vibrio parahaemolyticus* in shrimps (*Penaeus monodon*) and culture environment. *Int Food Res J* **16**, 89-95.

Suthienkul, O., Iida, T., Park, K. S., Ishibashi, M., Supavej, S., Yamamoto, K. & Honda, T. (1996). Restriction fragment length polymorphism of the *tdh* and *trh* genes in clinical *Vibrio parahaemolyticus* strains. *J Clin Microbiol* **34**, 1293-1295.

Tada, J., Ohashi, T., Nishimura, N., Shirasaki, Y., Ozaki, H., Fukushima, S., Takano, J., Nishibuchi, M. & Takeda, Y. (1992). Detection of the thermostable direct hemolysin gene (*tdh*) and the thermostable direct hemolysin-related hemolysin gene (*trh*) of *Vibrio parahaemolyticus* by polymerase chain reaction. *Mol Cell Probe* **6**, 477-487.

Takikawa, I. (1958). Studies on pathogenic halophillic bacteria. *Yokohama Med Bull* **9**, 313-322.

Tamber, S., Ochs, M. M., Robert, E. W. & Hancock, R. E. W. (2006). Role of the novel OprD family of porins in nutrient uptake in *Pseudomonas aeruginosa*. *J Bacteriol* **188**, 45-54.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA 4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596-1599.

Tanabe, T., Funahashi, T., Okajima, N., Nakao, H., Takeuchi, Y., Miyamoto, K., Tsujibo, H. & Yamamoto, S. (2011). The *Vibrio parahaemolyticus* *pvuA1* gene (formerly termed *psuA*) encodes a second ferric vibrioferrin receptor that requires tonB2. *FEMS Microbiol Lett* **324**, 73-79.

Taniguchi, H., Hirano, H., Kubomura, S., Higashi, K. & 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.

Taniguchi, H., Kubomura, S., Hirano, H., Mizue, K., Ogawa, M. & Mizugushi, Y. (1990). Cloning and characterization of a gene encoding a new thermostable hemolysin from *Vibrio parahaemolyticus*. *FEMS Microbiol Lett* **67**, 339-346.

**Tassin, M. G., Siebeling, R. J., Roberts, N. C. & Larson, A. D. (1983).** Presumptive identification of *Vibrio* species with H antiserum. *J Clin Microbiol* **18**, 400–407.

**Teh, C. S. J., Chua, K. H. & Thong, K. L. (2009).** Simultaneous differential detection of human pathogenic and nonpathogenic *Vibrio* species using a multiplex PCR based on *gyrB* and *pntA* genes. *J Appl Microbiol* **108**, 1940–1945.

**Terai, A. (1991).** Evidence for insertion sequence-mediated spread of the thermostable direct hemolysin gene among *Vibrio* Species. *Microbiology* **173**, 5036–5046.

**Terzi, G., Büyüktanir, O. & Yurdusev, N. (2009).** Detection of the *tdh* and *trh* genes in *Vibrio parahaemolyticus* isolates in fish and mussels from Middle Black Sea Coast of Turkey. *Lett Appl Microbiol* **49**, 757-763.

**Thanassi, D. G. & Hultgren, S. J. (2000).** Multiple pathways allow protein secretion across the bacterial outer membrane. *Curr Opin Cell Biol* **12**, 420-430.

**Theptaranon, Y. & Keratikorn, K. (2005).** *Kao-Man-Koong Thai agro-products*. National science and technology development agency (NSTDA), Thailand.

**Theunissen, S., Vergauwen, B., De Smet, L., Van Beeumen, J., Van Gelder, P. & Savvides, S. N. (2009).** The agglutination protein AggA from *Shewanella oneidensis* MR-1 is a TolC-like protein and forms active channels *in vitro*. *Biochem Bioph Res Co* **386**, 380-385.

**Thiennimitr, P., Winter, S. E. & Bäumler, A. J. (2012).** *Salmonella*, the host and its microbiota. *Curr Opin Microbiol* **15**, 108–114.

**Thompson, C. C., Thompson, F. L. & Vicente, A. C. P. (2008).** Identification of *Vibrio cholerae* and *Vibrio mimicus* by multilocus sequence analysis (MLSA). *Int J Syst Evol Microbiol* **58**, 617-621.

**Thompson, C. C., Vicente, A. C. P., Souza, R. C., Vasconcelos, A. T. R., Vesth, T., Jr, N. A., Ussery, D. W., Iida, T. & Thompson, F. L. (2009).** Genomic taxonomy of *Vibrios*. *BMC Evol Biol* **9**, 258.

**Thompson, F. L., Iida, T. & Swings, J. (2004).** Biodiversity of *Vibrios*. *Microbiol Mol Biol Rev* **68**, 403–431.

**Thompson, F. L., Gevers, D., Thompson, C. C., Dawyndt, P., Naser, S., Hoste, B., Munn, C. B. & Swings, J. (2005).** Phylogeny and molecular identification of *Vibrios* on the basis of multilocus sequence analysis. *Appl Environ Microbiol* **71**, 5107-5115.

**Thompson, F. L., Gomez-Gil, B., Vasconcelos, A. T. R. & Sawabe, T. (2007).** Multilocus sequence analysis reveals that *Vibrio harveyi* and *V. campbellii* are distinct species. *Appl Environ Microbiol* **73**, 4279-4285.

**Thompson, C, C., Thompson, F. L., Vandemeulebroecke, K., Hoste, B., Dawyndt, P. & Swings, J. (2004).** Use of *recA* as an alternative phylogenetic marker in the family *Vibrionaceae*. *Int J Syst Evol Microbiol* **54**, 919-924.

**Tokuda, H. & Matsuyama, S. I. (2004).** Sorting of lipoproteins to the outer membrane in *E. coli*. *Biochim Biophys Acta* **1693**, 5–13.

**Tookwinas, S., Chiyakum, K. & Somsueb, S. (2005).** Aquaculture of White Shrimp *Penaeus vannamei* in Thailand. *SEAFDEC / AQD Institutional Repository ( SAIR )*.

**Tookwinas. S. (1993).** *Shrimp culture in Thailand-Present status and future directions for research*. Coastal Aquaculture Division, Department of Fisheries, Thailand.

**Trias, J. & Nikaido, H. (1990).** Outer membrane protein D2 catalyzes facilitated diffusion of carbapenems and penems through the outer membrane of *Pseudomonas aeruginosa*. *Antimicrob Agents Ch* **34**, 52–57.

**Tsang, T. M., Felek, S. & Krukonis, E. S. (2010).** Ail binding to fibronectin facilitates *Yersinia pestis* binding to host cells and Yop delivery. *Infect Immun* **78**, 3358–3368.

**Turner, K. M. E. & Feil, E. J. (2007).** The secret life of the multilocus sequence type. *Int J Antimicro Ag* **29**, 129–135.

**Urwin, R. & Maiden, M. C. J. (2003).** Multi-locus sequence typing-a tool for global epidemiology. *Trends Microbiol* **11**, 479–487.

**Vogt, J. & Schulz, G. E. (1999).** The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence. *Structure* **7**, 1301–1309.

**Vongxay, K., Wang, S., Zhang, X., Wu, B., Hu, H., Pan, Z., Chen, S. & Fang, W. (2008).** Pathogenetic characterization of *Vibrio parahaemolyticus* isolates from clinical and seafood sources. *Int J Food Microbiol* **126**, 71–75.

**De Vriendt, K., Theunissen, S., Carpentier, W., De Smet, L., Devreese, B. & Van Beeumen, J. (2005).** Proteomics of *Shewanella oneidensis* MR-1 biofilm reveals differentially expressed proteins, including AggA and RibB. *Proteomics* **5**, 1308–1316.

**Vuddhakul, V., Chowdhury, A., Laohaprertthisan, V., Pungrasamee, P., Patararungrong, N., Thianmontri, P., Ishibashi, M., Matsumoto, C. & Nishibuchi, M. (2000).** Isolation of a pandemic O3:K6 clone of a *Vibrio parahaemolyticus* strain from environmental and clinical sources in Thailand. *Appl Environ Microbiol* **66**, 2685-2689.

**Wagley, S., Koofhethile, K., Wing, J. B. & Rangdale, R. (2008).** Comparison of *Vibrio parahaemolyticus* isolated from seafoods cases of gastrointestinal disease in the UK. *Int J Environ Health Res* **18**, 283-293.

**Wagley, S., Koofhethile, K. & Rangdale, R. (2009).** Prevalence and potential pathogenicity of *Vibrio parahaemolyticus* in Chinese mitten crabs (*Eriocheir sinensis*) harvested from the River Thames estuary, England. *Journal Of Food Protection* **72**, 60–66.

**Wagner, M. & Haider, S. (2012).** New trends in fluorescence in situ hybridization for identification and functional analyses of microbes. *Curr Opin Biotech* **23**, 96-102.

**Wagner, M., Horn, M. & Daims, H. (2003).** Fluorescence in situ hybridisation for the identification and characterisation of prokaryotes. *Curr Opin Microbiol* **6**, 302–309.

**Wang, D., Wang, H., Zhou, Y., Zhang, Q., Zhang, F., Du, P., Wang, S., Chen, C. & Kan, B. (2011a).** Genome sequencing reveals unique mutations in characteristic metabolic pathways and the transfer of virulence genes between *V. mimicus* and *V. cholerae*. *PloS One* **6**, e21299.

**Wang, R., Huang, J., Zhang, W., Lin, G., Lian, J., Jiang, L., Lin, H., Wang, S. & Wang, S. (2011b).** Detection and identification of *Vibrio parahaemolyticus* by multiplex PCR and DNA–DNA hybridization on a microarray. *J Genet Genomics* **38**, 129–135.

**Werner, J. & Misra, R. (2005).** YaeT (Omp85) affects the assembly of lipid-dependent and lipid-independent outer membrane proteins of *Escherichia coli*. *Mol Microbiol* **57**, 1450–1459.

**Whitaker, W. B., Parent, M. a, Boyd, A., Richards, G. P. & Boyd, E. F. (2012).** The *Vibrio parahaemolyticus* ToxRS regulator is required for stress tolerance and colonization in a novel orogastric streptomycin-induced adult murine model. *Infect Immun* **80**, 1834–1845.

**Whitaker, W. B., Parent, M. a., Naughton, L. M., Richards, G. P., Blumerman, S. L. & Boyd, E. F. (2010).** Modulation of responses of *Vibrio parahaemolyticus* O3:K6 to pH and temperature stresses by growth at different salt concentrations. *Appl Environ Microbiol* **76**, 4720–4729.

**WHO & FAO. (2011).** Risk assessement of *Vibrio parahaemolyticus* in seafood. *World Health Organization/Food and Agriculture Organization microbiological risk assessment series* **16**, 193pp. Rome.

**Wibbenmeyer, J. A., Provenzano, D., Candice, F., Klose, K. E., Delcour, A. H. & Landry, C. F. (2002).** *Vibrio cholerae* OmpU and OmpT porins are differentially affected by bile. *Infect Immun* **70**, 121–126.

**William H Barker, Jr., MD. and Eugene J. Gangarosa, M. (1974).** Food poisoning due to *Vibrio parahaemolyticus*. *Annu Rev Med* **25**, 75–81.

**Wong, H. & Lin, C. (2001).** Evaluation of typing of *Vibrio parahaemolyticus* by three PCR methods using specific primers. *J Clin Microbiol* **39**, 4233–4240.

**Wong, H., Liu, C., Pan, T., Lee, C., Shih, D. Y. & Wang, T. (1999).** Molecular typing of *Vibrio parahaemolyticus* isolates, obtained from patients involved in food poisoning outbreaks in Taiwan, by random amplified polymorphic DNA analysis. *J Clin Microbiol* **37**, 1809–1812.

**Wong, H., Liu, S., Wang, T., Chiou, C., Liu, D., Nishibuchi, M., Lee, B. & Lee, C. (2000).** Characteristics of *Vibrio parahaemolyticus* O3:K6 from Asia. *Appl Environ Microbiol* **66**, 3981–3986.

**Wooldridge, K. G. & Williams, P. H. (1993).** Iron uptake mechanisms of pathogenic bacteria. *FEMS Microbiol Rev* **12**, 325–348.

**Wootipoom, N., Bhoopong, P., Pomwised, R., Nishibuchi, M., Ishibashi, M. & Vuddhakul, V. (2007).** A decrease in the proportion of infections by

pandemic *Vibrio parahaemolyticus* in Hat Yai. *J Med Microbiol* **56**, 1630–1638.

**Wyban, J. (2007).** Thailand's white shrimp revolution. *Global aquaculture advocate* 56–58. Available at http://www.hihealthshrimp.com/.

**Xicohtencatl-Cortés, J., Lyons, S., Chaparro, A. P., Hernández, D. R., Saldaña, Z., Ledesma, M. a, Rendón, M. a, Gewirtz, A. T., Klose, K. E. & Girón, J. a. (2006).** Identification of proinflammatory flagellin proteins in supernatants of *Vibrio cholerae* O1 by proteomics analysis. *Mol Cell Proteomics* **5**, 2374-2383.

**Xie, J., Sun, X., Pan, Y. & Zhao, Y. (2012).** Combining basic electrolyzed water pretreatment and mild heat greatly enhanced the efficacy of acidic electrolyzed water against *Vibrio parahaemolyticus* on shrimp. *Food Control* **23**, 320-324.

**Xie, Z. Y., Hu, C.-Q., Chen, C., Zhang, L. P. & Ren, C. H. (2005).** Investigation of seven *Vibrio* virulence genes among *Vibrio alginolyticus* and *Vibrio parahaemolyticus* strains from the coastal mariculture systems in Guangdong, China. *Lett Appl Microbiol* **41**, 202-207.

**Xu, C., Ren, H., Wang, S. & Peng, X. (2004).** Proteomic analysis of salt-sensitive outer membrane proteins of *Vibrio parahaemolyticus*. *Res Microbiol* **155**, 835-842.

**Xu, C., Wang, S., Ren, H., Lin, X., Wu, L. & Peng, X. (2005).** Proteomic analysis on the expression of outer membrane proteins of *Vibrio alginolyticus* at different sodium concentrations. *Proteomics* **5**, 3142-3152.

**Yamaguchi, K., Yu, F. & Inouye, M. (1988).** A single amino acid determinant of the membrane localization of lipoproteins in *E. coli*. *Cell* **53**, 423–432.

**Yamamoto, A., Vuddhakul, V., Charernjiratragul, W., Vose, D., Osaka, K., Shigematsu, M., Toyofuku, H., Yamamoto, S. & Nishibuchi, M. (2008).** Quantitative modeling for risk assessment of *Vibrio parahaemolyticus* in bloody clams in southern Thailand. *Int J Food Microbiol* **124**, 70–78.

**Yamamoto, S., Hara, Y., Tomochika, K. & Shinoda, S. (1995a).** Utilization of hemin and hemoglobin as iron sources by *Vibrio parahaemolyticus* and identification of an iron-repressible hemin-binding protein. *FEMS Microbiol Lett* **128**, 195–200.

**Yamamoto, S., Akiyama, T., Okujo, N., Matsu-ura, S. & Shinoda, S. (1995b).** Demonstration of a ferric vibrioferrin-binding protein in the outer membrane of *Vibrio parahaemolyticus*. *Microbiol Immunol* **39**, 759–766.

**Yamamoto, S., Okujo, N., Yoshida, T., Matsuura, S. & Shinoda, S. (1994).** Structure and iron transport activity of vibrioferrin, a new siderophore of *Vibrio parahaemolyticus*. *J Biochem* **115**, 868-874.

**Yamashita, S., Lukacik, P., Barnard, T. J., Noinaj, N., Felek, S., Tsang, T. M., Krukonis, E. S., Hinnebusch, B. J. & Buchanan, S. K. (2011).** Structural insights into Ail-mediated adhesion in *Yersinia pestis*. *Structure* **19**, 1672–1682.

Yan, Y., Cui, Y., Han, H., Xiao, X., Wong, H., Tan, Y., Guo, Z., Liu, X., Yang, R. & 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., & other authors. (2010). Structure and functional characterization of *Vibrio parahaemolyticus* thermostable direct hemolysin. *J Biol Chem* **285**, 16267–16274.

Yang, L., Zhan, L., Han, H., Gao, H., Guo, Z., Qin, C., Yang, R., Liu, X. & Zhou, D. (2010). The low-salt stimulon in *Vibrio parahaemolyticus*. *International Journal of Food Microbiology* **137**, 49–54.

Yang, Z. R. & Thomson, R. (2005). Bio-basis function neural network for prediction of protease cleavage sites in proteins. *IEEE T Neural Networ* **16**, 263–274.

Yeung, P. S., Hayes, M. C., Depaola, A., Kaysner, C. A., Kornstein, L. & Boor, K. J. (2002). Comparative phenotypic, molecular, and virulence characterization of *Vibrio parahaemolyticus* O3:K6 isolates. *Appl Environ Microbiol* **68**, 2901–2909.

Yu, C. S., Lin, C. J. & Hwang, J. K. (2004). Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci* **13**, 1402–1406.

Yu, Y., Hu, W., Wu, B., Zhang, P., Chen, J., Wang, S. & Fang, W. (2011). *Vibrio parahaemolyticus* isolates from southeastern Chinese coast are genetically diverse with circulation of clonal complex 3 strains since 2002. *Foodborne Pathog Dis* **8**, 1169–1176.

Yuan, Y., Wang, X., Guo, S. & Qiu, X. (2011). Gene cloning and prokaryotic expression of recombinant outer membrane protein from *Vibrio parahaemolyticus*. *Chin J Oceanol Limn* **29**, 952–957.

Zen-Yoji, H., Hitokoto, H., Morozumi, S. & Le Clair, R. a. (1971). Purification and characterization of a hemolysin produced by *Vibrio parahaemolyticus*. *J Infect Dis* **123**, 665–667.

Zhang, G. Z. & Huang, D. S. (2004). Prediction of inter-residue contacts map based on genetic algorithm optimized radial basis function neural network and binary input encoding scheme. *J Comput Aided Mol Des* **18**, 797–810.

Zhou, X., Shah, D. H., Konkel, M. E. & Call, D. R. (2008). Type III secretion system 1 genes in *Vibrio parahaemolyticus* are positively regulated by ExsA and negatively regulated by ExsD. *Mol Microbiol* **69**, 747–764.

Zhou, X., Konkel, M. E. & Call, D. R. (2010a). 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.

Zhou, X., Konkel, M. E. & Call, D. R. (2010b). Regulation of type III secretion system 1 gene expression in *Vibrio parahaemolyticus* is dependent on interactions between ExsA, ExsC, and ExsD. *Virulence* **1**, 260–272.

**Zwirglmaier, K. (2005).** Fluorescence in situ hybridisation (FISH)-the next generation. *FEMS Microbiol Lett* **246**, 151–158.

# 7. APPENDIX 1
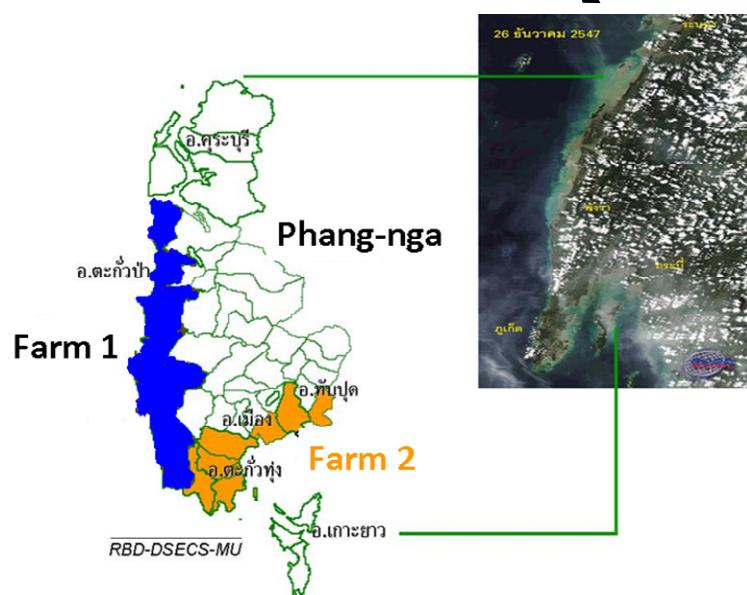


**Figure A1.** Locations of shrimp farms, 1 and 2, where isolates from shrimp tissue and water were collected. Both farms are located in the south of Thailand. Farm1 is situated at a coastal region that is open to the Andaman Sea whereas farm 2 is located inland and is isolated from the Andaman Sea.
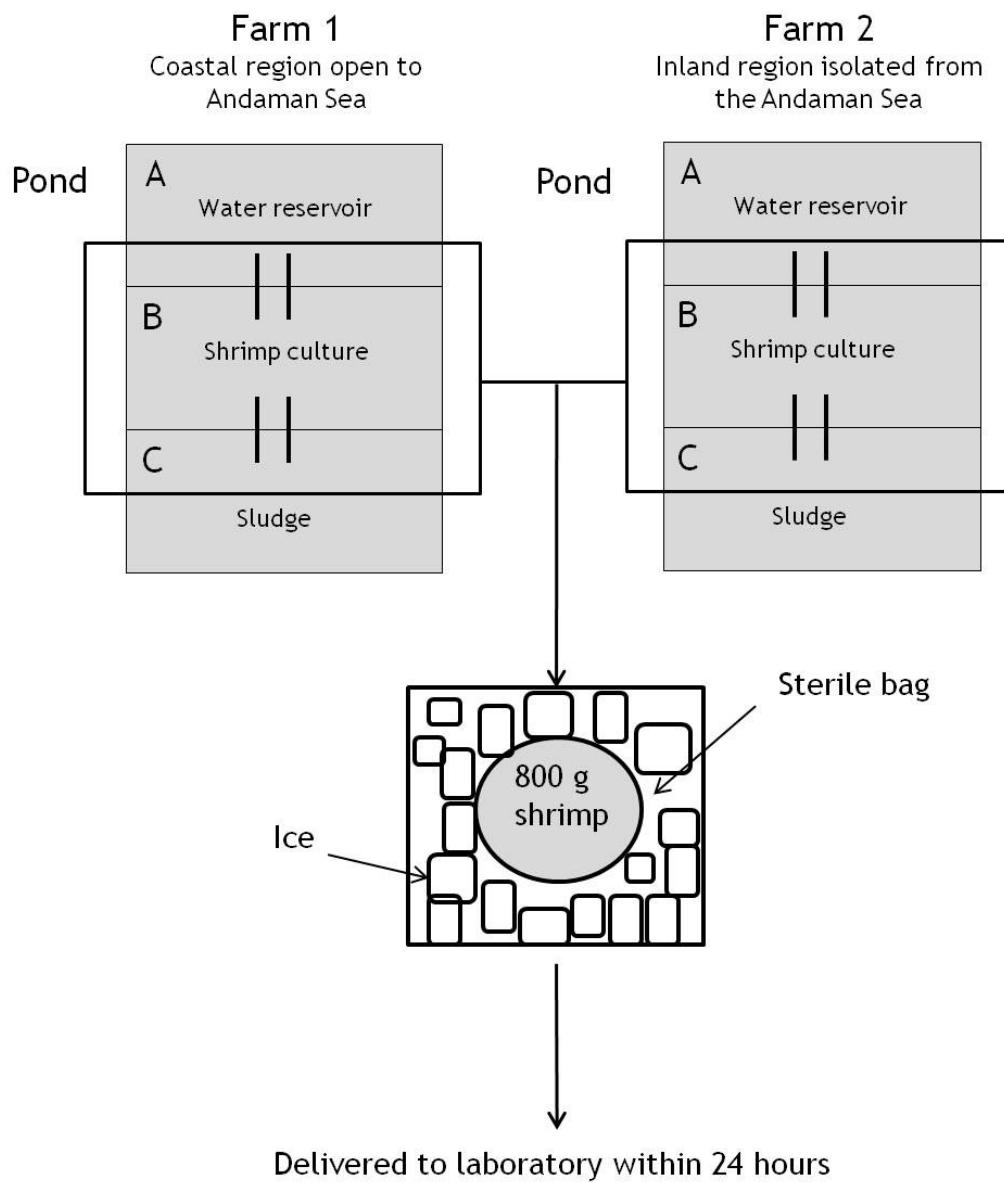
**Figure A2.** Shrimp sampling method from two shrimp farms, 1 and 2, in the south of Thailand. Both farms were each divided into three ponds, A, B, and C. Pond A is used for water preparation that allows adjustment of suitable water conditions (e.g. pH, salinity, dissolved oxygen concentration, etc.) for shrimp growing. Water of optimal conditions is transferred to pond B which will be used for growing the shrimps. Pond C is used to store sludge or biological waste from the shrimp growing pond after harvesting. Shrimps were collected from pond 2 and put on ice after harvesting. Shrimp samples were delivered from the south of Thailand to the laboratory in Bangkok within 24 h.
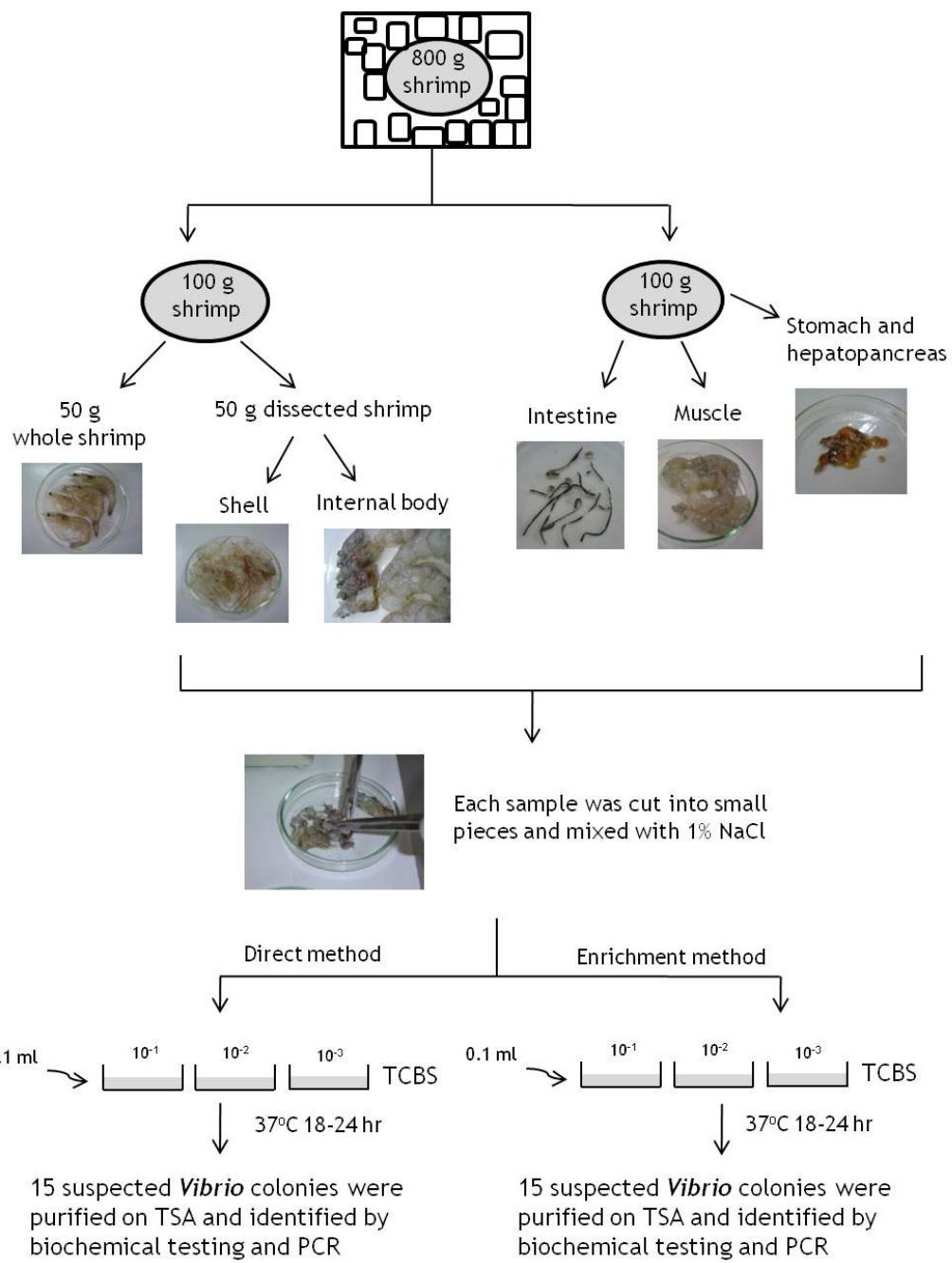
**Figure A3.** Bacterial sampling methods for *V. parahaemolyticus* from shrimp samples. A hundred grams (approximately 4-5 shrimps) of shrimps were dissected into different parts. Tissue of different parts from 100 g of shrimps were pooled and mixed with 1% NaCl solution. Dilutions of the mixture were plated using direct and enrichment methods. Suspected *Vibrio* colonies were further tested and *V. parahaemolyticus* isolates were subsequently identified.
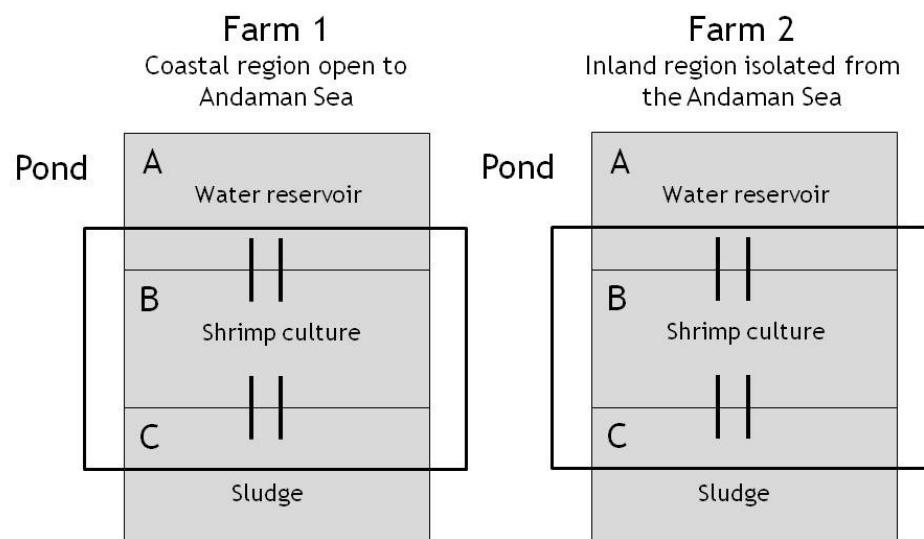
| Shrimp tissue | Farm 1 | Farm 2 |
|---|---|---|
| Muscle | No isolates were collected | 3 isolates (VP94, VP96, and VP98) 14 August 2007 |
| Intestine | 2 isolates (VP100 and VP102) 5 August 2007 | 2 isolates (VP104 and VP106) 14 August 2007 |
| Stomach and Hepatopancreas | No isolates were collected | 4 isolates (VP84, VP86, VP88, and VP90) 14 August 2007 |
| Shell | 2 isolates (VP116 and VP118) 5 August 2007 | 1 isolates (VP122) 14 August 2007 |
| Internal body | 4 isolates (VP108, VP110, VP112, and VP114) 5 August 2007 | No isolates were collected |

**Figure A4.** *V. parahaemolyticus* isolates recovered from different parts of shrimp at shrimp farms 1 and 2.

Farm 1
Coastal region open to
Andaman Sea

Farm 2
Inland region isolated from
the Andaman Sea

Pond

A
Water reservoir

B
Shrimp culture

C
Sludge

Pond

A
Water reservoir

B
Shrimp culture

C
Sludge

Sample farm 1

A     B     C

Sample farm 2

A     B     C

3 water samples were collected at 60 cm depth
(1 sample per pond from 2 farms)

dilution in 0.85% NaCl

Direct method

0.1 ml     10     $10^{-1}$     $10^{-2}$     TCBS

37°C 18-24 hr

15 suspected *Vibrio* colonies were
purified on TSA and identified by
biochemical testing and PCR

Enrichment
method

0.1 ml     10     $10^{-1}$     $10^{-2}$     TCBS

37°C 18-24 hr

15 suspected *Vibrio* colonies were
purified on TSA and identified by
biochemical testing and PCR

Selected *V. parahaemolyticus* isolates

Farm 1/A =2 isolates (VP44, VP46)
Farm 1/B = 7 isolates (VP48, VP50, VP52, VP54, VP56, VP58, and VP60)
Farm 1/C = 3 isolates (VP62, VP64, and VP66)
Farm 2/A = 3 isolates  (VP72, VP74, and VP76)
Farm 2/B = 1 isolates (VP80)

All water samples were collected on 21 January 2008

**Figure A5.** Sampling protocol of *V. parahaemolyticus* isolates from water at shrimp farms 1 and 2. Water samples were collected from ponds A, B, and C at farm 1 and ponds A and B at farm 2. Dilutions of the collected water were diluted with 0.85% NaCl and plated by the direct and enrichment methods. Suspected *Vibrio* colonies were further tested and *V. parahaemolyticus* isolates were subsequently identified.

**Table A1.** Polyvalent and monovalent K-antiserum of *V. parahaemolyticus*

| Polyvalent K antiserum | K antiserum |
| --- | --- |
| I | 1, 3, 4, 5, 6, 7, 8 |
| III | 9, 10, 11, 12, 13, 15, 17 |
| III | 18, 19, 20, 21, 22, 23, 24 |
| IV | 25, 26, 28, 29, 30, 31, 32 |
| V | 33, 34, 36, 37, 38, 39, 40 |
| VI | 41, 42, 43, 44, 45, 46, 47 |
| VII | 48, 49, 50, 51, 52, 53, 54 |
| VIII | 55, 56, 57, 58, 59, 60, 61 |
| IX | 63, 64, 65, 66, 67, 68, 69, 70, 71 |

DENKA SEIKEN Co., Ltd., Tokyo, Japan

## 8. APPENDIX 2

## Growth media composition

**Trytone Soya Broth (TSB) + 3% NaCl**

1 litre

| | |
|---|---|
| Tryptone Soya Broth, dehydrated (Oxoid) | 30.0 g |
| NaCl, AnalaR NORMAPUR (BDH PROLABO) | 30.0 g |
| Distilled $H_2O$ | to 1 litre |

Autoclaved at 121$^o$C for 15 min

**Trytone Soya Agar (TSA) + 3% NaCl**

1 litre

| | |
|---|---|
| Tryptone Soya Agar, dehydrated (Oxoid) | 40.0 g |
| NaCl, AnalaR NORMAPUR (BDH PROLABO) | 30.0 g |
| Distilled $H_2O$ | to 1 litre |

Autoclaved at 121$^o$C for 15 min

**50% (v/v) glycerol TSB + 3% NaCl**

50 ml

| | |
|---|---|
| Tryptone Soya Broth, dehydrated (Oxoid) | 1.5 g |
| NaCl, AnalaR NORMAPUR (BDH PROLABO) | 1.5 g |
| Glycerol, ≥99.0 % (Sigma) | 25.0 ml |
| Distilled $H_2O$ | 25.0 ml |

Autoclaved at 121$^o$C for 15 min

# 9. APPENDIX 3
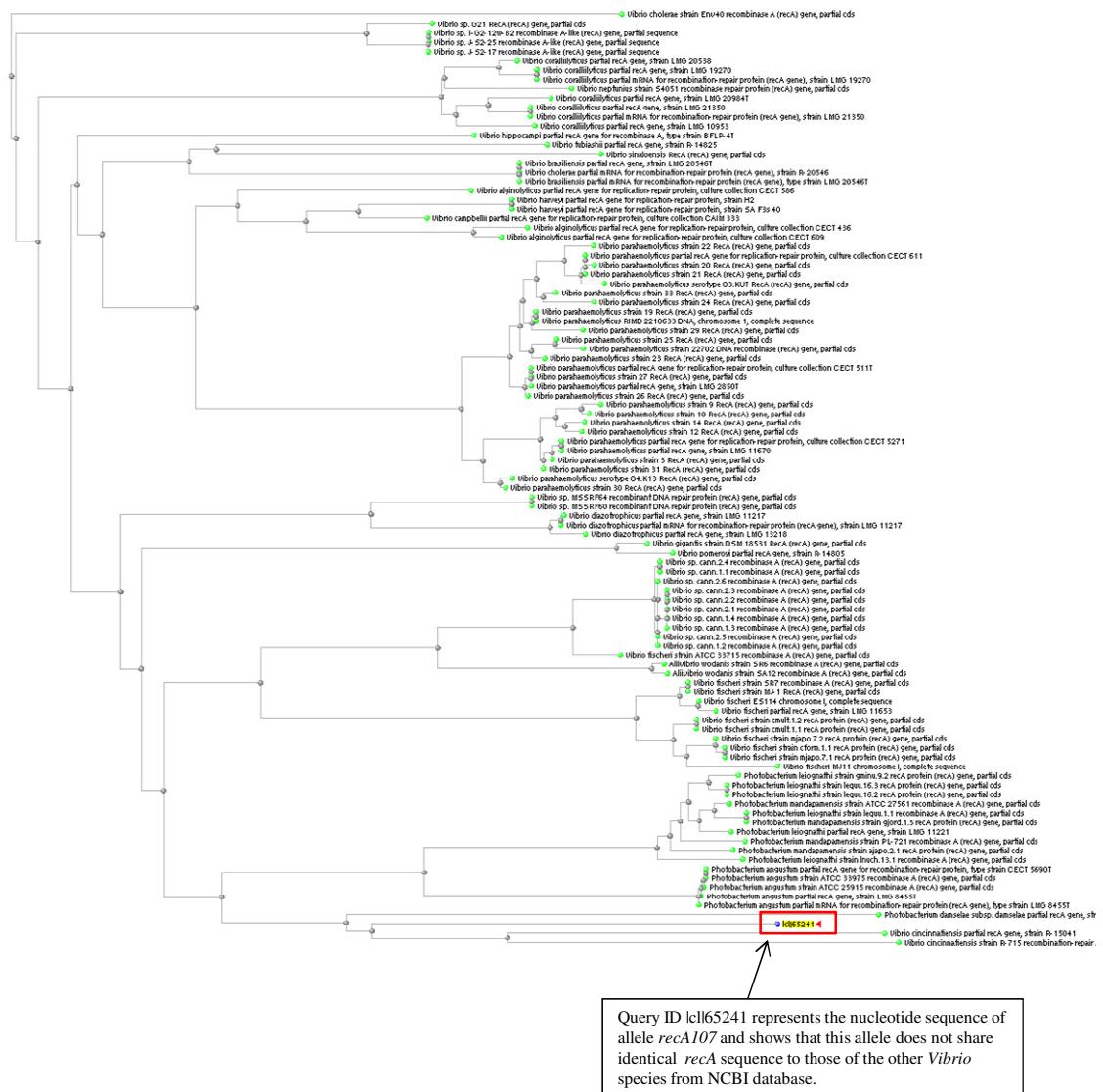


**Figure A6.** Neighbour-Joining of tree *recA* nucleotide sequences of sixteen *Vibrios* species and related species such as *Photobacterium* species from NCBI database. Query ID |cl|65241 represents the nucleotide sequence of allele *recA107* and shows that this allele has not been identified in *Vibrio* species by representing a distinct lineage in the pholygenetic tree of *recA* allele.
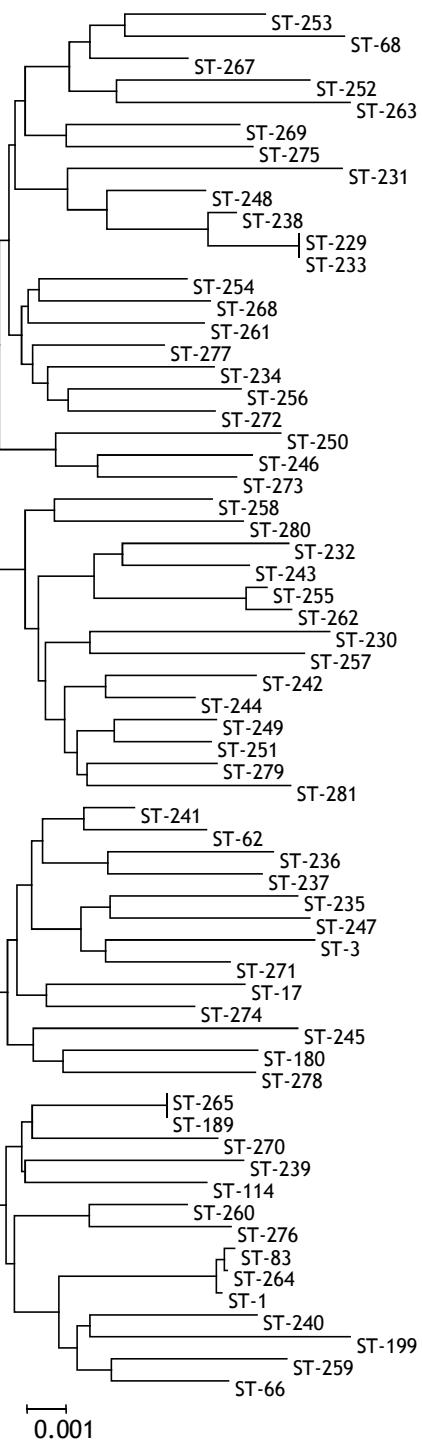
**Figure A7.** Neighbour-Joining tree of 63 STs representing the concatenated sequences of six housekeeping genes (with *recA* removed) of 102 *V. parahaemolyticus* isolates from multiple sources in Thailand. The branching pattern of this tree does not consist of the two main lineages A and B (as well as clades I and II) that are clearly differentiated in the phylogenetic tree of the concatenated sequences of all seven housekeeping genes (Fig. 2.43).
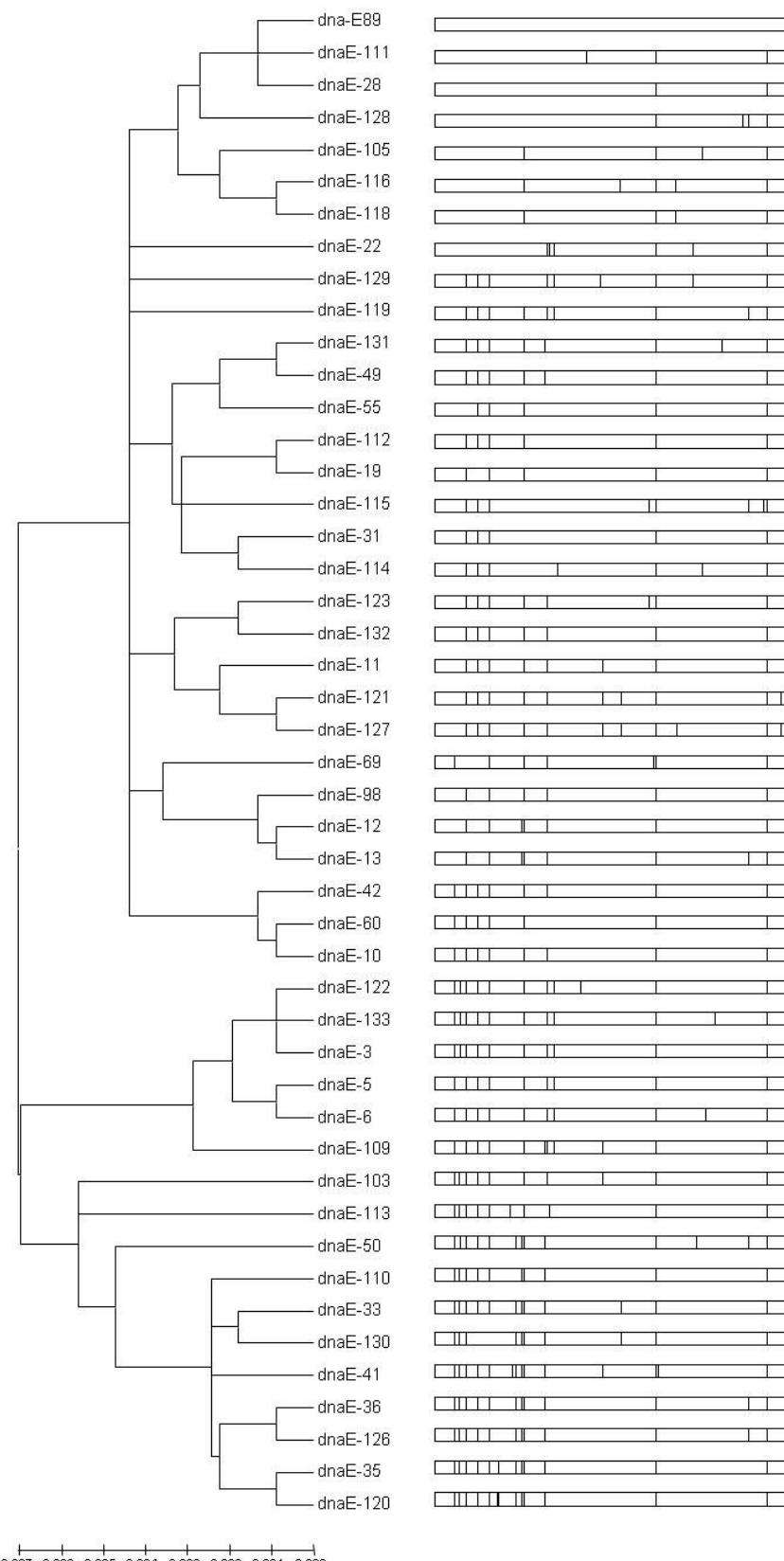
**Figure A8**. Neighbour-Joining tree and distribution of polymorphic nucleotide sites among 47 *dnaE* allele sequences. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, dnaE-89.
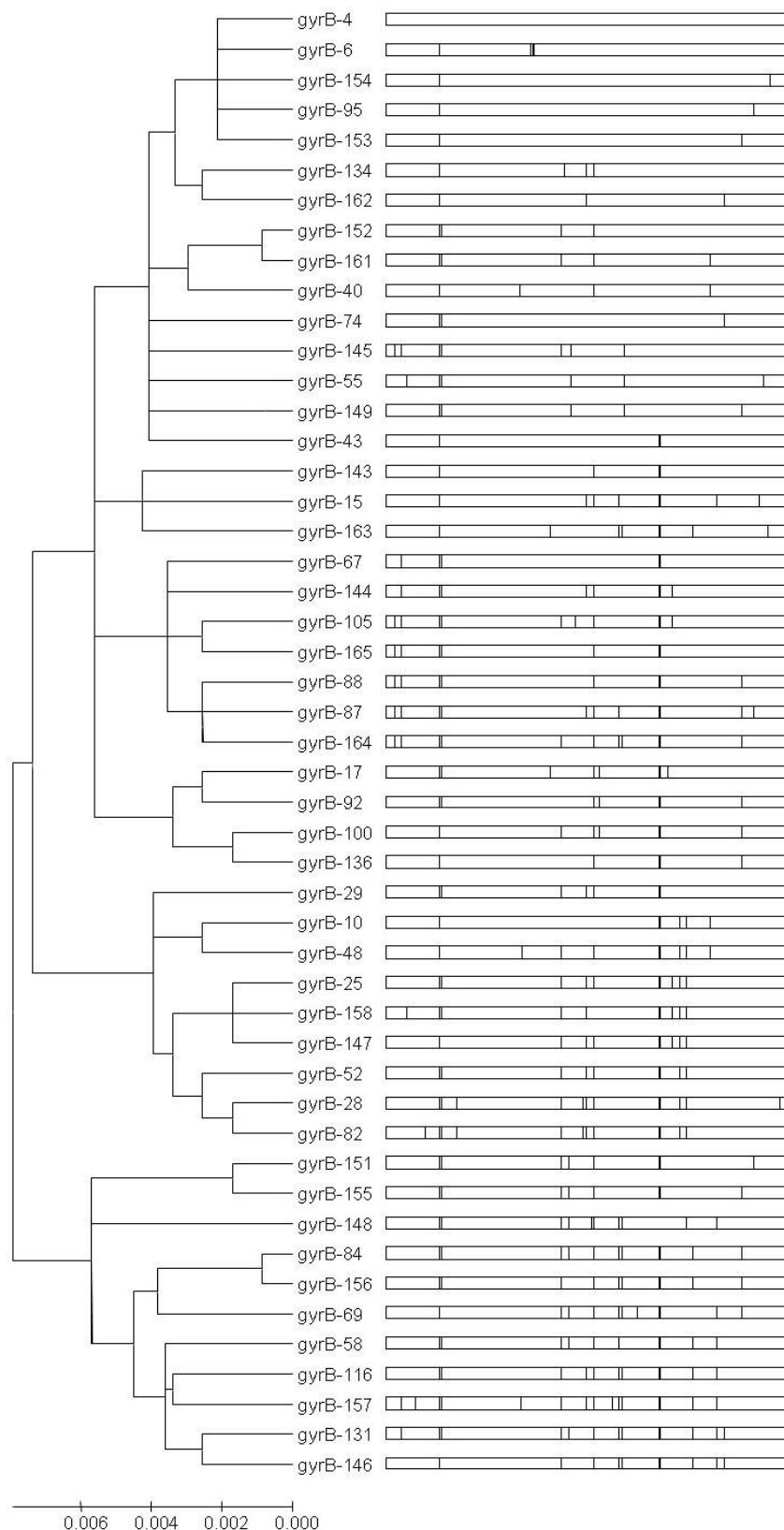
**Figure A9.** Neighbour-Joining tree and distribution of polymorphic nucleotide sites among 49 *gyrB* allele sequences. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, gyrB-4.
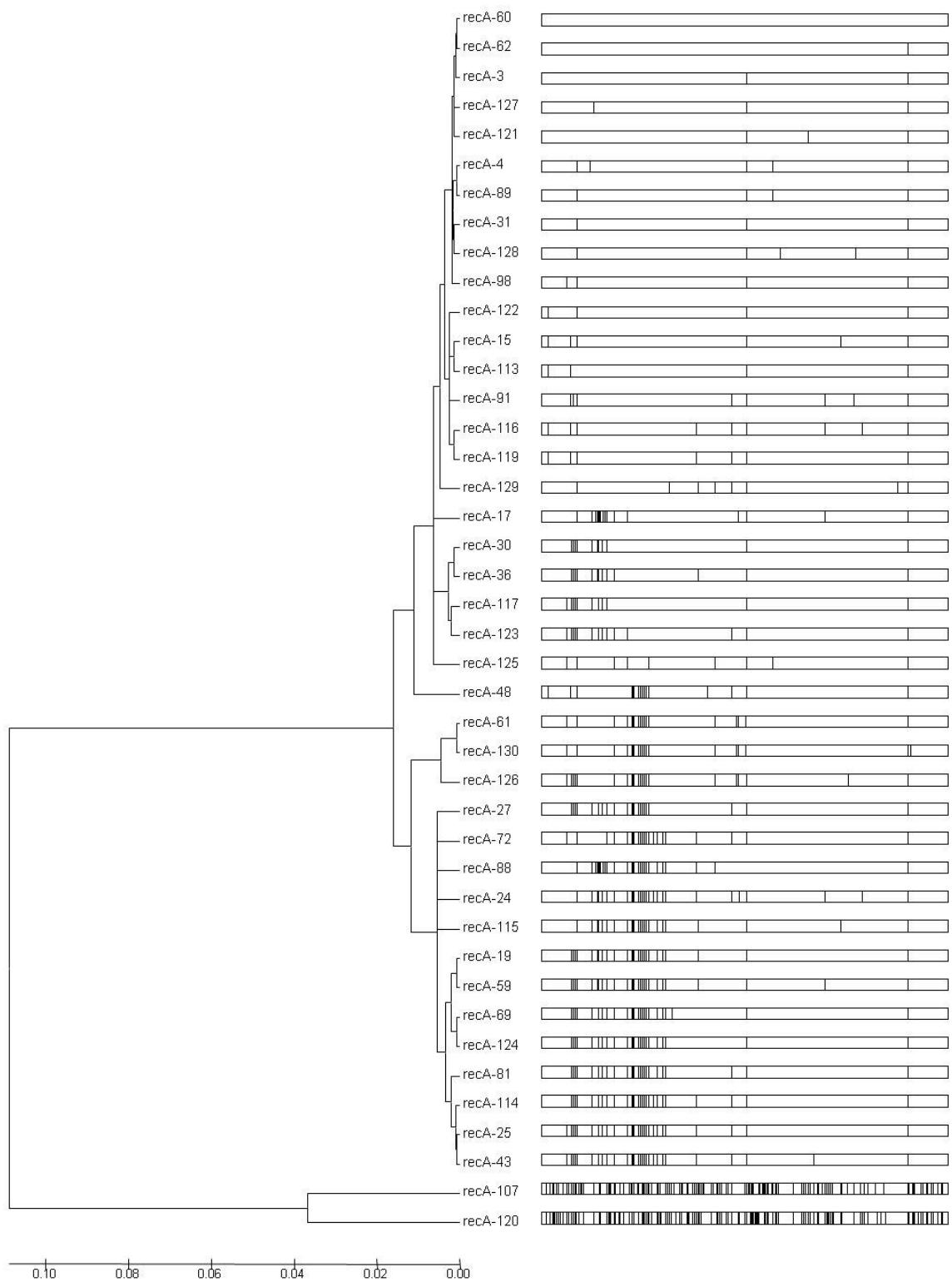
**Figure A10.** Neighbour-Joining tree and distribution of polymorphic nucleotide sites among 42 *recA* allele sequences. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, recA-60.
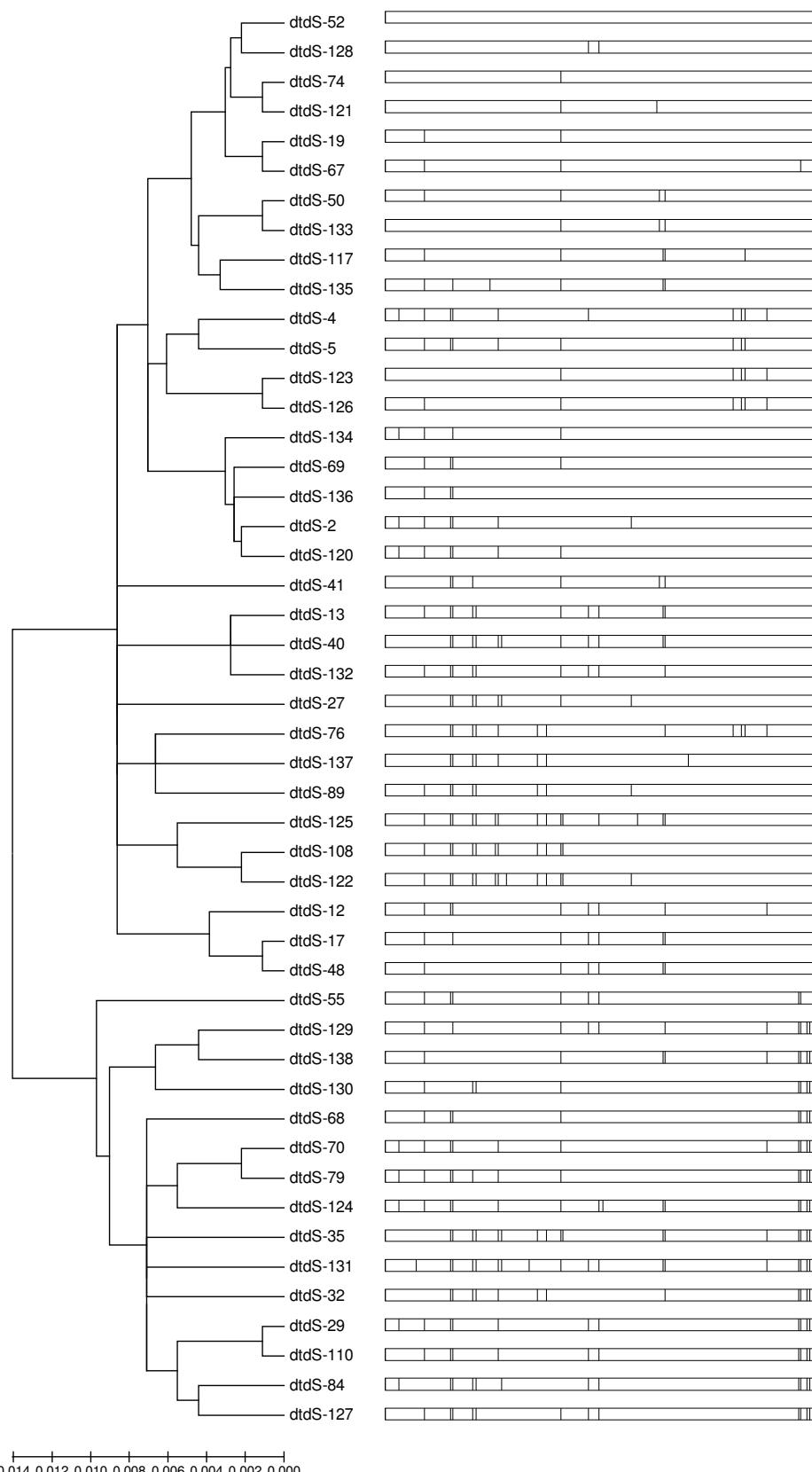
**Figure A11**. Neighbour-Joining tree and distribution of polymorphic nucleotide sites among 49 *dtdS* allele sequences. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, dtdS-52.
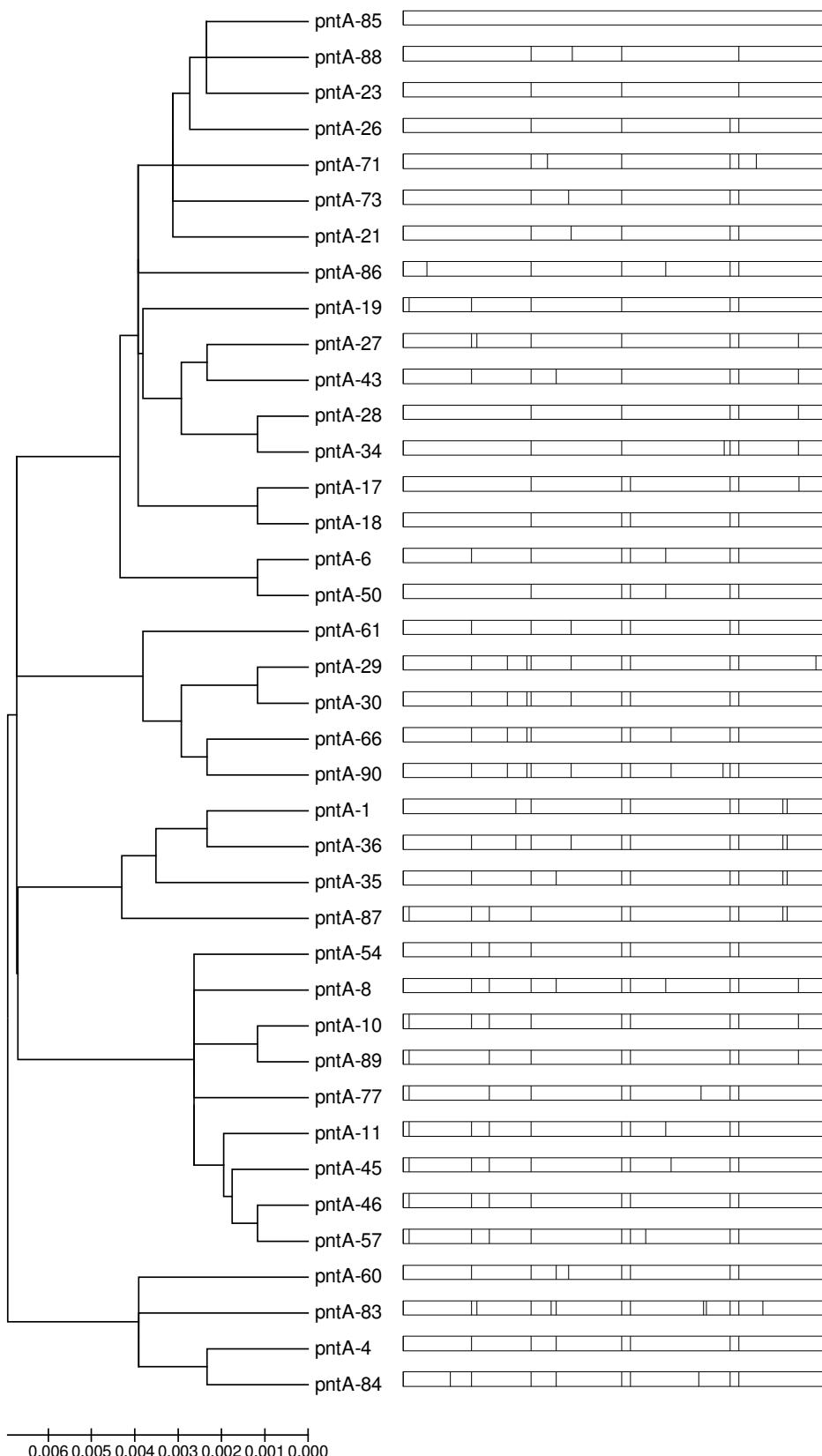
**Figure A12.** Neighbour-Joining tree and distribution of polymorphic nucleotide sites among 39 *pntA* allele sequences. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, pntA-85.
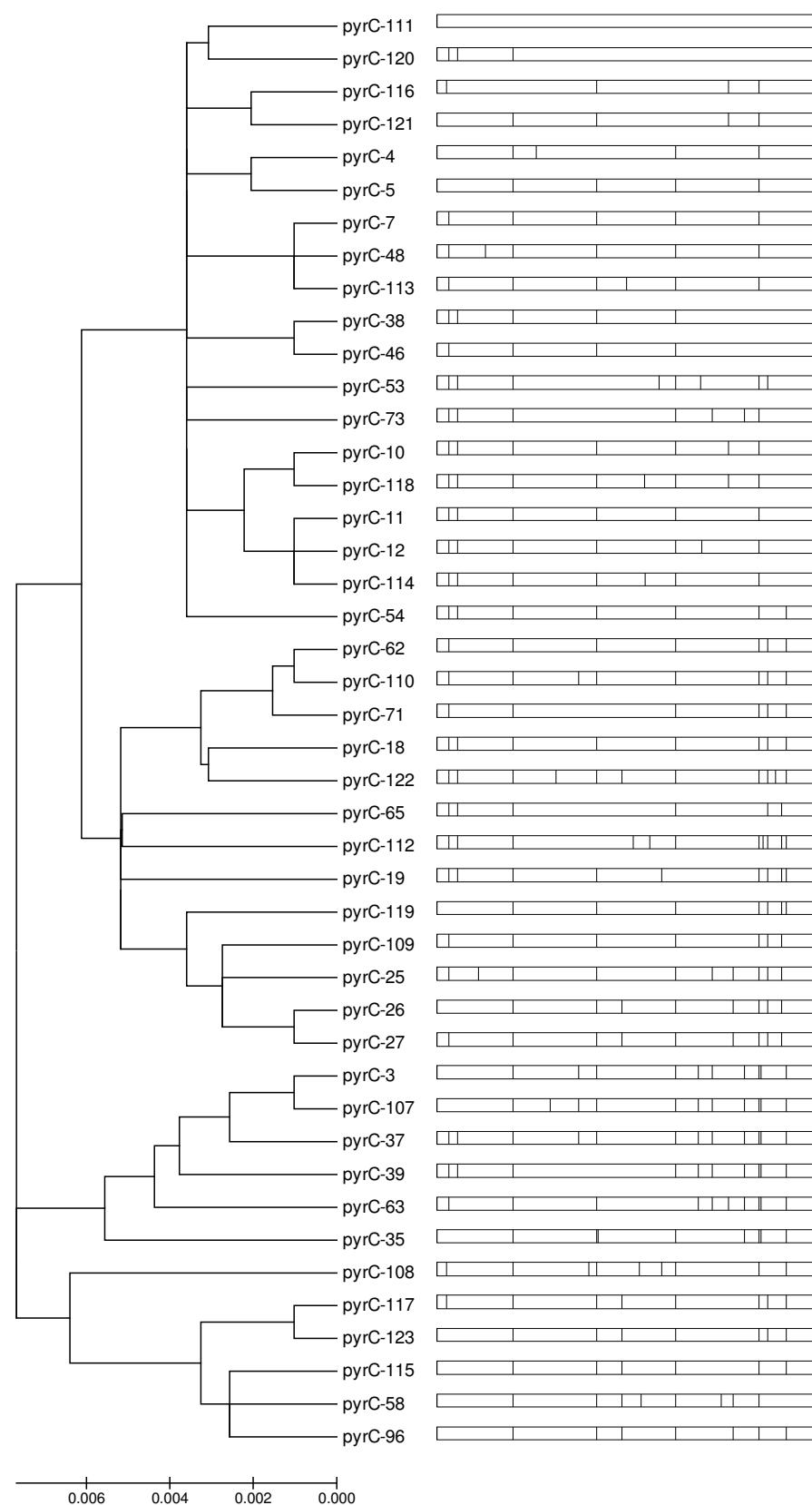
**Figure A13.** Neighbour-Joining tree and distribution of polymorphic nucleotide sites among 44 *pyrC* allele sequences. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, pyrC-111.
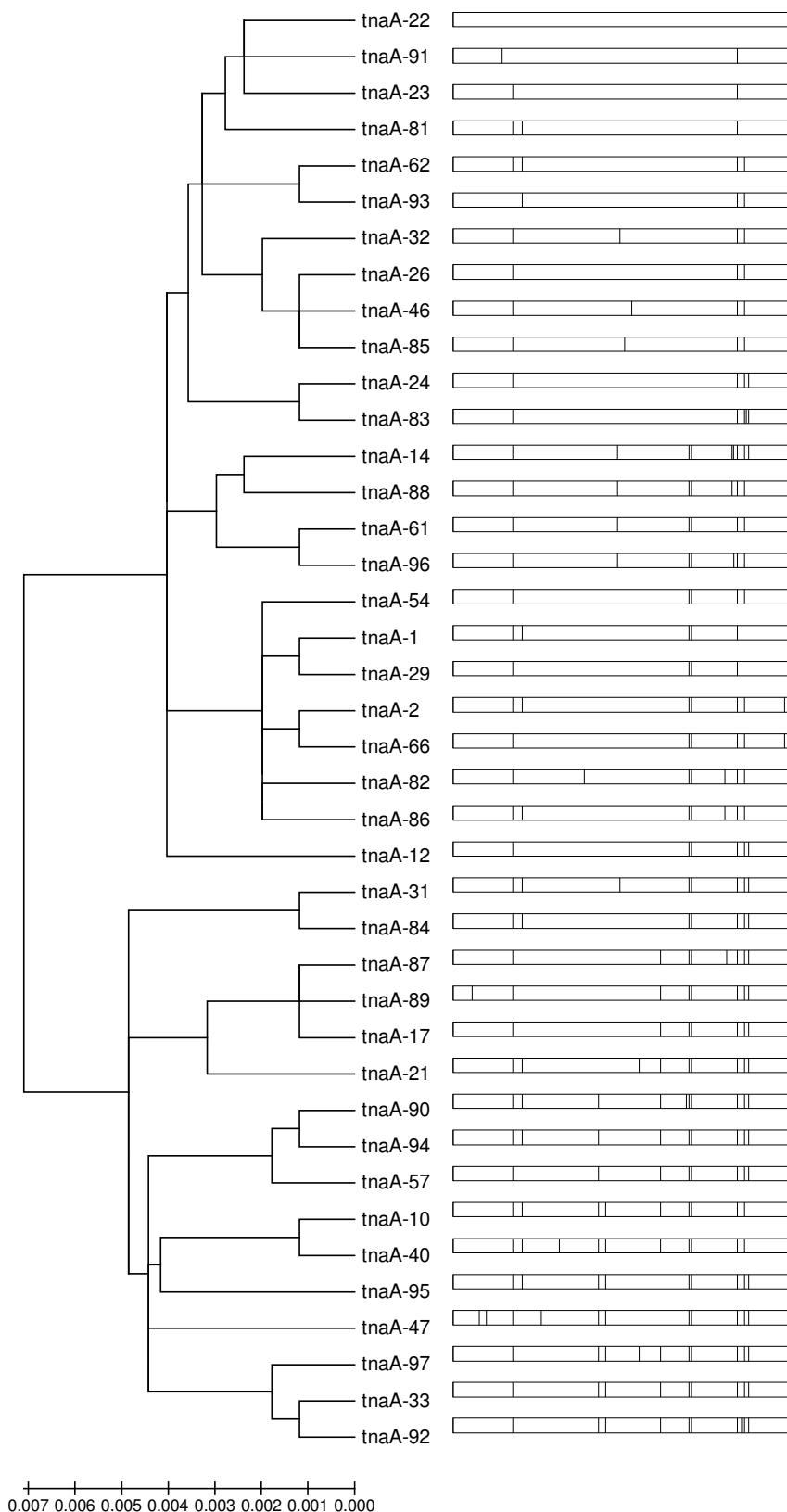
**Figure A14.** Neighbour-Joining tree and distribution of polymorphic nucleotide sites among 40 *tnaA* allele sequences. Vertical lines represent polymorphic nucleotide sites with respect to the top sequence, tnaA-22.

Population clustering of *V. parahaemolyticus* 61 STs
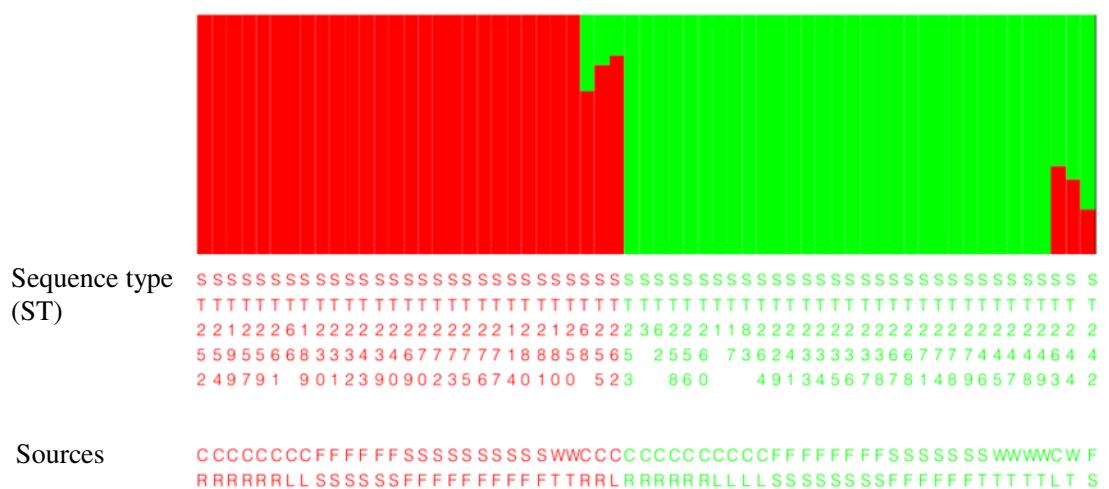


**Figure A15.** Bayesian clustering analysis inferred by BAPS for 61 STs indicate two distinct clusters (red and green). The colours represent different population clusters and the coloured segment is the fraction of genotype belonging to each cluster. The population structure was obtained using an admixture analysis model where K = 2. Each individual ST and sources of isolates are represented by a horizontal scale. CL represents clinical isolates, CR represents carrier isolates, SF represents seafood isolates, FS represents frozen shrimp isolates, ST represents shrimp tissue isolates, and WT represent water isolates.

**Figure A16.** Recombination events among 63 STs predicted by RDP3 analysis. Thirteen unique recombination events were identified from 175 recombination signals detected among 63 STs of *V. parahaemolyticus*. The figure shows the estimated recombination breakpoints in the nucleotide sequences of 63 STs. Predominant recombination events occuring at the *recA* locus are represented.
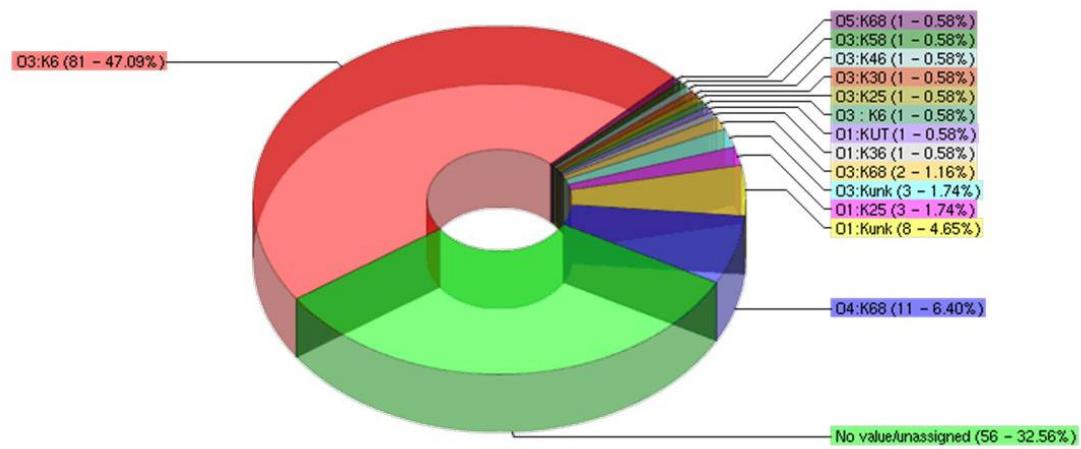
| Serotype | Number of strains | Percentage (%) |
|----------|-------------------|----------------|
| O3:K6 | 81 | 47.09% |
| Unassigned | 56 | 32.56% |
| O4:K68 | 11 | 6.40% |
| O1:KUT | 8 | 4.65% |
| O1:K25 | 3 | 1.74% |
| O3:KUT | 3 | 1.74% |
| O3:K68 | 2 | 1.16% |
| O1:K36 | 1 | 0.58% |
| O1:KUT | 1 | 0.58% |
| O3:K6 | 1 | 0.58% |
| O3:K25 | 1 | 0.58% |
| O3:K30 | 1 | 0.58% |
| O3:K46 | 1 | 0.58% |
| O3:K58 | 1 | 0.58% |
| O5:K68 | 1 | 0.58% |

**Figure A17.** Serotypic diversity of 172 *V. parahaemolyticus* isolates representing ST3 from MLST database (http://pubmlst.org/ vparahaemolyticus/)

**Table A2.** PCR conditions and primer pairs that were used to obtain housekeeping gene fragments

| Isolate | Housekeeping genes | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *dnaE* | | *gyrB* | | *recA* | | *dtdS* | | *pntA* | | *pyrC* | | *tnaA* | |
| | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T |
| VP2 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP4 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 56 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP6 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 56 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP8 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP10 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP12 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP14 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP16 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 56 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP18 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | - | - | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP20 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP22 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP24 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP26 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP28 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP30 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 56 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP32 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 56 | tnaAF3/R3 | 59 |
| VP34A | dnaE-F3/R1 | 60 | gyrBF1/R1 | 59 | - | - | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP34B | dnaE-F3/R1 | 60 | gyrBF1/R1 | 59 | - | - | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP36 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 56 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP38 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP40 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |

Table A2. (continued)

| Isolate | Housekeeping genes | | | | | | | | | | | | | |
|---------|-------------------|---|--------|---|--------|---|--------|---|--------|---|--------|---|--------|---|
| | *dnaE* | | *gyrB* | | *recA* | | *dtdS* | | *pntA* | | *pyrC* | | *tnaA* | |
| | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T |
| VP42 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | - | - | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP44 | dnaE-F3/R1 | 59 | gyrBF1/R2 | 56 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP46 | dnaE-F3/R1 | 59 | gyrBF1/R2 | 56 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP48 | dnaE-F3/R1 | 59 | gyrBF1/R2 | 56 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP50 | dnaE-F3/R1 | 59 | gyrBF1/R2 | 56 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP52 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP54 | dnaE-F3/R1 | 59 | gyrBF1/R2 | 56 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP56 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF3/R1 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP58 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP60 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP62 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 57 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 56 | tnaAF3/R3 | 59 |
| VP64 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP66 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP68 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 56 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP72 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP74 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP76 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP78 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | - | - | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP80 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP82 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | - | - | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP84 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 57 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP86 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 57 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |

Table A2. (continued)

| Isolate | Housekeeping genes | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *dnaE* | | *gyrB* | | *recA* | | *dtdS* | | *pntA* | | *pyrC* | | *tnaA* | |
| | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T |
| VP88 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 57 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP90 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 58 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP92A | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | - | - | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP92B | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | - | - | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP94 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 57 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP96 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 57 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP98 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 57 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP100 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP102 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP104 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 57 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP106 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 58 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP108 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP110 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP112 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP114 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP116 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP118 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP120 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP122 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 57 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP124 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 57 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP126 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP128A | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP128B | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |

Table A2. (continued)

| Isolate | Housekeeping genes | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | *dnaE* | | *gyrB* | | *recA* | | *dtdS* | | *pntA* | | *pyrC* | | *tnaA* | |
| | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T |
| VP130A | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 56 | tnaAF3/R3 | 59 |
| VP130B | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF3/R1 | 59 | pyrCF1/R1 | 56 | tnaAF3/R3 | 59 |
| VP132 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP134 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | - | - | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP136 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP138 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP140 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF3/R1 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP142 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 56 | pntAF3/R1 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP144A | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP144B | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP146 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP148 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP150 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP152 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP154A | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP154B | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP156 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP158 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP160 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP162 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP164 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP166 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |

Table A2. (continued)

| Isolate | Housekeeping genes | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *dnaE* | | *gyrB* | | *recA* | | *dtdS* | | *pntA* | | *pyrC* | | *tnaA* | |
| | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T |
| VP168 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP170 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP172A | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP172B | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP174 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP176 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP178 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP180 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP182 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP184 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP186 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP188 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP190 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF3/R2 | 57 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP192 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP194 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP196 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP198 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP200 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP202 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP204 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP206A | dnaE-F3/R1 | 60 | gyrBF1/R2 | 56 | recAF1/R3 | 59 | dtdSF3/R3 | 56 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP206B | dnaE-F3/R1 | 60 | gyrBF1/R2 | 56 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP208 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |

Table A2. (continued)

| Isolate | Housekeeping genes | | | | | | | | | | | | | |
|---------|-------|----|-------|----|-------|----|-------|----|-------|----|-------|----|-------|----|
| | *dnaE* | | *gyrB* | | *recA* | | *dtdS* | | *pntA* | | *pyrC* | | *tnaA* | |
| | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T | Primer | T |
| VP210 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP212 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP214 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP216 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP218 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 55 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP220 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP222 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP224 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP226 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 56 | tnaAF3/R3 | 59 |
| VP228 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP230 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP232 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP234 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF3/R3 | 56 | pntAF1/R3 | 59 | pyrCF1/R1 | 56 | tnaAF3/R3 | 59 |
| VP236 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 56 | tnaAF3/R3 | 59 |
| VP238 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP240 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| VP242 | dnaE-F3/R1 | 59 | gyrBF1/R1 | 59 | recAF1/R3 | 59 | dtdSF1/R1 | 59 | pntAF1/R3 | 59 | pyrCF1/R1 | 59 | tnaAF3/R3 | 59 |
| Total | 128 | | 128 | | 121 | | 128 | | 126 | | 128 | | 128 | |

T = annealing temperature used in PCR reactions

Grey shade highlight represents PCR conditions obtained from further optimizations of each isolate.

**Table A3.** Amino acid sequence types (aaSTs) correspond to nucleotide sequence types (mlstSTs) of *V. parahaemolyticus* 348 STs from MLST database (http://pubmlst.org/vparahaemolyticus/)

| aaST | mlstSTs |
|---|---|
| 1 | 19, **269**, 299 |
| 2 | **1,**4,5,9,10,13,16,**17**,18,22,23,24,25,26,87,98,102,109,110,122,130,132,133,153,157,158,167,170,183,217,219,227,228,**234,239,241,243,248,249,254,267**,**271**,293,300,309,322,330,348 |
| 3 | 191 |
| 4 | 209, **277**, 321 |
| 5 | 82,**83,264** |
| 6 | 112 |
| 7 | 2,**3**,21,27,71,72,159,176,192,196,201,211,220,**235**,266,295,305,343 |
| 8 | **66** |
| 9 | **68** |
| 10 | 146, 151, 155, 203, **260** |
| 11 | 63,140,208,323 |
| 12 | 6,7 |
| 13 | 14 |
| 14 | **256** |
| 15 | **258** |
| 16 | 288 |
| 17 | 12,20,74,76,85,113,116,121,124,166,172,202,210,301,310 |
| 18 | 311 |
| 19 | 86 |
| 20 | 64,328,332 |
| 21 | 126 |
| 22 | 15, **255** |
| 23 | 73 |
| 24 | 28,135 |
| 25 | 148,152 |
| 26 | 342 |
| 27 | 165 |
| 28 | 163 |
| 29 | 317 |
| 30 | 88, **189, 265,** 345 |
| 31 | 282 |
| 32 | 308 |

Table A3. (continued)

| aaST | mlstSTs |
|---|---|
| 33 | **229,233** |
| 34 | 8,30,31,37,46,47,49,50,53,55,56,57,60,**62**,67,70,79,80,84,91,92,97,100,101,104,105,106,115,117,119,120,123,125,131,134,139,141,142,147,156,161,164,171,173,174,178,**180,**188,194,195,198,200,207,213,214,215,216,218,223,224,226,**242,244,245,246,247,251,253,259,263,270,273,275,280**,289,291,304,312,315,316,319,325,326,327,329,335,336,341 |
| 35 | 48, 225, **268** |
| 36 | 75,78,181,284,340 |
| 37 | 205 |
| 38 | **232** |
| 39 | 36,38,39,40,41,42,51,52,59,107,111,129,137,138,187,193,**199,237** |
| 40 | 54, **230** |
| 41 | 96,143,185,283,287,200,000 |
| 42 | 169 |
| 43 | 128 |
| 44 | **238** |
| 45 | 90,108,168,318 |
| 46 | 11,221 |
| 47 | **257** |
| 48 | 204 |
| 49 | **250** |
| 50 | 285 |
| 51 | 297 |
| 52 | 29,99,136,145,154,179,190,**240,272,276**,337 |
| 53 | 94, 162, 197, **278,** 320 |
| 54 | 286 |
| 55 | 32,33,34,77,95,306,307,324,334 |
| 56 | 127, 186, 206, **236, 262** |
| 57 | **261** |
| 58 | 89 |
| 59 | 150 |
| 60 | 43,44,81,314,346 |
| 61 | 313 |
| 62 | 61 |
| 63 | 149 |
| 64 | 294 |
| 65 | 339 |

Table A3. (continued)

| aaST | mlstSTs |
|------|---------|
| 66 | 45 |
| 67 | 93,144,177,296,333,300 |
| 68 | 69,222 |
| 69 | 290,298 |
| 70 | 58 |
| 71 | 184 |
| 72 | 103 |
| 73 | 160 |
| 74 | 182 |
| 75 | 35 |
| 76 | **252** |
| 77 | 65 |
| 78 | 118 |
| 79 | 303 |
| 80 | **114** |
| 81 | 212, **274** |
| 82 | 175 |
| 83 | **231** |
| 84 | **281** |
| 85 | **279** |
| 86 | 331,344 |
| 87 | 338 |

*Numbers in bold represent STs that were found in Thai *V. parahaemolyticus* isolates in the present study.

**Table A4**. Identified non-OMPs from proteomic analysis of eight representative *V. parahaemolyticus* isolates

| No. | Gene name | Protein name | Subcellular location |
| --- | --- | --- | --- |
| 1 | *rpsB* | 30S ribosomal protein S2 | Cytoplasm |
| 2 | *rpsC* | 30S ribosomal protein S3 | Cytoplasm |
| 3 | *rpsD* | 30S ribosomal protein S4 | Cytoplasm |
| 4 | *rpsE* | 30S ribosomal protein S5 | Cytoplasm |
| 5 | *rpsF* | 30S ribosomal protein S6 | Cytoplasm |
| 6 | *rpsG* | 30S ribosomal protein S7 | Cytoplasm |
| 7 | *rpsH* | 30S ribosomal protein S8 | Cytoplasm |
| 8 | *rpsK* | 30S ribosomal protein S11 | Cytoplasm |
| 9 | *rpsL* | 30S ribosomal protein S12 | Cytoplasm |
| 10 | *rplA* | 50S ribosomal protein L1 | Cytoplasm |
| 11 | *rplB* | 50S ribosomal protein L2 | Cytoplasm |
| 12 | *rplD* | 50S ribosomal protein L4 | Cytoplasm |
| 13 | *rplE* | 50S ribosomal protein L5 | Cytoplasm |
| 14 | *rplF* | 50S ribosomal protein L6 | Cytoplasm |
| 15 | *rplM* | 50S ribosomal protein L13 | Cytoplasm |
| 16 | *rplO* | 50S ribosomal protein L15 | Cytoplasm |
| 17 | *rplP* | 50S ribosomal protein L16 | Cytoplasm |
| 18 | *rplQ* | 50S ribosomal protein L17 | Cytoplasm |
| 19 | *rplR* | 50S ribosomal protein L18 | Cytoplasm |
| 20 | *rplT* | 50S ribosomal protein L20 | Cytoplasm |
| 21 | *htpG* | Chaperone protein htpG | Cytoplasm |
| 22 | *VP0797* | Cysteine synthase | Cytoplasm |
| 23 | *tufA* | Elongation factor Tu | Cytoplasm |
| 24 | *VP0994* | Formate acetyltransferase | Cytoplasm |
| 25 | *cadA* | Lysine decarboxylase | Cytoplasm |
| 26 | *VP0404* | RNA polymerase sigma factor | Cytoplasm |
| 27 | *glyA1* | Serine hydroxymethyltransferase 1 | Cytoplasm |
| 28 | *VP1543* | Cytochrome c oxidase, subunit CcoO | Inner membrane |
| 29 | *VP1541* | Cytochrome c oxidase, subunit CcoP | Inner membrane |
| 30 | *VP1053* | Cytochrome d ubiquinol oxidase, subunit I | Inner membrane |
| 31 | *VP1054* | Cytochrome d ubiquinol oxidase, subunit II | Inner membrane |
| 32 | *VPA0628* | Cytochrome o ubiquinol oxidase, subunit I | Inner membrane |

Table A4. (continued)

| No. | Gene name | Protein name | Subcellular location |
|-----|-----------|--------------|---------------------|
| 33 | *VPA0627* | Cytochrome o ubiquinol oxidase, subunit II | Inner membrane |
| 34 | *atpB* | ATP synthase subunit a | Inner membrane |
| 35 | *atpA* | ATP synthase subunit alpha | Inner membrane |
| 36 | *atpF* | ATP synthase subunit b | Inner membrane |
| 37 | *VPA0297* | PTS system, fructose-specific IIBC component | Inner membrane |
| 38 | *VP2046* | PTS system, glucose-specific IIBC component | Inner membrane |
| 39 | *VP0831* | PTS system, N-acetylglucosamine-specific IIABC component | Inner membrane |
| 40 | *VP2662* | Putative ABC superfamily transport protein | Inner membrane |
| 41 | *VPA1642* | Putative CymC protein | Inner membrane |
| 42 | *VPA1726* | Putative proline dehydrogenase | Inner membrane |
| 43 | *VPA0922* | NAD(P) transhydrogenase, alpha subunit | Inner membrane |
| 44 | *crp* | Cyclic AMP receptor protein | Inner membrane |
| 45 | *rnfE* | Electron transport complex protein rnfE | Inner membrane |
| 46 | *oxaA* | Inner membrane protein oxaA | Inner membrane |
| 47 | *nqrA* | Na(+)-translocating NADH-quinone reductase subunit A | Inner membrane |
| 48 | *VP2437* | NupC family protein | Inner membrane |
| 49 | *VP0356* | Pyruvate kinase | Inner membrane |
| 50 | *VP0503* | Sodium/alanine symporter | Inner membrane |
| 51 | *VP1302* | Sodium/dicarboxylate symporter | Inner membrane |
| 52 | *VP2032* | SohB protein, peptidase U7 family | Inner membrane |
| 53 | *VP0843* | Succinate dehydrogenase, cytochrome b556 subunit | Inner membrane |
| 54 | *VP0443* | Ubiquinol-cytochrome c reductase, cytochrome c1 | Inner membrane |
| 55 | *VPA1643* | Putative maltose operon periplasmic protein | Periplasm |
| 56 | *VP3056** | Putative periplasmic protein | Periplasm |
| 57 | *VP1624* | Putative TEGT family carrier/transport protein | Periplasm |
| 58 | *VP2091** | Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein | Periplasm |

Table A4. (continued)

| No. | Gene name | Protein name | Subcellular location |
|-----|-----------|--------------|---------------------|
| 59 | *VP2385* | Glycerol uptake facilitator protein GlpF | Periplasm |
| 60 | *VP2724\** | Putative uncharacterized protein VP2724 | Periplasm |
| 61 | *ribH* | 6,7-dimethyl-8-ribityllumazine synthase | Unknown |
| 62 | *VP2990* | Putative uroporphyrin-III C-methyltransferase | Unknown |
| 63 | *VP2778* | Peptidyl-prolyl cis-trans isomerase | Unknown |
| 64 | *VPA0545* | Peptidyl-prolyl cis-trans isomerase | Unknown |
| 65 | *VP2121* | Alcohol dehydrogenase/acetaldehyde dehydrogenase | Unknown |
| 66 | *VP2618* | FkuB | Unknown |
| 67 | *VP2540* | Protease, insulinase family | Unknown |
| 68 | *VP2534* | Signal recognition particle protein | Unknown |
| 69 | *murA* | UDP-N-acetylglucosamine 1-carboxyvinyltransferase | Unknown |
| 70 | *VP0320* | Putative uncharacterized protein VP0320 | Unknown |
| 71 | *YajC* | Putative uncharacterized protein VP0589 | Unknown |
| 72 | *VP0672* | Putative uncharacterized protein VP0672 | Unknown |
| 73 | *VP1243* | Putative uncharacterized protein VP1243 | Unknown |
| 74 | *VP1326* | Putative uncharacterized protein VP1326 | Unknown |
| 75 | *VP1594* | Putative uncharacterized protein VP1594 | Unknown |
| 76 | *VP1789* | Putative uncharacterized protein VP1789 | Unknown |

\*Periplasm proteins that are predicted to be outer membrane proteins of

*V. parahaemolyticus* isolate RIMD2210633 by bioinformatics analysis