Genomic characterization of *Vibrio parahaemolyticus* from Pacific white shrimp and rearing water in Malaysia reveals novel sequence types and structural variation in genomic regions containing the *Photorhabdus* insect-related (Pir) toxin-like genes

Chrystine Zou Yi Yan[1,2] , Christopher M. Austin[1,2,3,4] , Qasim Ayub[1,2], Sadequr Rahman[1,2] and Han Ming Gan[1,2,3,4]*

[1] School of Science, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia.

[2] Monash University Malaysia Genomics Facility, Tropical Medicine and Biology Multidisciplinary Platform, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia.

[3] Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong, Victoria, Australia.

[4] Deakin Genomics Centre, Deakin University, Geelong, Victoria, Australia.

*Corresponding author

Name: Han Ming Gan

Email: han.gan@deakin.edu.au

Keywords: multi-locus sequence typing, shrimp, whole-genome sequencing, phylogenomics, comparative genomics

Abstract

The Malaysian and global shrimp aquaculture production has been significantly impacted by acute hepatopancreatic necrosis disease (AHPND) typically caused by *Vibrio parahaemolyticus* harboring the pVA plasmid containing the $pirA^{Vp}$ and $pirB^{Vp}$ genes which code for *Photorhabdus* insect-related (Pir) toxin. The limited genomic resource for *V. parahaemolyticus* strains from Malaysian aquaculture farms precludes an in-depth understanding of their diversity and evolutionary relationships. In this study, we isolated shrimp-associated and environmental (rearing water) *V. parahaemolyticus* from three aquaculture farms located in Northern and Central Malaysia followed by whole-genome sequencing of 40 randomly selected isolates on the Illumina MiSeq. Phylogenomic analysis and multilocus sequence typing (MLST) reveal distinct lineages of *V. parahaemolyticus* that harbor the $pirAB^{Vp}$ genes. The recovery of pVA plasmid backbone devoid of $pirA^{Vp}$ or $pirAB^{Vp}$ in some *V. parahaemolyticus* isolates suggests that the toxin genes are prone to deletion. The new insight gained from phylogenomic analysis of Asian *V. parahaemolyticus*, in addition to the observed genomic instability of pVa plasmid, will have implications for improvements in aquaculture practices to diagnose, treat or limit the impacts of this disease.

Introduction

Acute hepatopancreatic necrosis disease (AHPND), also known as early mortality syndrome, is an emerging shrimp disease that has caused more than 60% reduction in shrimp (*Litopenaeus vannamei*) production worldwide, with global economic losses estimated to be up to USD $1 billion annually (FAO, 2013). Initially reported in China in 2009, this disease continued to spread to other parts of Asia including Vietnam, Thailand, and Malaysia (Zorriehzahra & Banaederakhshan, 2015). To date, AHPND is still prevalent in Malaysia despite a lower frequency of occurrence in recent years (Chu *et al.*, 2016).

The causative agent of AHPND was successfully isolated through laboratory assays that fulfill the Koch's postulates and was subsequently identified as a specific strain of *V. parahaemolyticus* (Tran *et al.*, 2013). Comparative genomics of AHPND-causing and avirulent *V. parahaemolyticus* strains identified the presence of a 70-kb pVa plasmid containing the $pirAB^{Vp}$ genes in the virulent strain that code for binary toxins with structural similarity to the *Bacillus* Cry insecticidal toxin-like proteins (Lee *et al.*, 2015). This binary toxin induces cell

death through pore formation in shrimp cell membranes leading to tissue degradation and digestive organ dysfunction (Lai *et al.*, 2015, Lee *et al.*, 2015, Han *et al.*, 2016). Current genomic sampling suggests that most of AHPND-causing *Vibrio* spp. belong to the species *V. parahaemolyticus* that cluster within the *Vibrio harveyi* clade (Tran *et al.*, 2013). However, in recent years, the pVa plasmid has also been isolated from other *Vibrio* species thus indicating the mobility and compatibility of the virulent plasmid among members of the *Vibrio* genus (Kondo *et al.*, 2015, Han *et al.*, 2016, Dong *et al.*, 2017, Xiao *et al.*, 2017, Liu *et al.*, 2018).

The increasing genomic representation of *Vibrio* sequences in public database enables a more comprehensive analysis and understanding of the evolutionary relationships among different species and strains (Osama *et al.*, 2012, Jun *et al.*, 2013, Gomez-gil *et al.*, 2014, Gomez-Jimenez *et al.*, 2014, Kondo *et al.*, 2014, Yang *et al.*, 2014, Castillo *et al.*, 2015, Yang *et al.*, 2015). The generation of high-quality draft genomes also allows the extraction of full-length house-keeping genes used for multilocus sequencing typing (MLST), precluding the laborious steps associated with amplification and sequencing of multiple PCR fragments (Osama *et al.*, 2012, Joensen *et al.*, 2014, Kavousi *et al.*, 2016). Successful MLST of the *V. parahaemolyticus* genomes has been recently demonstrated in a comparative genomic analysis of 93 publicly available *V. parahaemolyticus* genomes and 9 newly sequenced Chinese *V. parahaemolyticus* isolates (Fu *et al.*, 2017). A recent study identified substantial genomic diversity between outbreak countries and indicated at least 12 independent AHPND-related genomic clones based on whole-genome SNP analysis (Fu *et al.*, 2017). Despite the comprehensive genomic sampling, nearly half of the strains analyzed are not associated with AHPND and of the 9 newly sequenced Chinese isolates, only 2 strains were associated with AHPND (Fu *et al.*, 2017). Samples size calculations based on Bayesian method of inference recommended that 2 to 20 isolates per sample source should be characterized to obtain maximal information on strain heterogeneity, which is crucial for accurate prediction of outbreak sources and transmission pathways (Döpfer *et al.*, 2008). Therefore, it remains unclear the degree to which insufficient sampling may have affected the accuracy of AHPND epidemiological predictions. To date, such a sample size requirement remain difficult to fulfill in genome-based epidemiology of aquaculture-associated pathogens despite the decreasing cost associated with microbial genome sequencing (Land *et al.*, 2015, Tang *et al.*, 2017).

Genomic resources for bacterial isolates from Malaysia are scarce and were generally presented as individual genome report without a strong focus on comparative genomics (Hong *et al.*, 2012, Yap *et al.*, 2012, Gan *et al.*, 2017). For example, in 2017, only five genomes of Malaysian *V. parahaemolyticus* associated with shrimp aquaculture environment were reported (Foo *et al.*, 2017) followed closely by only two additional pVa-harbouring *V. parahaemolyticus* genomes a year later, each presented as a single genome report (Devadas *et al.*, 2018, Devadas *et al.*, 2018). Such a small dataset ($n < 10$) limits the inference of the genomic diversity and evolutionary relationships of *V. parahaemolyticus* in Malaysia. In this study, we attempt to overcome this shortcoming by sequencing an additional 40 *V. parahaemolyticus* isolates from deceased shrimps and rearing water collected from three Malaysian shrimp aquaculture farms. From the *de-novo* assembled genomes, we characterized their genetic diversity based on *in-silico* MLST and determined their evolutionary relationships with other Asian *V. parahaemolyticus* isolates. The genetic and structural analysis of the pVa plasmids was also investigated, providing new insights into their prevalence and diversity among Asian *V. parahaemolyticus* strains.

Materials and methods

Field Sampling

Sampling was performed in three Malaysian shrimp aquaculture farms with suspected AHPND outbreaks. Two of these were located in Negeri Sembilan, Central Peninsular Malaysia and one in Terengganu, Northern Peninsular Malaysia (Figure 1). In all three aquaculture farms, pond rearing water was collected for strain isolation. Two deceased shrimps exhibiting the symptoms of AHPND such as empty stomach, pale hepatopancreas, and empty midgut were also provided by the farm manager of the Terengganu (Northern Malaysia) aquaculture farm (Table 1). At the request of the aquaculture farm managers, the exact sampling locations were not disclosed in this study to protect the identity of the farms.

Strain isolation and DNA extraction

Enrichment of *Vibrio* spp. from homogenized shrimp hepatopancreas and pond water was performed on thiosulfate-citrate-bile salts-sucrose (TCBS) agar (Oxoid Ltd, UK). We chose the hepatopancreas as the isolation source for shrimp as it is usually targeted by AHPND *V. parahaemolyticus* as previously determined based on PCR and histopathology (Khimmakthong &

Sukkarun, 2017). After incubation overnight at 30°C, blue-green colonies were isolated and sub-cultured on a fresh batch of TCBS agar followed by similar incubation setting, to verify their identity as *V. parahaemolyticus*. The confirmed blue-green colonies were sub-cultured onto nutrient agar (Merck, Kenilworth, NJ, USA) supplemented with 3% (w/v) NaCl. These culture plates were incubated overnight at 30°C and DNA from pure single colonies were extracted using a modified salting-out method (Sokolov, 2000).

Whole-genome sequencing

DNA from each isolate was normalized to 0.2 ng/uL based on Qubit2 measurement (Thermo Fisher Scientific, Waltham, MA, USA) and used to prepare NexteraXT sequencing library (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Each constructed library contains a unique combination of Nextera dual-index barcodes that allow pooling and sequencing on the same MiSeq flowcell. Equal molarity of each barcoded library was pooled followed by preparation for sequencing on the Illumina MiSeq sequencer located at Monash University Malaysia (2×250 bp paired-end read setting).

***De novo* assembly and *in-silico* taxonomic validation**

Raw MiSeq read files were filtered for adapter sequences and low-quality base calls using Trimmomatic version 0.36 (Bolger *et al.*, 2014) with the parameters PE, ILLUMINACLIP:Trimmomatic/adapters/NexteraPE.fa:2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:20, MINLEN:20. The trimmed reads were then assembled into contigs using SPAdes version 3.10.0 (Bankevich *et al.*, 2012) with the default parameters. The assembled contigs were subsequently scaffolded and gap-closed using SSPACE version 3.0 (Boetzer *et al.*, 2011) and GapFiller v1.10 (Boetzer *et al.*, 2012) respectively. QUAST- version 4.4 (Gurevich *et al.*, 2013) was used to assess the assembly using *V. parahaemolyticus* ATCC 17802^{T} as a reference genome. To confirm species identity, JSpecies version 1.2.1 (Richter *et al.*, 2015) was used to calculate the pairwise average nucleotide identity (ANI) of the assembled genomes with *V. parahaemolyticus* ATCC 17802^{T} as the positive control and *Vibrio cholera* ATCC 14035^{T} as the negative control.

Genome annotation and multi-locus sequence typing

Identification of pVa plasmid, *pirA*^{Vp} and *pirB*^{Vp} genes was performed using local BLASTN version 2.7.1+. Gene prediction and annotation was performed in NCBI and *in-silico* MLST was based on seven housekeeping genes in the MLST database (https://pubmlst.org/vparahaemolyticus/: *dnaE, dtdS, gyrB, pyrC, pntA, recA*, and *tnaA*) using three approaches e.g. local BLASTN version 2.7.1+, SRST-2 version 0.2.0, and MLST 1.8 (Camacho *et al.*, 2009, Larsen *et al.*, 2012, Inouye *et al.*, 2014).

Whole-genome phylogeny

Publicly available *V. parahaemolyticus* genomes were downloaded from NCBI Assembly in October 2018. Genomes with poor assembly metrics (> 400 contigs), inconsistent sample origin or erroneous taxonomic assignment were removed (Supplementary Table 1). For the identification of pan-genome, core-genome and accessory genome from the filtered and forty newly assembled *V. parahaemolyticus* genomes (Table 1), we used Roary version 3.12.0 (Page *et al.*, 2015) and decreased the minimum BLASTP percentage identity from 95% to 90%. In addition, the *V. cholerae* ATCC 14035^{T} genome was also included in the Roary analysis as the outgroup. A maximum-likelihood tree was then constructed based on the alignment of the core-genomes using FastTree version 2.1.8 (Price *et al.*, 2010) with generalized time-reversible (GTR) model. Visualization of the reconstructed maximum likelihood tree used FigTree v1.4.3.

Plasmid and gene neighborhood visualization

The presence of pVA among the *V. parahaemolyticus* genome assemblies was assessed based on BLASTN search (e-value of 1e-5) against the reference pVa from *V. parahaemolyticus* 13-028/A3 (GenBank Accession Number: NC_025152). Nucleotide alignment and visualization of the selected pVa-harbouring *V. parahaemolyticus* genomes against reference pVa used BRIG version 0.95 (Alikhan *et al.*, 2011). Visualization of the gene neighborhood surrounding the *pirB*^{Vp} among *pirA*$^{Vp-}$ *pirB*$^{Vp+}$ isolates used EasyFig version 2.1 with BLASTN search (e-value of 0.001, minimum length coverage of 100 bp) (Sullivan *et al.*, 2011).

Results

Genomic characterization and *in-silico* taxonomic validation of sequenced strains

We sequenced and assembled forty whole-genomes of *V. parahaemolyticus* isolates from three farms in Peninsular Malaysia using the Illumina MiSeq platform. The assembled whole-genomes of these *V. parahaemolyticus* strains showed an average cumulative length of 5 Mb and 45% GC content, consistent with the assembly statistics of most published genomes of *V. parahaemolyticus* in the NCBI database (Supplementary Table 1) (Jun *et al.*, 2013, Gomez-gil *et al.*, 2014, Gomez-Jimenez *et al.*, 2014, Kondo *et al.*, 2014, Yang *et al.*, 2014, Castillo *et al.*, 2015, Kondo *et al.*, 2015, Yang *et al.*, 2015). All newly sequenced isolates exhibited more than 95% pairwise average nucleotide identity (ANI) to *V. parahaemolyticus* ATCC 17802 (Supplementary Figure 1).

High genomic diversity of *Vibrio parahaemolyticus* associated with Malaysian aquaculture farms

Phylogenomic clustering of isolates corresponds broadly with their MLST profiles instead of their country of isolation. The 40 Malaysian *V. parahaemolyticus* isolates sequenced in this study as well as 5 from our previous work (Foo et al., 2017) formed 8 distinct genomic clusters of which 6 consist entirely of rearing water isolates (Figure 1). Based on the current genomic sampling, with the exception for the genomic cluster corresponding to ST-970, the remaining 7 genomic clusters are unique to Malaysian aquaculture farm isolates with each cluster consisting of only isolates that are exclusive to a specific farm (Figure 1).

New sequence types of *Vibrio parahaemolyticus* revealed by *in-silico* multilocus sequence typing

MLST profiling indicates that that the *dtdS* gene coding for threonine 3-dehydrogenase is the most variable among the 7 housekeeping genes of currently sequenced Asian *V. parahaemolyticus* with 45 alleles identified (Supplementary Table 1). Conversely, the *pntA* gene coding for NAD(P) transhydrogenase subunit alpha subunit has the least number of observed alleles (n=28, 5 are found in the isolates of this study) (Supplementary Table 1). A total of 33 known and 18 new sequence types (STs) were identified from the currently sequenced Asian *V.*

parahaemolyticus strains of which 7 (ST-1911 to 1915, ST-2012 and ST-2013) were exclusive to the *V. parahaemolyticus* isolates in this study (Figure 1). All currently sampled *V. parahaemolyticus* isolates that were assigned to ST-390, ST-1911, ST-1912, ST-1914, ST-1915 contain no detectable $pirAB^{Vp}$ and pVa plasmid backbone in their genome assemblies. Among the pVa-harbouring Malaysian isolates, a majority of them were assigned to ST-970 (n=9) followed by ST2013 (n=8), ST-1913 (n=3), ST-2012 (n=1) and ST-392 (n=1) which together formed four major genomic clusters in the phylogenomic tree (Figure 1).

Three major structural variants of the pVA plasmid among Asian *V. parahaemolyticus*

Of the 52 sequence types of *V. parahaemolyticus* identified in this current genomic sampling, $pirAB^{Vp}$ could only be detected among isolates from 8 validated (ST-2013, ST-114, ST-150, ST-392, ST-413, ST-970, ST-1166, and ST-1913) and 1 undetermined sequence types (Figure 1). Strain v110 with the undetermined sequence type has an unassigned *pntA* allele with the closest match to *pntA48* (Supplementary Table 1). Whole-genome alignment of pVA-harbouring Asian *V. parahaemolyticus* isolates against the reference pVA plasmid showed near-complete sequence conservation of the plasmid except for the pir^{Vp} gene region that exhibits three major structural variants e.g. $pirA^{Vp+}B^{Vp+}$, $pirA^{Vp-}B^{Vp+}$ and $pirA^{Vp-}B^{Vp-}$ (Figure 2 and Supplementary Figure 2). A closer inspection of the *pirB* gene neighborhood in both $pirA^{-}B^{+}$ NA2 and ND19 mutants revealed a complete deletion of the $pirA^{Vp}$ gene and a slight truncation at the 5' end of $pirB^{Vp}$. Immediately upstream of the 5' truncated $pirB^{Vp}$ gene is a small fragment of a transposase gene showing significant nucleotide homology to that of the reference pVa plasmid (Figure 3). The observed gene organization in NA2 and ND19 contrasted with that of *V. parahaemolyticus* mutant isolate XN87 that appears to have a transposase inserted into the $pirA^{Vp}$ gene (Figure 3).

Discussion

We report the whole-genome sequencing of 40 *V. parahaemolyticus* isolated from Pacific white shrimp hepatopancreas and aquaculture rearing water in Malaysia. The phylogenomic clustering indicates at least four distinct origins of $pirAB^{Vp}$-harbouring *V. parahaemolyticus* in Malaysia pending future sampling and genome sequencing. Multiple origins of $pirAB^{Vp}$+ *V. parahaemolyticus* were also observed in a previous MLST study in Thailand indicating high genetic diversity among AHPND *V. parahaemolyticus* isolates (Zorriehzahra &

Banaederakhshan, 2015). Furthermore, we have substantially expanded the genomic representation within the ST-970 clade, which now consists of *pirA*$^{Vp+}$*B*$^{Vp+}$, *pirA*$^{Vp-}$*B*$^{Vp+}$ and *pirA*$^{Vp-}$*B*$^{Vp-}$ isolates.

In this study, despite the detection of several STs among the *V. parahaemolyticus* isolates with some even sharing the same isolation source, the pVa plasmid could only be detected in isolates from a certain sequence type, suggesting that the distribution of virulence factor for AHPND among culturable *V. parahaemolyticus* maybe ST-specific. Given that a majority of the isolates sequenced in this study were collected from the rearing water, it is also possible that the pVa plasmids were lost due to the lack of selection pressure or had not been acquired and propagated among certain sequence types. A successful AHPND strain must first have genomic makeup that permits replication and maintenance of the pVa plasmid (Petersen, 2011). Secondly, it must be able to colonize and persist in its shrimp host as well as exhibit timely expression of the *pir* genes during pathogenesis (Gode□Potratz *et al.*, 2011, Liu & Chen, 2015, Soonthornchai *et al.*, 2015, Li *et al.*, 2017, Ng *et al.*, 2018). Further, given that AHPND leads to nearly 100% host mortality, it is also essential that the genomic make-up of an AHPND strain supports persistence outside its animal host post-death e.g. pond rearing water, which will have a diversity of microbial, chemical and physical compositions (Chen *et al.*, 2017, Cornejo-Granados *et al.*, 2017, Dittmann *et al.*, 2017, Zoqratt *et al.*, 2018). That said, the presence of pVa-harbouring Asian *V. parahaemolyticus* in multiple distantly-related clades suggests that many existing *V. parahaemolyticus* may already have the genomic potential to transform into an AHPND strain after the acquisition of the virulent pVa plasmid presumably through conjugal transfer.

The expression of both PirA^{Vp} and PirB^{Vp} toxins are required to cause the onset of AHPND (Hirono *et al.*, 2016), but it has been hypothesized that PirA^{Vp} is not expressed until later in the infection stage (Lai *et al.*, 2015, Lee *et al.*, 2015, Sirikharin *et al.*, 2015, Han *et al.*, 2016). Interestingly, despite the inability of a *pirA* mutant *V. parahaemolyticus* XN87 to produce the PirA^{Vp} and PirB^{Vp} toxins or to cause AHPND lesions, it still causes 50% mortality in shrimps presumably from the expression of other putative virulence genes located in the pVA plasmid (Phiwsaiya *et al.*, 2017). Instead of a transposon-inserted *pirA*^{Vp}, we observed a complete *pirA*^{Vp} deletion in NA2 and ND19 isolated from the aquaculture farm in Northern Malaysia. Despite the notable structural dissimilarity observed in *pir* region between XN87 and NA2/ND19, the close proximity of the transposase to *pirB*^{Vp} in NA2 and ND19 is likely going to disrupt the expression

of *pirB*^{Vp} due to a polar mutation effect as reported in XN87 and various transposon mutants (Harayama *et al.*, 1983, Richardson *et al.*, 1990, Gan *et al.*, 2011, Gan *et al.*, 2012, Phiwsaiya *et al.*, 2017).

The *pirAB*^{Vp} genes are usually flanked by 2 identical transposons with 18 bp of terminal inverted repeats with significant homology to the IS903 in *Escherichia coli* (Xiao *et al.*, 2017). Minor structural variations observed within the pVa region containing the *pirAB*^{Vp} genes suggest that they are prone to deletion or transposition. This is supported by previous studies reporting the natural loss of *pirAB*^{Vp} genes in the virulent strains of *V. parahaemolyticus* (strain E1) and *V. owensii* after laboratory sub-culturing with an estimated 22% loss rate of *pirAB*^{Vp} in *V. owensii* (Tinwongger *et al.*, 2016, Liu *et al.*, 2018). Given that all *V. parahaemolyticus* isolates in this study have gone through at least two sub-culturing to ensure culture purity prior to whole-genome sequencing, it is possible that the *pir*^{Vp} gene(s) were lost from the initially *pir*+ isolates in the laboratory rather than in their natural environment, due to the lack of selection pressure (Smith & Bidochka, 1998). Future study investigating the stability of *pirA*^{Vp} and *pirB*^{Vp} gene via PCR among our isolates under different growth conditions will be required. In addition, whole metagenome sequencing will be instructive to more accurately estimate the fraction of naturally occurring pVa plasmid with *pirA*^{Vp} and/or *pirB*^{Vp} deletion in the shrimp pond water and affected hepatopancreas (Jitwasinkul *et al.*, 2016, Krawczyk *et al.*, 2018).

Conclusions

We report the whole-genome sequencing on 40 *V. parahaemolyticus* isolates from deceased shrimps and rearing water collected from three Malaysian shrimp aquaculture farms. The number of new sequence types, MLST alleles and lineages of *V. parahaemolyticus* identified from only three sampling sites in this study underscores the overlooked genomic diversity of culturable *V. parahaemolyticus* associated with shrimp aquaculture environment. In addition to expanding the genomic resources of *pirAB*^{Vp}-bearing *V. parahaemolyticus*, especially for sequence types without genomic representation, it is recommended that future studies also include sequencing of avirulent *V. parahaemolyticus* isolates from APNHD-affected shrimp aquaculture environment to gain a better understanding of *V. parahaemolyticus* genomic diversity that may provide new insights into future pathogen management in shrimp aquaculture.

Acknowledgments

This study is supported by the Ministry of Higher Education under the grant FRGS/1/2016/STG05/MUSM/03/1. We would like to thank Lab-Ind Resource Sdn Bhd for their kind assistance in the sampling process.

Conflict of interest

The authors have no conflict of interest to declare

References

Alikhan NF, Petty NK, Ben Zakour NL & Beatson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**: 402.

Bankevich A, Nurk S, Antipov D, *et al.* (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19**: 455-477.

Boetzer M, Henkel CV, Jansen HJ, Butler D & Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578-579.

Boetzer M, Pirovano W, Zerbino D, *et al.* (2012) Toward almost closed genomes with GapFiller. *Genome Biology* **13**: R56.

Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K & Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Castillo D, Jun JW, D'Alvise P, Middelboe M, Gram L, Liu S & Katharios P (2015) Draft Genome Sequence of *Vibrio parahaemolyticus* VH3, Isolated from an Aquaculture Environment in Greece. *Genome announcements* **3**: e00731-00715.

Chen W-Y, Ng TH, Wu J-H, Chen J-W & Wang H-C (2017) Microbiome Dynamics in a Shrimp Grow-out Pond with Possible Outbreak of Acute Hepatopancreatic Necrosis Disease. *Scientific Reports* **7**: 9395.

Chu KB, Ahmad I, Zahrah S, *et al.* (2016) Current status of acute hepatopancreatic necrosis disease (AHPND) and other transboundary diseases of farmed shrimps in Indonesia Current Status of Acute Hepatopancreatic Necrosis. p.^pp. 37-43. Makati City, Philippines.

Cornejo-Granados F, Lopez-Zavala AA, Gallardo-Becerra L, Mendoza-Vargas A, Sánchez F, Vichido R, Brieba LG, Viana MT, Sotelo-Mundo RR & Ochoa-Leyva A (2017) Microbiome of Pacific Whiteleg shrimp reveals differential bacterial community composition between Wild, Aquacultured and AHPND/EMS outbreak conditions. *Scientific Reports* **7**: 11783.

Devadas S, Bhassu S, Christie Soo TC, Yusoff FM & Shariff M (2018) Draft Genome Sequence of the Shrimp Pathogen *Vibrio parahaemolyticus* ST17.P5-S1, Isolated in Peninsular Malaysia. *Microbiology Resource Announcements* **7**: e01053-01018.

Devadas S, Bhassu S, Christie Soo TC, Mohamed Iqbal SN, Yusoff FM & Shariff M (2018) Draft Genome Sequence of a *Vibrio parahaemolyticus* Strain, KS17.S5-1, with Multiple Antibiotic Resistance Genes, Which Causes Acute Hepatopancreatic Necrosis Disease in Penaeus monodon in the West Coast of Peninsular Malaysia. *Microbiology Resource Announcements* **7**: e00829-00818.

Dittmann KK, Rasmussen BB, Castex M, Gram L & Bentzon-Tilia M (2017) The aquaculture microbiome at the centre of business creation. *Microbial Biotechnology* **10**: 1279-1282.

Dong X, Wang H, Xie G, Zou P, Guo C, Liang Y & Huang J (2017) An isolate of *Vibrio campbellii* carrying the *pirVP* gene causes acute hepatopancreatic necrosis disease. *Emerging Microbes & Infections* **6**: e2.

Döpfer D, Buist W, Soyer Y, Munoz MA, Zadoks RN, Geue L & Engel B (2008) Assessing Genetic Heterogeneity within Bacterial Species Isolated from Gastrointestinal and Environmental Samples: How Many Isolates Does It Take? *Applied and Environmental Microbiology* **74**: 3490-3496.

FAO (2013) Report of the FAO/MARD Technical Workshop on Early Mortality Syndrome (EMS) or Acute Hepatopancreatic Necrosis Syndrome (AHPNS) of Cultured Shrimp (under TCP/VIE/3304). *FAO Fisheries and Aquaculture Report* **1053**: 1-52.

Foo SM, Eng WWH, Lee YP, Gui K & Gan HM (2017) New Sequence Types of *Vibrio parahaemolyticus* Isolated from a Malaysian Aquaculture Pond, as Revealed by Whole-Genome Sequencing. *Genome Announcements* **5**: e00302-00317.

Fu S, Tian H, Wei D, Zhang X & Liu Y (2017) Delineating the Origins of *Vibrio parahaemolyticus* Isolated from Outbreaks of Acute Hepatopancreatic Necrosis Disease in Asia by the Use of Whole Genome Sequencing. *Frontiers in Microbiology* **8**.

Gan HM, Shahir S & Yahya A (2012) Cloning and functional analysis of the genes coding for 4-aminobenzenesulfonate 3,4-dioxygenase from *Hydrogenophaga* sp. PBC. *Microbiology* **158**: 1933-1941.

Gan HM, Ibrahim Z, Shahir S & Yahya A (2011) Identification of genes involved in the 4-aminobenzenesulfonate degradation pathway of *Hydrogenophaga* sp. PBC via transposon mutagenesis. *FEMS microbiology letters* **318**: 108-114.

Gan HM, Rajasekaram G, Eng WWH, Kaniappan P & Dhanoa A (2017) Whole-genome sequences of two carbapenem-resistant *Klebsiella quasipneumoniae* strains isolated from a tertiary hospital in Johor, Malaysia. *Genome Announc* **5**: e00768-00717.

Gode-Potratz CJ, Kustusch RJ, Breheny PJ, Weiss DS & McCarter LL (2011) Surface sensing in *Vibrio parahaemolyticus* triggers a programme of gene expression that promotes colonization and virulence. *Molecular Microbiology* **79**: 240-263.

Gomez-gil B, Soto-rodríguez S, Lozano R & Betancourt-lozano M (2014) Draft Genome Sequence of *Vibrio parahaemolyticus* Strain M0605, Which Causes Severe Mortalities of Shrimps in Mexico. **2**: 2104.

Gomez-Jimenez S, Noriega-Orozco L, Sotelo-Mundo RR, *et al*. (2014) High-Quality Draft Genomes of Two *Vibrio parahaemolyticus* Strains Aid in Understanding Acute Hepatopancreatic Necrosis Disease of Cultured Shrimps in Mexico. *Genome announcements* **2**: e00800-00814.

Gurevich A, Saveliev V, Vyahhi N & Tesler G (2013) QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**: 1072-1075.

Han JE, Tang KFJ, Tran LH & Lightner (2016) Photorhabdus insect-related (Pir) toxin-like genes in a plasmid of Vibrio parahaemolyticus, the causative agent of acute hepatopancreatic necrosis disease (AHPND) of shrimp. *Dis Aquat Organ* **21**: 4062-4072.

Harayama S, Bollinger J, Iino T & Hazelbauer GL (1983) Characterization of the mgl operon of *Escherichia coli* by transposon mutagenesis and molecular cloning. *J Bacteriol* **153**: 408-415.

Hirono I, Tinwongger S, Nochiri Y & Kondo H (2016) Latest research on acute hepatopancreatic necrosis disease (AHPND) of penaeid shrimps. *SEAFDEC/AQD Institutional Repository (SAIR)*.

Hong KW, Gan HM, Yin W-F & Chan K-G (2012) Whole-genome sequence of *Cupriavidus* sp. strain BIS7, a heavy-metal-resistant bacterium. p.^pp. Am Soc Microbiol.

Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, Zobel J & Holt KE (2014) SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Medicine* **6**: 90.

Jitwasinkul T, Suriyaphol P, Tangphatsornruang S, Hansen MA, Hansen LH, Sørensen SJ, Permpikul C, Rongrungruang Y & Tribuddharat C (2016) Plasmid metagenomics reveals multiple antibiotic resistance gene classes among the gut microbiomes of hospitalised patients. *Journal of Global Antimicrobial Resistance* **6**: 57-66.

Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM & Aarestrup FM (2014) Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic *Escherichia coli*. *Journal of Clinical Microbiology* **52**: 1501-1510.

Jun JW, Kim JH, Choresca CH, Shin SP, Han JE & Park SC (2013) Draft Genome Sequence of *Vibrio parahaemolyticus* SNUVpS-1 Isolated from Korean Seafood. *Genome announcements* **1**: e00132-00112.

Kavousi N, Eng WWH, Lee YP, Tan LH, Thuraisingham R, Yule CM & Gan HM (2016) First High-Quality Draft Genome Sequence of *Pasteurella multocida* Sequence Type 128 Isolated from Infected Bone. *Genome Announcements* **4**.

Khimmakthong U & Sukkarun P (2017) The spread of *Vibrio parahaemolyticus* in tissues of the Pacific white shrimp Litopenaeus vannamei analyzed by PCR and histopathology. *Microbial pathogenesis* **113**: 107-112.

Kondo H, Van PT, Dang LT & Hirono I (2015) Draft Genome Sequence of Non-*Vibrio parahaemolyticus* Acute Hepatopancreatic Necrosis Disease Strain KC13.17.5, Isolated from Diseased Shrimp in Vietnam. *Genome Announc* **3**: e00978-00915.

Kondo H, Tinwongger S, Proespraiwong P, Mavichak R, Unajak S, Nozaki R & Hirono I (2014) Draft Genome Sequences of Six Strains of *Vibrio parahaemolyticus* Isolated from Early Mortality Syndrome/Acute Hepatopancreatic Necrosis Disease Shrimp in Thailand. *Genome announcements* **2**: e00221-00214.

Krawczyk PS, Lipinski L & Dziembowski A (2018) PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Research* **46**: e35-e35.

Lai HC, Ng TH, Ando M, *et al.* (2015) Pathogenesis of acute hepatopancreatic necrosis disease (AHPND) in shrimp. *Fish and Shellfish Immunology* **47**: 1006-1014.

Land M, Hauser L, Jun S-R, *et al.* (2015) Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics* **15**: 141-161.

Larsen MV, Cosentino S, Rasmussen S, *et al.* (2012) Multilocus sequence typing of total-genome-sequenced bacteria. *Journal of Clinical Microbiology* **50**: 1355-1361.

Lee C-T, I-Tung C, Yang Y-T, *et al.* (2015) The opportunistic marine pathogen *Vibrio parahaemolyticus* becomes virulent by acquiring a plasmid that expresses a deadly toxin. *Proceedings of the National Academy of Sciences* **112**: E5445-E5445.

Li P, Kinch LN, Ray A, Dalia AB, Cong Q, Nunan LM, Camilli A, Grishin NV, Salomon D & Orth K (2017) Acute Hepatopancreatic Necrosis Disease (AHPND)-causing *Vibrio parahaemolyticus* strains maintain an antibacterial Type VI Secretion System with versatile effector repertoires. *Applied and Environmental Microbiology*.

Liu L, Xiao J, Zhang M, Zhu W, Xia X, Dai X, Pan Y, Yan S & Wang Y (2018) A *Vibrio owensii* strain as the causative agent of AHPND in cultured shrimp, *Litopenaeus vannamei*. *Journal of Invertebrate Pathology* 0-1.

Liu M & Chen S (2015) A novel adhesive factor contributing to the virulence of *Vibrio parahaemolyticus*. *Scientific Reports* **5**: 14449.

Ng TH, Lu CW, Lin SS, Chang CC, Tran LH, Chang WC, Lo CF & Wang HC (2018) The Rho signalling pathway mediates the pathogenicity of AHPND-causing *V. parahaemolyticus* in shrimp. *Cellular Microbiology* **20**: e12849.

Osama A, Gan HM, Teh CSJ, Yap K-P & Thong K-L (2012) Genome sequence and comparative genomics analysis of a *Vibrio cholerae* O1 strain isolated from a cholera patient in Malaysia. *Journal of bacteriology* **194**: 6933-6933.

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA & Parkhill J (2015) Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**: 3691-3693.

Petersen J (2011) Phylogeny and compatibility: plasmid classification in the genomics era. *Arch Microbiol* **193**: 313-321.

Phiwsaiya K, Charoensapsri W, Taengphu S, *et al.* (2017) A Natural *Vibrio parahaemolyticus ⊿ pirA Vp pirB Vp+* Mutant Kills Shrimp but Produces neither Pir Vp Toxins nor Acute Hepatopancreatic Necrosis Disease Lesions. *Applied and Environmental Microbiology* **83**: e00680-00617.

Price MN, Dehal PS & Arkin AP (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* **5**.

Richardson K, Nixon L, Mostow P, Kaper JB & Michalski J (1990) Transposon-induced non-motile mutants of *Vibrio cholerae*. *J Gen Microbiol* **136**: 717-725.

Richter M, Rossello-Mora R, Oliver Glockner F & Peplies J (2015) JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **32**: 929-931.

Sirikharin R, Taengchaiyaphum S, Sanguanrut P, Chi TD, Mavichak R, Proespraiwong P, Nuangsaeng B, Thitamadee S, Flegel TW & Sritunyalucksana K (2015) Characterization and PCR detection of binary, pir-like toxins from *Vibrio parahaemolyticus* isolates that cause acute hepatopancreatic necrosis disease (AHPND) in shrimp. *PLoS One* **10**: 1-17.

Smith MA & Bidochka MJ (1998) Bacterial fitness and plasmid loss: the importance of culture conditions and plasmid size. *Can J Microbiol* **44**: 351-355.

Sokolov EP (2000) An improved method for DNA isolation from mucopolysaccharide-rich molluscan tissues. *Journal of Molluscan Studies* **66**: 573-575.

Soonthornchai W, Chaiyapechara S, Jarayabhand P, Söderhäll K & Jiravanichpaisal P (2015) Interaction of *Vibrio* spp. with the Inner Surface of the Digestive Tract of *Penaeus monodon*. *PLOS ONE* **10**: e0135783.

Sullivan MJ, Petty NK & Beatson SA (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* **27**: 1009-1010.

Tang P, Croxen MA, Hasan MR, Hsiao WWL & Hoang LM (2017) Infection control in the new age of genomic epidemiology. *American Journal of Infection Control* **45**: 170-179.

Tinwongger S, Nochiri Y, Thawonsuwan J, Nozaki R, Kondo H, Awasthi SP, Hinenoya A, Yamasaki S & Hirono I (2016) Virulence of acute hepatopancreatic necrosis disease PirAB-like relies on secreted proteins not on gene copy number. *Journal of Applied Microbiology* **121**: 1755-1765.

Tran L, Nunan L, Redman RM, Mohney LL, Pantoja CR, Fitzsimmons K & Lightner DV (2013) Determination of the infectious nature of the agent of acute hepatopancreatic necrosis syndrome affecting penaeid shrimp. *Diseases of Aquatic Organisms* **105**: 45-55.

Xiao J, Liu L, Ke Y, Li X, Liu Y, Pan Y & Yan S (2017) Shrimp AHPND-causing plasmids encoding the PirAB toxins as mediated by *pirAB* -Tn903 are prevalent in various Vibrio species. *Scientific Reports* 1-11.

Yang N, Liu M, Luo X & Pan J (2015) Draft genome sequence of Strain ATCC 17802T, the type strain of *Vibrio parahaemolyticus*. *Marine Genomics* **24**: 203-205.

Yang Y-t, Chen I-T, Lee C-T, *et al*. (2014) Draft Genome Sequences of Four Strains of *Vibrio parahaemolyticus* , Three of Which Cause Early Mortality Syndrome / Acute. *Genome* **2**: 2-3.

Yap K-P, Gan HM, Teh CSJ, Baddam R, Chai L-C, Kumar N, Tiruvayipati SA, Ahmed N & Thong K-L (2012) Genome sequence and comparative pathogenomics analysis of a *Salmonella enterica* serovar Typhi strain associated with a typhoid carrier in Malaysia. *Journal of bacteriology*.

Zoqratt MZHM, Eng WWH, Thai BT, Austin CM & Gan HM (2018) Microbiome analysis of Pacific white shrimp gut and rearing water from Malaysia and Vietnam: implications for aquaculture research and management. *PeerJ* **6**: e5826.

Zorriehzahra MJ & Banaederakhshan R (2015) Review article - Early Mortality Syndrome (EMS) as new Emerging Treat in Shrimp Industry. *Advances in Animal and Veterinary Science* **3**: 64.

Figure Legends

Figure 1: Maximum likelihood tree based on core genome alignment depicting the evolutionary relationships of 135 Asian *Vibrio parahaemolyticus* isolates. Approximate sampling location was indicated in the peninsular Malaysia map located at the top left corner. Sequence types are labeled accordingly with ST followed by their designated profile number based on PubMLST database (https://pubmlst.org/vparahaemolyticus). The tree was rooted with *V. cholera* as the outgroup and for clarity, the root has been shortened by approximately 80% of the actual length shown. Branch lengths and node values indicate the number of substitutions per site and SH-like support values, respectively. Two-letter codes in brackets next to the tip label indicate sample origin. MY, Malaysia; HK, Hong Kong, CN, China; KR, Korea; IN, India; VN, Vietnam; TH, Thailand.

Figure 2: BRIG visualization of *V. parahaemolyticus* pVa plasmid. Using BLASTN analyses, pVa plasmid from clade-representative strains of *V. parahaemolyticus* throughout the Asian region were aligned against plasmid pVPA3-1 and visualized using BRIG. Black arcs indicate hypothetical proteins with no InterProScan hits. The innermost circle represents the reference sequence of plasmid pVPA3-1 harbored by *V. parahaemolyticus* strain 13-028/A3. Outer rings illustrate shared identity with pVa plasmids found in some samples of this study and NCBI database. The varying saturation of the ring colors is based on 100%, 90% and 70% identity. Clear white gaps indicate no sequence similarity in that area. Full visualization of all *V. parahaemolyticus* strains harboring pVa plasmid can be found in Supplementary Figure 2.

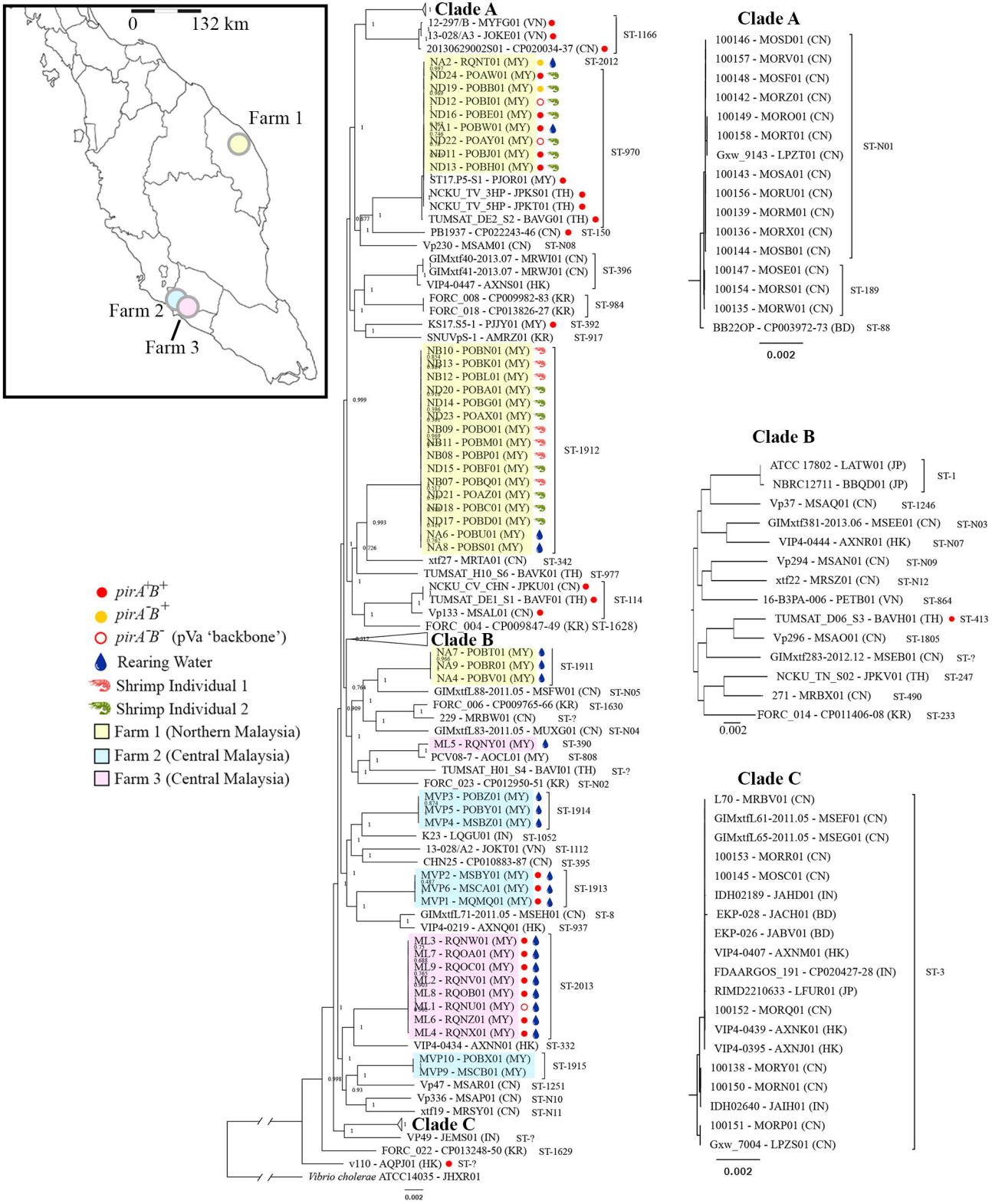
Figure 3: Gene organization of $pirB^{Vp}$-containing contigs among selected *V. parahaemolyticus* isolates. Vertical blocks between sequences indicate regions of homology shaded according to BLASTn. Left and right labels on each sequence are isolate names and GenBank accession numbers, respectively. Blue arrows below the XN87 sequence indicate the location of

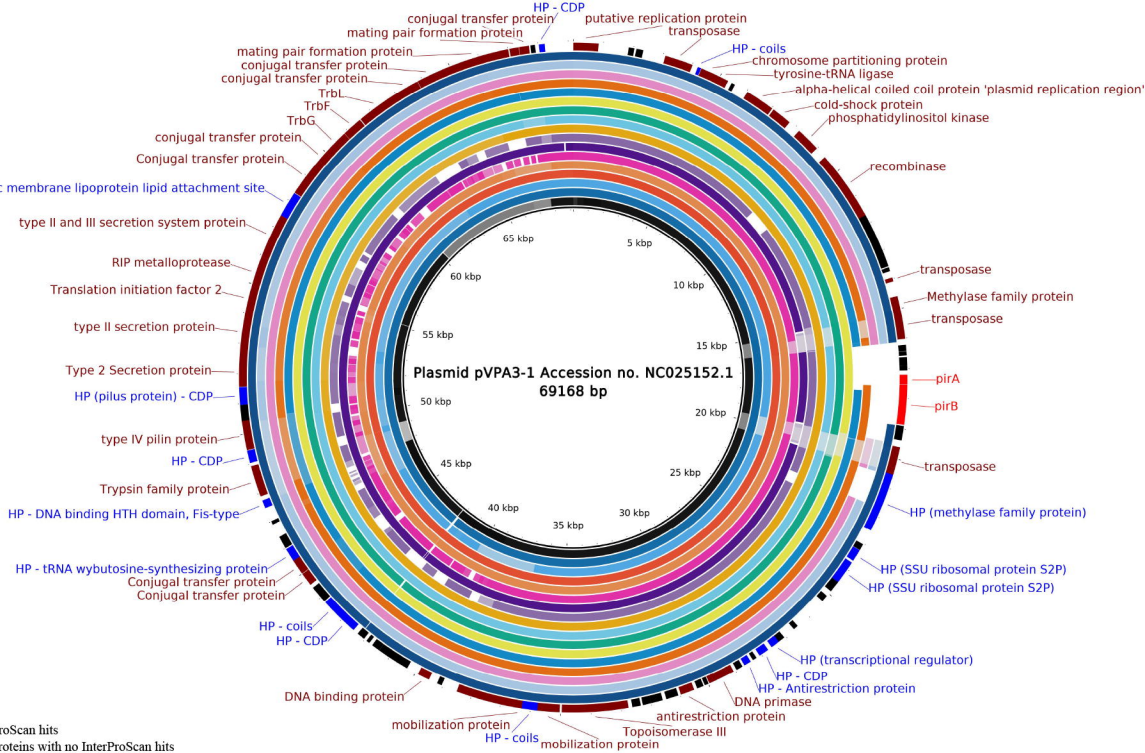
transposase-inserted pirA genes (*pirA* fragment 1 and *pirA* fragment2) that were not annotated in the original GenBank file. The direction of arrow indicates gene transcription orientation.

Supplementary Figure 1: JSpecies results based on ANIm calculation with *V. parahaemolyticus* ATCC 17802 as positive control and *V. cholerae* ATCC 14035 as the negative control.

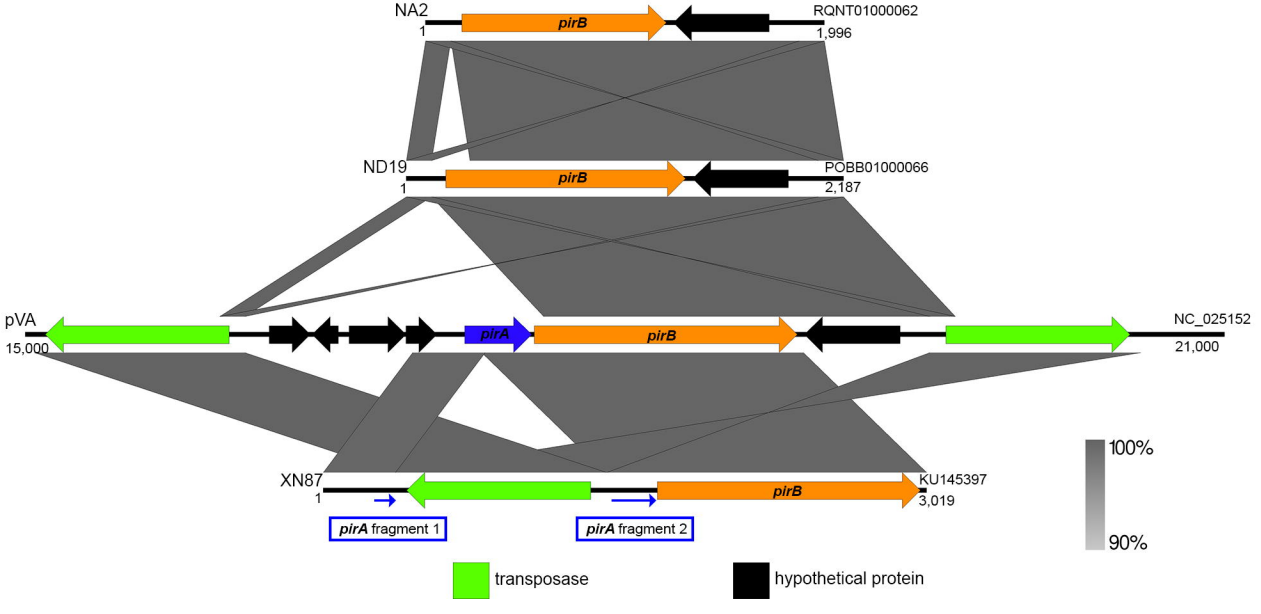
Supplementary Figure 2: Compilation of *V. parahaemolyticus* pVa plasmid. Using blastn analyses, *pir* plasmid from strains of *V. parahaemolyticus* throughout the Asian region were aligned against pVPA3-1 and visualized using BRIG. Black arcs indicate hypothetical proteins with no InterProScan hits. The innermost circle represents the reference sequence of plasmid pVPA3-1 identified in *V. parahaemolyticus* strain 13-028/A3. Outer rings illustrate shared identity with Pir plasmids found in some samples of this study and NCBI database.

Supplementary Table 1: Summary of strains used in this study with accession number, origin, genome statistics, MLST profiles, and presence of $pirA^{Vp}$ and $pirB^{Vp}$.


0
132 km
Farm 1
Farm 2
Farm 3
Clade A
12-297/B - MYFG01 (VN)
13-028/A3 - JOKE01 (VN)
20130629002S01 - CP020034-37 (CN)
ST-1166
NA2 - RQNT01 (MY)
ST-2012
ND24 - POAW01 (MY)
ND19 - POBB01 (MY)
ND12 - POBI01 (MY)
ND16 - POBE01 (MY)
NA1 - POBW01 (MY)
ND22 - POAY01 (MY)
ND11 - POBJ01 (MY)
ND13 - POBH01 (MY)
ST17.P5-S1 - PJOR01 (MY)
NCKU_TV_3HP - JPKS01 (TH)
NCKU_TV_5HP - JPKT01 (TH)
TUMSAT_DE2_S2 - BAVG01 (TH)
ST-970
PB1937 - CP022243-46 (CN)
ST-150
Vp230 - MSAM01 (CN)
ST-N08
GIMxtf40-2013.07 - MRWI01 (CN)
GIMxtf41-2013.07 - MRWJ01 (CN)
VIP4-0447 - AXNS01 (HK)
ST-396
FORC_008 - CP009982-83 (KR)
FORC_018 - CP013826-27 (KR)
ST-984
KS17.S5-1 - PJJY01 (MY)
ST-392
SNUVpS-1 - AMRZ01 (KR)
ST-917
NB10 - POBN01 (MY)
NB13 - POBK01 (MY)
NB12 - POBL01 (MY)
ND20 - POBA01 (MY)
ND14 - POBG01 (MY)
ND23 - POAX01 (MY)
NB09 - POBO01 (MY)
NB11 - POBM01 (MY)
NB08 - POBP01 (MY)
ND15 - POBF01 (MY)
NB07 - POBQ01 (MY)
ND21 - POAZ01 (MY)
ND18 - POBC01 (MY)
ND17 - POBD01 (MY)
NA6 - POBU01 (MY)
NA8 - POBS01 (MY)
ST-1912
xtf27 - MRTA01 (CN)
ST-342
TUMSAT_H10_S6 - BAVK01 (TH)
ST-977
NCKU_CV_CHN - JPKU01 (CN)
TUMSAT_DE1_S1 - BAVF01 (TH)
Vp133 - MSAL01 (CN)
ST-114
FORC_004 - CP009847-49 (KR)
ST-1628
Clade B
NA7 - POBT01 (MY)
NA9 - POBR01 (MY)
NA4 - POBV01 (MY)
ST-1911
GIMxtfL88-2011.05 - MSFW01 (CN)
ST-N05
FORC_006 - CP009765-66 (KR)
ST-1630
229 - MRBW01 (CN)
ST-?
GIMxtfL83-2011.05 - MUXG01 (CN)
ST-N04
ML5 - RQNY01 (MY)
ST-390
PCV08-7 - AOCL01 (MY)
ST-808
TUMSAT_H01_S4 - BAVI01 (TH)
ST-?
FORC_023 - CP012950-51 (KR)
ST-N02
MVP3 - POBZ01 (MY)
MVP5 - POBY01 (MY)
MVP4 - MSBZ01 (MY)
ST-1914
K23 - LQGU01 (IN)
ST-1052
13-028/A2 - JOKT01 (VN)
ST-1112
CHN25 - CP010883-87 (CN)
ST-395
MVP2 - MSBY01 (MY)
MVP6 - MSCA01 (MY)
MVP1 - MQMQ01 (MY)
ST-1913
GIMxtfL71-2011.05 - MSEH01 (CN)
ST-8
VIP4-0219 - AXNQ01 (HK)
ST-937
ML3 - RQNW01 (MY)
ML7 - RQOA01 (MY)
ML9 - RQOC01 (MY)
ML2 - RQNV01 (MY)
ML8 - RQOB01 (MY)
ML1 - RQNU01 (MY)
ML6 - RQNZ01 (MY)
ML4 - RQNX01 (MY)
ST-2013
VIP4-0434 - AXNN01 (HK)
ST-332
MVP10 - POBX01 (MY)
MVP9 - MSCB01 (MY)
ST-1915
Vp47 - MSAR01 (CN)
ST-1251
Vp336 - MSAP01 (CN)
ST-N10
xtf19 - MRSY01 (CN)
ST-N11
Clade C
VP49 - JEMS01 (IN)
ST-?
FORC_022 - CP013248-50 (KR)
ST-1629
v110 - AQPJ01 (HK)
ST-?
Vibrio cholerae ATCC14035 - JHXR01
0.002
pirA+B+
pirA-B+
pirA-B- (pVa 'backbone')
Rearing Water
Shrimp Individual 1
Shrimp Individual 2
Farm 1 (Northern Malaysia)
Farm 2 (Central Malaysia)
Farm 3 (Central Malaysia)
Clade A
100146 - MOSD01 (CN)
100157 - MORV01 (CN)
100148 - MOSF01 (CN)
100142 - MORZ01 (CN)
100149 - MORO01 (CN)
100158 - MORT01 (CN)
Gxw_9143 - LPZT01 (CN)
100143 - MOSA01 (CN)
100156 - MORU01 (CN)
100139 - MORM01 (CN)
100136 - MORX01 (CN)
100144 - MOSB01 (CN)
ST-N01
100147 - MOSE01 (CN)
100154 - MORS01 (CN)
100135 - MORW01 (CN)
ST-189
BB22OP - CP003972-73 (BD)
ST-88
0.002
Clade B
ATCC 17802 - LATW01 (JP)
NBRC12711 - BBQD01 (JP)
ST-1
Vp37 - MSAQ01 (CN)
ST-1246
GIMxtf381-2013.06 - MSEE01 (CN)
ST-N03
VIP4-0444 - AXNR01 (HK)
ST-N07
Vp294 - MSAN01 (CN)
ST-N09
xtf22 - MRSZ01 (CN)
ST-N12
16-B3PA-006 - PETB01 (VN)
ST-864
TUMSAT_D06_S3 - BAVH01 (TH)
ST-413
Vp296 - MSAO01 (CN)
ST-1805
GIMxtf283-2012.12 - MSEB01 (CN)
ST-?
NCKU_TN_S02 - JPKV01 (TH)
ST-247
271 - MRBX01 (CN)
ST-490
FORC_014 - CP011406-08 (KR)
ST-233
0.002
Clade C
L70 - MRBV01 (CN)
GIMxtfL61-2011.05 - MSEF01 (CN)
GIMxtfL65-2011.05 - MSEG01 (CN)
100153 - MORR01 (CN)
100145 - MOSC01 (CN)
IDH02189 - JAHD01 (IN)
EKP-028 - JACH01 (BD)
EKP-026 - JABV01 (BD)
VIP4-0407 - AXNM01 (HK)
FDAARGOS_191 - CP020427-28 (IN)
RIMD2210633 - LFUR01 (JP)
100152 - MORQ01 (CN)
VIP4-0439 - AXNK01 (HK)
VIP4-0395 - AXNJ01 (HK)
ST-3
100138 - MORY01 (CN)
100150 - MORN01 (CN)
IDH02640 - JAIH01 (IN)
100151 - MORP01 (CN)
Gxw_7004 - LPZS01 (CN)
0.002



Notes:
CDP - consensus disorder prediction
HP - hypothetical proteins with InterProScan hits
Unlabelled black arcs - hypothetical proteins with no InterProScan hits


NA2
1
pirB
RQNT01000062
1,996
ND19
1
pirB
POBB01000066
2,187
pVA
15,000
pirA
pirB
NC_025152
21,000
XN87
1
pirB
KU145397
3,019
pirA fragment 1
pirA fragment 2
100%
90%
transposase
hypothetical protein