

Horizontal gene transfer contributes to virulence and antibiotic resistance of *Vibrio harveyi* 345 based on complete genome sequence analysis

Yiqin Deng

South China Sea Fisheries Research Institute

Haidong Xu

South China Sea Fisheries Research Institute

Youlu Su

South China Sea Fisheries Research Institute

Songlin Liu

South China Sea Institute of Oceanology Chinese Academy of Sciences

Liwen Xu

South China Sea Fisheries Research Institute

Zhixun Guo

South China Sea Fisheries Research Institute

Jinjun Wu

South China Sea Fisheries Research Institute

Changhong Cheng

South China Sea Fisheries Research Institute

Juan Feng (✉ juanfeng@scsfri.ac.cn)

Chinese Academy of Fishery Sciences <https://orcid.org/0000-0001-8933-5365>

Research article

Keywords: *Vibrio harveyi*, virulence, antibiotic resistance, comparative genomics, horizontal gene transfer

Posted Date: September 24th, 2019

DOI: <https://doi.org/10.21203/rs.2.13151/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on October 22nd, 2019. See the published version at <https://doi.org/10.1186/s12864-019-6137-8>.

Abstract

Background Horizontal gene transfer (HGT), which is affected by environmental pollution and climate change, promotes genetic communication, changing bacterial pathogenicity and drug resistance. However, few studies have been conducted on the effect of HGT on the high pathogenicity and drug resistance of the opportunistic pathogen *Vibrio harveyi*. Results *V. harveyi* 345 that was multidrug resistant and infected *Epinephelus oanceolus* was isolated from a diseased organism in Shenzhen, Southern China, an important and contaminated aquaculture area. Analysis of the entire genome sequence predicted 5,678 genes including 487 virulence genes contributing to bacterial pathogenesis and 25 antibiotic-resistance genes (ARGs) contributing to antimicrobial resistance. Five ARGs (*tetm* , *tetb* , *qnrS* , *dfra17* , and *sul2*) and one virulence gene (CU052_28670) on the pAQU-type plasmid p345-185, provided direct evidence for HGT. Comparative genome analysis of 31 *V. harveyi* strains indicated that 217 genes and 7 gene families, including a class C beta-lactamase gene, a virulence-associated protein D gene, and an OmpA family protein gene were specific to strain *V. harveyi* 345. These genes could contribute to HGT or be horizontally transferred from other bacteria to enhance the virulence or antibiotic resistance of 345. Mobile genetic elements in 71 genomic islands encoding virulence factors for three type III secretion proteins and 13 type VI secretion system proteins, and two incomplete prophage sequences were detected that could be HGT transfer tools. Evaluation of the complete genome of *V. harveyi* 345 and comparative genomics indicated genomic exchange, especially exchange of pathogenic genes and drug-resistance genes by HGT contributing to pathogenicity and drug resistance. Climate change and continued environmental deterioration are expected to accelerate the HGT of *V. harveyi* , increasing its pathogenicity and drug resistance. Conclusion This study provides timely information for further analysis of *V. harveyi* pathogenesis and antimicrobial resistance and developing pollution control measurements for coastal areas.

Background

With the rapid expansion of aquaculture, vibriosis has become one of the most serious diseases to endanger sustainable aquaculture development [1, 2]. *Vibrio harveyi* is a halophilic, luminescent Gram-negative γ -proteobacteria, as an important pathogen of vibriosis [4-5]. *V. harveyi* has been extensively studied for more than a decade and is reported to be a serious pathogen for a range of marine vertebrates and invertebrates [4-7]. The earliest report of *V. harveyi* causing serious infections was to common snook (*Centropomus undecimalis*) in Florida, USA [8]. Subsequently, *V. harveyi* was shown to cause disease in *Penaeus monodon* and *Penaeus japonicas* in Thailand [10], brown-spotted grouper in Kuwait [9], common dentex (*Dentex dentex*) in Spain [10], *Holothuria scabra* (Holothuroidea, Echinodermata) in Toliara, Madagascar [11], Asian seabass (*Lates calcarifer*) in Malaysia [12], and *Epinephelus spp.* in China [13-15].

The wide use of antibiotics, increased environmental pollution and global climate change are leading to the enhancement of drug resistance of *Vibrio spp.*, including *V. harveyi*. Several studies show that drug resistance and multidrug resistance (MDR) are occurring in aquaculture worldwide, including in shrimp

ponds in Thailand, Malaysia, India, and China [16-19]; shellfish farms in Malaysia, Korea, USA, Poland and China [20-24]; and fish farms in Italy, Korea, and China [25-27]. Enhanced drug resistance leads to stronger virulence resulting in difficulty in preventing and treating *V. harveyi* infection [15, 16]. For example, Nakayama et al. [28] found that gradually increasing antibiotic concentration and frequent subculturing enhances *V. harveyi* antibiotic resistance, elevating the toxicity of *V. harveyi*. The pathogenic and drug-resistant genes of *V. harveyi* are a key to the fundamental cause of pathogenicity and drug resistance. Therefore, studying the pathogenic and drug-resistance genes of *V. harveyi* will provide an important foundation for determining pathogenic and drug-resistance mechanisms.

Although *V. harveyi* a recognized pathogen of marine animals, different strains vary in their ability to cause disease [29]. The main virulence factors of *V. harveyi* are extracellular proteases, outer membrane proteins, hemolysins, esterases, phospholipases, exotoxin and secretion systems [29, 30]. Production of antimicrobial enzymes that inactivate antibacterial drugs is an important resistant mechanism [31]. Inactivating enzymes can be expressed by genes on plasmids or chromosomes [32]. Virulence factors and antibiotic-resistance genes (ARGs) can also be vertically transferred and spread via horizontal gene transfer (HGT) through mobile genetic elements (MGEs) such as plasmids, bacteriophages, transposons, integrative and conjugative elements (ICEs), and genetic islands (GIs) [33]. Park et al. [34] found that MGEs that are prophages, GIs and pathogenicity islands carry different combinations of virulence factors that promote immune evasion and superantigens that contribute to serious *Staphylococcus aureus* infection. Le et al. [35] reported that 0.92% (36 of 38,895) analyzed proteins from 31 *Rickettsiales* genomes are associated with strong bootstrap support for HGT with function as ATPases, aldolases, transporter activities, cystathionine beta-lyases, sugar phosphate permeases, growth inhibitors and antitoxin activities. For horizontal transfer, exogenous DNA must evade the bacterial immune system such as restriction modification systems (RM systems) and CRISPR-Cas systems [36, 37].

Serious infection and multidrug resistance lead to the emergency of the availability of whole genome sequences for different *V. harveyi* strains to determine pathogenic and antimicrobial-resistance mechanisms. Evidence is growing that HGT is an important driving force for prokaryotic evolution affecting pathogenicity and drug resistance [34, 38]. Coastal urbanization has recently intensified, resulting in the production of a large amount of antibiotics, heavy metals, and nutrients pollutants (Table S1), which may increase the pathogenicity and drug resistance of *V. harveyi* by affecting HGT [39, 40]. *V. harveyi* is the dominant species that causes serious infection and mortality of farmed fish in Guangdong, Southern China [14, 15]. However, little information is available on the mechanism of *V. harveyi* pathogenesis and drug resistance.

The predominant strain *V. harveyi* 345 is multidrug resistant to ampicillin, rifampicin, tetracycline, pediatric compound sulfamethoxazole tablets, vancomycin, doxycycline, trimethoprim, streptomycin, kanamycin, sulfamethoxazole, furazolidone, cefixime, and chloramphenicol. This strain was isolated from a diseased *E. oanceolustus* in Shenzhen, Southern China. *V. harveyi* 345 is suspected to lead to kidney enlargement and softening, spleen enlargement, and anal bleeding of *E. oanceolustus* and has a median lethal dose (LD₅₀) of 9.83×10^5 CFU·g⁻¹ [41]. Therefore, its complete genome information and HGT

events are helpful for clarifying *V. harveyi* pathogenesis and drug resistance, controlling *V. harveyi* disease and reducing economic losses. In this study, we present the entire genome sequence of *V. harveyi* 345 with comparative genomics analysis of its pathogenesis, antimicrobial resistance and genome expansion caused by HGT.

Results

General features of *V. harveyi* 345

A total of 1,189 megabases clean data and 98,602 subreads were generated assembling into two contigs with average genome coverage of 41.71×. After error correction, GATK analysis and gap filling, a genome of 6,185,822 bp was generated, which was smaller than comparable strains KC13.17.5, E385, and 74F and larger than 27 other comparable strains (Table S6). The chromosomal G + C content (44.76%) of 345 agreed with values from 30 comparable strains (44.60%-45.60%) (Table S6). Combined sequencing analysis revealed that the complete genome of *V. harveyi* 345 contained two circular chromosomes, named ChI (3,713,225 bp, 44.81% G + C, CP025537) and ChII (2,220,396 bp, 44.81% G + C, CP025538), and two circular megaplasmids named p345-185 (185,327 bp, 43.48% G + C, CP025539) and p345-67 (66,874 bp, 43.48% G + C, CP025540) (Table 1 and Fig. 1).

Gene annotation

A total of 5,678 genes (fewer than ZJ0603 and more than 29 other strains) (Table S6) were predicted to be an average 929 bp with a mean GC content of 45.68%, accounting for 85.29% of the genome. The three highest levels of gene length distribution were 400–500 bp, 600–700 bp, and 300–400 bp and the three lowest were 0–100 bp, 1,900–2,000 bp, and 1,800–1,900 bp (Fig. 2A). In addition, 128 tRNA, 31 rRNA and 13 sRNA genes were identified with confidence. A total of 89 tandem repeats, including 44 minisatellite DNA and 35 microsatellite DNA, were predicted (Table 1 and Fig. 1).

Predicted open reading frames (ORFs) were further classified into Clusters of Orthologous Groups (COG) (4,586 ORFs in total contenting 80.77%, Fig. 2B). According to COG categorization analysis, the top five groups in abundance were amino acid transport and metabolism (361 ORFs), transcription (350 ORFs), general function prediction only (343 ORFs), signal transduction mechanisms (312 ORFs) and carbohydrate transport and metabolism (312 ORFs) (Fig. 2B). In addition, 3,247 genes were classified into 40 functional Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Additional File 1: Figure S1). Some genes were involved in more than one KEGG pathway and included the top five pathways for replication and repair (784 ORFs), carbohydrate metabolism (633 ORFs), signal transduction (506 ORFs), infectious diseases (463 ORFs) and membrane transport (432 ORFs). Furthermore, 3,453 genes were classified into 40 functional Gene Ontology (GO) classifications (Additional File 2: Figure S2). Some genes were involved in more than one GO classifications and included the top five classifications of metabolic process (1,914 ORFs), cellular process (1,881 ORFs), catalytic activity (1,785 ORFs), single-organism process (1,526 ORFs) and binding (1,494 ORFs).

Twelve RM system genes were annotated in *V. harveyi* strain 345 (Table 2). Ten belonged to Type I RM systems, encoding three R subunits (*hsdR*), two S subunits (*hsdS*) and five M subunits (*hsdM*). The other two belonged to type III RM systems, encoding a methyltransferase (*mod*) and a restriction enzyme (*res*).

Virulence factors

A total of 487 putative virulence factors were predicted (Table 3 and Additional File 3: Table S2), involving adherence (type IV pili, lipooligosaccharide LOS, and OmpU), motility (polar flagellar proteins, chemotaxis proteins, lateral flagellin proteins, and motor proteins), regulation (Autoinducer-2 and cholerae autoinducer-1), secretion system (type II/III/IV/VI secretion system proteins, T2/3/4/6SS proteins), and iron uptake. Many genes were identical to genera other than *Vibrio* such as *Escherichia*, *Pseudomonas*, *Salmonella*, *Aeromonas*, and *Mycobacterium*. T3SS1, which is widely reported to be involved in *Vibrio* pathogenesis, was found on ChI. T6SSs are implicated in cell targeting and virulence, and are believed to mediate antibacterium [42]. Six T6SS proteins on Hcp secretion island-1 were detected. The *tlh* gene encoding a thermolabile hemolysin (TLH) was on ChII. In addition, four virulence genes (CU052_28670, CU052_00025, CU052_00190, and CU052_00400) were on plasmid with CU052_28670 on p345-185, and CU052_00025, CU052_00190, and CU052_00400 on p345-67.

Antimicrobial-resistance genes

We identified 25 genes predicted to have >40% identity to well-characterized ARGs (Table 4), including genes for aminoglycoside (*golc*, *adeb*, *emre*), penicillin (*pbp2*, *pbp1a*, *pbp1b*, *pbp2*), tetracycline (*tet34*, *tetm*, *tetb*), chloramphenicol (*adeb*, *ceob*, *mdtI*) and trimethoprim resistance (*dfra26*). In addition, five genes (CU052_28095, CU052_28120, CU052_28525, CU052_28540, and CU052_29140) were on the plasmid p345-185. These results were consistent with antibiotic-susceptibility assay data for strain 345 showing resistance to streptomycin, ampicillin, tetracycline, chloramphenicol, trimethoprim, and sulfamethoxazole.

Genomic islands, prophages and CRISPR-Cas systems

Forty-seven GIs were detected on ChI (Additional File 4: Table S3) and 24 on ChII in *V. harveyi* 345 (Additional File 5: Table S4). In the GIs of ChI, 40 transposases were encoded belonging to the IS110, IS21, IS256, IS3, IS5/IS1182, IS66, IS91, ISL3, and ISNCY families and seven belonging to two GIs. Another eight genes were identified that encoded IS66 family insertion sequence hypothetical proteins. Three integrases and one serine recombinase were encoded. Three genes were predicted to encode type III secretion proteins and one was predicted to encode another virulence factor (CU052_RS12915). In the GIs of ChII, nine genes encoding transposases belonging to the IS3 and ISNCY families were detected with four belonging to more than one GI. We found three genes encoding integrases and 13 genes on three GIs predicted to encode type VI secretion system (T6SS) proteins.

Two intact prophage sequences were identified (Additional File 6: Table S5) and no CRISPR elements were predicted.

Phylogenetic analysis, pan-core genes, dispensable and strain-specific genes

Phylogenetic trees showed that *V. harveyi* 345 closely related to other *V. harveyi* strains, especially *V. harveyi* VHJR7 and *V. harveyi* CAIM463 (Fig. 3).

The pan-gene refers to the total number of genes in all 31 strains, while core-gene represents orthologous genes shared among them. Dispensable genes were not included in at least one of the 31 strains and strain-specific genes were included in only one strain. Using 31 genomes (Additional File 7: Table S6), 9,878 pan-genes with a total size of 2,764,797 amino acid (aa) and 3,338 core-genes with a total size of 1,133,465 aa were identified (Fig. 4A-B-C). A dispensable gene heatmap showed that each strain contained strain-specific genes (Fig. 4D). A total of 217 genes were specific to *V. harveyi* 345 with 10 on p345-67, 94 on ChI, 17 on ChII, and 96 on p345-185 (Additional File 8: Table S7, Fig. 5A). Twenty-nine genes were annotated by Non-Redundant (NR) database and the others were hypothetical protein-encoding genes (Additional File 8: Table S7, Fig. 5B). Four annotated genes encoded MGEs, including two type IV conjugative transfer proteins (CU052_28715 and CU052_28750) and two recombinases (CU052_29120 and CU052_00050) on p345-185 or p345-67. Four (CU052_13455, CU052_13535, CU052_13525, and CU052_13445) were phage-related genes on ChI. Four genes (CU052_14535, CU052_14240, CU052_23640, and CU052_13020) on Chr I encoded membrane proteins. One beta-lactamase class C-encoding gene (CU052_00920) was on Chr I. Two DNA methylase-encoding genes (CU052_28385 and CU052_28885) belonging to the bacterial immune system were on p345-185. One virulence-associated protein D (CU052_13320) was on Chr I (Additional File 8: Table S7).

Strain-specific gene family

A strain-specific gene family was present in one of the 31 strains. Seven gene families were strain specific to *V. harveyi* 345 with > 50% identity to *Pseudovibrio* spp., *Photobacterium* spp., *Escherichia* spp., and other *Vibrio* spp. (*V. parahaemolyticus*) and were involved in recombination, peptide synthesis, conjugation and integration (Table 5).

HGT candidates on chromosomes

CU052_00920 in *V. harveyi* 345 encoded a beta-lactamase class C consisting of 486 amino acids. It shared no similarity with any protein of *V. harveyi*, *Vibrio* spp. or even γ -proteobacteria presently known. Instead, it shared more than 40% identity with the protein in four *Pseudovibrio* spp. (58.76–59.21% identity) suggesting recent acquisition of this gene from a *Pseudovibrio* strain (Fig. 6A and D). Two proteins were annotated, one as a serine hydrolase and the other as a class A beta-lactamase-related serine hydrolase.

CU052_13320 encoded a virulence-associated protein D of 94 amino acids. It shared 40–75.53% identity with *Pseudomonas* spp., *Aeromonas* spp., *Neisseria* spp., *Haemophilus* spp., *Rodentibacter* spp., *Alysiella* spp., *Kingella* spp., *Haemophilus* spp., and *Halomonas* spp. No similarity with any *V. harveyi* or known *Vibrio* spp. proteins was detected (data not shown). The top 15 similar strains were selected for muti-

alignment and phylogenetic analysis (Fig. 6B and E). CU052_13320 was closest to a virulence factor in *Pseudomonas aeruginosa* (73.40% identity) and *Aeromonas sobria* (75.53% identity), suggesting HGT of this gene from *P. aeruginosa* or *A. sobria* (Fig. 6B and E).

The CU052_14535 gene, encoding a 167 amino acid membrane protein in *V. harveyi* 345, was identified as an OmpA family protein in many other homologs. CU052_14535 showed more than 40% identity with many strains of *Photobacterium* spp., *Salinivibrio* spp., *Aeromonas* spp., *Ferrimonas* spp., *Grimontia* spp., *Aliivibrio* spp., and other *Vibrio* species, but not with *V. harveyi* (data not shown). The top 15 similar strains were selected for multi-alignment and phylogenetic analysis (Fig. 6C and F). CU052_14535 was close to an outer membrane beta-barrel protein in *V. splendidus* (97.04% identity), and an OmpA family protein in *V. crassostreae* (97.04% identity), suggesting recent acquisition of this gene from *V. splendidus* or *V. crassostreae* (Fig. 6C and F).

HGT candidates on plasmids

Five antibiotic resistance genes and one virulence gene, CU052_28670, involved in immune evasion were detected on plasmid p345-185. Basic local alignment search tool (BLAST) of nucleotide (Blastn) results showed high homolog (query cover > 40% and identity > 40%) of p345-185 with plasmids pVPS62, V36, pVPS114, pVP2, and pVPS91 from *V. parahaemolyticus*; pVAS114 and pVAS19 from *V. alginolyticus*; and pAQU1 from *Photobacterium damsela*, but no homology with other *V. harveyi* plasmids (Fig. 7A). Plasmid p345-185 showed 99.97% homolog (query cover = 99%) with pVPS62. CU052_29140 (*sul2*) was not found on 345 chromosomes and showed 100% amino acid identity to *sul2* in *Acinetobacter baumannii*, *Escherichia coli*, *Actinobacillus pleuropneumoniae*, *Histophilus somni*, and *Salmonella enteric* (Fig. 7B).

Nucleotide sequence accession numbers

The complete genome sequence of *V. harveyi* 345 was deposited in GenBank with accession numbers CP025537 (ChI), CP025538 (ChII), CP025539 (plasmid p345-185) and CP025540 (plasmid p345-67).

Discussion

V. harveyi 345 was isolated and studied because of its multidrug resistance and serious virulence to *E. oanceolustus* in Shenzhen, Southern China. The complete genome sequence of *V. harveyi* 345 was determined and compared with 30 *V. harveyi* strains. *V. harveyi* 345 was assembled into two circular chromosomes and two plasmids. It had a larger genome than 27 strains other than KC13.17.5, E385, and 74F and had a consistent G + C content with the 30 others strains. Compared to the 30 *V. harveyi* strains, *V. harveyi* 345 had more predicted genes than the others strains except for ZJ0603. In total, 487 virulence genes encoding proteins in flagella, iron uptake, pili, LPS, CPS, chemotaxis and type II/III/IV/VI secretion involved in fish disease in clinics, were identified in *V. harveyi* 345. In addition, 25 ARGs including genes for resistance to aminoglycoside (*golc*, *adeb*, *emre*), penicillin (*pbp2*, *pbp1a*, *pbp1b*, *pbp2*), tetracycline (*tet34*, *tetm*, *tetb*), chloramphenicol (*adeb*, *ceob*, *mdtI*) and trimethoprim (*dfra26*) were

found in *V. harveyi* 345, consistent with its resistance to streptomycin, ampicillin, tetracycline, chloramphenicol, trimethoprim, and sulfamethoxazole.

Virulence genes and antimicrobial-resistance genes are acquired by bacterial replication and by HGT. Horizontal transfer is the exchange of genetic material within species without any sexual mechanism [43]. This phenomenon is widely documented in bacteria and has a role in bacterial evolution and adaptation [43]. Comparative genomics of 31 *V. harveyi* strains was conducted to analyze HGT events by identifying strain-specific genes and strain-gene families. Considering the average gene number of 5,087 for the 31 *V. harveyi* strains, the 3,338 core genes represented approximately 66% of the total genome, meaning that approximately two-thirds of the genome was conserved among all strains. However, flower plots and dispensable heatmaps showed that each strain contained strain-specific genes, probably obtained by HGT. A total of 217 genes were specific to *V. harveyi* 345, which was more than other strains except for ZJ0603. This result suggested that a large number of HGT events happened in 345. In addition, seven gene families were specific to *V. harveyi* 345. Most (80.65%, 175/217) of the strain-specific genes had >40% identity to other *Vibrio* species. The remaining came from other genera such as *Shewanella*, *Photobacterium*, *Pseudovibrio*, and *Escherichia*. We focused on the characterization of three strain-specific genes, CU052_00920, CU052_13320, and CU052_14535. CU052_00920 encoded a class C beta-lactamase and was transferred from *Pseudovibrio* spp. Beta-lactamase enzymes are reported to inactivate beta-lactam antibiotics by hydrolyzing the peptide bond of the characteristic four-membered beta-lactam ring, rendering the antibiotic ineffective [44]. This result suggested that CU052_00920 contributed to antibiotic resistance, especially ampicillin resistance of *V. harveyi* 345. CU052_13320 encoded a virulence-associated protein D and closes to a virulence factor in *Pseudomonas aeruginosa* (73.40% identity) and *Aeromonas sobria* (75.53% identity). CU052_13320 probably contributed to the virulence of 345. CU052_14535 encoded an OmpA family protein and was acquired from *V. splendidus* or *V. crassostreae*. Among pathogenic bacteria, OmpA proteins are important for pathogenesis including bacterial adhesion, invasion, or intracellular survival and evasion of host defenses or stimulation of proinflammatory cytokine production [45]. This result suggested that CU052_14535 regulates the virulence of *V. harveyi* 345. DNA can also be horizontally transmitted between bacteria through plasmids, phages or uptake of naked DNA from environment [46]. Homology analysis showed that p345-185 had 99.97% homology (query cover = 99%) to pVPS62 in *V. parahaemolyticus*. PVPS62 is reported to be a pAQU-type plasmid and emerge MDR conjugative plasmid among important pathogens [47]. Five antimicrobial-resistance genes (*tetM*, *tetB*, *qnrS*, *dfra17*, and *sul2*) involved in resistance to tetracycline, fluoroquinolone, trimethoprim, and sulfonamide, and one virulence gene (CU052_28670) involved in immune evasion were located on plasmid p345-185. These genes could be horizontally transferred, promoting drug resistance and virulence. Especially, CU052_29140 (*sul2*) showing 100% amino acid identity to *sul2* in *Acinetobacter baumannii*, *Escherichia coli*, *Actinobacillus pleuropneumoniae*, *Histophilus somni*, and *Salmonella enteric*, should contribute to the resistance of pediatric compound sulfamethoxazole tablets and sulfamethoxazole. Similarly, Klein et al. [48] found that chloramphenicol, oxytetracycline and chlortetracycline could be successful transferred by R-plasmids.

Except for plasmids, HGT encompasses a variety of genetic units, collectively known as MGEs [49] and including phages, GIs, and integrating conjugative elements (ICEs). A total of 47 GIs, with encoding many transposases, integrases, and recombinases, and two incomplete prophage sequences were identified in *V. harveyi* 345. These genetic units probably acted as HGT delivery tools, contributing to pathogenesis, drug resistance and environmental adaptation of *V. harveyi* 345. Several type III/VI secretion proteins along with IS family transposases were predicted on the same GI that could be easily transmitted. Dissemination of these genes could further compromise *Vibrio* infections, limiting treatment options. These results further indicated that HGT contributed to virulence and antibiotic resistance of *V. harveyi* 345.

The 345 strain was isolated from a diseased *E. oanceolustus* in Shenzhen, Southern China. Recent increasing aquaculture density has led to frequent breeding diseases and increased use of antibiotics in Southern China, resulting in serious antibiotic residues and pollution [50]. Intensification of coastal urbanization has led to massive discharge of pollutants (heavy metals, nutrients, biocides), and intensification of human activities has caused environmental climate change, especially global warming [51]. Many studies show that the wide use of antibiotics; pollution by heavy metals, nutrients and biocides; and global climate change regulate HGT by affecting plasmid replication, changing phage activity, adjusting HGT-related enzyme activity, and damaging immune systems. These factors regulate bacterial resistance and pathogenicity [52, 53, 54]. For example, Beaber et al. [53] found that ciprofloxacin induces transfer of SXT, an ICE derived from *V. cholera* encoding genes that confer resistance to chloramphenicol, sulphamethoxazole, trimethoprim and streptomycin. The mechanism is increasing expression of SXT activators, enhancing drug resistance. Similarly, high environmental temperatures in combination with UV irradiation accelerate the spread of *stx* (Shiga toxin) genes by enhancing Stx prophage induction and Stx phage-mediated gene transfer [52]. Therefore, environmental pollution and climate change may increase the communication of virulence genes and drug resistance genes by affecting HGT. The result will enhance the toxicity and drug resistance of *V. harveyi* and affect the ecological safety of aquaculture ecosystems.

Conclusions

We present the first complete genome of the serious disease-causing *V. harveyi* 345 which has a multidrug-resistant phenotype. We did comparative genomics between this strain and 30 other *V. harveyi* strains. The genome was determined to study *V. harveyi* 345 virulence and antimicrobial resistance. Multiple virulence factors and resistance genes were predicted in this strain, consistent with results of Jiang et al [38]. We searched for evidence of HGT and evaluated genomic traits relating to HGT. The high quality complete genome sequence generated in this study will aid further studies for a deeper understanding of the mechanisms of *Vibrio* pathogenesis and antibiotic resistance. These studies could improve seafood quality and reduce economic loss. We recommend that more research be done on the response mechanism of HGT relative to environmental change and antibiotic use. This research will be an important scientific basis for predicting outbreaks and controlling *V. harveyi* disease. As environmental pollution including use of antibiotics and climate change enhance the virulence and drug resistance of

pathogens, research should explore management approaches, for example by bacteriophages, to manage the occurrence of *V. harveyi* in the environment [55].

Methods

DNA extraction

V. harveyi 345 was grown overnight in 2216E medium (BD, USA) at 28°C with vigorous shaking. Overnight cultures were inoculated with a 1:1,000 dilution into 20 ml fresh 2216E medium and grown until OD₆₀₀ = 0.5. Cells were collected by centrifugation and genomic DNA extracted using the CTAB method [56]. The quantity and quality of genomic DNA were evaluated using a Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA) and 1% agarose gel electrophoresis. DNA was stored at -80°C and prepared for genome sequencing.

Whole genome sequencing and assembly

A PacBio RSII 20-kb template library was constructed from at least 10 µg genomic DNA at the Beijing Genomics Institute (BGI). The library was subjected to quality control (QC) with Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA) and Nanodrop 2000 (Thermo Fisher Scientific, Wilmington, DE, USA) to check the concentration, purity and integrity of library templates. Libraries were sequenced on Illumina HiSeq 4000 and Pacbio RSII sequencing platforms. Genomes were primarily assembled with a variety of software for short sequence assembly based on clean data from the Illumina HiSeq 4000 platform. RS_HGAP Assembly3 of SMRT Analysis v.2.3.0 was used to assemble data from the Pacbio RSII platform. Contigs were error corrected successively by soapSNP and soapIndel software and GATK analysis. Gaps between contigs were filled by PCR.

General annotation

ORFs were predicted by Glimmer v.3.02 [57], and genome annotation was completed using the NCBI prokaryotic genome automatic annotation pipeline with identity $\geq 40\%$ and E_value $\leq 10^{-5}$ (Blast and annotation for all study sets used the same identities and E_values). COG, KEGG, GO and NR databases were used to search domain architecture. The tRNA genes were identified by tRNAscan-SE v.1.23 [58] and rRNA genes with RNAmmer v.1.2 [59]. The sRNAs were predicted by blasting nucleic acid against Rfam database [60]. Tandem repeats were predicted by Tandem Repeat Finder v.4.04 [61].

Identification of virulence factors and ARGs

Putative virulence factors of *V. harveyi* 345 were predicted by blasting against the Virulence Factor Database (VFDB) (<http://www.mgc.ac.cn/VFs/main.htm>) [62] and ARGs by blasting against the Antibiotic Resistance Genes Database (ARDB) (<http://ardb.cbcb.umd.edu/>) [63].

Prediction of GIs, prophages, and CRISPR-Cas systems

GIs, prophages, and CRISPR-Cas systems were respectively identified using online tools IslandViewer 4, PHAge Search Tool (PHAST), and CRISPR Finder software [64-66].

Comparative genomics

Comparative genome analysis was conducted among 31 *V. harveyi* strains with *V. harveyi* VH2 as the reference strain and an average gene number of 5,087 (Additional File 7: Table S6) to assess *V. harveyi* evolution. A phylogenetic tree was constructed based on the single nucleotide polymorphism (SNP) matrix of the 31 strains with the maximum-likelihood method, bootstrapped 1,000 times via Treebest-1.9.2 software [67].

Extraction of pan-, core-, dispensable and strain-specific genes

Core-pan genes were analyzed using a BLAST method [68]: Genes were taken from reference genome *V. harveyi* VH2 as a gene pool. Genes predicted in one strain among other 30 strains (Query samples) were BLAST with the gene pool, and results filtered by length and identity (>40%). BLAST coverage ratios (BCR, reference BCR = match/reference length x 100%, query BCR = match/query length x 100%, and match stands for the length used for BLAST) of genes from the gene pool and Query samples were calculated separately. If BCR values from the reference and Query sample were smaller than the setting value (40%), the reference gene did not have homology with the Query gene. The nonhomologous gene from the Query genome was added to the gene pool. Query samples were processed and the final gene pool used as the pan gene pool. The non-homolog genes from the reference genome were the strain specific genes against the Query sample. Nonhomologous genes were BLAST with the gene pool of another Query sample, repeated for all samples, and the final nonhomologous gene pool was used as the strain-specific gene pool.

After aligning to genes from samples, BCR values of genes from the pan gene pool were calculated for each sample. The coverage array was generated for this pool. If the BCR value of a gene in each sample was larger than the setting value, the gene was a core gene. If the gene was predicted from an assembled result, BLAST results were filtered and the sequence was removed if the number N (N means uncertain nucleotides in gene) was larger than the setting (30%) for the gene. Dispensable genes were included in the pan-gene pool but not the core-gene pool.

Construction of gene families

Gene families were constructed using genes from the reference strain and the target bacterium. Protein sequences were aligned with basic local alignment search tool of protein (Blastp) and redundancy was eliminated. Gene family TreeFam clustering was carried out with alignment results and Hcluster_sg software. Alignment results for proteins were converted into multiple sequence amino acids in the CDS area after multiple sequence alignments with the clustered gene family using Muscle software [69]. Gene family tree construction analysis was carried out for multiple sequence alignment results using Muscle and neighborhood-joining (NJ) method with Treebest software [67].

Characterization of putative HGT

Of the strain-specific genes, we extracted CU052_00920, encoding a beta-lactamase class C; CU052_13320, encoding a virulence-associated protein D; and CU052_14535 encoding an OmpA family protein for further analysis on contribution to the antibiotic resistance or virulence of *V. harveyi* 345. The genes were compared to the NR protein database of NCBI using the Blastp tool. The results were filtered manually to find significant hits to proteins belonging to species other than *V. harveyi*. Multiple sequence alignment was conducted using ClustalW [70]. An alignment phylogenetic tree was constructed from aligned sequences using the Kimura 2-parameter model [71] with the neighbor-joining method, bootstrapped 1000 times, with MEGA6.0 software [72].

In addition, the entire sequence of p345-185 was compared to the NR nucleotide database of NCBI using a Blastn tool. A phylogenetic tree was constructed with the neighbor-joining method using max seq difference = 0.75 and BLAST pairwise alignments. The characterization method was the same for CU052_29140 (*sul2*) on p345-185 as for CU052_00920, CU052_13320, and CU052_14535.

List Of Abbreviations

HGT: Horizontal gene transfer; ATG: Antibiotic-resistance gene; MDR: Multidrug resistance; MGE: Mobile genetic element; ICEs: Integrative and conjugative elements; GIs: Genetic islands; RM: restriction modification; LD50: Median lethal dose; ORFs: Open reading frames; COG: Clusters of Orthologous Groups; KEGG: Kyoto Encyclopedia of Genes and Genomes, GO: Gene Ontology; NR: Non-Redundant; TLH: Thermolabile hemolysin; VFDB: Virulence Factor Database; ARDB: Antibiotic Resistance Genes Database; T6SS: Type VI secretion system; BGI: Beijing Genomics Institute; QC: Quality control; PHAST: PHAge Search Tool; SNP: Single nucleotide polymorphism; BLAST: Basic Local Alignment Search Tool, NJ: Neighborhood-joining.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Data generated or analysed during this study are included in this published article and its supplementary information files. Genomic data are available from the GenBank (accession numbers: CP025537, CP025538, CP025539 and CP025540).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the Hainan Provincial Natural Science Foundation of China (319QN336), the Natural Science Fund of Guangdong (2018A030310695), the National Natural Science Foundation of China (31902415, 41806147), Guangdong YangFan Innovative & Entrepreneurial Research Team Program (2016YT03H038), and the China Agriculture Research System (CARS-48).

Authors' contributions

LX isolated the strain. YD, HX and JF designed the project. JW extracted the DNA. YD, YS and ZG performed the computational analyses. YD interpreted the data and wrote the manuscript. SL and CC critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank International Science Editing (<http://www.internationalscienceediting.com>) for editing this manuscript.

References

1. Egidius E. Vibriosis: pathogenicity and pathology. A review. *Aquaculture*. 1987;67(1-2):15-28.
2. Osunla CA, Okoh AI. *Vibrio* pathogens: a public health concern in rural water resources in Sub-Saharan Africa. *Int J Environ Res Public Health*. 2017;14(10):1188.
3. Jiravanichpaisal P, Miyazaki T, Limsuwan C. Histopathology, biochemistry, and pathogenicity of *Vibrio harveyi* infecting black tiger prawn *Penaeus monodon*. *J Aquat Anim Health*. 1994;6(1):27-35.
4. Zorrilla I, Arijó S, Chabrillon M, Diaz P, Martínez-Manzanares E, Balebona MC, et al. *Vibrio* species isolated from diseased farmed sole, *Solea senegalensis* (Kaup), and evaluation of the potential virulence role of their extracellular products. *J Fish Dis*. 2003;26(2):103-8.
5. Karunasagar I, Pai R, Malathi GR, Karunasagar I. Mass mortality of *Penaeus monodon* larvae due to antibiotic-resistant *Vibrio harveyi* *Aquaculture*. 1994;128(3-4):203-9.
6. Nishimori E, Hasegawa O, Numata T, Wakabayashi H. *Vibrio carchariae* causes mass mortalities in Japanese abalone, *Sulculus diversicolor supratexta*. *Fish Pathol*. 1998;33:495-502.
7. Luna GM, Bongiorno L, Gili C, Biavasco F, Danovaro R. *Vibrio harveyi* as a causative agent of the White Syndrome in tropical stony corals. *Environ Microbiol Rep*. 2010;2(1):120-7.
8. Kraxberger-Beatty T, McGarey DJ, Grier HJ, Lin DV. *Vibrio harveyi*, an opportunistic pathogen of common snook, *Centropomus undecimalis* (Bloch), held in captivity. *J Fish Dis*. 1990;13:557-60.

9. Saeed MO. Association of *Vibrio harveyi* with mortalities in cultured marine fish in Kuwait. *Aquaculture*. 1995;136(1-2):21-9.
10. Company R, Sitj A, Bobadilla, Pujalte MJ, Garay E, Alvarez-Pellitero P, et al. Bacterial and parasitic pathogens in cultured common dentex, *Dentex dentex* *J Fish Dis*. 1999;22(4):299-309.
11. Becker P, Gillan D, Lanterbecq D, Jangoux M, Rasolofonirina R, Rakotovao J, et al. The skin ulceration disease in cultivated juveniles of *Holothuria scabra* (Holothuroidea, Echinodermata). *Aquaculture*. 2004(1-4);242:13-30.
12. Ransangan J, Lal T M, Al-Harbi A H. Characterization and experimental infection of *Vibrio harveyi* isolated from diseased Asian seabass (*Lates calcarifer*). *Malaysian Journal of Microbiology*, 2012, 8(2): 104-15.
13. Yingxue Q, Jun W, Yongquan S, Dexiang Wang, Xinzhong Chen. Studies on the pathogenic bacterium of ulcer disease in *Epinephelus awoara*. *Acta Oceanologica Sinica*. 2018,(1):154-59.
14. Shen GM, Shi CY, Fan C, Jia D, Wang SQ, Xie GS et al. Isolation, identification and pathogenicity of *Vibrio harveyi*, the causal agent of skin ulcer disease in juvenile hybrid groupers *Epinephelus fuscoguttatus* × *Epinephelus lanceolatus*. *J Fish Dis*. 2017;40:1351-62.
15. Xu X, Liu K, Wang S, Guo W, Xie Z, Zhou Y. Identification of pathogenicity, investigation of virulent gene distribution and development of a virulent strain-specific detection PCR method for *Vibrio harveyi* isolated from Hainan Province and Guangdong Province, China. *Aquaculture*. 2017;468:226-34.
16. Yano Y, Hamano K, Satomi M, Tsutsui I, Ban M, Aue-Umneoy D. Prevalence and antimicrobial susceptibility of *Vibrio* species related to food safety isolated from shrimp cultured at inland ponds in Thailand. *Food Control*. 2014;38(4):30-6.
17. Letchumanan V, Yin WF, Lee LH, Chan KG. Prevalence and antimicrobial susceptibility of *Vibrio parahaemolyticus* isolated from retail shrimps in Malaysia. *Front Microbiol*. 2015;6:33.
18. Devi R, Surendran PK, Chakraborty K. Antibiotic resistance and plasmid profiling of *Vibrio parahaemolyticus* isolated from shrimp farms along the southwest coast of India. *World J Microb Bio*. 2009;25(11):2005-12.
19. Ji H, Chen Y, Guo Y, Liu X, Wen J, Liu H. Occurrence and characteristics of *Vibrio vulnificus* in retail marine shrimp in china. *Food Control*, 2011;22(12):1935-40.
20. Letchumanan V, Pusparajah P, Tan LTH, Yin WF, Lee LH, Chan KG. Occurrence and antibiotic resistance of *Vibrio parahaemolyticus* from shellfish in Selangor, Malaysia. *Front Microbiol*. 2015;6:1417.
21. Yano Y, Yokoyama M, Satomi M, Oikawa H, Chen SS. Occurrence of *Vibrio vulnificus* in fish and shellfish available from markets in china. *J Food Prot*. 2004;67(8):1617-23.
22. Baker-Austin C, McArthur JV, Tuckfield RC, Najarro M, Lindell AH, Gooch J, et al. Antibiotic resistance in the shellfish pathogen *Vibrio parahaemolyticus* isolated from the coastal water and sediment of georgia and south carolina, USA. *J Food Protect*. 2008;71:2552-8.

23. Kang CH, Shin YJ, Kim WR, Kim YG, Song K C, Oh E G, et al. Prevalence and antimicrobial susceptibility of *Vibrio parahaemolyticus*, isolated from oysters in Korea. *Environ Sci Pollut R*. 2016;23(1):918-26.
24. Lopatek M, Wieczorek K, Osek J. Prevalence and antimicrobial resistance of *Vibrio parahaemolyticus* isolated from raw shellfish in Poland. *J Food Protect*. 2015;78(5):1029-33.
25. Scarano C, Spanu C, Ziino G, Pedonese F, Dalmasso A, Spanu V, et al. Antibiotic resistance of *Vibrio* species isolated from *Sparus aurata* reared in Italian mariculture. *New Microbiol*. 2014;37(3):329-37.
26. Oh EG, Son KT, Yu H, Lee TS, Lee HJ, Shin S, et al. Antimicrobial resistance of *Vibrio parahaemolyticus* and *Vibrio alginolyticus* strains isolated from farmed fish in Korea from 2005 through 2007. *J Food Protect*. 2011; 74(3):380-6.
27. He Y, Jin L, Sun F, Hu Q, Chen L. Antibiotic and heavy-metal resistance of *Vibrio parahaemolyticus* isolated from fresh shrimps in Shanghai fish markets, China. *Environ Sci Pollut R*. 2016;23(15):15033-40.
28. Nakayama T, Ito E, Nomura N, Nomura N, Matsumura M. Comparison of *Vibrio harveyi* strains isolated from shrimp farms and from culture collection in terms of toxicity and antibiotic resistance. *FEMS Microbiol Lett*. 2010;258(2):194-9.
29. Austin B, Zhang XH. *Vibrio harveyi*: a significant pathogen of marine vertebrates and invertebrates. *Lett Appl Microbiol*. 2006;43(2):119-24.
30. Natrah FM, Ruwandepika HA, Pawar S, Karunasagar I, Sorgeloos P, Bossier P, et al. Regulation of virulence factors by quorum sensing in *Vibrio harveyi*. *Vet Microbiol*. 2011;154(1-2):124-9.
31. De GP, Wright GD. Antibiotic resistance by enzyme inactivation: from mechanisms to solutions. *Chembiochem*. 2010;11(10):1325-34.
32. Martins A, Spengler G, Molnár J, Amaral L. Bacterial Antibiotic Resistance. 2014;eLS.
33. Sobecky PA, Hazen TH. Horizontal gene transfer and mobile genetic elements in marine systems. *Methods Mol Biol*. 2009;532:435-53.
34. Park J, Zhang Y, Bouboltz AM, Zhang XQ, Schuster SC, Ahuja U, et al. Comparative genomics of the classical *Bordetella* subspecies: the evolution and exchange of virulence-associated diversity amongst closely related pathogens. *BMC Genomics*. 2012;13(1):545.
35. Le PT, Ramulu HG, Guijarro L, Paganini J, Gouret P, Chabrol O, et al. An automated approach for the identification of horizontal gene transfers from complete genomes reveals the rhizome of *Rickettsiales*. *Bmc Evol Biol*. 2012;12(1):243-243.
36. Jiang W, Marraffini LA. CRISPR-Cas: New tools for genetic manipulations from bacterial immunity systems. *Annu Rev Microbiol*. 2015;69(1):209-28.
37. Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol R Mmbr*. 2013;77(1):53-72.
38. Karakulska J, Kaczyn´Ska M, Furowicz AJ, Fijałkowski K. Mobile genetic elements as carriers of antimicrobial resistance genes and transfer of drug resistance in *Acinetobacter*-selected problems.

- Advan Agricul Sci. 2010;5(1):73-81.
39. Hu HW, Wang JT, Li J, Li JJ, Ma YB, Chen D, et al. Field-based evidence for copper contamination induced changes of antibiotic resistance in agricultural soils. *Environ Microbiol.* 2016;18:3896-909.
 40. Sun M, Ye M, Wu J, Feng Y, Wan J, Tian D, et al. Positive relationship detected between soil bioaccessible organic pollutants and antibiotic resistance genes at dairy farms in Nanjing, Eastern China. *Environ Pollut.* 2015;206:421-8.
 41. Jiang K, Xu LW, Su YL, Ma HL, Liu GF, Guo ZX, et al. Isolation and identification of two pathogenetic *Vibrio harveyi* strains from pearl gentian. *Ecol Sci.* 2017;36:16-24.
 42. Coulthurst SJ. The Type VI secretion system-a widespread and versatile cell targeting system. *Res Microbiol.* 2013;164(6):640-54.
 43. Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. *Nat Rev Genet.* 2015;16(8):472-82.
 44. Majiduddin FK, Materon IC, Palzkill TG. Molecular analysis of beta-lactamase structure and function. *Int J Med Microbiol.* 2002;292(2):127-37.
 45. Confer AW, Ayalew S. The OmpA family of proteins: Roles in bacterial pathogenesis and immunity. *Vet Microbiol.* 2013;163(3-4):207-22.
 46. Calamari D, Zuccato E, Castiglioni S, Bagnati R, Fanelli R. Strategic survey of therapeutic drugs in the rivers Po and Lambro in northern Italy. *Environ Sci Technol.* 2016;37:1241-8.
 47. Li R, Ye L, Wong MHY, Zheng Z, Chan EWC, Chen S. Evolution and comparative genomics of paqu-like conjugative plasmids in *Vibrio* J Antimicrob Chemoth. 2017;72:2503-6.
 48. Klein BU, Siesenop U, Boehm K. Investigations on transferable antibiotic resistance through R-plasmids between obligate and facultative fish pathogenic bacteria. *B Eur Assoc Fish Pat.* 1996;16:138-42.
 49. Lang A, Beatty JT, Rice PA. Guest editorial: Mobile genetic elements and horizontal gene transfer in prokaryotes. *Curr Opin Microbiol.* 2017;38:v-vii.
 50. Liu X, Steele JC, Meng XZ. Usage, residue, and human health risk of antibiotics in Chinese aquaculture: a review. *Environ. Pollut.* 2017;223:161-69.
 51. Yu M, Carmichael GR, Zhu T, Cheng Y. Sensitivity of predicted pollutant levels to urbanization in China. *Atmos Environ.* 2012; 60(none):544-54.
 52. Tamang MD, Sunwoo H, Jeon B. Phage-mediated dissemination of virulence factors in pathogenic bacteria facilitated by antibiotic growth promoters in animals: a perspective. *Anim Health Res Rev.* 2017;18:160-6.
 53. Beaber JW, Hochhut B, Waldor MK. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature.* 2004;427(6969):72-4.
 54. Mizuno T, Mizushima S. Signal transduction and gene regulation through the phosphorylation of two regulatory components: the molecular basis for the osmotic regulation of the porin genes. *Mol Microbiol.* 1990;4(7):1077-82.

55. Letchumanan V, Chan KG., Pusparajah P, Saokaew S, Duangjai A, Goh BH, et al. Insights into bacteriophage application in controlling *Vibrio* Front Microbiol, 2016;7:1114.
56. Chen WP, Kuo TT. A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. Nucleic Acids Res. 1993;21(9):2260.
57. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics. 2007(6);23:673-9.
58. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955-64.
59. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35(9):3100-8.
60. Griffiths-Jones S. Rfam: an RNA family database. Nucleic Acids Res. 2003;31(1):439-41.
61. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573-80.
62. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 2005;33:D325-8.
63. Bo L, Mihai P. ARDB-Antibiotic resistance genes database. Nucleic Acids Res. 2009;37:D443-7.
64. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 2017;45:W30-5.
65. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011;39:W347-52
66. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35:W52-7.
67. Tannistha N, Catherine O, Pratap SA, Justin B, Timothy A, Mitali ST, et al. A genomic survey of positive selection in *Burkholderia pseudomallei* provides insights into the evolution of accidental virulence. Plos Pathog. 2010;6(4):e1000845.
68. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464:59-65.
69. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5(1):113-20.
70. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using clustalw and clustalx. Curr Protoc Bioinformatics. 2002;Chapter 2(Unit 2), Unit 2.3.
71. Srivathsan A, Meier R. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics. 2012;28(2):190-4.
72. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0 . Mol Biol Evol . 2013;30:2725-9.

Tables

Table 1 Genome features of *V. harveyi* 345

Features	ChI	ChII	Plasmid p345-185	Plasmid p345-67
Length (bp)	3,713,225	2,220,396	185,327	66,874
G + C content (%)	44.81	44.81	43.48	43.48
Genes	3,345	2,044	208	81
rRNA gene	28	3	0	0
tRNA genes	113	15	0	0
sRNA genes	6	7	0	0
Minisatellite DNA	26	15	2	1
Microsatellite DNA	22	11	1	1
GenBank accession no.	CP025537	CP025538	CP025539	CP025540

Table 2 The RM system of *V. harveyi* 345

Gene_id	Gene name	Products
CU052_00675	<i>hsdR</i>	type I restriction enzyme, R subunit
CU052_10560	<i>hsdR</i>	type I restriction enzyme, R subunit
CU052_10570	<i>hsdS</i>	type I restriction enzyme, S subunit
CU052_10575	<i>hsdM</i>	type I restriction enzyme, M protein
CU052_17245	<i>hsdM</i>	type I restriction enzyme, M protein
CU052_17265	<i>hsdR</i>	type I restriction enzyme, R subunit
CU052_17270	<i>hsdS</i>	type I restriction enzyme, S subunit
CU052_17275	<i>hsdM</i>	type I restriction enzyme, M protein
CU052_17285	<i>hsdM</i>	type I restriction enzyme, M protein
CU052_27190	<i>hsdM</i>	type I restriction enzyme, M protein
CU052_16835	<i>mod</i>	adenine-specific DNA-methyltransferase
CU052_16840	<i>res</i>	type III restriction enzyme

Table 3 Virulence factors predicted with VFDB database

Virulence factor	Annotation	Chromosome	Identity genus
Adherence			
<i>acfA</i>	ACF	II	Vibrio
<i>llpA</i>	llpA	I	Vibrio
<i>VP1611</i>	MAM7	I	Vibrio
<i>mshH, mshI, mshJ, mshK, mshL, mshM, mshN, mshE,</i>	MSHA type IV pilus	I, II	Vibrio
<i>mshG, mshF, mshA, mshD, mshA, mshA, mshA, mshA,</i>			
<i>ompU</i>	OmpU	I	Vibrio
<i>aatC, aatC</i>	TCP	II	Escherichia
<i>flp-1, flp-1, tadB, tadC, tadA, flp, rcpC, rcpA, tadZ, tadA,</i>	Flp type IV pili	I, II	Vibrio, Aggregatibacter
<i>tadB, tadC, tadD, tadE, tadF, tadG, rcpA, tadA</i>			
<i>vfr, rpoN, rpoS, pilW, fimT, pilE, pilM, pilN, pilO, pilP,</i>	type IV pili	I, II	Vibrio, Pseudomonas, Salmonella,
<i>pilQ, pilT, pilU, pilF, pilT, pilR, pilR, pilR, pilW, pilA,</i>			Dichelobacter, Yersinia, Haemophilus
<i>pilR, pilR, pilR, pilD, pilC, pilB</i>			
<i>CJJ81176_1161, CFF8240_1412, CFF8240_1412,</i>	LOS	I, II	Campylobacter, Haemophilus
<i>C8J_1080, kdsB, lpxK, msbA, kdsA, lpxC, gmhA/lpcA,</i>			
<i>htrB, lgtF, lgtF, opsX/rfaC, kdkA, msbB, rfaD, orfM, htrB,</i>			
<i>galE, lpxA, lpxB, gmhA/lpcA, htrB, lpxH,</i>			

yhxB/manB,

galE, msbA, rfaE, neuC1, lpxD

tuf, tuf

EF-Tu

I

Mycoplasma

htpB

Hsp60

I

Legionella

lap

Listeria adhesion
protein

I

Listeria

CT396, CT396

MOMP

I

Chlamydia

Motility

*flaD, flaA, flaL, flaK, flaJ, flal, flaH, flaG,
flaF, flaE,*

Flagella

I, II

Vibrio,
Pseudomonas,
Salmonella,

*flaD, flaC, flaB, cheR, cheA, cheB, cheV,
cheV, cheW,*

Legionella,
Aeromonas,
Pseudomonas

*cheW, cheY, cheZ, filM, flaB, flaD, flaD, flaE,
flaG, flal,*

*fleQ, fleQ, fleR, flgA, flgA, flgB, flgC, flgD,
flgE,*

*flgF, flgG, flgH, flgl, flgJ, flgK, flgL, flgM,
flgM flgN,*

*flgN, flgO, flgP, flgT, flhA, flhA, flhB, flhB,
flhF, flhG,*

*fliA, fliD, fliE, fliE, fliF, fliF, fliG, fliG, fliH, fliH,
fliI,*

*fliI, fliJ, fliJ, fliK, fliL, fliM, fliN, fliN, fliO, fliP,
fliP,*

*fliQ, fliQ, fliR, fliR, fliS, fliY, flmH, flmH, flmH,
flrA,*

flrB, flrC, lafA, lafB, lafC, lafD, lafE, lafF,

lafK,

lafS, lafU/motB, lafW, motA, motA, motB,

motX, motY,

motY, scrA, scrB, scrC, scrG, B565_1123

Regulation

luxS, adeG AI-2 I, II Vibrio, Acinetobacter

cqsA CAI-1 II Vibrio

Secretion system

epsA, epsB, epsC, epsE, epsF, epsG, epsH, epsI, epsJ, EPS T2SS I Vibrio

epsK, epsL, epsM, epsN, gspD

exsA, exsD, sycN, tyeA, vcrD, vcrG, vcrH, vcrR, vcrV, T3SS1 I Vibrio

vecA, virG, vopB, vopD, vopN, vopQ, vopR, vscB, vscC,

vscD, vscF, vscG, vsch, vscl, vscJ, vsck, vscl, vscN,

vscO, vscQ, vscR, vscS, vscT, vsxU, vsxX, vscY, vscC

hopAJ2, hopJ1, mlr6326, ati2, ati1 Other T3SS I, II Pseudomonas, Mesorhizobium, Aeromonas

lpg2936, trwD, virB1, virB10, CbuK_1823, CbuG_0575, T4SS I, II Legionella, Bartonella, Brucella, Coxiella

CbuG_1738, COXBURSA331_A0369

PFLU6009, Psyr_2628, clpV1, pppA, clpV1, Psyr_2628, T6SS I, II Pseudomonas, Burkholderia, Citrobacter,

<i>tssH/clpV/bscF, cts2A, lip2, impB, impB, impC, impC</i>			Agrobacterium
Toxin			
<i>secA2</i>	Ace	I	Mycobacterium
<i>tth</i>	Thermolabile hemolysin	II	Vibrio (<i>Vibrio parahaemolyticus</i>)
<i>clbQ, clbM, clbL, clbI, clbH, clbF, clbE, clbD, clbC</i>	Colibactin	I	Escherichia
<i>vctA</i>	Enterobactin	II	Vibrio
<i>ast</i>	Enterotoxin	I	Aeromonas
Immune evasion			
<i>ABK1_0097, ABK1_0097, BC5270, BC5275, BC5275,</i>	Capsule	I, II	Vibrio, Staphylococcus, Mycoplasma,
<i>BCE_5384, cap8J, cpsA, cpsB, cpsB, cpsC, cpsD, cpsE,</i>			Enterococcus, Acinetobacter, Neisseria,
<i>cpsF, cpsG, cpsJ, eps9, eps9, kpsF, kpsF, oppF, oppF,</i>			Enterococcus, Campylobacter, Burkholderia,
<i>oppF, rmlB, rmlD, sicC/synD, uppS, wbfB, wbfV/wcvB,</i>			Bacillus
<i>wcbN, wecA, wza, wzb, wzc</i>			
<i>fabZ, acpXL, kbtB, waaZ, waaF, gtrB</i>	LPS	I	Brucella, Helicobacter, Pseudomonas, Shigella
<i>ACICU_00876, ASA_0824, ASA_3329, cbsB, ccmA, ccmB,</i>	Iron uptake	I, II	Vibrio, Acinetobacter, Aeromonas,

<i>ccmC, ccmE, ccmF, ccmF, chuY, ciuD, fur,</i>			Burkholderia,
<i>hemA, hemB,</i>			Dickeya, Escherichia,
<i>hemC, hemE, hemG, hemH, hemL, hemN,</i>			Haemophilus,
<i>hemN, hitA, hitB,</i>			Legionella,
			Mycobacterium,
<i>hitC, hitC, hmuU, hutA, huvB, huvC, huvD,</i>			Pseudomonas,
<i>huvX, huvZ,</i>			Salmonella,
			Shigella,
<i>iutA, mbtI, orbG, orbH, pchH, phuV,</i>			Yersinia
<i>PSEEN2499, psuA, pvdH,</i>			
<i>pvsA, pvsB, pvsC, pvsD, pvsE, pvuA, pvuB,</i>			
<i>pvuC, pvuD,</i>			
<i>pvuE, sitA, sitB, sitC, sitD, vabC, vctC, vctC,</i>			
<i>vctD, vctD,</i>			
<i>vctG, vctG, vctP, viuC</i>			
Antiphagocytosis			
<i>algB, algW, algW, algU, mucP, algR</i>	Alginate	I, II	Pseudomonas
<i>pgi, mrsA/glmM</i>	Exopolysaccharide	I	Haemophilus
<i>hasD, hasB, hasF</i>	Hyaluronic acid capsule	I, II	Pseudomonas, Serratia, Yersinia
<i>kata, katA, clpC, clpE, clpP, ctpV, ibpA,</i>	Stress protein	I, II	Neisseria, Listeria,
<i>msrA/B (pilB),</i>			Mycobacterium,
<i>regX3, regX3, sodB, sodB, sodC1</i>			Legionella, Salmonella
Exoenzyme			
<i>icl</i>	Isocitratase	I	Mycobacterium
<i>mip, mip</i>	Mip	I, II	Legionella
<i>Regulation</i>			Mycobacterium, Legionella,

			Salmonella
<i>mprA, mprA, mprA, relA, rpoS, csrA, sigA</i>	Regulation	I	
Efflux pump			
<i>mtrD</i>	MtrCDE	I	Neisseria
<i>panC</i>	Metabolic adaptation		Mycobacterium
Others			
<i>csgD, csgD, glnA1, ML1683, ML1683, leuD, MGA_1142,</i>	Others		
<i>ipaH, ECS88_3547, ndk, pdhB, ppsE, cysC1, argK, eno, plr/gapA,</i>		I, II	Vibrio, Acinetobacter, Escherichia,
<i>hpt, sugC, sugC, sugC, sugC, sugC, trpD, bfmR, bfmR, bfmR, bfmR,</i>			Listeria, Mycobacterium, Mycoplasma,
<i>vibD, vibE, vibA, VVA1298, VVA1298, VVA1298, vctP</i>			Pasteurella, Pseudomonas, Salmonella,
			Shigella, Streptococcus

Table 4 Antimicrobial resistance genes predicted with the ARDB database

Gene_id	Identity (%)	Resistance_Type	Antibiotic_Resistance	Description
CU052_02205	47.53	<i>pbp2</i>	penicillin	penicillin-binding protein 2
CU052_02455	99.35	<i>tet34</i>	tetracycline	oxytetracycline resistance phosphoribosyltransferase domain-containing protein Tet(34)
CU052_03895	47.65	<i>tolc</i>	aminoglycoside, glycyclcycline, macrolide, beta_lactam,acriflavin	outer membrane channel protein TolC
CU052_04185	67.79	<i>ksga</i>	kasugamycin	16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase RsmA
CU052_04200	40.24	<i>dfra26</i>	trimethoprim	type 3 dihydrofolate reductase
CU052_05755	49.16	<i>mexw</i>	-	multidrug efflux RND transporter permease subunit
CU052_07910	46.51	<i>pbp1a</i>	penicillin	PBP1A family penicillin-binding protein
CU052_09355	42.63	<i>pbp1b</i>	penicillin	penicillin-binding protein 1B
CU052_09890	49.11	<i>pbp2</i>	penicillin	penicillin-binding protein 2
CU052_13950	51.93	<i>adeb</i>	chloramphenicol, aminoglycoside	multidrug efflux RND transporter permease subunit
CU052_15205	98.25	<i>norm</i>	tigecycline, streptomycin, kanamycin, ciprofloxacin, norfloxacin	MATE family efflux transporter
CU052_15655	43.59	<i>emre</i>	aminoglycoside	QacE family quaternary

				ammonium compound efflux SMR transporter
CU052_16225	48.04	<i>bcr</i>	–	Bcr/CflA family drug resistance efflux transporter
CU052_21770	40.74	<i>ykkc</i>	na_antimicrobials	QacE family quaternary ammonium compound efflux SMR transporter
CU052_23115	43.06	<i>ceob</i>	chloramphenicol	multidrug efflux RND transporter permease subunit
CU052_25305	44.41	<i>emrd</i>	–	multidrug transporter EmrD
CU052_26580	43.94	<i>catb5</i>	chloramphenicol	antibiotic acetyltransferase
CU052_27055	58.33	<i>qnra</i>	fluoroquinolone	Qnr family pentapeptide repeat protein
CU052_27855	62.23	<i>mdtl</i>	chloramphenicol	multidrug transporter MdtL
CU052_28095	100	<i>tetm</i>	tetracycline	tetracycline resistance ribosomal protection protein Tet(M)
CU052_28120	100	<i>tetb</i>	tetracycline	tetracycline efflux MFS transporter Tet(B)
CU052_28525	67.43	<i>qnrs</i>	fluoroquinolone	quinolone resistance pentapeptide repeat protein QnrVC6
CU052_28540	64.47	<i>dfra17</i>	Trimethoprim	trimethoprim-resistant dihydrofolate reductase DfrA
CU052_29140	100	<i>sul2</i>	sulfonamide	sulfonamide-resistant dihydropteroate synthase Sul2

Table 5 Strain-specific gene families of *V. harveyi* 345

Gene	Locus	Identity	E_value	Subject	Subject_description
Family				ID	
ID					
1	CU052_06110	91.98	1.00E-125	gi 752301544 ref WP_041157164.1	hypothetical protein, partial [Vibrio mytili]
1	CU052_29120	100	0.00E+00	gi 446775679 ref WP_000852935.1	recombinase [Escherichia coli]
2	CU052_00970	62	0.00E+00	gi 359344462 gb AEV37836.1	Putative peptide synthetase [Pseudovibrio sp. FO-BEG1]
2	CU052_00980	55.33	0.00E+00	gi 504051870 ref WP_014285864.1	non-ribosomal peptide synthetase [Pseudovibrio sp. FO-BEG1]
3	CU052_00975	61.49	2.00E-125	gi 504051870 ref WP_014285864.1	non-ribosomal peptide synthetase [Pseudovibrio sp. FO-BEG1]
3	CU052_00965	62.42	0.00E+00	gi 359344462 gb AEV37836.1	Putative peptide synthetase [Pseudovibrio sp. FO-BEG1]
4	CU052_28750	100	8.00E-84	gi 760459713 gb AJP18243.1	MULTISPECIES: type IV conjugative transfer system protein TraA [Vibrio parahaemolyticus]
4	CU052_28755	100	3.00E-84	gi 691549048 ref WP_032073022.1	hypothetical protein [Vibrio sp. 04Ya090]
5	CU052_05575	75.81	6.00E-93	gi 647339569 ref WP_025768281.1	integrase [Vibrio parahaemolyticus]
5	CU052_29060	100	2.00E-107	gi 504199683 ref WP_014386785.1	hypothetical protein [Photobacterium damsela]
6	CU052_00925	50.69	0.00E+00	gi 359344454 gb AEV37828.1	Polyketide synthase [Pseudovibrio sp. FO-BEG1]

6	CU052_01010	56.65	0.00E+00	gi 359344469 gb AEV37843.1	Polyketide synthase [Pseudovibrio sp. FO-BEG1]
7	CU052_28405	96.77	2.00E-80	gi 504199762 ref WP_014386864.1	hypothetical protein [Photobacterium damsela]
7	CU052_28260	100	6.00E-96	gi 504199733 ref WP_014386835.1	hypothetical protein [Photobacterium damsela]

Additional Files

Additional File 1: Figure S1. KEGG pathway classifications of *V. harveyi* 345 coding sequences. tif file

Additional File 2: Figure S2. GO functional classifications of *V. harveyi* 345 coding sequences. tif file

Additional File 3: Table S1 Global comparison of pollutant (antibiotics, heavy metals, and nutrients) concentrations of water samples. doc file

Additional File 4: Table S2 Details of virulence factors predicted with the VFDB database. xls file

Additional File 5: Table S3 Genomic islands in a single chromosome (CP025537) of *V. harveyi* 345. xls file

Additional File 6: Table S4 Genomic islands in a single chromosome (CP025538) of *V. harveyi* 345. xls file

Additional File 7: Table S5 Prophage sequences present in *V. harveyi* 345 chromosomes. xls file

Additional File 8: Table S6 Gene information and gene statistics for 31 *V. harveyi* strains. xls file

Additional File 9: Table S7 Strain-specific genes of *V. harveyi* 345. xls file

Figures



Figure 1

Circular representation of the *V. harveyi* 345 genome. Inner to outer: genome size, all annotated genes, forward-strand genes, reverse-strand genes, tRNA, rRNA, sRNA, GC and GC-skew.

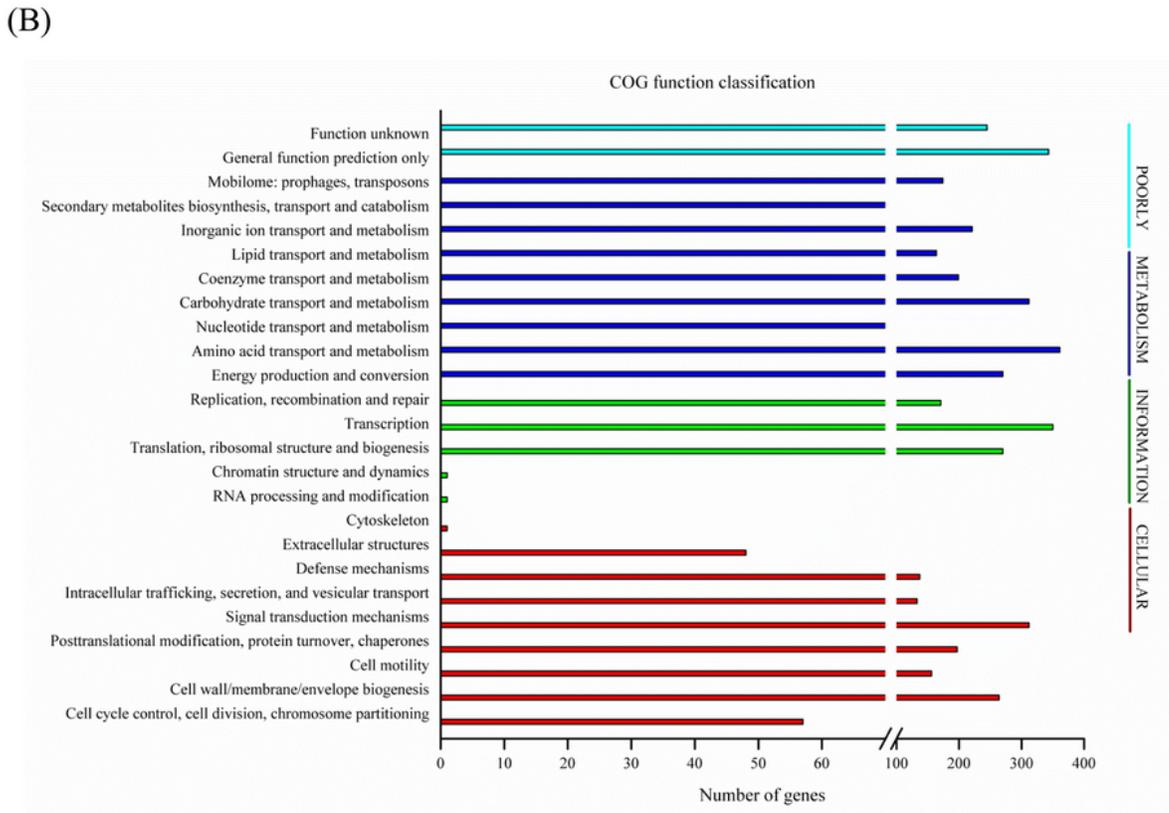
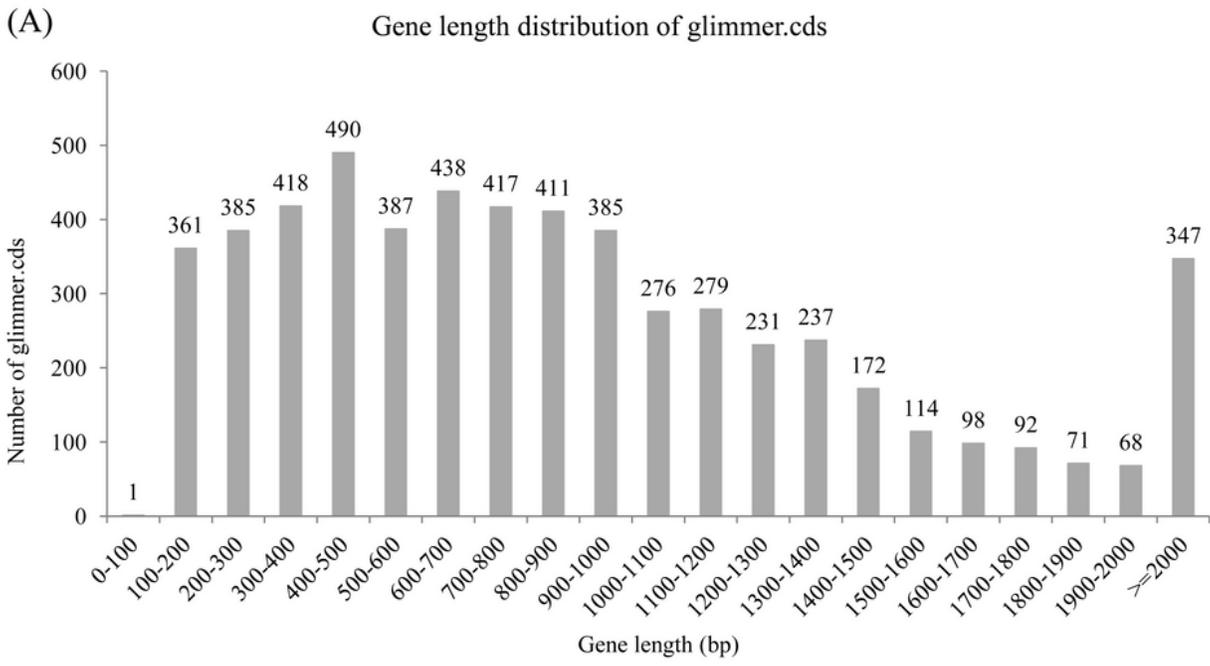


Figure 2

Coding sequence length distribution and COG functional classifications. Coding sequence length distribution of *V. harveyi* 345 (A). COG functional classifications of *V. harveyi* 345 coding sequences (B).

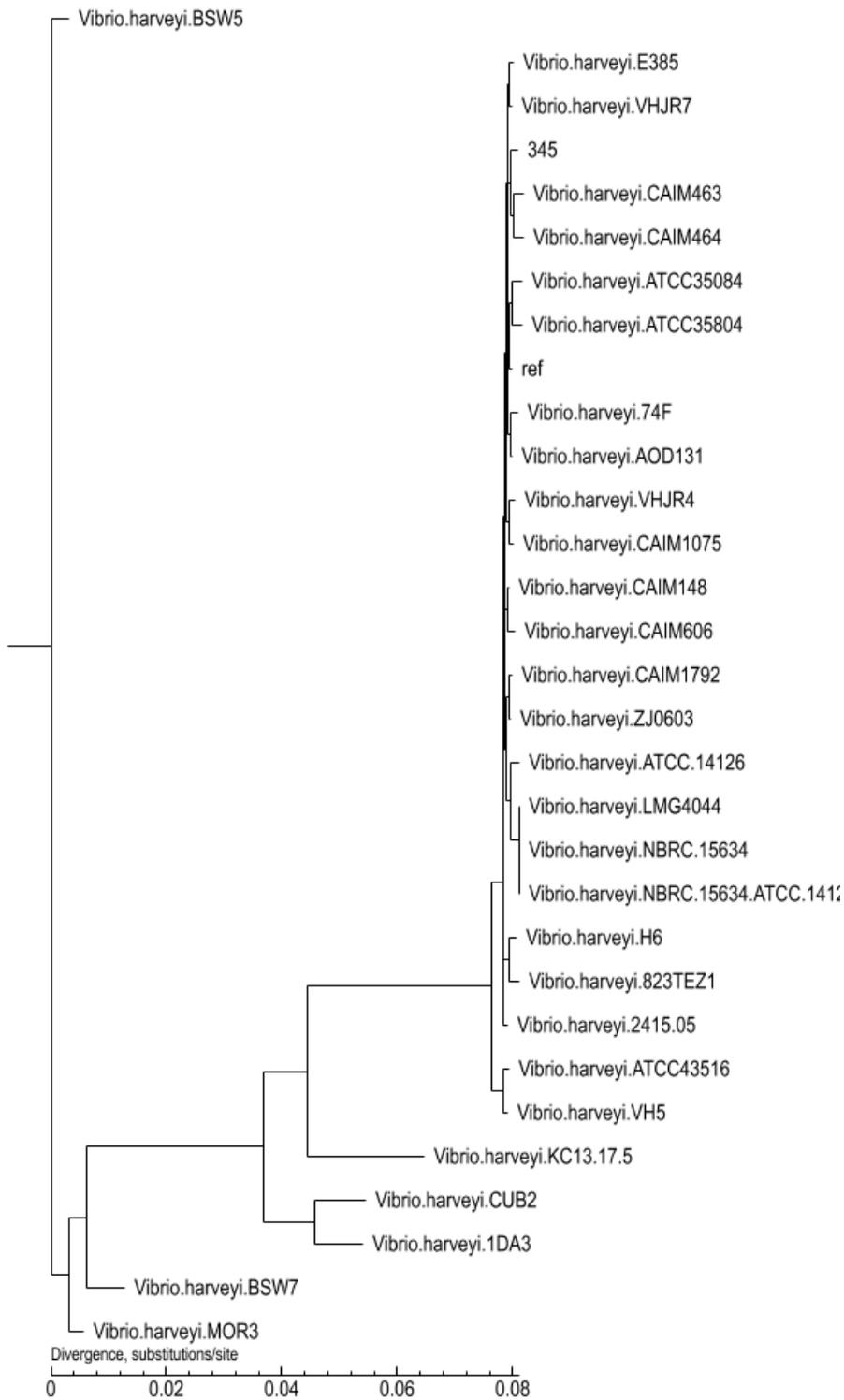


Figure 3

Phylogenetic relationship of *V. harveyi* 345 and 30 compared *V. harveyi* strains. Results are based on the SNP matrix of the 31 strains and Treebest-1.9.2 software and maximum-likelihood method.

Strain	Description
1	<i>V. harveyi</i> VH2 (the reference)
2	<i>V. harveyi</i> 1DA3
3	<i>V. harveyi</i> CAIM606
4	<i>V. harveyi</i> NBRC.15634.ATCC.14126
5	<i>V. harveyi</i> CAIM464
6	<i>V. harveyi</i> CAIM463
7	<i>V. harveyi</i> ZJ0603
8	<i>V. harveyi</i> KC13.17.5
9	<i>V. harveyi</i> 823TEZ1
10	<i>V. harveyi</i> VHS
11	<i>V. harveyi</i> E385
12	<i>V. harveyi</i> NBRC.15634
13	<i>V. harveyi</i> BSW5
14	<i>V. harveyi</i> ATCC35804
15	<i>V. harveyi</i> CUB2
16	<i>V. harveyi</i> CAIM1075
17	<i>V. harveyi</i> AOD131
18	<i>V. harveyi</i> CAIM148
19	<i>V. harveyi</i> MOR3
20	<i>V. harveyi</i> 2415.05
21	<i>V. harveyi</i> ATCC43516
22	<i>V. harveyi</i> 74F
23	<i>V. harveyi</i> VHJR4
24	<i>V. harveyi</i> VHJR7
25	<i>V. harveyi</i> H6
26	<i>V. harveyi</i> BSW7
27	<i>V. harveyi</i> ATCC.14126
28	<i>V. harveyi</i> ATCC35084
29	<i>V. harveyi</i> CAIM1792
30	<i>V. harveyi</i> LMG4044
31	<i>V. harveyi</i> 345

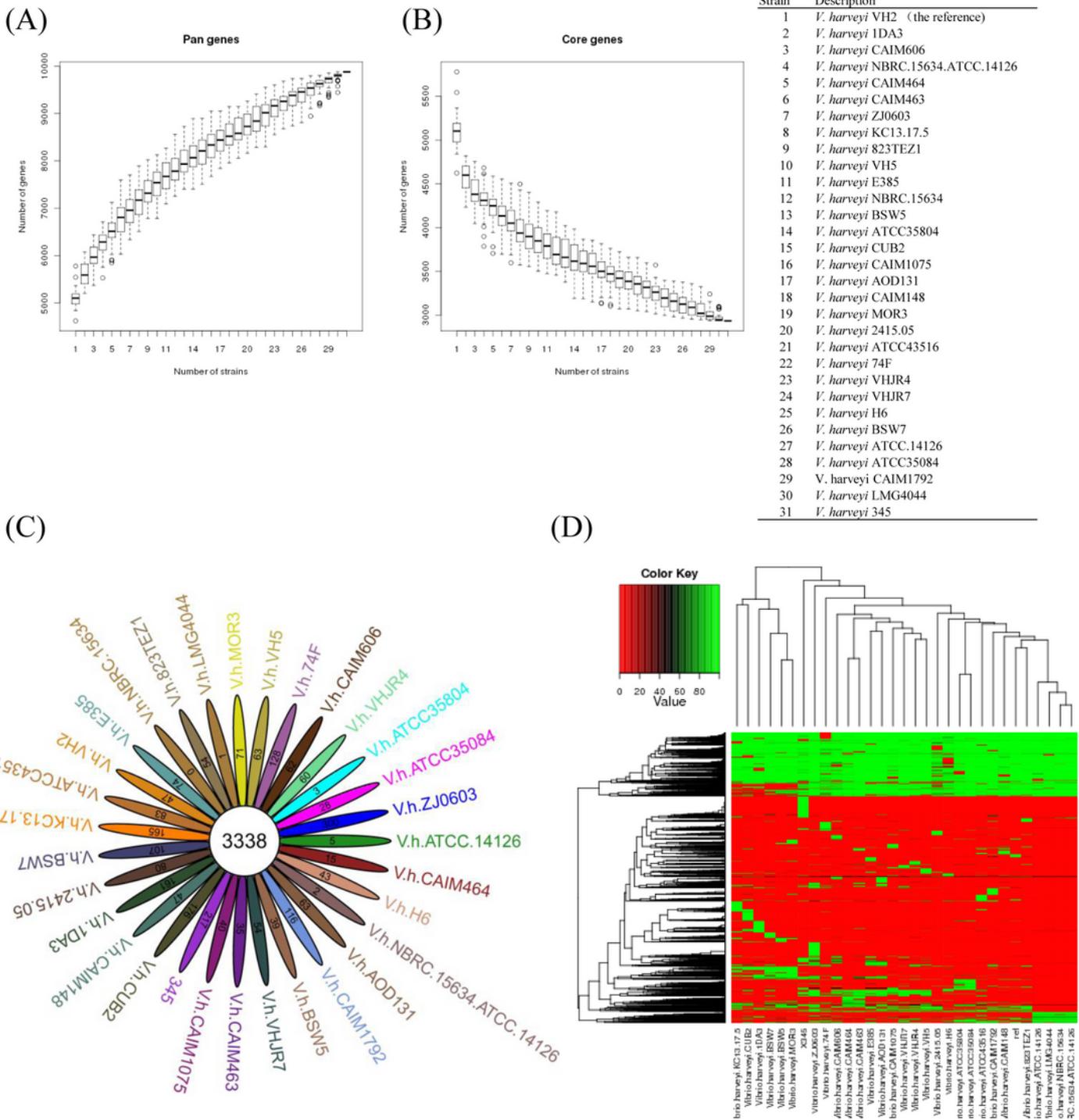


Figure 4

Pan-core genes, strain-specific and dispensable genes. Pan-gene dilution curve of 31 *V. harveyi* strains (A), core-gene dilution curve of the 31 strains (B), flower plot of strain-specific genes and core-genes (C), and heat map after removal of core-gene (dispensable heatmap) (D).

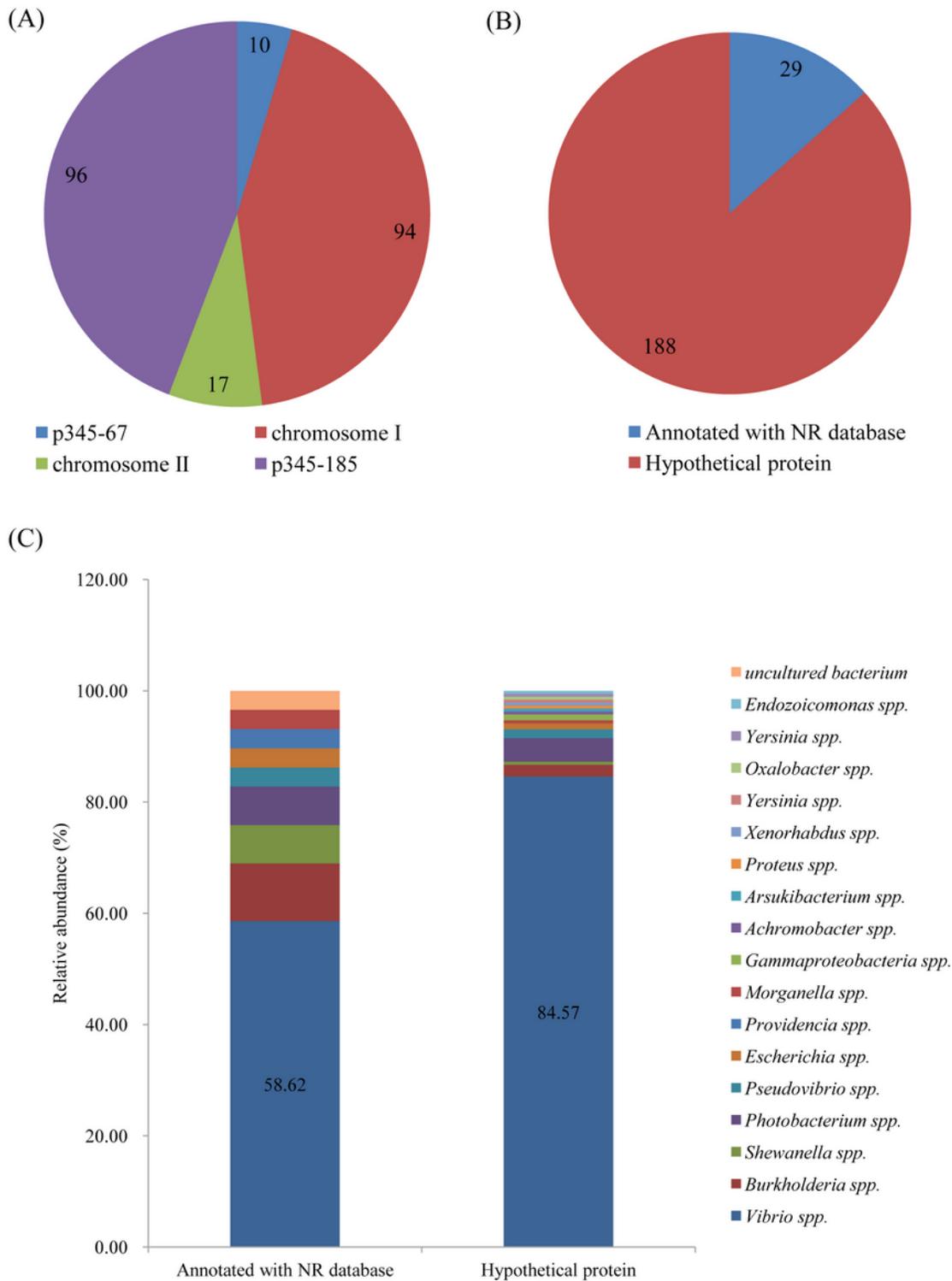


Figure 5

Strain-specific genes of *V. harveyi* 345. Strain-specific genes locations (A). Gene number of hypothetical proteins and proteins annotated by NR Database (B). Composition and relative abundance of strain-specific genes (C).

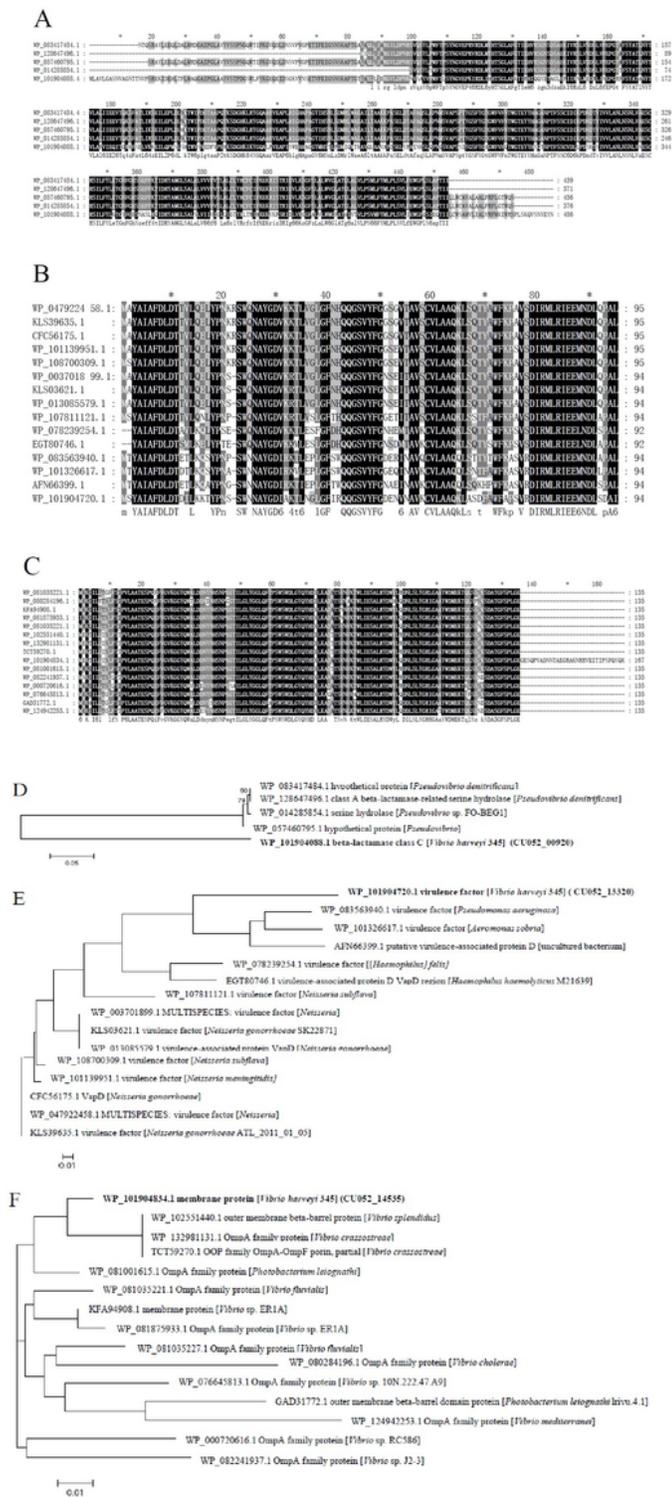


Figure 6

Evolutionary analysis of CU052_00920, CU052_13320, and CU052_14535. Multisequence alignment (A-C) and phylogenetic analysis (D-F) of CU052_00920, CU052_13320, and CU052_14535 with amino acids. Multisequence alignment was by ClustalW and phylogenetic tree was constructed with MEGA 6. Bar substitutions per sequence position: 0.05 (D), 0.01(E), and 0.01(F).

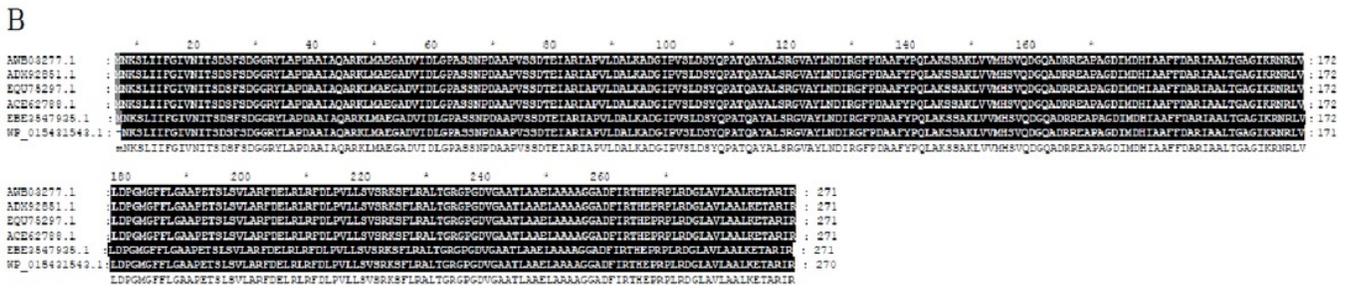
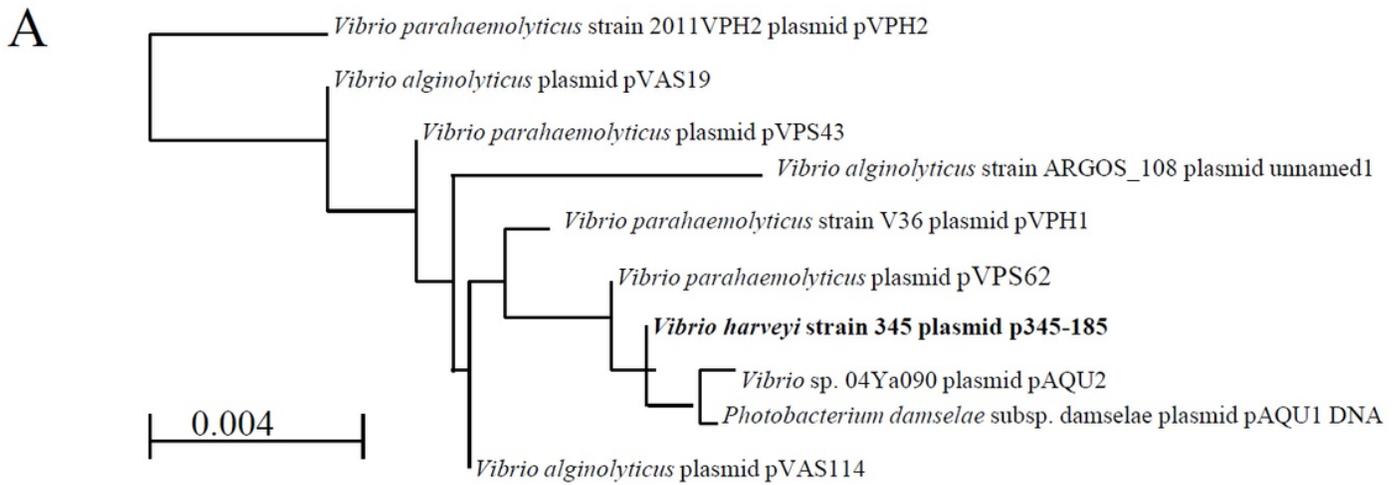


Figure 7

Evolutionary analysis of p345-185 and CU052_29140. Nucleotides phylogenetic analysis of p345-185 (A) and amino acids multisequence alignment of CU052_29140 (B). Phylogenetic tree was constructed with the neighbor-joining method, max seq difference = 0.75 using BLAST pairwise alignments. Multisequence alignment was by ClustalW.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS3.xlsx](#)
- [TableS4.xls](#)
- [TableS2.xls](#)
- [TableS7.xls](#)
- [TableS1.pdf](#)
- [FigureS2.jpg](#)
- [FigureS1.jpg](#)

- [TableS5.xls](#)
- [TableS6.xlsx](#)