Data in Brief

Draft genome sequence of pathogenic bacteria *Vibrio parahaemolyticus* strain Ba94C2, associated with acute hepatopancreatic necrosis disease isolate from South America

Leda Restrepo [a,*], Bonny Bayot [a], Irma Betancourt [a], Andres Pinzón [b]

[a] *Escuela Superior Politécnica del Litoral, ESPOL, Centro Nacional de Acuicultura e Investigaciones Marinas, CENAIM, Campus Gustavo Galindo Km 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador*
[b] *Grupo de Investigación en Bioinformática y Biología de Sistemas, Instituto de Genética, Universidad Nacional de Colombia, Bogotá, Colombia*

ARTICLE INFO

ABSTRACT

Vibrio parahaemolyticus is a pathogenic bacteria which has been associated to the early mortality syndrome (EMS) also known as hepatopancreatic necrosis disease (AHPND) causing high mortality in shrimp farms. Pathogenic strains contain two homologous genes related to insecticidal toxin genes, PirA and PirB, these toxin genes are located on a plasmid contained within the bacteria. Genomic sequences have allowed the finding of two strains with a divergent structure related to the geographic region from where they were found. The isolates from the geographic collection of Southeast Asia and Mexico show variable regions on the plasmid genome, indicating that even though they are not alike they still conserve the toxin genes. In this paper, we report for the first time, a pathogenic *V. parahaemolyticus* strain in shrimp from South America that showed symptoms of AHPND. The genomic analysis revealed that this strain of *V. parahaemolyticus* found in South America appears to be more related to the Southeast Asia as compared to the Mexican strains. This finding is of major importance for the shrimp industry, especially in regards to the urgent need for disease control strategies to avoid large EMS outbreaks and economic loss, and to determine its dispersion in South America. The whole-genome shotgun project of *V. parahaemolyticus* strain Ba94C2 have been deposited at DDBJ/EMBL/GenBank under the accession PRJNA335761.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications	
Organism	*Vibrio parahaemolyticus*
Strain	Ba94C2
Sequencer	Illumina MiSeq
Data format	assembled
Experimental factors	microbial strain originally isolated from necrotizing lesions on the hepatopancreas of shrimps
Experimental features	whole genome shotgun sequencing followed by genome assembly and gene description
Consent	N/A
Sample source location	South America

1. Direct link to deposited data

Deposited data can be found here: http://www.ncbi.nlm.nih.gov/bioproject/335761.

* Corresponding author.
E-mail address: ledarestrepoc@gmail.com (L. Restrepo).

2. Experimental design, materials and methods

Vibrio parahaemolyticus is a member of the family Vibrionaceae and can be found in diverse marine habitats, which is typically known to cause food-borne gastroenteritis in humans. The pathogen causing EMS or AHPND has been identified as a *V. parahaemolyticus* strain by Tran et al. [1]. One bacterial strain was isolated from shrimp hepatopancreas obtained in cultured *Penaeus (Litopenaeus) vannamei*. The bacterial DNA sample was prepared according to the phenol-chloroform method of Sambrook and Russell [2]. The genome of *V. parahaemolyticus* strain was generated using an Illumina TruSeq DNA PCR-Free Library (Illumina Cambridge Ltd., UK) was performed on the Illumina MiSeq PE300 platform. We obtained 1,235,827 reads, for a total of 280.9 Mbp. After trimming and quality analysis, one hundred and fifteen contigs were *de novo* assembled using SPAdes [3], for a genome size of 2.07 Mbp (46.0 coverage average). We also assembled the sequenced genome by reference, by means of Bowtie using *Vibrio parahaemolyticus* RIMD 2210633 genome as template (5.16 Mbp size, with a GC content of 45.40%). Gene prediction was performed by GenePRIMP [4]. Structural and functional annotation was performed over coding sequences using RAST server [5]. Briefly, sequences were annotated using BLAST against non-

redundant protein databases from NCBI (*E*-value ≤ 1e − 7 and minimal alignment length percentage ≥ 50%), Swiss-Prot [6], Clusters of Orthologous Groups (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG) [7,8] and Gene Ontology (GO) [9].

The *V. parahaemolyticus* strain Ba94C2 *de novo* final assembly consisted of 40 scaffolds, 2.07 Mbp in size, with 113 untranslated RNA sequences (Table 1) and 1932 predicted coding regions, represented by two chromosomes (Ch) of approximately 3.28 Mbp (Ch I) and 1.877 Mbp (Ch II). We also found one large extrachromosomal plasmid, previously reported in all AHPND strains [10–13]. The contig number 16 (69.9 kbp length with a 66.9 coverage average) aligned significantly to *V. parahaemolyticus* strain 13–028/A3 plasmid from Vietnam [14]. The putative virulence genes hereby annotated were also identified and used to diagnose AHPND by PCR.

A homologue of the insecticidal Photorhabdus insect-related binary toxin PirA/B was also identified in this study. This one is flanked by an arrangement of SWAT-3 type transposases (VSWAT3 like transposases). This might suggest that the infection cluster could be a mobile element, which implies that it is interchangeable between different *Vibrio* populations. Nevertheless, in this region, we were not able to identify either resistance genes or secretion systems pivotal for the infection process. Therefore, this virulence cluster should be further tested to evaluate if it can only be transmitted between populations where organisms have a pre-adaptation to the acquisition of insecticidal production genes.

Most coding gene functions were predicted by the functional categorization of COGs and GO. Gene functional distribution showed three pathogenicity mechanism systems of invasion and intracellular resistance involved in the protein synthesis (SSU ribosomal proteins): SSU ribosomal protein S2p (SAe), SSU ribosomal protein S15p (S13e), SSU ribosomal protein S16p, SSU ribosomal protein S18p (zinc-independent), SSU ribosomal protein S6p, SSU ribosomal protein S7p (S5e), SSU ribosomal protein S12p (S23e) and translation elongation factor G. We also found at least twelve systems that contribute to the resistance of antibiotics and toxic compounds, related to copper homeostasis and tolerance, and multidrug resistance efflux pumps. Likewise, we found sections with this kind of system also described in previous studies for *V. cholerae*, which are: tetracycline resistance and ribosome protection type (translation elongation factor G paralog), cobalt-zinc-cadmium resistance protein, CopG protein, copper-translocating and copper homeostasis protein CutE, magnesium and cobalt efflux protein CorC, acriflavin resistance protein, multi antimicrobial extrusion protein (Na (+)/drug antiporter) and MATE family of MDR efflux pumps. We were also able to annotate at least three secretion systems (T1SS, T2SS and T4SS).

In conclusion, in this study, we sequenced a *V. parahaemolyticus* strain isolated from a shrimp pond in South America; this being a pathogenic bacteria related to AHPND. This bacterial strain exhibited the putative toxin genes (PirA/B) as well as a plasmid sequence similar to the one previously reported in Southeast Asia isolates, but less related to the strain from Mexico. The results of this study warrant a more profound analysis of this emergent disease and the implementation of disease

control strategies for cultured shrimp to avoid future disease outbreaks. This may also contribute to progress fundamental knowledge on pathogens to develop new disease control strategies.

3. Nucleotide sequence accession number

The whole-genome shotgun project of *V. parahaemolyticus* Ba94C2 have been deposited at DDBJ/EMBL/GenBank under the accession PRJNA335761.

References

[1] L. Tran, L. Nunan, R.M. Redman, L.L. Mohney, C.R. Pantoja, K. Fitzsimmons, D.V. Lightner, Determination of the infectious nature of the agent of acute hepatopancreatic necrosis syndrome affecting penaeid shrimp. Dis. Aquat. Org. 105 (1) (2013) 45–55.

[2] J. Sambrook, D.W. Russell, Purification of nucleic acids by extraction with phenol: chloroform. Cold Spring Harb. Protoc. 1 (2006) (pdb-prot4455).

[3] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (5) (2012) 455–477.

[4] A. Pati, N.N. Ivanova, N. Mikhailova, G. Ovchinnikova, S.D. Hooper, A. Lykidis, N.C. Kyrpides, GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat. Methods 7 (6) (2010) 455–457.

[5] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, et al., The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9 (1) (2008) 1.

[6] B. Boeckmann, A. Bairoch, R. Apweiler, M.C. Blatter, A. Estreicher, E. Gasteiger, et al., The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31 (1) (2003) 365–370.

[7] R.L. Tatusov, N.D. Fedorova, J.D. Jackson, A.R. Jacobs, B. Kiryutin, E.V. Koonin, et al., The COG database: an updated version includes eukaryotes. BMC Bioinf. 4 (1) (2003) 1.

[8] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, M. Hattori, The KEGG resource for deciphering the genome. Nucleic Acids Res. 32 (Suppl. 1) (2004) D277–D280.

[9] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, et al., Gene Ontology: tool for the unification of biology. Nat. Genet. 25 (1) (2000) 25–29.

[10] H. Kondo, S. Tinwongger, P. Proespraiwong, R. Mavichak, S. Unajak, R. Nozaki, I. Hirono, Draft genome sequences of six strains of *Vibrio parahaemolyticus* isolated from early mortality syndrome/acute hepatopancreatic necrosis disease shrimp in Thailand. Genome Announc. 2 (2) (2014), e00221–14.

[11] B. Gomez-Gil, S. Soto-Rodríguez, R. Lozano, M. Betancourt-Lozano, Draft genome sequence of *Vibrio parahaemolyticus* strain M0605, which causes severe mortalities of shrimps in Mexico. Genome Announc. 2 (2) (2014), e00055–14.

[12] Y.T. Yang, I.T. Chen, C.T. Lee, C.Y. Chen, S.S. Lin, L.I. Hor, et al., Draft genome sequences of four strains of *Vibrio parahaemolyticus*, three of which cause early mortality syndrome/acute hepatopancreatic necrosis disease in shrimp in China and Thailand. Genome Announc. 2 (5) (2014), e00816–14.

[13] J.E. Han, K.F. Tang, L.H. Tran, D.V. Lightner, 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. Org. 113 (1) (2015) 33.

[14] Y.T. Yang, I.T. Chen, C.T. Lee, C.Y. Chen, S.S. Lin, L.I. Hor, et al., Draft genome sequences of four strains of *Vibrio parahaemolyticus*, three of which cause early mortality syndrome/acute hepatopancreatic necrosis disease in shrimp in China and Thailand. Genome Announc. 2 (5) (2014), e00816-14.

Table 1
Genome characteristics of *V. parahaemolyticus* strain Ba94C2 draft genome.

Attributes	Values
Assembly size (bp)	2,070,435
Total number of contigs	115
Contig N50 (bp)	34,478
Total number of scaffolds	40
Scaffold N50 (bp)	294,884
L50	3
GC content %	45.7
CDS	1932
rRNAs	12
tRNAs	101