<table>
<thead>
<tr>
<th>Title</th>
<th>Analysis of Vibrio seventh pandemic island II and novel genomic islands in relation to attachment sequences among a wide variety of Vibrio cholerae strains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Nguyen, Hai Tuan</td>
</tr>
<tr>
<td>Citation</td>
<td>Nagasaki University (長崎大学), 博士(医学) (2018-03-20)</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2018-03-20</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10069/38594">http://hdl.handle.net/10069/38594</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2018 The Authors. Microbiology and Immunology published by The Societies and John Wiley &amp; Sons Australia, Ltd. This is an open access article under the terms of the Creative Commons Attribution‐NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.</td>
</tr>
</tbody>
</table>
Analysis of *Vibrio* seventh pandemic island II and novel genomic islands in relation to attachment sequences among a wide variety of *Vibrio cholerae* strains

Tuan Hai Nguyen1, Tho Duc Pham1, Naomi Higa2, Hanako Iwashita2, Taichiro Takemura3, Makoto Ohnishi4, Kouichi Morita1,5 and Tetsu Yamashiro2

1Leading Program, Graduate School of Biomedical Sciences, Nagasaki University, 1-12-4 Sakamoto Nagasaki city, Nagasaki 852-8523, Japan, 2Department of Bacteriology, Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan, 3Department of Tropical Microbiology, Nagasaki University Institute of Tropical Medicine, Nagasaki, Japan, 4Department of Bacteriology I, National Institute of Infectious Diseases, Tokyo, Japan and 5Department of Virology, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan

**ABSTRACT**

*Vibrio cholerae* O1 El Tor, the pathogen responsible for the current cholera pandemic, became pathogenic by acquiring virulent factors including *Vibrio* seventh pandemic islands (VSP)-I and –II. Diversity of VSP-II is well recognized; however, studies addressing attachment sequence left (*attL*) sequences of VSP-II are few. In this report, a wide variety of *V. cholerae* strains were analyzed for the structure and distribution of VSP-II in relation to their attachment sequences. Of 188 *V. cholerae* strains analyzed, 81% (153/188) strains carried VSP-II; of these, typical VSP-II, and a short variant was found in 36% (55/153), and 63% (96/153), respectively. A novel VSP-II was found in two *V. cholerae* non-O1/non-O139 strains. In addition to the typical 14-bp *attL*, six new *attL*-like sequences were identified. The 14-bp *attL* was associated with VSP-II in 91% (139/153), and a short variant was found in 36% (55/153), and 63% (96/153), respectively. A novel VSP-II was found in two *V. cholerae* non-O1/non-O139 strains. In addition to the typical 14-bp *attL*, six new *attL*-like sequences were identified. The 14-bp *attL* was associated with VSP-II in 91% (139/153), whereas the remaining six types were found in 9.2% (14/153) of *V. cholerae* strains. Of note, six distinct types of the *attL*-like sequence were found in the seventh pandemic wave 1 strains; however, only one or two types were found in the wave 2 or 3 strains. Interestingly, 86% (24/28) of *V. cholerae* seventh pandemic strains harboring a 13-bp *attL*-like sequence were devoid of VSP-II. Six novel genomic islands using two unique insertion sites to those of VSP-II were identified in 11 *V. cholerae* strains in this study. Four of those shared similar gene clusters with VSP-II, except integrase gene.

**Key words** attachment sequence, genomic island, *Vibrio cholerae*, VSP-II.

*Vibrio cholerae*, the causative agent of the severe watery diarrheal disease cholera, has been spreading globally as a result of its high adaptation and ability to cause explosive outbreaks. Cholera still remains a significant public health concern in many areas of the world because of its high morbidity and mortality. Seven cholera pandemics have been described in human history since 1817 (1). The classical biotype is considered the cause of the first six cholera pandemics, whereas the seventh and current, which emerged from the Indonesian island of Sulawesi in 1961, is mainly caused by strains of the El Tor biotype. This current seventh pandemic is the longest pandemic that has ever been recorded, lasting more than half a century (2). It is the most disastrous pandemic in terms of geographical area and number of people infected (3). Analysis of representative historical and recent *V. cholerae* isolates indicated eight distinct phyletic lineages L1-L8, the L1 being classical isolates.

**Correspondence**

Tetsu Yamashiro, Department of Bacteriology, Graduate School of Medicine, University of the Ryukyus, 207 Uehara, Nishihara, Okinawa 903-0215, Japan. Tel: +81-98-895-1125, fax: +81-98-895-1408; email: tyamashi@med.u-ryukyu.ac.jp

Received 31 October 2017; revised 9 December 2017; accepted 26 December 2017.

**List of Abbreviations:** *attL* and *attR*, attachment sequence left and right; CI, circular intermediate; CTX, cholera toxin phage; GI, genomic island; NGS, next generation sequencing; ORF, open reading frame; SNP, single nucleotide polymorphism; VSP, *Vibrio* seventh pandemic island.
the L2 representing El Tor isolates of the seventh pandemic, which was further subdivided into three independent overlapping waves based on SNP in the genome and the type of CTX they harbored (4). Wave 1, wave 2 and wave 3 strains originated from a single ancestor in the Bay of Bengal (4, 5) and subsequently have spread to Asia, Africa and Latin America. The current seventh pandemic is believed to have emerged from a non-pathogenic strain after acquisition of important virulent factors: VSP-I, and —II, and El Tor type of CTX (6). VSP-I and —II were reported for the first time and described to be unique to the seventh pandemic El Tor strains by Dziejman et al. (7). VSP-II was originally identified as a 7.5-kb region which was subsequently described by O’Shea et al. to be part of a 26.9-kb region spanning from VC0490 to VC0516 (7, 8). VSP-I and —II were described to be received by the seventh pandemic strains by lateral gene transfer event and hypothesized to increase the fitness advantage of the isolates (7, 9, 10). The insertion site of VSP-II is at the tRNA-methionine locus, VC0516.1. The island is integrated between two attachment sequences, attL (14 bp) and attR (16 bp) (11, 12). Significant diversity of VSP-II was observed in previous studies with the main focus on the genomic variation of the island (8, 12, 13). Here we analyzed a wide variety of strains, presenting different biotypes and serogroups collected from numerous geographical locations and covering a wide time period. The present study aims to investigate the broader genetic variation of VSP-II, identify variation of VSP-II attachment site sequences, and identify VSP-II relevant GI. We categorized V. cholerae strains with SNP and invariable sites analyses, and pathogenicity and virulence profiles were estimated by determining species-specific gene (ompW), serogroup-specific genes (wbe O1, wbf O139), biotype-specific genes (tcpA, rstR, ctxB), virulence genes (ctxB, ctxA, tcpA). The analysis showed a novel VSP-II variant, six new categories of VSP-II attachment sequences, and four new GI sharing conserved gene clusters with VSP-II. Significantly, the 13-bp attachment sequence was found associated with the absence of VSP-II in the seventh pandemic lineage.

Materials and Methods

V. cholerae strains and DNA extraction

Out of 188 V. cholerae strains analyzed, 178 strains were of our laboratory collection, isolated from Asian and African countries. Colonies grown on nutrient agar were inoculated in Heart Infusion Broth (Nissui, Tokyo, Japan) and cultured at 37°C overnight with shaking. Genomic DNA extraction was carried out by QIAGEN DNeasy Blood & Tissue Kits (QIAGEN, Valencia, CA, USA) according to the manufacturer’s instructions. Information on 10 other strains was downloaded from GenBank and used for the analyses. All V. cholerae strains each with an accession number are listed in Supplementary Table S1. C-genome SNP and invariable sites analyses were carried out for categorizing all V. cholerae strains into five groups as follows: non-O1/non-O139, pre-seventh pandemic, seventh pandemic wave 1, seventh pandemic wave 2, and seventh pandemic wave 3 (Supplementary Table S1).

Whole genome sequencing data

Purified genomic DNA extracts were subjected to whole genome random sequencing analysis with NGS by HiSeq 2000 and HiSeq 2500 instruments (Illumina, San Diego, CA, USA). The library was prepared by Nextera XT kit (Illumina) and TruSeq Nano DNA Sample preparation kit (Illumina) according to the manufacturer’s instructions. Paired-end reads were checked using FastQC version 0.11.5, sequencing reads were sampled at 72-fold genome size to reduce computational resources. Quality-based trimming was done using Sickle version 1.33. De novo assemblies were carried out using different parameters in CLC Genomics Workbench 8.5.1, and Velvet version 1.2.09 (14, 15). The resulting contigs with short length <200 bp were filtered out with home customized-script. The assemblies were further improved by Metassembler version 1.5 (16), and MeGA-Merge-1.0 (17). Mapping of the reads to the references using CLC Genomics Workbench is a second approach in addition to the de novo assemblies. Strains O395, N16961, and MJ-1236 were used as references for the mapping.

Genome annotation and comparative genomic analysis

Complete and draft genomes were annotated by Prokka version 1.12 (18) and ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder), if necessary. Comparison files that contain blast-hits with sequence identity were generated using BLAST+ 2.2.31 (19), and visualized with genoPlotR package in R (20).

Identification of VSP-II and novel GI

Identification of the VSP-II region was based on mapping of the reads to the reference strain N16961. The novel VSP-II was identified as a GI sharing a similar integrase gene (annotated as integrase and having ≥80% similarity over ≥70% sequence length) inserted at the site between VC0489 and VC0517 and was not described
previously. In this study, the VSP-II relevant GI was defined as any GI that shares at least one similar gene with VSP-II, but not a similar integrase gene.

Pathogenicity and virulence profiling

DNA sequence identification using a BLAST search against a prepared local BLAST database in the CLC Genomics Workbench was carried out to determine the presence of the species-specific gene (ompW), O-antigen biosynthesis serogroup specific genes (wbe O1 and wbf O139), biotype-specific genes (tcpA, rstR and ctxB), and virulence gene ctxA (21, 22). Type of ctxB and rstR genes was determined by nucleotide variation described elsewhere (23).

RESULTS

Categorization of V. cholerae strains

All 188 V. cholerae strains were confirmed to carry ompW. Alignment of 1763 core genes with a length of 1,693,138 base pairs identified out of a predicted 8535 genes containing 105,941 core genome SNP was used for categorization of V. cholerae strains. Analyses indicated that 154 strains were V. cholerae O1 biotype El Tor, 23 were V. cholerae O139, and 11 strains were non-O1/non-O139 (Supplementary Table S1).

Distribution of VSP-II among V. cholerae strains

Any type of VSP-II identified in the study was inserted at the site between genes VC0489 and VC0517 and harbored an identical integrase gene. Among the 188 V. cholerae analyzed, 81% (153/188) of strains carried any type of VSP-II; of these, typical VSP-II, and a short variant with VC0495-VC0498 deletion was found in 36% (55/153), and 63% (96/153), respectively. In the present study, four ORF, VC0502b, VC0511b, VC0511c, and VC0512b in the typical VSP-II were newly annotated as ZnuA precursor, transposition protein of transposon Tn7, a transposase, and uncharacterized protein, respectively (Supplementary Table S2 and Fig. 1).

Fig. 1. Comparative analysis of VSP-II in 188 Vibrio cholerae strains. (a) Genetic organization of the typical seventh pandemic VSP-II of strain N16961 was used as reference. (b,c) Genetic organizations of VSP-II with VC0495-VC0498 deletion and the novel VSP-II variant are shown, respectively. (d) Schematic of chromosomal flanking region of VSP-II negative strains is indicated. Number of strain carriers is presented (n). Genes are shown by arrows, with the direction indicating the coding strand. Chromosomal flanking genes are in black. Conserved gene clusters are highlighted in grey. Integrase gene is in green. Attachment left and right sites (attL and attR) are shown by tiny vertical black bars. Similar regions based on sequence identity are indicated by dye blocks between genomes. Blast hits identity is represented by color intensity shown in the color scale.
Variation of attL sequence and distribution of VSP-II in relation to attL-like sequences

In general, VSP-II was integrated between two attachment sequences attL (14 bp) and attR (16 bp). In addition to the typical 14-bp attL, six new categories of attL-like sequences were identified in all 153 V. cholerae strains analyzed. An alignment of attL-like sequences showed that the last nine nucleotides of the sequences were identical (Fig. 2). The 16-bp attR sequence remained identical in all 153 strains carrying any type of VSP-II.

Our data indicated that VSP-II associated at different frequencies with the type of attL-like sequence. The 14-bp attL was found associated with any type of VSP-II at the frequency of 139/153 (91%), whereas the remaining six attL-like sequences were found in 9.2% (14/153) of V. cholerae strains carrying any type of VSP-II (Table 1). Of note, six distinct categories of the attL-like sequence were found in 47 wave 1 strains carrying VSP-II; however, only two categories were found in eight wave 2 strains, and one category in all 96 wave 3 strains (Table 1). Remarkably, 86% (24/28) of V. cholerae seventh pandemic strains harboring the 13-bp attL-like sequence were devoid of VSP-II (Table 1). The remaining 11 V. cholerae strains that did not carry VSP-II were categorized as non-seventh pandemic V. cholerae strains and did not carry an attL-like motif (Table 1).

DISCUSSION

The seventh pandemic strain was estimated to be initiated in 1954 by acquisition of important virulence factors: VSP-I and −II, El Tor type of CTX phage.

Table 1. Distribution of VSP-II in relation to attL-like sequences among 188 Vibrio cholerae strains

<table>
<thead>
<tr>
<th>attL-like</th>
<th>Wave 1 (n=70)</th>
<th>Wave 2 (n=9)</th>
<th>Wave 3 (n=96)</th>
</tr>
</thead>
<tbody>
<tr>
<td>attL</td>
<td>T S V</td>
<td>T S V</td>
<td>T S V</td>
</tr>
<tr>
<td>14-bp</td>
<td>36 0 0 0 6 0 0 0 0 96 0 0 0 0 0 0 0 0 1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-bp</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-bp</td>
<td>4 0 0 23 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12-bp</td>
<td>1 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-bp</td>
<td>1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-bp</td>
<td>3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-bp</td>
<td>2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>47 0 0 23 8 0 0 0 1 0 96 0 0 0 0 0 0 0 2 11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

†V. cholerae O1 pre-seventh pandemic (n=2), non-O1/non-O139 (n=11). ‡attL-like sequence was absent at the site between VC0489 and VC0517. §V. cholerae strain N16961 used as reference. ¶Novel GI variant identified in this study; −, no genomic island was found at the site between VC0489 and VC0517.
and additional mutations before it became the seventh pandemic in 1961 (6). The typical VSP-II was found to be dominant in wave 1 and 2 strains, but was not found in all 96 strains of wave 3. It was replaced by a short VSP-II with VC0495-VC0498 deletion (Fig. 1; Supplementary Table S1). In wave 3, 95 strains isolated from the Vietnam 2007–2010 outbreaks were identified to be of a highly clonal population (Supplementary Table S1). The original strain of these outbreaks, presumably introduced from the Bay of Bengal, was described in previous studies (24, 25). After 2010, molecular investigations confirmed that no strain has been reported to carry the short VSP-II. This is a result of the fact that cholera outbreaks have abated in Vietnam (25–27). Our data indicated that this short VSP-II was found dominant in wave 3 strains in Vietnam during the 2007–2010 outbreaks (Fig. 1; Supplementary Table S1). The typical VSP-II was displaced by CIRS101 VSP-II and the short VSP-II in many endemic areas in Nepal, Pakistan, Bangladesh and Haiti (13, 28–31).

Table 2. Distribution of new GI among 188 Vibrio cholerae strains

<table>
<thead>
<tr>
<th>Novel GI</th>
<th>Wave 1</th>
<th>Wave 2</th>
<th>Wave 3</th>
<th>O1†</th>
<th>NAG‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI-VC0154-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4§</td>
</tr>
<tr>
<td>GI-VC0154-2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>GI-VC0154-3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>GI-VC0209-1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GI-VC0209-2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>GI-VC0209-3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4§</td>
</tr>
<tr>
<td>Negative</td>
<td>66</td>
<td>9</td>
<td>96</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>9</td>
<td>96</td>
<td>2</td>
<td>11</td>
</tr>
</tbody>
</table>

†O1 pre-seventh pandemic. ‡non-O1/non-O139 group is indicated. §These are the same strains.

Fig. 3. Comparative analysis of six novel genomic islands. (a) Genetic organizations of three genomic islands inserted at the site between VC0153 and VC0154 were compared. V1†, V2† and V3† indicate genomic islands GI-VC0154-1, –2 and –3, respectively. (b) Genetic organizations of three genomic islands inserted at the site between VC0208 and VC0209 were compared. V1‡, V2‡ and V3‡ indicate genomic islands GI-VC0209-1, –2 and –3, respectively. Number of carrier strains is presented (n). Putative attachment left and right sites (attL and attR) are shown by tiny vertical black bars. Genes are shown by arrows, with the direction indicating the strand. Chromosomal flanking genes are in black. Integrase gene is in green. Conserved gene clusters are highlighted in grey. Similar regions based on sequence identity are indicated by dye blocks between genomes. Blast hits identity is represented by color intensity shown in the color scale.
endemic areas of India (12). Continuous surveillance of these short VSP-II in endemic and epidemic areas is critical to track its circulation in the global trend of cholera epidemics.

A novel VSP-II was identified in two non-O1/non-O139 V. cholerae strains, but was not found in any pandemic V. cholerae strain (Fig. 1; Supplementary Table S1), indicating its limited distribution in local sporadic outbreaks. The finding of the novel VSP-II indicated a larger genetic variation of VSP-II in V. cholerae non-O1/non-O139 strains which are believed to serve as natural reservoirs for GI (9, 31, 32). Of the 35 V. cholerae strains lacking VSP-II, 11 non-seventh pandemic strains were found to carry a 16-bp sequence identical to attR at the insertion site of VSP-II; however, a sequence identical to 13-bp attL-like sequence was found in the remaining 24 seventh pandemic strains at the site (Table 1).

Typically, VSP-II was flanked by two attachment sequences, attL (14 bp) and attR (16 bp) as a consequence of site-specific integration (11, 12). Significantly, seven categories of attL sequence, including six newly identified, were found associated at different frequencies with VSP-II island in this study. VSP-II linked to the typical 14-bp attL at a high frequency, and to the other categories in a small proportion (Table 1). Interestingly, 86% (24 out of 28 strains) of the 13-bp attL-like sequence was found in the absence of VSP-II in 24 V. cholerae seventh pandemic strains (Table 1). Murphy and Boyd showed that VSP-II can be excised from chromosomes and can form a CI and, after excision, the insertion site of VSP-II contained a vacant site in all 24 seventh pandemic strains (Table 1). Moreover, the strains with or without VSP-II were found circulated in the seventh pandemic wave 1 and wave 2, indicating no selection disadvantage has been conferred by the loss of VSP-II among the V. cholerae population of clinical isolates (Supplementary Table S1). These observations allow us to raise the hypotheses that the 13-bp attL-like sequence might be involved in promoting excision of VSP-II from the bacterial genome or in inhibiting re-integration of CI of VSP-II. Further, the detection of a 13-bp attL-like sequence region could be used as a molecular marker for monitoring the absence of VSP-II among V. cholerae seventh pandemic strains.

Two gene clusters, VC0495–VC0498 and VC0504–VC0510, typically found in VSP-II, were shared by the novel VSP-II, and also shared by four novel GI (Fig. 3; Supplementary Table S2). Previous studies reported that the two conserved gene clusters were carried by GI-81 and GI-118 (31), indicating that those conserved gene clusters are commonly shared by many GI.

The integrase encoded in five out of six novel GI, as well as in previously reported GI-81 and GI-118, is different from that encoded in any type of VSP-II. We observed that none of the 11 V. cholerae strains that carried any type of novel GI did not carry any type of VSP-II, simultaneously (Supplementary Table S1). This indicates that novel GI described in the study may have a similar function that possessed by VSP-II. The entire 96 V. cholerae wave 3 strains analyzed in the present study were determined to carry the short VSP-II with the VC0495–VC0498 deletion (Fig. 1; Supplementary Table S1). It could be hypothesized that V. cholerae strains with the short VSP-II may have a selection advantage over those with typical VSP-II, as we observed that strains with the short VSP-II had been circulated and become predominant in the 2007–2010 cholera endemic seasons and epidemic areas (13, 28–31).

Association of VSP-II with epidemic potential should not be dismissed concerning the history of the seventh cholera pandemic. The occurrence of different VSP-II variants coincided with altered CTX prophage, which was observed in cholera outbreaks and had cryptically changed the epidemiology of cholera (4, 23, 31, 33, 34). Several distinct shifts in CTX prophages have been shown to be evidence of population changes of V. cholerae (23, 34). However, the role of VSP-II in the pathogenesis and dynamic nature of epidemic V. cholerae remains largely to be clarified. Variants, such as CIRS101 VSP-II, a short VSP-II with VC0495–VC0498 deletion, and WASA VSP-II were found associated with epidemic tendencies in Haiti, South Asia countries, Western Africa and South America, respectively (4, 12, 29, 31, 35).

ACKNOWLEDGMENTS

Authors are grateful to the Program for Nurturing Global Leaders in Tropical and Emerging Communicable Diseases, Graduate School of Biomedical Sciences, Nagasaki University, and to Japan Initiative for Global Research Network on Infectious Diseases (J-GRID) from Ministry of Education, Culture, Sport, Science & Technology in Japan, and Japan Agency for Medical Research and Development (AMED) for supporting the study.

DISCLOSURE

The authors have no conflicts of interest to declare.
REFERENCES


**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of this article at the publisher’s web-site.

**Table S1.** List of strains used in the present study

**Table S2.** Similarity between typical VSP-II, novel VSP-II and two novel genomic islands