Infection, Genetics and Evolution xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

# Population structural analysis of O1 El Tor *Vibrio cholerae* isolated in China among the seventh cholera pandemic on the basis of multilocus sequence typing and virulence gene profiles

Haijian Zhou [a,b,1], Xuan Zhao [a,1,2], Rui Wu [a], Zhigang Cui [a,b], Baowei Diao [a], Jie Li [a], Duochun Wang [a,b], Biao Kan [a,b,*], Weili Liang [a,b,*]

[a] National Institute for Communicable Disease Control and Prevention, State Key Laboratory for Infectious Disease Prevention and Control, Chinese Centre for Disease Control and Prevention, 155, Changbai Road, Changping, Beijing 102206, People's Republic of China
[b] Collaborative Innovation Centre for Diagnosis and Treatment of Infectious Diseases, 866 Yuhangtang Road, Hangzhou 310003, People's Republic of China

ARTICLE INFO

*Article history:*
Received 13 September 2013
Received in revised form 12 December 2013
Accepted 16 December 2013
Available online xxxx

*Keywords:*
*Vibrio cholerae*
Multilocus sequence typing
Virulence gene
Serogroup O1
Population structure
CTXΦ

ABSTRACT

Serogroup O1 *Vibrio cholerae* is the most common agents to cause epidemic and pandemic cholera disease. In this study, multilocus sequence typing (MLST) was performed on 160 serogroup O1 strains (including 42 toxigenic and 118 non-toxigenic), and the virulence/fitness gene profiles of 16 loci were further analysed for 60 strains of these. Eighty-four sequence types (STs) with 14 clonal complexes were distinguished, and 29 STs were unique. Except SD19771005, all toxigenic strains were well-separated from the non-toxigenic strains. While a group of non-toxigenic strains clustered closer to the toxigenic strains compared to the other strains. Overall the examined gene loci showed higher presence rates in the toxigenic strains compared to the non-toxigenic strains. It is worth noting that the presence rates of VPI, TLC, VSP-I and VSP-II in the non-toxigenic strains that were clustered closer to the toxigenic strains were much higher compared to the other non-toxigenic strains. Our study indicated the complex population structure of O1 strains, and parts of non-toxigenic strains are genetically more closely related to toxigenic strains than other non-toxigenic strains, suggesting that these strains may have a higher potential for infection with CTXΦ in the environment or host intestine and is more efficient to become new pathogenic or epidemic clones.



## 1. Introduction

Cholera is a historically severe dehydrating diarrheal disease and remains a public health problem worldwide, particularly in Africa and Asia. The causative organism is *Vibrio cholerae*, which is divided into serogroups on the basis of the somatic O antigen. The O1 serogroup is known to be the most common strain to cause this epidemic and pandemic disease. Humans are the transient host for *V. cholerae*. Aquatic systems, such as river, marine, and brackish water are environmental reservoirs where the bacteria multiply in association with zooplankton (Colwell, 1996; Faruque et al., 1998). The organism is transmitted to contaminate food and water mainly via the faecal-oral route.

The evolutionary genetic relationships among *V. cholerae* strains have been examined using different methods. In these studies, the epidemic *V. cholerae* isolates form a lineage that is separate from the nonepidemic strains (Beltran et al., 1999; Byun et al., 1999; Karaolis et al., 1995; Wachsmuth et al., 1993). One common characteristic of epidemic and pandemic cholera isolates is that these strains harbour genes encoding the cholera toxin (CT) and the toxin co-regulated pilus (TCP), which are two of the most important virulence factors of *V. cholerae*. CT is encoded by *ctxAB* genes, which are integral components of the filamentous phage CTXΦ (Waldor and Mekalanos, 1996). The TCP biosynthesis genes are located on the *Vibrio* pathogenicity island 1 (VPI-1) (Karaolis et al., 1998). A number of other genes or gene clusters have also been identified predominantly among epidemic *V. cholerae* isolates and may play a role in the environmental fitness of the isolates and in cholera pathogenesis. These genes or gene clusters include filamentous phage RS1Φ (Davis et al., 2002), toxin-linked cryptic plasmid (TLC) (Rubin et al., 1998), RTX toxin gene cluster(Lin et al., 1999), *hapA* and *hapR* (Jobling et al., 1997), VPI-2 (Jermyn and Boyd, 2002), hemolysin (*hlyA*), genes encoding type VI secretion system (T6SS) (Pukatzki et al., 2006), Chitin-regulated pilus (ChiRP)

* Corresponding authors at: Department of Diarrheal Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Centre for Disease Control and Prevention, 155, Changbai Road, Changping, Beijing 102206, People's Republic of China. Tel./fax: +86 010 58900744.
*E-mail addresses:* kanbiao@icdc.cn (B. Kan), liangweili@icdc.cn (W. Liang).
[1] These authors contributed equally to this work.
[2] Present address: Tianjing Centre for Disease Control and Prevention, Tianjing 300171, People's Republic of China.

1567-1348/$ - see front matter © 2014 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.meegid.2013.12.016

(Meibom et al., 2004), mannose-sensitive hemolysin agglutination pilin (MSHA) (Jonson et al., 1991), Vibrio polysaccharide (VPS) (Fong et al., 2010), and Vibrio seventh pandemic island I (VSP-I) and VSP-II, which consists of VC0175 to VC0185 and VC0490 to VC0516, respectively. These two islands were found exclusively among the El Tor biotype isolates in comparative genomic studies using a *V. cholerae* DNA microarray (Dziejman et al., 2002). Recently study focusing on genomic diversity of 2010 Haitian cholera outbreak strains showed that CTXΦ, RS1Φ, TLC, VPI-1, VPI-2, VSP-I and VSP-II were found in almost all clinical O1 isolates but not in clinical or environmental *V. cholerae* non-O1/O139 isolates; meanwhile, the non-O1/O139 populations in Haiti harbor a genomic backbone similar to that of toxigenic *V. cholerae* O1 (Hasan et al., 2012). Revealing the virulence gene profiles and elucidating the steps and significance of virulence gene acquisition in the evolution of *V. cholerae* is helpful to understand the underlying phylogenetic relationships among strains.

Multilocus sequence typing (MLST) is a useful molecular subtyping method that plays an important role in epidemiological and phylogeny studies. MLST has been previously used to study molecular subtyping, population structure and lateral gene transfer of *V. cholerae* (Keymer and Boehm, 2011; Lee et al., 2006; Octavia et al., 2013; Salim et al., 2005). In this study, we used MLST to study the population structure of O1 serogroup *V. cholerae* strains isolated in China among the seventh cholera pandemic and examined the presence and distribution of gene regions associated with virulence or survival and persistence in the environment among strains to elucidate the role of the gene regions involved in the emergence of epidemic isolates.

## 2. Materials and methods

### 2.1. Bacterial strains

One hundred and sixty *V. cholerae* serogroup O1 strains, including 42 toxigenic (*ctxAB*-positive) and 118 non-toxigenic (*ctxAB*-negative) strains, isolated in China were investigated in this study (Supplementary Table 1). All of the strains were isolated in different years (1961–2010) and were involved in three nationwide cholera epidemics in the 1960s, 1980s and 1990s, and the epidemic-interval periods in China. The strains were isolated using gentamicin-selective agar or thiosulfate citrate bile salt sucrose (TCBS) agar and identified using the *V. cholerae* diagnostic sera and biochemical tests. Additionally, six O1 strains (N16961, O395, MJ-1236, M66-2, 2010EL-1786 and IEC224) from other countries were used as reference strains in MLST analysis.

### 2.2. MLST

The DNA genomes were extracted using the DNeasy Blood & Tissue Kit (Qiagen, Dusseldorf, Germany) according to the instructions provided by the manufacturer. Seven loci (*dtds*, *gyrB*, *mdh*, *pntA*, *pyrC*, *recA* and *tnaA*) were used for the MLST analysis. These loci were used in the MLST schemes for other *Vibrio* species, such as *V. parahaemolyticus* and *V. vulnificus* (Bisharat et al., 2005; González-Escalona et al., 2008). Both chromosomes are represented with three (*gyrB*, *mdh*, *recA*) and four (*dtds*, *pntA*, *pyrC*, *tnaA*) genes from chromosomes I and II, respectively, and the genes were evenly distributed around the chromosomes. The primers for the seven housekeeping genes are shown in Supplementary Table 2. The PCR products were sequenced in both directions using the Big Dye cycle sequencing kit (ABI) according to the manufacturer's instructions. Sequencing was performed on an ABI 3770 automatic sequencer. The gene sequences were analysed using the Molecular Evolutionary Genetics Analysis (MEGA) suite of programmes, version 5.1 (Kumar et al., 2001) or BioNumerics version 5.10 software (Applied Maths, Kortrijk, Belgium). Phylogenetic gene trees were constructed using the neighbour-joining method with the Jukes–Cantor distance method. Bootstrap values were calculated for 1000 trees.

Sequences of each locus were compared with each other to determine the allele numbers and MLST types (STs). The STs were analysed using eBURST to determine the presence of clonal complexes (CCs) (Feil et al., 2004). BioNumerics version 5.10 software was used to create the minimum spanning tree. In the minimum spanning tree, the founder ST was defined as the ST with the greatest number of single-locus variants. Types were represented by circles and the size of a circle indicated the number of strains with this particular type.

### 2.3. Pulsed-field gel electrophoresis (PFGE)

We used the PulseNet 1-day standardised PFGE protocol for *V. cholerae* (Cooper et al., 2006). Cell suspensions were placed in polystyrene tubes (Falcon; 12 × 75 mm), and their optical density was adjusted to 4.0–4.2 using the Densimat photometer (bioMérieux, Marcy l'Etoile, France). *V. cholerae* slices were digested using 20 U per slice of *NotI* (New England BioLabs, Ipswich, MA, USA) for 4 h at 37 °C. Electrophoresis was performed using a CHEF-DRIII system (Bio-Rad Laboratories, Hercules, CA, USA). Images were captured on the Gel Doc 2000 system (Bio-Rad) and converted into TIFF files for computer analysis using the BioNumerics version 5.10 software (Applied Maths, Kortrijk, Belgium). The similarity between two patterns was expressed as the Dice coefficient ($S_D$) (Dice, 1945) and dendrogrammes were clustered and constructed using the unweighted pair group method with arithmetic averages (UPGMA).

### 2.4. PCR analysis

PCR was used to assay 60 O1 serogroup *V. cholerae* isolates (including 20 toxigenic and 40 non-toxigenic isolates) for the presence of 16 loci associated with virulence or survival and persistence in the environment of *V. cholerae*. Of the 16 virulence loci examined, 12 loci consisted of multiple genes (CTXΦ, VPI-1, T6SS locus, VPI-2, VSP-I, VSP-II, RTX toxin cluster, TLC, VPS, MSHA, RS1Φ, ChiRP) and 4 loci had single gene (*hlyA*, *hapA*, *hapR*, *vpsR*). A total of 26 primer pairs were used to determine the distribution of the 16 loci among the 60 *V. cholerae* isolates. Five primer pairs were used to assay for the presence of VPI-1, two primer pairs (each) were used to assay for T6SS, VSP-I, VSP-II, VPI-2, VPS, RTX, TLC, and one primer pair (each) was used to assay for the presence of CTXΦ, ChiRP, *hlyA*, *hapA*, *hapR*, RS1Φ, and MSHA. The PCR were prepared in a reaction volume of 50 μl with 5 μl of 10× PCR buffer (Takara, Dalian, China), 1 unit *Taq* polymerase (Takara), 200 μM of dNTPs (Takara), 0.4 μM of each primer set, 20 ng of the DNA template and filtered sterile water. PCR was performed with an initial denaturation step at 94 °C for 5 min followed by 33 cycles each consisting of an initial denaturation at 94 °C for 40 s followed by annealing and extension steps. The primers and corresponding annealing temperatures are listed in Supplementary Table 2.

### 2.5. Data analysis

Nei's indices (Malorny et al., 2008) of each locus and Simpson's diversity index (Hunter and Gaston, 1988) of the typing methods were calculated. The formulas were as follows: Nei's index = $1-\sum$(allele frequency)$^2$; Simpson's diversity index $(D) = 1-\sum[n_j(n_j-1)]/[N(N-1)]$, where $n_j$ is the number of strains belonging to the $j$th pattern, and $N$ is the number of strains in the population.



Please cite this article in press as: Zhou, H., et al. Population structural analysis of O1 El Tor *Vibrio cholerae* isolated in China among the seventh cholera pandemic on the basis of multilocus sequence typing and virulence gene profiles. Infect. Genet. Evol. (2014), http://dx.doi.org/10.1016/j.meegid.2013.12.016