Infection, Genetics and Evolution 26 (2014) 241–249

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

The purifying trend in the chromosomal integron in *Vibrio cholerae* strains during the seventh pandemic

Cuicai Zhang [a,b,1], Bo Pang [a,b,1], Zhemin Zhou [c,1], Haiyin Wang [a,b], Haijian Zhou [a,b], Xin Lu [a,b], Pengcheng Du [a,b], Lijuan Zhang [a,b], Jie Li [a,b], Zhigang Cui [a,b], Chen Chen [a,b], H.W. Stokes [d,*,1], Biao Kan [a,b,*,1]

[a] *State Key Laboratory for Infectious Disease Prevention and Control, Department of Diarrheal Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, People's Republic of China*
[b] *Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou, People's Republic of China*
[c] *The University of Warwick, Coventry CV4 7AL, United Kingdom*
[d] *The ithree Institute, University of Technology, Sydney, New South Wales, Australia*

ARTICLE INFO

Article history:
Received 9 December 2013
Received in revised form 14 May 2014
Accepted 17 May 2014
Available online 3 June 2014

Keywords:
Cholera
Chromosomal integron
Evolution
Gene cassettes
Vibrio cholerae

ABSTRACT

Chromosomal integron (CI) arrays in *Vibrio* spp. are generally large and display great variation. Here we determined the sequence of CI array in a toxigenic O139 *Vibrio cholerae* strain and compared it with the arrays from the genome of different O1 biotypes available in GenBank. Then PCR scanning was used to determine the CI array variations in 83 epidemic O139 strains and subsequently these variations were compared with that found in toxigenic O1 El Tor strains in our previous work. Few differences were observed in the cohort of toxigenic O139 strains compared to the toxigenic O1 El Tor strains. On the basis of CI arrays, the toxigenic O1 El Tor and O139 strains isolated concurrently in recent years appear to be more similar to each other than to the O1 strains isolated in previous decades, suggesting a closer evolutionary relationship between them. Comparison of CI arrays in toxigenic O1 El Tor and O139 *V. cholerae* strains isolated between 1961 and 2009 revealed a purifying trend in the CI arrays in the chronological order during the seventh pandemic.

1. Introduction

Cholera is a life-threatening disease, particularly in developing countries with poor sanitation and limited healthcare. More than 200 serogroups of *Vibrio cholerae* have been documented, although only serogroups O1 and O139 are responsible for cholera epidemics and pandemics (Kaper et al., 1995). The sixth cholera pandemic was caused by *V. cholerae* serogroup O1 strains defined as the classical biotype. In contrast, the seventh and ongoing cholera pandemic is caused by *V. cholerae* O1 strains possessing the El Tor biotype, which has replaced classical strains from the 1960s. Further variants of El Tor known as hybrid strains because they possess properties also found in classical strains, began to appear in the 1990s (Nair et al., 2002). Coincident with this appearance, a second serogroup, O139, capable of causing epidemics, appeared in India and Bangladesh in 1992 (Bhattacharya et al., 1993; Cholera Working Group, 1993). Serogroup O139 *V. cholerae* strains appeared soon after in other countries in Southeast Asia (Sack et al., 2004) and first emerged in China in 1993. Since then cases have been reported every year in that country (Qu et al., 2003). Comparative genomic analysis of O1 and O139 *V. cholerae* strains indicated that the major differences between these two serogroups clustered in three regions of the genome. These regions are the lipopolysaccharide (LPS) synthesis gene cluster, the *Vibrio* pathogenicity island II and the chromosomal integron (CI) (Dziejman et al., 2002; Labbate et al., 2007; Pang et al., 2007), highlighting the importance of lateral gene transfer (LGT) in the evolution of pathogenic *V. cholerae* strains. Phylogenetic analysis demonstrated that all of the seventh pandemic *V. cholerae* (toxigenic O1 and O139) strains are genetically closely related (Mutreja et al., 2011) with O139 strains probably being derived from O1 strains by a LGT-mediated serogroup switch (Bik et al., 1995).

Integrons are genetic elements that include a site-specific recombination (SSR) system (Martinez and de la Cruz, 1990; Stokes and Hall, 1989). This SSR system can integrate mobilizable elements known as gene cassettes (Hall et al., 1991; Hall and Collis, 1995). Gene cassettes normally contain a single gene, and the insertion of a cassette into an integron potentially leads to the expression of this gene via an integron-associated promoter

* Corresponding authors. Address: P.O. Box 5, Changping, Beijing, People's Republic of China. Tel.: +86 10 58900744; fax: +86 10 58900742 (B. Kan).
E-mail address: kanbiao@icdc.cn (B. Kan).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.meegid.2014.05.019
1567-1348/

(Collis and Hall, 1995). Integrons were first identified in the context of the spread of antibiotic resistance where pathogens commonly carry integrons on other mobile elements with the associated cassettes including antibiotic resistance genes (Hall, 2012; Partridge et al., 2009). However, integrons are also commonly found in the chromosomes of diverse bacteria (Boucher et al., 2007; Mazel, 2006). CIs were first identified in the chromosomes of the *Vibrio* genus (Mazel et al., 1998) where arrays can be particularly large, often possessing over 100 cassettes and comprising several percent of the genome of any given strain (Cambray et al., 2010; Mazel, 2006). The gene cassette arrays in the *Vibrio* genus are also highly diverse and include many novel gene functions. Cassette arrays comprise the single largest region of diversity in the strains of this genus (Boucher et al., 2007; Mazel, 2006; Rowe-Magnus et al., 2001, 2003). The large arrays in *Vibrio* genus CIs essentially comprise unique coding regions separated by the cassette recombination site, *attC*, recognized by the integron-associated site-specific recombinase (Rowe-Magnus et al., 1999). *attC* sites can be quite variable in sequence and length; however, they show a high degree of conservation in the *Vibrio* genus (Mazel et al., 1998), so that PCR primers can be designed that specifically target almost any of the many such sites in large arrays (Labbate et al., 2007).

To date detailed whole-genome sequencing analysis has been widely used to study the evolution of bacterial species and has assisted the understanding of how pandemic *V. cholerae* strains evolve (Mutreja et al., 2011). However, many single nucleotide polymorphism (SNP) -based analyses of the core genome excluded events that are important in pathogen evolution involve LGT acquisition or loss of gene or gene clusters. Apart from the gain or loss of overtly important virulence factors, other regions, such as the gene arrays associated with CIs, can serve as important markers in the high-resolution discrimination of otherwise identical, or near identical, strains (Gao et al., 2011; Labbate et al., 2007). Since integron arrays evolve at a very fast rate, PCR screening that targets these arrays can be used for the phylogenetic analysis of closely related strains (Chowdhury et al., 2010; Labbate et al., 2007; Tokunaga et al., 2010). In a previous work, we have employed an "overlap" PCR scanning approach to explore the structural diversity of the CIs in toxigenic *V. cholerae* O1 El Tor strains (Gao et al., 2011) and found a correlation between the CI array structure and the isolation time of the strains. Studies to date involving the CIs of toxigenic *V. cholerae* have mainly focused on O1 serogroup strains and no extensive analysis has been undertaken in O139 strains. Although a draft sequence of O139 has been available for several years, it was not definitively clear that the CI array was within a definitive single contig because of the tandem duplications in this region. Therefore, we cannot rely on this draft for CI analysis. Here we report the complete sequence of the CI in the toxigenic *V. cholerae* O139 strain JX20062026 for the first time. This was compared to the CI arrays from the strains O1 El Tor strain N16961, O1 El Tor hybrid strain MJ1236 and O1 classical strain O395 available in public databases. We further used PCR scanning to explore the genetic diversity in 83 O139 strains isolated between 1993 and 2009 and compared them with the diversity in toxigenic O1 El Tor strains as defined in our previous study (Gao et al., 2011).

2. Materials and methods

2.1. Strains

The CI data analyzed in this study were obtained from 142 *V. cholerae* strains. The information of these strains is listed in Table S1. Within them 59 toxigenic (carrying cholera toxin genes *ctxAB*) O1 El Tor strains were analyzed in our previous study (Gao et al., 2011) and were used for the comparison in this study. Eighty-three serogroup O139 strains, except for MO45 from India, were isolated from different regions in China in different years since 1993. Sixty-five of these O139 strains were isolated from cholera patients and six from asymptomatic carriers. Six O139 strains were isolated from environment water during routine environmental surveillance (Table S1). The O139 serogroup was confirmed by agglutination with a specific antiserum (Denka Seiken, Japan) and by PCR amplification of a genomic sequence involved in O139 lipopolysaccharide (LPS) biosynthesis (Falklind et al., 1996). The *ctxAB* genes were detected with PCR using published primers (Liu et al., 2004). *ctxAB*-positive strains were defined as toxigenic and strains negative for these genes as nontoxigenic. All 142 *V. cholerae* strains were assigned to four groups according to the isolation years and serogroups (Table S1). Group 1 includes 12 toxigenic O1 El Tor strains isolated between 1961 and 1966, Group 2 includes 24 toxigenic O1 El Tor strains isolated between 1973 and 1989, Group 3 includes 23 toxigenic O1 El Tor strains isolated between 1990 and 2008 and Group 4 includes 79 toxigenic O139 strains isolated between 1993 and 2009.

2.2. Chromosomal DNA preparation

The strains were cultured at 37 °C in 5 ml of Luria–Bertani (LB) broth to an optical density at 600 nm of 0.8. Genomic DNA (gDNA) was extracted using NucleoSpin Tissue kits (Macherey-Nagel, Germany) according to the manufacturer's protocol.

2.3. Characterization of the CI in the toxigenic O139 strain JX20062026 and O1 El Tor hybrid strain MJ1236

The complete genome of JX20062026, a toxigenic O139 *V. cholerae* strain, was sequenced in our *V. cholerae* genome sequencing project. Its complete CI sequence has been submitted to the GenBank database under the accession No. KF680548. The complete genome sequence of MJ1236, a toxigenic O1 El Tor hybrid *V. cholerae*, was retrieved from the GenBank database (NC012667/NC012668). A *V. cholerae* integron integrase gene and a region that shows multiple sequence repeats representing obvious *Vibrio*-like *attC* sites were identified in the small chromosomes of both JX20062026 and MJ1236. These two regions were then analyzed with the annotation of cassette and integron data (ACID) (Joss et al., 2009). The *attI* sites were identified with the default parameters. The *attC* sites and gene cassettes were determined with a PERL script based on the BLAST program. The CI information in N16961 and O395 were obtained directly from previous studies (Feng et al., 2008; Labbate et al., 2007).

2.4. Primer design for PCR scanning

The principles of primer design have been described elsewhere (Gao et al., 2011; Pang et al., 2011). Briefly, every two adjacent ORFs were amplified by a pair of primers designed inside the two ORFs. Every other adjacent amplicon overlapped with the next one. More than two ORFs were included in one amplicon if no primers could be designed with this principle. A 1427-bp fragment that included the *intI* gene was also recovered. This amplicon was derived from primers that span the point adjacent to the *intI* gene at which the CI is found in the chromosome in toxigenic O139 *V. cholerae* strain JX20062026. Based on the JX20062026 CI and array sequence, 68 primer pairs were designed using Primer3 (Untergasser et al., 2012) in the first round of screening. Another seven primer pairs were designed in the second round. The sizes of overlapping segments between every two adjacent amplicons ranged up to 1958 bp. For amplicons from other strains whose