Original Article

Structural Variation of the Superintegron in the Toxigenic *Vibrio cholerae* O1 El Tor*

GAO Yan[§,$], PANG Bo[§], WANG Hai Yin, ZHOU Hai Jian, CUI Zhi Gang, and KAN Biao[#]

State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China

Abstract

Objective To understand the genetic structures and variations of the superintegron (SI) *in Vibrio cholerae* isolated in the seventh cholera pandemic.

Methods Polymerase chain reaction scanning and fragment sequencing were used. Sixty toxigenic *V. cholerae* O1 El Tor strains isolated between 1961 and 2008 were analyzed.

Results Some variations were found, including insertions, replacements, and deletions. Most of the deletions were probably the result of recombination between *V. cholerae* repeat sequences. The majority of the variations clustered together. The SIs of the strains isolated in the 1960s and 1970s showed more diversity, whereas SI cassette variations in strains isolated in the 1990s and after were lower, with ~24 kb signature sequence deletion. This indicates the predominant SI in the host during the epidemic in the 1990s and after. The insertion cassettes suggested the mobilization from the SIs of other *V. cholerae* serogroups and *Vibrio mimicus*.

Conclusion The study revealed that structural variations of SIs were obvious in the strains isolated in epidemics in different decades, whereas the divergence was based on syntenic structure of SIs in these El Tor strains. Also, the continuing cassette flows in the SIs of the host strains during the seventh cholera pandemics were displayed.

Key words: Superintegron; Cassette; *Vibrio cholerae*

Biomed Environ Sci, 2011; 24(6):579-592 doi:10.3967/0895-3988.2011.06.001 ISSN:0895-3988

www.besjournal.com(full text) CN:11-2816/Q

INTRODUCTION

Integrons were first discovered in the characterization of multiple-resistance-encoding plasmids and transposons[1-2]. The major components of an integron are an *attC* site (associated with gene cassette), *intI* gene (encoding an integrase), *attI* site, and a promoter[3]. The integrase can integrate and excise gene cassettes by catalyzing recombination between *attI* and *attC*, and recombination between *attC*s[4]. A distinct type of integron, superintegron (SI), was first identified in a *Vibrio cholerae* strain, with the characteristics of a large number of cassettes that are clustered, spaced by *attC* sites, and specific integrase[5]. Comparison of SI structures among the *Vibrio* species and the different isolates within the same species may provide valuable data for the analyses of SI nature, gene flow, and evolution of SIs and hosts. It has been

*This work was supported by the National Natural Science Foundation of China (30800987), National Basic Research Priorities Program (2009CB522604) and Priority Project on Infectious Disease Control and Prevention (2008ZX10004-009).

#Correspondence should be addressed to KAN Biao. Tel: 86-10-58900703. Fax: 86-10-58900742. E-mail: kanbiao@icdc.cn

§GAO Yan and PANG Bo contributed equally to this work.

$Present address: Chaoyang Center for Disease Control and Prevention, Chaoyang District, Beijing 100021, China.

Biographical note of the first authors: GAO Yan, female, born in 1971, Ph. D candidate, majoring in pathogen biology; PANG Bo, male, born in 1974, associate professor, majoring in pathogen biology.

Received: May 27, 2011; Accepted: June 13, 2011

shown that a low number of cassette counterparts are shared among different *Vibrio* species, which suggests a wide range of species source for the entrapped genes and an active cassette assembly process[6-8]. SI structures have been found in many bacterial genomes, including gammaproteobacteria, betaproteobacteria, and deltaproteobacteria, besides *Vibrionaceae*[6-7]. SI is a potential gene capture system and may play a role in bacterial adaptation and evolution[9]. Although the functions of most of the encoded genes in SI are unknown, some of the SI open reading frames (ORFs) encode adaptive functions including pathogenicity and antibiotic resistance determinants[7,10-12].

Historically seven cholera pandemics have been recorded. Toxigenic *V. cholerae* serogroup O1 of biotype El Tor caused the seventh cholera pandemic since 1961[13], and O139 cholera emerged in Bangladesh and India in 1992 and caused epidemics in Southeast Asia[14]. In *V. cholerae* N16961, an SI is located in the small chromosome and carries 216 ORFs[15]. *attC* is called the *Vibrio cholerae* repeat sequence (VCR) in the SI of *V. cholerae*[6]. The comparative genome hybridization (CGH) and whole genome PCR scanning have suggested that the content variance of SIs in different lineages of *V. cholerae* is distinct[16-17]. With three whole genome sequences of *V. cholerae* (including the toxigenic/notoxigenic O1 El Tor strains and the toxigenic O1 classical strain), substantial changes in the SI of these strains have been revealed[18]. Studies based on PCR and hybridization have also revealed plastic structures of SIs between different serogroups[7,19-21].

However, CGH and other PCR-based genotyping analyses have not revealed the structural details of the SI. In some studies, PCR was performed with primers based on the conservative sequence of VCR, and then the amplicons were analyzed with electrophoresis, Southern hybridization and sequencing to compare the ORF content of the SI in different *V. cholerae* strains[19-21]. The ORFs arrangement in SIs, and deletion and insertion of the ORFs in SIs remain unclear. Many repeat sequences exist in SIs, which also make it impossible to assemble the SI sequence and to study the gene arrangement in SIs. PCR scanning[17], which uses the overlapping amplicons to study the arrangement of genes, is valuable for the genome fragment assembly and especially for those containing many repeat sequences.

In this study, a detailed PCR scanning strategy combined with sequencing was used to analyze the strain-to-strain genetic organization variance of the SI in 60 toxigenic *V. cholerae* O1 El Tor strains during the seventh cholera pandemic in China. The structure of the SIs in the test strains was basically syntenic; however, diversity and decadal signatures were also observed, characterized by successive ORF deletion, ORF insertion, and a few potential ORF translocations. Homologous recombination based on repeat sequence and VCR played roles in the gene flows of SIs.

MATERIALS AND METHODS

Strains

From 1961 to 2008, three epidemiologically defined cholera epidemics occurred in China. In this study, 60 toxigenic O1 El Tor strains isolated in different epidemic periods and inter-epidemic periods in different geographic regions were selected for analysis (Figure 1). The details of experimental strains are shown in Table 1.

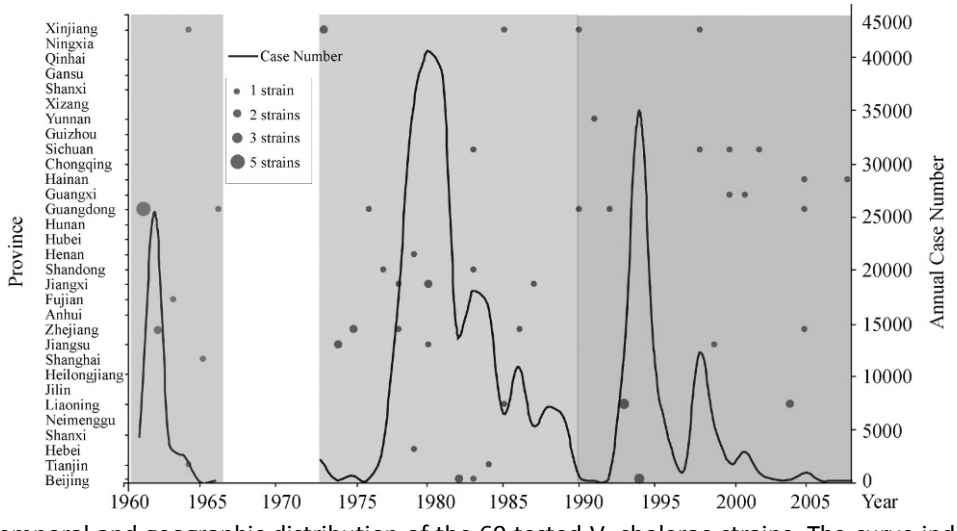

Figure 1. Temporal and geographic distribution of the 60 tested V. cholerae strains. The curve indicates the cholera cases reported in China between 1961 and 2008. The spots denote the number of strains: the small to big spots denote 1, 2, 3, and 5 strains, respectively.