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.

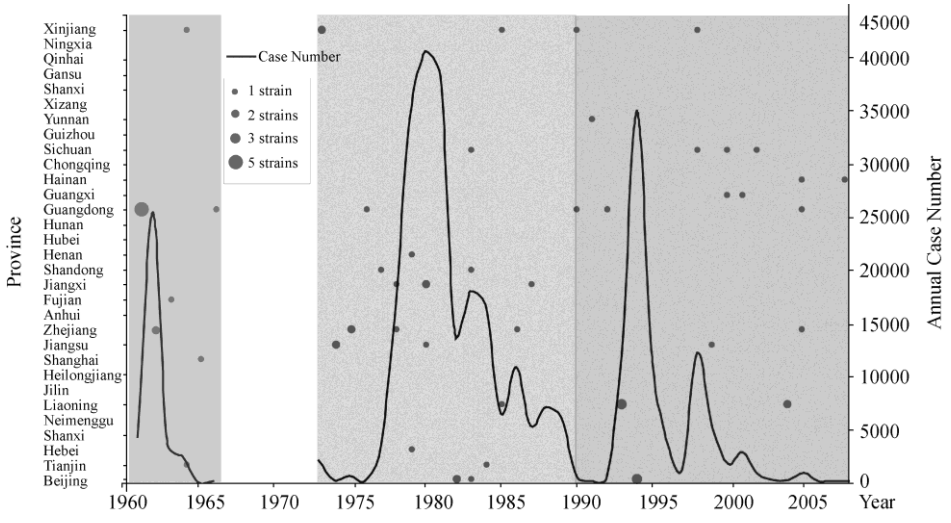


Figure 1. Temporal and geographic distribution of the 60 tested V. cholerae strains. The curve indicates the cholera cases reported in China between 1961 and 2008. The spots denote the number of strains: the small to big spots denote 1, 2, 3, and 5 strains, respectively.

Table 1. Characteristics of the Toxigenic O1 El Tor Strains Used in This Study

Strains	Serotype	Year Isolated	Source	Site Isolated	Number of the ORFs in SI
61226*	Ogawa	1961	Patient	Guangdong	211
1961119*	Ogawa	1961	Patient	Guangdong	206+
19612533	Ogawa	1961	Patient	Guangdong	140
19612540*,#	Ogawa	1961	Patient	Guangdong	157
196153*	Ogawa	1961	Patient	Guangdong	213+
62110*	Ogawa	1962	Patient	Zhejiang	210+
62048*	Ogawa	1962	Patient	Zhejiang	214
63244*	Ogawa	1963	Unknown	Fujian	211
64193*,#	Ogawa	1964	Patient	Tianjin	172
642345	Ogawa	1964	Patient	Xinjiang	209
65930*	Inaba	1965	Screw	Shanghai	214
661673*	Ogawa	1966	Patient	Guangdong	212
73364*	Ogawa	1973	Unknown	Xinjiang	199
73448*	Ogawa	1973	Unknown	Xinjiang	199
7470*	Ogawa	1974	Patient	Jiangsu	210
74449*	Ogawa	1974	Unknown	Jiangsu	203
75714*	Ogawa	1975	Unknown	Zhejiang	211
751130*	Ogawa	1975	Unknown	Zhejiang	212
GDL-ETV760	Inaba	1976	Patient	Guangdong	216
7751	Ogawa	1977	Water	Shandong	141+
78599*	Ogawa	1978	Unknown	Jiangxi	198+
781079*	Ogawa	1978	Patient	Zhejiang	212
7925*	Ogawa	1979	Unknown	Henan	214
79192*	Ogawa	1979	Patient	Hebei	214
80954*	Ogawa	1980	Patient	Jiangsu	214
801290	Ogawa	1980	Patient	Jiangxi	203
801298	Inaba	1980	Water	Jiangxi	216
829	Ogawa	1982	Water	Beijing	140
8212	Ogawa	1982	Water	Beijing	140
83101	Ogawa	1983	Unknown	Shandong	216
83535	Inaba	1983	Unknown	Sichuan	215+
83795	Inaba	1983	Unknown	Beijing	216
84159*,#	Ogawa	1984	Unknown	Tianjin	172
8593	Inaba	1985	Unknown	Liaoning	216
85079	inaba	1985	Unknown	Xinjiang	216
19861071	Inaba	1986	Patient	Zhejiang	216
8788	Inaba	1987	Unknown	Jiangxi	216
90–26*,#	Ogawa	1990	Unknown	Xinjiang	172
1990016*,#	Inaba	1990	Unknown	Guangdong	167
1991225*,#	Ogawa	1991	Unknown	Yunnan	172
1992031*,#	Ogawa	1992	Patient	Guangdong	172
93014*,#	Ogawa	1993	Patient	Liaoning	172
1993005*,#	Ogawa	1993	Patient	Liaoning	172
1993050*,#	Ogawa	1993	Unknown	Liaoning	172
94068*,#	Ogawa	1994	Patient	Beijing	172
1994400*,#	Ogawa	1994	Patient	Beijing	172
1994A20*,#	Ogawa	1994	Patient	Beijing	172
1998101#	Ogawa	1998	Environment	Sichuan	171
1998146*,#	Ogawa	1998	Patient	Xinjiang	172
1999001	Inaba	1999	Patient	Jiangsu	216
2000031*,#	Ogawa	2000	Patient	Guangxi	170
2000180*,#	Ogawa	2000	Environment	Sichuan	167+
2001058*,#	Inaba	2001	Patient	Guangxi	172
2002001*,#	Inaba	2002	Patient	Guizhou	172
2004152*,#	Ogawa	2004	Patient	Liaoning	172
2004DD534*,#	Inaba	2004	Unknown	Liaoning	172
2005125*,#	Ogawa	2005	Patient	Guangdong	172
2005122*,#	Inaba	2005	Patient	Hainan	172
2005035*,#	Ogawa	2005	Patient	Zhejiang	172
2008066*,#	Ogawa	2008	Patient	Hainan	171

Note. *Strains with deletion of VCA0291-VCA0293; #strains with deletion of VCA0395-VCA0436; +minimum number of ORFs in the SI was uncertain, because some possible insertion fragments were not amplified and the exact ORFs could not be determined.

Preparation of Chromosomal DNA

All strains were grown in 5 mL Luria–Bertani broth to OD_{600} of 0.8. Chromosomal DNA was prepared with NucleoSpin Tissue kit (Macherey–Nagel, Duren, Germany), according to the manufacturer's protocol.

Primer Design

Genome sequence of strain N16961 was used as the template. From VCA0290 (gene coding is followed with N16961 genome), every two adjacent ORFs were amplified by a pair of primers designed in the two ORFs respectively. Every two adjacent amplicons overlapped with each other. More ORFs were included in one amplicon if no primers could be designed with this principle. Based on the published genome sequence of N16961[15], 213 pairs of primers were designed using Oligo6.0 primer analysis software (Molecular Biology Insights, Cascade, CO, USA) (Supplemental Table). In these primers, 175, 32,

five and one pairs of primers were used to amplify two, three, four and five adjacent ORFs, respectively. Primers ranged from 18 to 22 nt (most being 22 nt), and the average expected size of the amplicons was 885 bp. The maximum and minimum amplicon sizes obtained were 1998 and 153 bp, respectively. Overlaps between adjacent segments ranged from 1 to 625 bp, with the average size being 69 bp. When an amplification reaction failed, the primers flanking this negative fragment were used, and new primers locating the flanking positive regions were redesigned once the tentative PCR was still negative. If an amplicon with different size from strain N16961 was obtained, then the amplicon was sequenced and the ORF and VCR were determined.

PCR Scanning

PCR reactions were performed on a DNA Engine Tetrad® 2 Peltier Thermal Cycler (Bio-Rad, Hercules, CA, USA). N16961 was used as the positive control. Amplification reactions were carried out in a 20-μL volume using *rTaq* DNA polymerase system (TaKaRa, Dalian, China) according to the manufacturer's instructions. PCR was performed under the following conditions: initial denaturation at 94 °C for 5 min, 30 cycles of denaturation at 94 °C for 30 s, annealing at 55-58 °C for 30 s, extension at 72 °C for 1 min; and a single final extension at 72 °C for 7 min. The PCR products were analyzed by electrophoresis through a 1% agarose gel in a Bio-Rad Wide Mini-Sub Cell GT system. Repeated PCR reactions were performed for the negative amplification reaction to assure the reliability of non-amplifiable samples.

Multiple Loci Variable Number of Tandem Repeats (VNTRs) Analysis (MLVA)

Five previously described VNTR markers (VC0147, VC0437, VC1650, VCA0171, and VCA0283) were used[22]. The primers used for amplification of these five loci were the same as those reported initially[22], except for VC0147, for which a new primer set (VC0147-F: 5'-CAA ACG CAG GAT GAA CCA-3', VC0147-R: 5'-AAG AAG CCA GCG CCA ATA-3') was designed to yield bigger PCR products, thus allowing better distinction from PCR products of other loci when analyzed by capillary separation. The PCR products were analyzed by capillary separation along with an internal size standard (GeneScan® ROX-500 size standard, PerkinElmer Applied Biosystems, San Jose, CA, USA.) on a PE Applied Biosystems ABI Prism® 3 730 instrument. The data were conserved as *.fsa files and processed by GeneMarker version 1.71 software (SoftGenetics LLC, State College, PA, USA). The size bins had an error range of ±0.5 bp. If any product sizes were situated outside this interval, an error message was returned. The products sizes were exported and transformed to allele profiles in Excel files, and were entered into BioNumerics software (Applied Maths, Sint-Martens-Latem, Belgium) as character values. Dendrograms were clustered and constructed by using the unweighted pair group method using arithmetic averages (UPGMA).

Pulsed-field Gel Electrophoresis (PFGE)

PFGE was performed according to the PulseNet standardized PFGE protocol for *V. cholerae* subtyping[23]. Genomic DNA from all isolates was prepared in agarose plugs and digested with the restriction enzyme *Not*I, separated in a 1% agarose gel in 0.5× Tris-borate-EDTA at 14 °C using a CHEF-DRIII apparatus (Bio-Rad, Hercules, CA, USA). The pulse time ranged from 2 to 10 s for 13 h, and from 20 to 25 s for 6 h, both at 6 V/cm. After visualization, the PFGE patterns were analyzed using BioNumerics. The similarity between two patterns was expressed as a Dice coefficient[24]. Dendrograms were clustered and constructed by using the UPGMA with a tolerance of 1.0%.

RESULTS

Structure Map of SIs in Test V.cholerae El Tor Strains

In total, 12 780 PCR scanning reactions covering the SIs of 60 toxigenic El Tor strains were performed, including N16961 as the control. All amplicons were obtained with N16961 genome DNA as the template. Of all the reactions, 11 070 were positive and 1 710 (13.4%) were negative, suggesting that there was not much variation among these strains. In the negative reactions, at least 97.4% (1 665/1 710) were successive in two or more overlapping amplicons (SI of N16961 as the reference). For each negative reaction, primers were redesigned in the positive flanking regions to determine whether the negative results were caused by deletion or new fragment insertion. In this way, ORFs which existed in the test strains but not in N16961 and deletions were obtained by PCR and sequencing (Table 2). However, for some regions in certain strains (e.g. fragments VCA0292-VCA0324 and VCA0327-VCA0329 in strain 7751), redesigned primers located in the positive flanking boundaries still failed to yield amplicons,

which suggested that these regions were replaced by new fragments, but were too long to be amplified in our PCR tests. We did not obtain the sequence information for these regions. This group included strains 83535, 196153, 1961119, 62110, 2000180, and 78599.

Table 2. Sequence Analysis of Continuous Negative Results of PCR

Strain	Continuous Negative Result of PCR	The Primers Beside Continuous Negative Result of PCR	The Scope in N16961 Blast with the Amplicons	Deletion Scope	Deletion Size	Deleted ORF	Flanking Sequences Character of the Deleted Fragment	Translocated ORFs
43 Strains*	Primer pair 2-primer pair3	primer2F, primer3R	310072-312881	310943-312311	1368 bp	VCA0292 -VCA0293	attI	
25 Strains#	Primer pair 74-primer pair 116	Primer pair A	369708-396673	370682-394648	23966 bp	VCA0395 -VCA0436	552bp repeat sequence	VCA0404 VCA0405
829, 8212 19612533	Primer pair 2-primer pair 24	Primer2F,26R	329148-310111	328731-310940	17791 bp	VCA0292 -VCA0322	attI	VCA0323 VCA0324
	Primer pair 36-primer pair 61	Primer36F,62R	336648-357644	337154-345275 345927-357176	8121 bp 11249 bp	VCA0338 -VCA0351 VCA0354 -VCA0369	VCR Transpose	Replace 4364bp
	Primer pair 65-primer pair (35),	Primer63F,69R	358385-366412	359903-364709	6806 bp	VCA0375 -VCA0382	VCR	VCA0381 VCA0382
	Primer pair 128-primer pair 133	Primer128F,134R	409856-403422	408385-403792	4593 bp	VCA0452 -VCA0457	VCR	
	Primer pair 135-primer pair (37)	Primer134F,141R	408731-415517	410230-414767	4520bp	VCA0460 -VCA0467	VCR	VCA0462-VCA0464
73364 73448	Primer pair (11) - primer pair 48	Primer34F, 49R	334858-345981	335997-345368	9371 bp	VCA0337 -VCA0351	562 bp Repeat sequence	
7751	Primer pair 65-primer pair (35)	Primer63F, 69R	358385-366412	359903-364709	6806 bp	VCA0375 -VCA0382	VCR	VCA0381-VCA0382
	Primer pair 36-primer pair 61	Primer36F, 62R	336648-357644	337154-345275 345927-357176	8121 bp 11249 bp	VCA0338 -VCA0351 VCA0354 -VCA0369	VCR Transpose	Replace 5474 bp
801290	Primer pair 56-primer pair 61	Primer56F, 66R	350943-363007	351098-359784 361065-361474	8686 bp 409 bp	VCA0362 -VCA0374	VCR	VCA0369-VCA0371
19612540 78599	Primer pair 10-primer pair (8)	Primer9F, 23R	315671-327487	316488-325743	10261 bp	VCA0303 -VCA0317	VCR	Replace 9255 bp

Note. *and #as in Table 1.

Within the test strains, the PCR scanning results of nine test strains were similar to those of N16961, which indicated that 216 ORFs also existed in those strains (Supplemental Figure 1, from the bottom, lines 1-4 and 6-11). Only ORF deletion (compared to N16961) was detected in the SIs of the 45 test strains if translocation was not taken into account (Supplemental Figure S1, from the bottom, lines 12-55). The number of deleted ORFs ranged from 1 to 49, which indicated that 167-215 ORFs existed in those strains (Supplemental Figure S1, from the bottom, lines 12-55). In two test strains (78599 and 19612540), inserted fragments were obtained by PCR with primers located in boundary regions and

sequenced, and were found to comprise 15 ORFs, which replaced the corresponding 15 ORFs in N16961. In the left four test strains (8212, 829, 19612533, and 7751), both ORF deletions and new fragment insertions were detected. A total of 140 ORFs in each were found in the SIs of strains 8 212, 829, and 19612533. The number of ORFs in strain 7751 was uncertain because of PCR failure of insertions in some SI regions. In summary, the number of ORFs in the SIs of the test strains ranged from 140 to 216 (Supplemental Figure S1 and Table 1).

SI gene maps of the 60 test toxigenic El Tor strains were constructed, except for strains 7751, 83535, 196153, 1961119, 62110, 2000180, and 78599, due to undetermined insertions, based on the PCR scanning and sequencing (Supplemental Figure S1). Most of the variations in the SIs of the test strains clustered in the first two-thirds of the SI (Figure 2). The variation regions did not randomly distribute in the SI, which indicated that certain sites within the SI were hotspots for gene gain or loss. The structure of SIs in the test strains seemed to be syntenic, except for ORFs deletion and replacement.

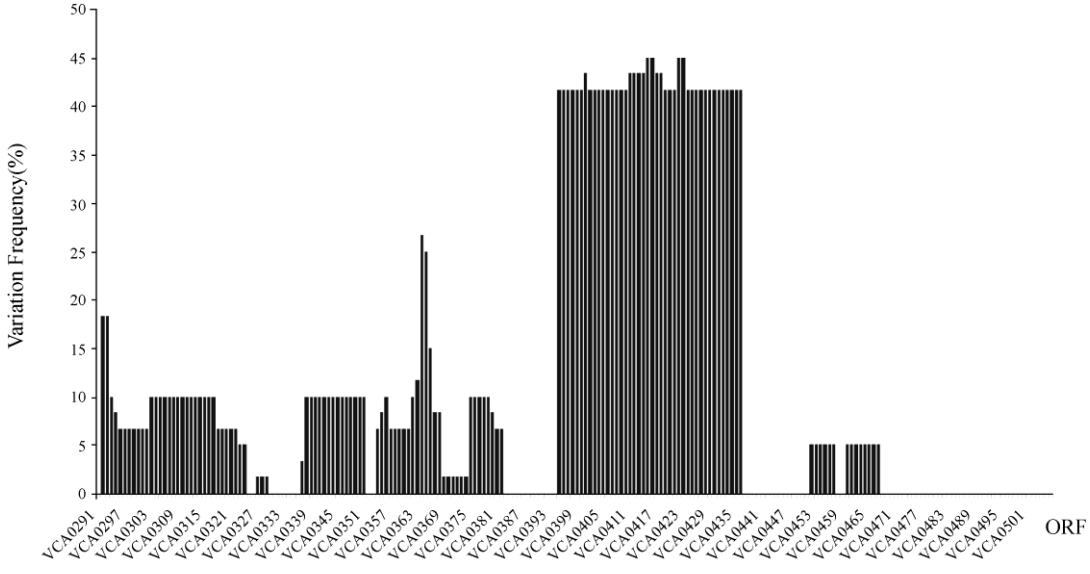


Figure 2. Variation frequency of ORFs in the SIs of the 60 tested *V. cholerae* strains based on the result of PCR scanning, compared to N16961. Most of the variation regions clustered in the first two-thirds of the SI (5' to 3').

Clustering Analysis of SIs from Test Strains Based on Gene Content

Based on the gene presence and absence in each SI, a minimum spanning tree was reconstructed for the test strains (Figure 3). The undetermined insertion genes were ignored. We distinguished the test strains into three groups: 1960s, 1970s-1980s, and 1990s and after, according to the epidemic curves and years since the start of the seventh cholera pandemic. SIs with the same content were grouped together. Different SI clustering characteristics for each epidemic period were found. In the tree, the strains isolated in the 1990s and after were grouped closely, whereas the 1970s-1980s strains formed a large dispersed cluster. No distinct clustering was found even if these strains were divided into two decades. The strains isolated in the 1960s were much more dispersed than the others, suggesting their divergence in SI content.

Gene Deletion, Insertion and Uncertain Transfer in SIs of the Test V. cholerae El Tor Strains

VCA0292-0293 Fragment Insertion No amplicons were obtained in 49 of the 60 test strains (81.7%) with primer pair 2F/3R (Supplemental Table), which detected VCA0292 and VCA0293. These strains spanned from 1961 to 2008, whereas all but one 1980s strain had these two genes. Amplicons of 1.4 kb were obtained in all of those 49 strains with primer pair 2F/3R, and sequencing confirmed the absence of a fragment spanning nt 310943-312311 (in N16961 genome), which included ORFs of VCA0292 and VCA0293 and the adjacent sequence (Figure 4a). These two genes are the closest genes to

IntI4 of SI. *attI*[10] was found in the upstream region of VCA0292. It has been reported that another integrase, *IntI1*, can catalyze recombination (*attI*×*attC*, *attC*×*attC*, and *attI*×*attI*), whereas recombination between *attI* and *attC* is the most efficient[25]. Moreover, it has been shown that the deletion frequencies at which *IntI1* and the integrase of Nitrosomonas europaea delete cassettes in the *attI* sites are low[26-27]. Therefore, we deduced that the fragment of VCA0292-0293 was integrated into the 11 test strains compared to those 49 strains, rather than deleted. The integration was possibly a recombination between *attI* and VCR mediated by integrase.

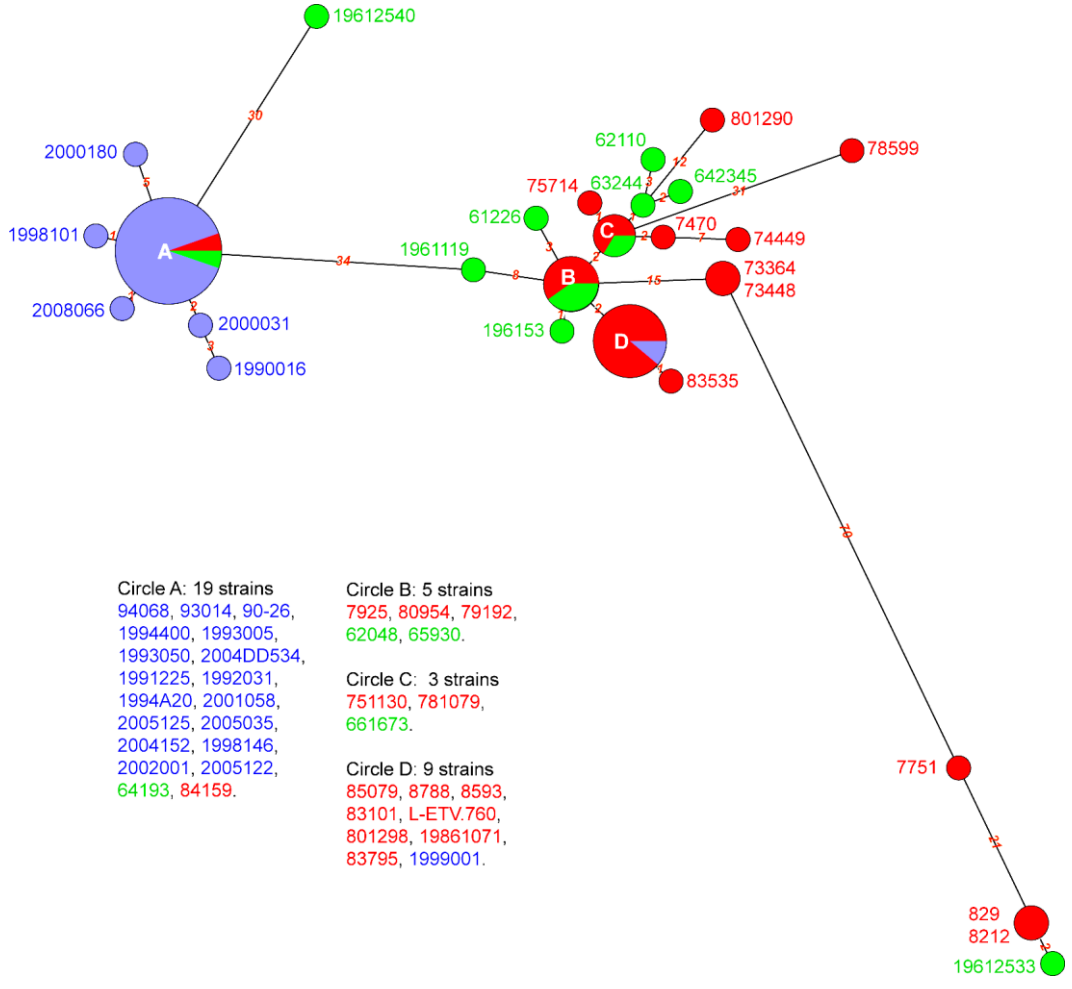


Figure 3. Minimal spanning tree of 26 subtypes of SIs based on the presence and absence data of the ORFs in the SIs of 60 toxigenic O1 El Tor *V. cholerae*. Subtypes are indicated by circles, whose diameter increases as the number of strains increases. The numbers on the black lines which connect two circles denote the number of ORFs difference between the two connected SI subtypes. The smallest and the second smallest circles denote subtypes which were only represented by one or two SIs. The strain names in which the SIs existed are shown. The circles with capital letters denote subtypes in which three or more SIs were included. The different capital letters represent strains in which different SI subtypes existed: A, including 94068, 93014, 90-26, 1994400, 1993005, 1993050, 2004DD534, 84159, 2005125, 2005035, 2004152, 1998146, 64193, 1991225, 1992031, 1994A20, 2001058, 2002001, and 2005122; B, including 7925, 80954, 79192, 62048, and 65930; C, including 751130, 781079, and 661673; D, including 85079, 8788, 8593, 83101, GDL-ETV.760, 801298, 19861071, 1999001, and 83795.

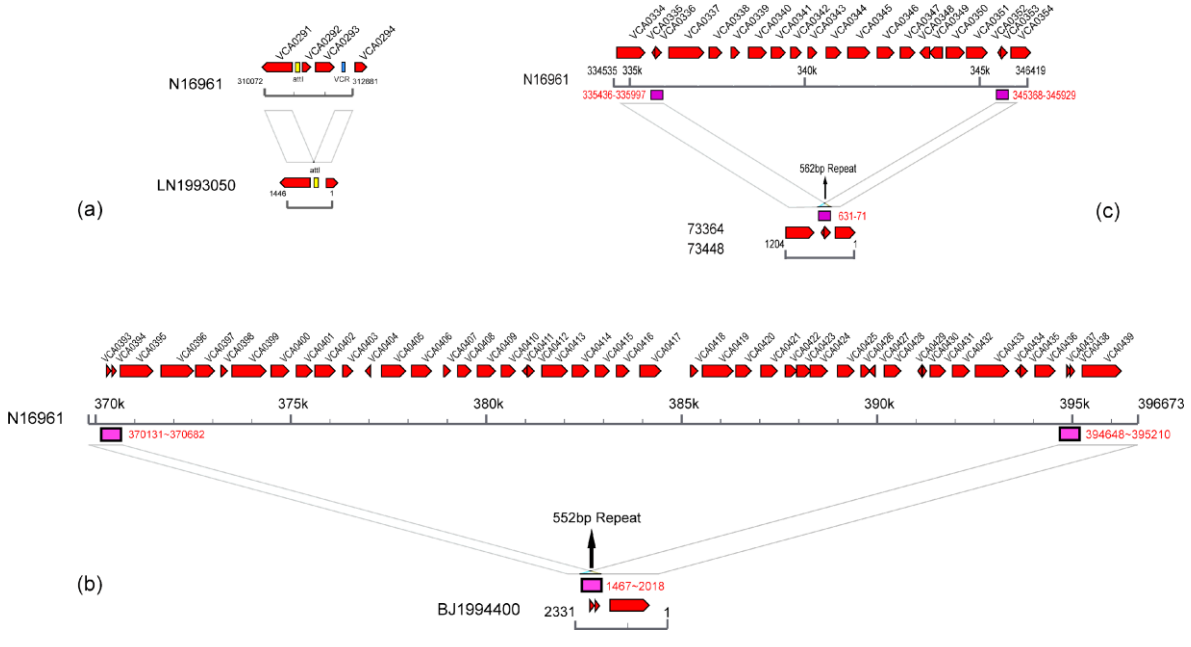
All the strains except for 801290, which possessed VCA0292 and VCA0293, had an identical SI genetic structure (named 80s-SI). We also performed MLVA and PFGE for these test strains.

Most of the 11 strains possessing VCA0292 and VCA0293 (10 and 9 strains, respectively, in these two analyses) clustered together (Supplemental Figures S2 and S3), which suggested that these strains have evolved from one common clone. Some divergence might have occurred, for example, in strain 801290, fragment VCA0362-0374 was probably deleted. Within this group, one exceptional strain, 1999001, had the same SI structure and similar MLVA and PFGE patterns, suggesting re-emergence of this clone in the 1990s.

VCA0395-0436 Deletion In the majority (95.7%, 22/23) of the test strains isolated in 1990 and after, and three strains isolated before 1990 (19612540, 64193, and 84159), most of the ORFs between VCA0395 and VCA0436 could not be detected (Supplemental Figure S1, from the top, lines 6-30). A previous study also has shown that an ~24-kb fragment deletion occurred in some O1 El Tor and O139 strains[20]. Fragments of 2.4 kb were obtained in those strains in our study with the primer pair A. (same to primer pair MGC86F/MGC128R in ref. 20). Sequencing revealed that 23 967 bp were deleted compared to the corresponding region in N16961. We called these SIs with absence of a 24-kb fragment 90s-SI. In N16961, 552-bp direct repeat sequences were detected in the both of the up and down stream flanking sequences of the 23 966-bp fragment; therefore, it is most likely that absence of this ~24-kb fragment was a deletion event caused by recombination between 552-bp repeats (Figure 4b). Additionally, the same deletion was also detected in three strains (19612540, 64193, and 84159) isolated before 1990 (Supplemental Figure S1).

In 96% (24/25) of the strains with 90s-SI amplicons could be obtained with primer pair 86 (detecting VCA0404 and VCA0405), although the result indicated that VCA0404 and VCA0405 were missing. No repeat copy of VCA0404 and VCA0405 was found in N16961 genomes. Therefore, we speculated these two genes translocated to other positions in the genomes of those strains. Similarly, some other dispersed translocations of genes within 90s-SIs were observed, such as VCA0396-0397 in strain 64193, VCA0406-0408 in 2002001, and VCA0363-0364 in 2000031. The absence of other scattered gene clusters was also found in several strains (Supplemental Figure S1).

In the minimum spanning tree of SIs, the 1990s strains with 90s-SI structure, which had the same or similar genetic components clustered together, and strains 64193, 84159, and 19612540 were also included (Figure 3). Most of the 1990s isolates clustered closely in MLVA and PFGE dendrogram trees (two clusters presented in Supplemental Figures S2 and S3), suggesting their discriminatory clones with other strains. Furthermore, strains 19612540 and 84159 had close similarity for PFGE patterns with the 1990s strains, and 84159 also had a similar MLVA pattern, suggesting the possible clone that appeared in the epidemics before 1990.



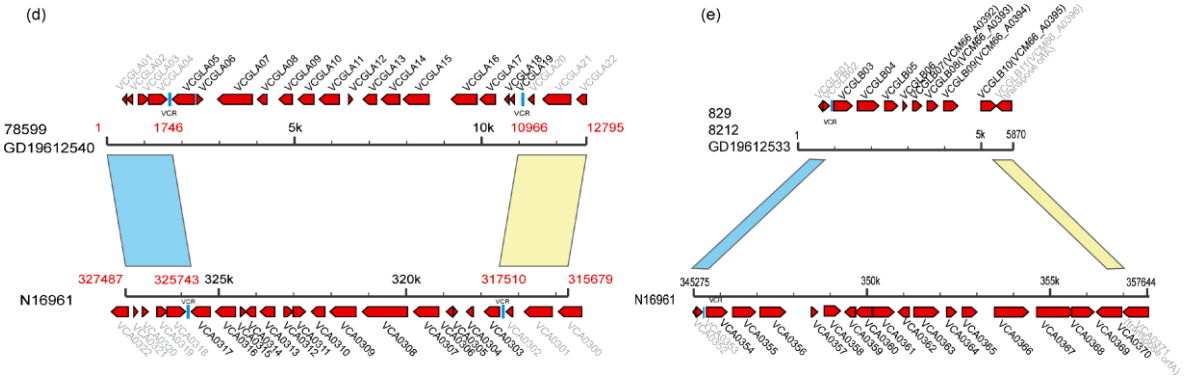


Figure 4. Schematic representation of sequence comparison between the tested strains and N16961. (a) Partial sequence comparison between strain with VCA0292-VCA0293 (N16961) and that without VCA0292-VCA0293 (LN1993050). The 1351-bp deleted sequence in LN1993050 included complete VCA0292, VCA0293, and the intergenic sequence between VCA0292 and VCA0293 and the adjacent sequence. The upstream region of the 1351-bp sequence was 223 bp away from *IntI4* (VCA0291), which is the position of the *attI* site (indicated as a yellow rectangle). (b) Partial sequence comparison between strain with VCA0395-VCA0436 (N16961) and that without VCA0395-VCA0436 (BJ1994400). In total, 23 966 bp were not detected in strain BJ1994400 in this region. The 9372-bp sequence included the complete sequence from VCA0395 to VCA0436 and the adjacent intergenic sequence. A 552-bp direct repeat sequence which was indicated as a purple rectangle was found in both of the flanking sequences of the up- and downstream region of fragment VCA0395-VCA0436. Fragment VCA0395-VCA0436 was not detected in most of the tested strains isolated after 1990 and three strains isolated before 1990. (c) Partial sequence comparison between strain with VCA0337-VCA0351 (N16961) and those without VCA0337-VCA0351 (73364 and 73448). The 9372-bp deletion sequence in those 3 test strains included the complete sequence from VCA0337 to VCA0351 and the adjacent intergenic sequence. A 562-bp direct repeat sequence which was indicated as purple rectangle was found in both of the flanking sequences of the up- and downstream regions of fragment VCA0337-VCA0351. (d) Partial sequence comparison between strain with fragment VCA0303-VCA0317 (N16961) and those in which fragment VCA0303-VCA0317 was replaced by a new sequence (GD19612540 and 78599). In those strains, fragment VCA0303-VCA0317 was replaced by a ~9-kb sequence that included 14 ORFs. VCRs were found in both of the flanking sequences of the up- and downstream regions of fragment VCA0303-VCA0317 in N16961. (e) Partial sequence comparison between strain with fragment VCA0354-VCA0370 (N16961) and those in which fragment VCA0354-VCA0370 was replaced by a new sequence (829, 8212 and GD19612533). In strains 829, 8212 and GD19612533, fragment VCA0354-VCA0370 was replaced by a ~4-kb sequence that included eight ORFs. Fragment VCGLB07-VCGLB11 in strains 829, 8212 and GD19612533 was identical to fragment VCM66A0392-VCM66A0396 in strain M66-2.

Deletions of VCA0292-0324, VCA0338-0351, VCA0375-0382, VCA0452-0457, and VCA0460-0467 All these mutations, possibly caused by deletion, occurred in the SIs of strains 19612533, 829, and 8212 [VCA0323-0324 was found to translocate to other positions in the genome but not within the SI (Supplemental Figure S1)]. Genes VCA0381 and VCA0382 in fragment VCA0375-0382, and genes VCA0462-VCA0464 in fragment VCA0460-0467 were still amplified individually, suggesting their translocation in the chromosome out of the SI. The same SI contents existed in these strains (Supplemental Figure S1 and Figure 3). These three strains also formed a tight cluster in MLVA and PFGE UPGMA trees (Supplemental Figures S2 and S3), suggesting their high clonality. We speculated that this was a non-predominant clone that caused a cholera epidemic, but it presented itself for >20 years.

Similar to GD19612533, 829 and 8212, VCA0338-0351 and VCA0375-0382 were also absent in strain 7751 (Supplemental Figure S1). Additionally, VCA0452-0457 and VCA0460-0467 were detected in strain 7751, although they were absent in strains 19612533, 829 and 8212. In the PFGE or MLVA UPGMA tree, 7751 was far removed from the cluster formed by 19612533, 829, and 8212.

VCA0337-0351 Deletion Negative results were obtained with primer pairs 11-48 (Supplemental Table 1) in strains 73364 and 73448, which covered VCA0337 to VCA0351. 19612533, 829, and 8212 also had similar fragment absence with one gene difference. A fragment of 1 204 bp was obtained in these two strains with primer pair 34F/49R. Sequencing confirmed a 9371-bp fragment was absent compared to SI of N16961. A 562-bp direct repeat sequence existed in both the up- and downstream flanking sequence of the absent fragment, suggesting the deletion event was mediated by homologous recombination of the 562-bp repeat sequences (Figure 4c). These two strains had the same SI composition (Figure 3), similar MLVA pattern (Supplemental Figure S2) but different PFGE patterns (Supplemental Figure S3), showing the possible difference in genome structure.

Cassettes Replacement The corresponding sequence of VCA0303-0317 (in N16961) in strains 19612540 and 78599 was replaced by a new fragment consisting of 15 ORFs (Figure 4d). The counterparts of these ORFs were also found in the genomes of several serogroups of *V. cholerae* and *V. mimicus* strains (Table 3), whereas in the genomes of these strains, these ORFs existed individually. VCA0395-0436 deletion in the 90s-SI was also seen in strain 19612540. The MLVA and PFGE patterns of these two strains differed greatly, which suggested that the VCA0303-0317 deletion happened independently in these two strains.

The counterpart of fragment VCA0354-0370 in N16961 was replaced by a fragment consisting of eight ORFs (Figure 4e) in strains GD19612533, 829, and 8212. Some of these ORFs also existed in *V. cholerae* classical strain O395 and El Tor strain M66-2. VCGLB03-VCGLB10 in this fragment had a corresponding counterpart in strain O395 (VC395_A0382-0389), which was also clustered (Figure 4e). VCGLB07-VCGLB11 also had corresponding counterparts in El Tor strain M66-2 (VCM66_A0392-0396). At the same site, a different fragment containing 10 ORFs replaced VCA0354-0370 in strain 7751 (Supplemental Figure S1).

However, although fragments VCA0292-0324, VCA0327-0329 and VCA0415-0416 were not detected in strain 7751, no amplicons were obtained with primers located in the flanking sequences of these fragments, suggesting large fragments might have been inserted into these regions. Such examples also included VCA0411-0418 in strain GD1961119, VCA0355-0356 in 62110, and VCA0375-0379 in 2000180.

Table 3. Function and the Best Hit of the ORFs in the 9-kb Insertion Sequence

ORF	Length	Function	Strains	Serogroup
VCGL000005	198	DinB family	*V. cholerae* AM-19226	O39
VCGL000006	55	conserved hypothetical protein	*V. cholerae* RC385	O135
VCGL000007	308	hypothetical protein A59_A0556	*V. cholerae* 623-39	non-O1/ non-O139
VCGL000008	86	Acetyltran-sferase, putative	*V. cholerae* MZO-3	O37
VCGL000009	119	glyoxalase family protein	*V. cholerae* 623-39	non-O1/ non-O139
VCGL000010	141	ORF-SIK7-1341	*V. cholerae*	
VCGL000011	179	Phosphogl-ycerate mutase family protein	*V. cholerae* MZO-3	O37
VCGL000013	121	SH3 domain protein	*V. cholerae* 1587	O12
VCGL000014	162	hypothetical protein A59_A0581	*V. cholerae* 623-39	non-O1/ non-O139
VCGL000015	231	conserved hypothetical protein	*V. cholerae* RC385	O135
VCGL000016	229	hypothetical protein VMA_000713	*V. mimicus* VM223	
VCGL000017	137	glyoxalase family protein	*V. cholerae* TMA 21	non-O1/ non-O139
VCGL000018	41	hypothetical protein VCA0435	*V. cholerae* N16961	O1
VCGL000019	37	hypothetical protein A5A_B0010	*V. cholerae* MZO-2	O14

Other Dispersive Absence of Gene Fragments Some other dispersed gene deletions compared to N16961 were also observed, such as: VCA0364-0365 absence in strains 751130, 75714, 781079, 74449, and 661673, where all the strains except for 74449 had identical or similar SI structure, 751130 and 75714 had the same MLVA and similar PFGE patterns, whereas others had similar MLVA but different PFGE patterns; VCA0364-0366 absence in strains 63244 and 642345, which had similar SI structures and MLVA patterns but different PFGE patterns; and VCA0365-0366 absence in strains 62110 and 61226, which also had similar SI structures and MLVA patterns, but different PFGE patterns. Other dispersive deletions included VCA0362-0374 in strain

801290 and VCA0294-0295 in 642345, and VCA0356 absence in 74449. These absences suggest sporadic genetic events occurred in different strains, but some clones with the same or different SI genetic composition and similar or different MLVA and PFGE patterns may also be observed, suggesting divergent genetic clones within these 1960s and 1970s strains.

Some insertions were also observed, although others still could not be discovered by PCR (Supplemental Figure S1), probably because the fragment length was much longer. It is interesting that the insertions were often accompanied by gene deletions in the integrating sites.

SI Content Types of Strains in Different Decades

Based on the cassette content sketching, we calculated the types of the SIs in the strains from different decades, on which the different gene contents were based. As shown in Table 4, the ratios of types/strains in different decades decreased from the 1960s to 1990s and after. Even the strains from the 1970s and 1980s were merged; the ratio was 0.52 (13/25) in the 1970s-1980s group. It seems that the gradually simplified diversity of SIs was observed, which may represent the SI diversity of toxigenic El Tor isolates from China, a regional cholera epidemic area.

Table 4. Ratios (No. of types/No. of strains) of the SIs in the Strains from Different Decades*

Decade	Number of Strains	Number of SI Types	Ratio
1960s	12	11	0.91
1970s	11	8	0.73
1980s	14	7	0.50
1990s and after	23	7	0.30

Note. *Ratio was 0.60 (15/25) when the strains from 1970s and 1980s were grouped together.

Gene Flows in SIs

Based on the cassettes content and repeat sequences analysis, a sketch map of hypothetical gene flows among the different SI clades in the minimum spanning tree was constructed, which showed possible generation of the SIs in the different strains by gene deletion, insertion and replacement (Figure 5). The ancestor SI (or even the

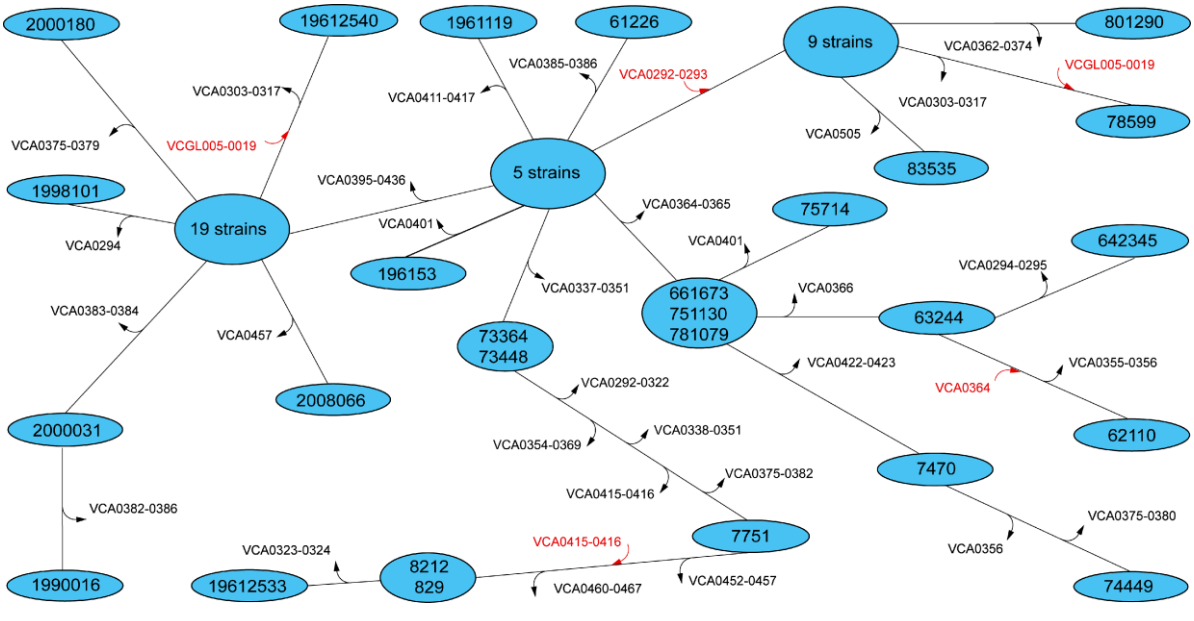


Figure 5. Proposed hypothetical gene flows in SIs in the 60 tested toxigenic O1 El Tor *V. cholerae* strains. Probable insertions and deletions of ORFs found in 60 tested *V. cholerae* strains are indicated by red and black arrows, respectively, along the minimum spanning tree based on the PCR scanning data. Hypothetical keypoint SIs are indicated by yellow circles. The names of the strains in which the corresponding SIs existed are shown in the circles. Nineteen strains comprised 94068, 93014, 90-26, 1994400, 1993005, 1993050, 2004DD534, 84159, 2005125, 2005035, 2004152, 1998146, 64193, 1991225, 1992031, 1994A20, 2001058, 2002001, and 2005122; five strains comprised 7925, 80954, 79192, 62048, and 65930; and nine strains comprised 85079, 8788, 8593, 83101, L-ETV.760, 801298, 19861071, 1999001, and 83795.

host strain) could not be predicted because of its complicated structure and active recombination, and the sampling of the strains used in this study, therefore no arrow was used in the link line between two SIs in Figure 5. However, frequent lateral gene transfer could be seen among the SIs with different gene content. The largest number of strains used in this study possessed 90s-SI, and some gene indels still occurred within the 1990s strains. Twelve strains isolated in the 1960s had the most variant SI cassette contents, whereas the 23 1990s strains had only seven SI patterns with different cassette contents (Figure 3).

DISCUSSION

The SI is inclined to integrate and excise ORFs frequently, which results in variance of its content and structure[28]. In this study, we analyzed the SI components based on PCR scanning and sequencing in isolates from the seventh cholera pandemic. By using this strategy, the deletions, insertions and rearrangement of cassettes were found within the different SIs. We showed that the SIs in the strains isolated from a span of nearly 50 years exhibited syntenic character but still certain diversity, especially those isolated before 1990.

An integron is an active gene capture system for its distinctive structure, especially the SI, and confers upon bacteria new abilities, such as survival, and drives bacterial evolution by integrating exogenous genes and converting them to expression[28]. With the collection of epidemic El Tor isolates between 1961 and 2008, multiple deletions/insertions and replacements were found to contribute to the diversity of SIs. It was particularly interesting that successive ORF variations were common in SI structural mutations, while deletions or insertions of single ORFs were rare, which shows clearly the mobilization of the cassettes in clusters. The SI integrase can integrate and excise ORFs through recombination between *attI* sites and VCRs or between different VCRs[29]. In the SIs of the test strains, VCRs were present in the flanking sequences of most of the variation regions, which suggested that these gene flows were caused by recombination between VCRs, which was catalyzed by integrase. Homologous recombination between the 550-bp repeat sequences flanking the VCA0395-0436 fragment possibly resulted in the characteristic ~24-kb deletion, which generated the 90s-SI structure. It also resulted in VCA0337-0351 deletion in two 1970s strains. Whether loss of these genes is advantageous for environmental and host adaption, or just a random deletion event during clone development, needs further functional studies. Within the SI of N16961, 24 copies of ~550-bp repeat sequences exist[10], and the 550-bp repeat has high identity to the 552 bp that mediated deletion of VCA0395-436 and the 562 bp that mediated deletion of VCA0337-0351. Thus, we speculate that 24 copies of ~550 bp may provide possible sites of cassette flow in clusters mediated by homologous recombination.

The ORFs in the SI of bacteria may be transferred from virus, bacteria and even eukaryotes[10], which is one of the sources of genome diversity and novel phenotypes in bacteria[30]. Within three different insertion fragments in the SIs of some strains, we identified with PCR scanning and sequencing some genes that were also found in different serogroups of *V. cholerae* and even *V. mimicus*, which is more closely related to *V. cholerae* than other *Vibrios*. The SI integrases of these two species and the VCRs and *V. mimicus* repeats were closer than those between *V. cholerae* and other *Vibrios*[31]. Thus, it seems that the genes captured by SIs of El Tor strains were preferentially those within the same species and the closest phylogenetic species. These autochthonous bacteria in the coastal and estuarine environment share common ecosystems, and lateral gene transfer may occur in their environmental niches[13,31-32]. The genes shared with *V. mimicus* in these insertion fragments are clustered in *V. cholerae*, whereas in *V. mimicus* SI, they are located at different sites. This suggests that SI cassettes can transfer both in individual and cluster manners among different hosts.

In our study, although the strains were analyzed at a regional level, the distinct structural variance of SIs in the host strains was observed for the time span of the seventh cholera pandemic. The SIs from the strains of the 1960s epidemic wave showed multiplicity in their contents and much dispersal in the minimum spanning tree. Continuous epidemic waves occurred in the 1970s and 1980s. The SIs in the strains from the 1970s and 1980s showed lower dispersal than those from the 1960s, whereas it seems that complex clones still contemporaneously existed. When the strains from these two decades were separated into two groups, some SIs from the 1980s strains showed a similar SI structure, however multiplex SIs were continuously dispersed in the 1970s and 1980s strains. In comparison, a highly similar SI structure with a characteristic ~24-kb

deletion existed in the strains from the 1990s and after, although some epidemic peaks appeared in this decade. Interestingly, the same ~24-kb deletion in 90s-SI in the El Tor strains has also been observed in the SI of O139 strain[20], which emerged and caused a cholera epidemic in 1992[14]. Some cassette contents of SIs and even their host strains are probably sustained for decades, because the same SI structure and similar MLVA and PFGE patterns were obtained from the strains from different decades. The 1990s clone with 90s-SI might have appeared in the 1980s, because strain 84159 had the 90s-SI and similar PFGE and MLVA patterns. Whether strain 64193 isolated in 1964 belongs to the 1990s clone with 90s-SI or not, needs further evolutionary analysis based on whole genome information. A similar situation was also observed for some 1960s, 1970s and 1980s strains, and further studies are required.

The gene content change in SIs is still obvious among strains from epidemics in different decades; however, the divergence is based on syntenic structure of SIs in these El Tor strains. From this study, it seems that the SI clades of the pandemic strains are undergoing simplification: the number of SI structural patterns in the strains grouped by decade decreased. Whether this is only the drift in clones in regional epidemics, or the result of natural stress selection needs to be further determined. However, it should be mentioned that SI diversity analysis may rely on extensive sampling. Clones of *V. cholerae* appearing in epidemics may be influenced by many factors, including their different environmental survival and their opportunity to invade and spread in the human population. Additionally, health care, economic factors and social development can influence the clonal spread by affecting epidemic severity. Therefore, the strains used in the evolutionary analysis may be only a part of the clones in the environment. The simplification may be only the SI structure dynamic character of the *V. cholerae* isolated in China. In addition, the panorama of gene flow and evolution of SIs could be described much clearly if the non-toxigenic El Tor strains are included, although it is difficult to obtain good samples from the strains surviving in the environment. In this study, only *V. cholerae* strains isolated in China were used. The use of strains from other countries and regions may give a global view of SI variation.

Drift in the genomes of the seventh pandemic *V. cholerae* strains is derived mainly from lateral gene transfer and is ongoing[30], and the diversity of SI structure observed in this study may provide proof of genome diversity. As a special genomic structure in *V. cholerae*, the SI confers genome diversity and probably environmental adaptation for evolution of this bacterium. Such a variable genome structure can be used in the strain evolution analysis in different scales and even in clone definition, and may be used in the molecular epidemiological surveillance and source tracing in outbreak investigations. Here, the detailed gene content of SIs from the different *V. cholerae* El Tor strains was analyzed. With the whole genome sequences of many strains, the interaction between SIs and the genome in *V. cholerae* evolution can also be conducted, and SI gene function may even be discovered.

ACKNOWLEDGEMENTS

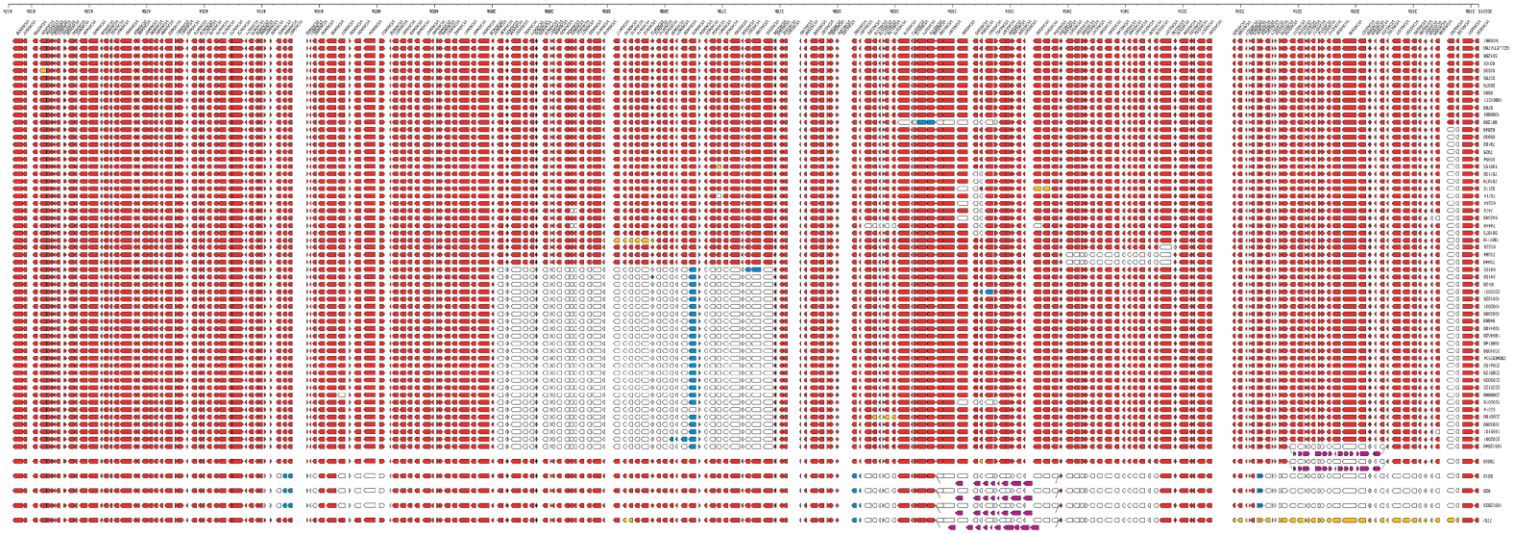
We sincerely thank Professor LIANG Wei Li for helpful comments and discussions.

REFERENCES

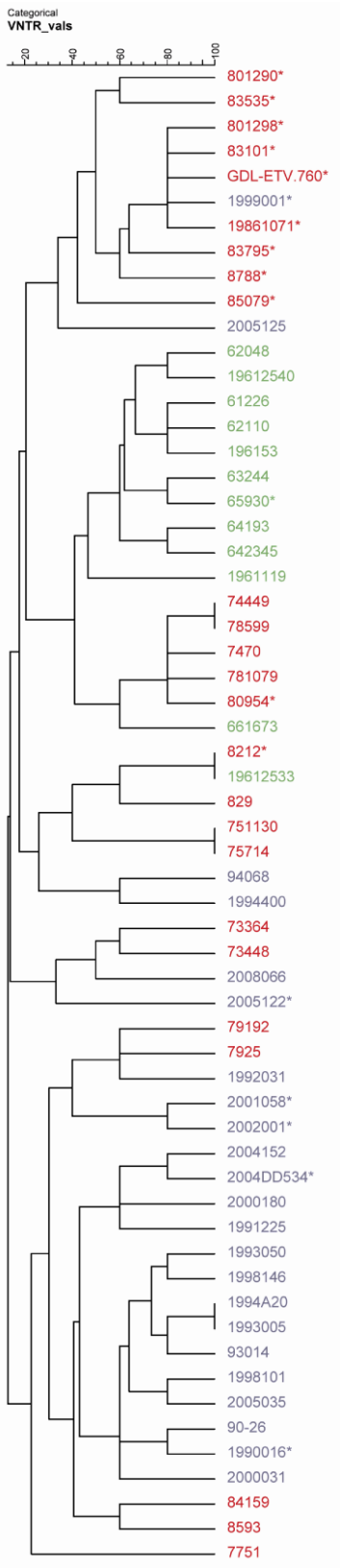
1. Martinez E, de la Cruz F. Genetic elements involved in Tn21 site-specific integration, a novel mechanism for the dissemination of antibiotic resistance genes. The EMBO journal, 1990; 9, 1275-81.
2. Stokes HW, Hall RM. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. Mol Microbiol, 1989; 3, 1669-83.
3. Hall RM, Collis CM. Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. Mol Microbiol, 1995; 15, 593-600.
4. Sundstrom L. The potential of integrons and connected programmed rearrangements for mediating horizontal gene transfer. APMIS Suppl, 1998; 84, 37-42.
5. Mazel D, Dychinco B, Webb VA, et al. A distinctive class of integron in the *Vibrio cholerae* genome. Science, 1998; 280, 605-8.
6. Barker A, Clark CA, Manning PA. Identification of VCR, a repeated sequence associated with a locus encoding a hemagglutinin in *Vibrio cholerae* O1. J Bacteriol, 1994; 176, 5450-8.
7. Dean A. Rowe-Magnus MZ, and Didier Mazel. The adaptive genetic arsental of pathogenic *Vibrio* species: the role of integrons. Washington, DC, ASM Press, 2006; 95-111.
8. Chen CY, Wu KM, Chang YC, et al. Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. Genome Res, 2003; 13, 2577-87.
9. Mazel D. Integrons: agents of bacterial evolution. Nat Rev Microbiol, 2006; 4(8), 608-20.
10. Rowe-Magnus DA, Guerout AM, Mazel D. Super-integrons. Res Microbiol, 1999; 150, 641-51.
11. Ogawa A, Takeda T. The gene encoding the heat-stable enterotoxin of *Vibrio cholerae* is flanked by 123-base pair direct repeats. Microbiol Immunol, 1993; 37, 607-16.
12. Franzon VL, Barker A, Manning PA. Nucleotide sequence encoding the mannose-fucose-resistant hemagglutinin of *Vibrio cholerae* O1 and construction of a mutant. Infect Immun, 1993; 61, 3032-7.

13. Kaper JB, Morris JG, Jr., Levine MM. Cholera. Clin Microbiol Rev, 1995; 8, 48-86.
14. Ramamurthy T, Garg S, Sharma R, et al. Emergence of novel strain of *Vibrio cholerae* with epidemic potential in southern and eastern India. Lancet, 1993; 341, 703-4.
15. Heidelberg JF, Eisen JA, Nelson WC, et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature, 2000; 406, 477-83.
16. Pang B, Yan M, Cui Z, et al. Genetic diversity of toxigenic and nontoxigenic *Vibrio cholerae* serogroups O1 and O139 revealed by array-based comparative genomic hybridization. J Bacteriol, 2007; 189, 4837-49.
17. Pang B, Zheng X, Diao B, et al. Whole genome PCR Scanning reveals the syntenic genome structure of toxigenic *Vibrio cholera* strains in the O1/O139 population. PLoS ONE, 2011; 6, e24267
18. Feng L, Reeves PR, Lan R, et al. A recalibrated molecular clock and independent origins for the cholera pandemic clones. PLoS ONE, 2008; 3, e4053.
19. Castaneda NC, Pichel M, Orman B, et al. Genetic characterization of *Vibrio cholerae* isolates from Argentina by *V. cholerae* repeated sequences-polymerase chain reaction. Diagn Microbiol Infect Dis, 2005; 53, 175-83.
20. Labbate M, Boucher Y, Joss MJ, et al. Use of chromosomal integron arrays as a phylogenetic typing system for *Vibrio cholerae* pandemic strains. Microbiology , 2007; 153, 1488-98.
21. Clark CA, Purins L, Kaewrakon P, et al. The *Vibrio cholerae* O1 chromosomal integron. Microbiology , 2000; 146, 2605-12.
22. Danin-Poleg Y, Cohen LA, Gancz H, et al. *Vibrio cholerae* strain typing and phylogeny study based on simple sequence repeats. J Clin Microbiol, 2007; 45, 736-46.
23. Cooper KL, Luey CK, Bird M, et al. Development and validation of a PulseNet standardized pulsed-field gel electrophoresis protocol for subtyping of Vibrio cholerae. Foodborne Pathog Dis, 2006; 3, 51-8.
24. Dice LR. Measures of the amount of ecological association between species. Ecology, 1945; 26, 6.
25. Collis CM, Recchia GD, Kim MJ, et al. Efficiency of recombination reactions catalyzed by class 1 integron integrase *IntI1*. J Bacteriol, 2001; 183, 2535-42.
26. Hansson K, Sundstrom L, Pelletier A, et al. *IntI2* integron integrase in Tn7. J Bacteriol, 2002; 184, 1712-21.
27. Leon G, Roy PH. Excision and integration of cassettes by an integron integrase of *Nitrosomonas europaea*. J Bacteriol, 2003; 185, 2036-41.
28. Rowe-Magnus DA, Guerout AM, Biskri L, et al. Comparative analysis of superintegrons: engineering extensive genetic diversity in the *Vibrionaceae*. Genome Res, 2003; 13, 428-42.
29. MacDonald D, Demarre G, Bouvier M, et al. Structural basis for broad DNA-specificity in integron recombination. Nature, 2006; 440, 1157-62.
30. Holmes AJ, Gillings MR, Nield BS, et al. The gene cassette metagenome is a basic resource for bacterial genome evolution. Environ Microbiol, 2003; 5, 383-94.
31. Rowe-Magnus DA, Guerout AM, Ploncard P, et al. The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. Proc Natl Acad Sci USA, 2001; 98, 652-57.
32. Chun J, Grim CJ, Hasan NA, et al. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. Proc Natl Acad Sci USA, 2009; 106, 15442-7.

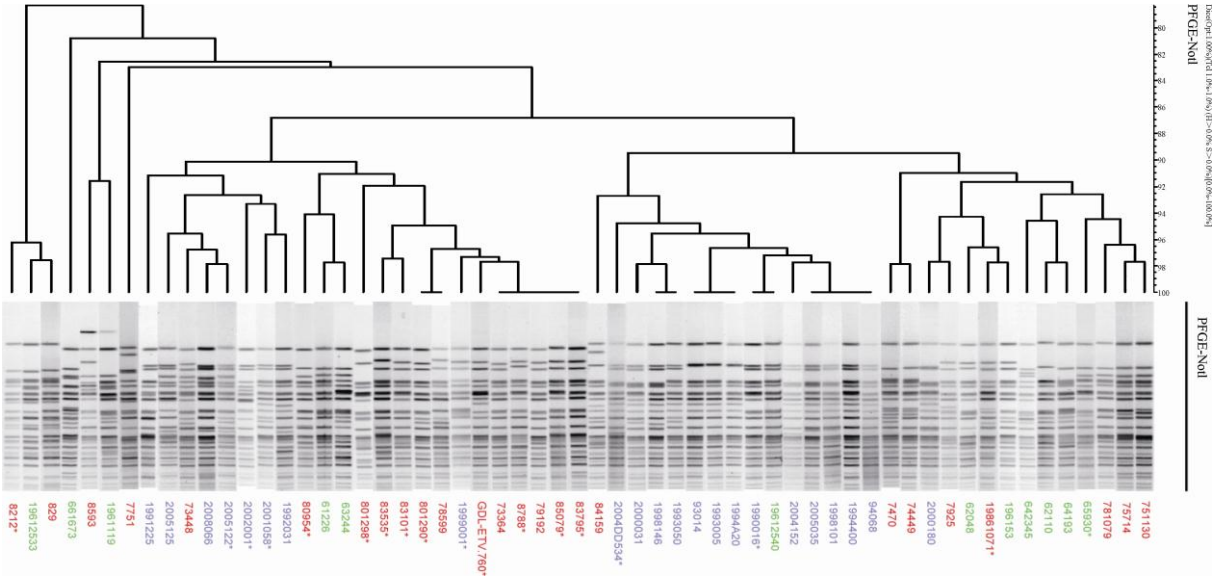
attention: Supplemental Table and Supplemental Figure S1, S2, and S3 can be found in the whole text on www.besjournal.com, 2011; 24(6).



Supplemental Figure S1. Composition map of the SIs in the 60 tested toxigenic O1 El Tor V. cholerae strains, drawn according to the results of PCR scanning and sequencing. The ORFs in N16961 are listed according to the position in the genome and are at the bottom of the figure. The strain names are listed on the left. Red arrows denote that sequences in the tested strains were the same as that in N16961. White arrows denote sequences that were not detected in the tested strains. Purple arrows denote sequences that were in the tested strains but not in N16961. Blue arrows denote potential sequence transfers. Orange arrows denote sequences that were not detected by PCR scanning and not confirmed by sequencing.



Supplemental Figure S2. MLVA pattern cluster of 60 test strains. The color of the strain name indicates the isolation time. Green, red and blue represents 1960s, 1970s–1980s and 1990s and after, respectively. An asterisk indicates that the serotype of the strain is Inaba.



Supplemental Figure S3. *Not*I-PFGE cluster of 60 test strains. The test strains are indicated as those in the MLVA cluster.

Supplemental Table. Primers in This Study

Primer pair	Product Size*	Left Primer	Right Primer	Primer pair	Product Size*	Left Primer	Right Primer
1	654	taaagctggtcaatacgcttac	agtatccaaatgcaccttatga	82	390	agtttggggaaaagataatgag	aacaactgtcccatactcttca
2	1060	tcataaggtgcatttggatact	cgtttcttctgctgttcactat	83	613	ctaaagtcgttcaacagcaaat	gttaaggactctgcaaacaaga
(1)	1965	tcataaggtgcatttggatact	ttccacccatttttacttcacta	84	637	ttgcagagtccttaacaaagtc	aatcctaatagtgatgccacaa
3	1150	ggtgtttgatcacaagagaaac	ttttgtctgcttggactatttc	85	582	ttggcagctttaatctttactg	ctttttgtgaaaccaagttgac
4	488	gaaatagtccaagcagacaaaa	taacagcctctgaaacagaaag	86	896	gtcaacttggtttcacaaaaag	tttcacgttacctttgagagag
5	462	tttctgtttcagaggctgttag	acactttctaccaatgcaacat	87	634	ctctctcaaaggtaacgtgaaa	acggtctaactcaaacactacg
(2)	743	atgttgcattggtagaaagtgt	gggcagtagcctctactaattt	88	472	ttcatagaaatgtcgaaagagc	ttaagttggtcacggttaagag
6	542	ggtcagtttggtttcacaaata	ctgcgcataaacttaatcactt	89	408	caacttaaccttgatcgttttg	cgactccaagctgtagataaaa
7	511	gaacgaagctgaagtgattaag	caccgacgcctatatatgtaac	90	353	ttttcgttgctggagtagttag	caaaatcaattcctgtcacttc
8	858	atgctggtgcaatgattactat	tgtctgatgaacacagagactg	91	586	gaagtgacaggaattgattttg	cccatgtataaatgctgttctc
9	853	caattattggtcatgatgtgtg	agttgacattcgtttcactgtt	92	363	ctaagcctgcgttaagtgtaag	tcgagacttgctagttttcatc
(3)	1409	aacagtgaaacgaatgtcaact	tacgatctgacgaataggttgt	93	167	ttcaaagtcgatgaaaactagc	ctgccaaaacttaaaacgatct
10	630	caacttaaccttgaatgctttg	tacgatctgacgaataggttgt	94	901	aacactcaaagtttggtttcac	agaaatgatagagcccaaatct
(4)	845	tcagatcgtacgaatggtaaat	gaaaccactgaactttcaacac	95	414	agatttgggctctatcatttct	tccttgaacctgtatgtcactt
11	465	acaaaatacagcatcatcactg	gaaaccactgaactttcaacac	96	355	tggactcataaagatcctcaaa	ctgcatattcaggtccatagtc
12	239	gtgttgaaagttcagtggtttc	tatgcaaaagcatctctcaaac	97	480	gtaatgaatttgctgtttggtc	gacaactcaatgaggcaataac
(5)	635	gtgttgaaagttcagtggtttc	ctcaccaactgtagccttagaa	98	576	gttattgcctcattgagttgtc	gctttctttacactcgatacca
13	1435	ttctaaggctacagttggtgag	tttgtgcctctttagcaatatc	99	1044	gaatagaattgaccgcatattg	ctcaaaacgatcaaggttaagt
14	1643	ttttcgacgtctacaaggtaag	tttaccgagatctcttttgttg	100	879	caacttaaccttgatcgttttg	aaaggtatcgtgcctacttttt
15	403	ttaaagacctaaagccattgtg	ccagaagaataagcatcgaata	101	808	ctagtctcgctgcattctttag	ctcatcgtcaccatctaatttc
16	483	gatgcttattcttctggtgcta	ggaaacacatggcttattacac	102	491	gtcattgattctgaaggttttg	acccactaaagcaatcattaca
(6)	960	gaagtttgcttctttggtgtaa	attctcgttgatggacttatga	103	630	tttgcaagaccttaatcttgac	acacctttacctcgttgtgtaa
17	585	aaagtccttaatggacatgaca	attctcgttgatggacttatga	104	209	ttacacaacgaggtaaaggtgt	gtttttgtaacgctagaggtga
18	495	tcataagtccatcaacgagaat	ctaaacaacaatgacgaacaca	105	466	gcagagtttattgctttggata	tacattgctgcttcattagaca
(7)	527	tgtgttcgtcattgttgtttag	agcagataagctgaattgtgtc	106	671	gttattgcctcattgagttgtc	ccattcgtattcaccataaaac
19	280	gcatcagaaacacacacataac	agcagataagctgaattgtgtc	107	398	tggactcataaagatcctcaaa	cgacatatctcgtcaatagtcc

(Continued)

Primer pair	Product Size*	Left Primer	Right Primer	Primer pair	Product Size*	Left Primer	Right Primer
20	654	gaatagaattgaccgcatattg	acggtctaactcaaacactacg	(22)	730	tgtgttcgtcattgttgtttag	acagtaaaatggcgtatgttgt
(8)	956	ttcatagaaatgtcgaaagagc	gcttatatcatcggcttttaca	108	426	cgacatgaagcatcttagtctt	tacgaaaccgaatagcatttac
21	557	ttcatagaaatgtcgaaagagc	cgctagagttgatgtcgatact	109	735	acaacatacgccattttactgt	gacaactcaatgaggcaataac
22	568	attgagttttgctctcataacg	ccaacttgtgaaatttggatac	(23)	831	acaacatacgccattttactgt	aatttcacctatgaactgacca
(9)	529	aaaacgcagtagtttcaaagtc	cttcgtatatttcgtgcagagt	110	386	aggtgaaatttagccagttct	caaaaccttcagaatcaatgac
23	360	agttttggtagcttgcttcac	cttcgtatatttcgtgcagagt	111	556	gtcattgattctgaaggttttg	acaccgtcttggtttatgtaga
24	433	agacgttatgagaaagcgtaaa	gaagtcggcaatgttatttcta	112	1078	tctacataaaccaagacggtgt	gttgaagtctcaagaggtagca
25	219	cgttgcagcttagaaataacat	cctaatgaggttactgcatctg	113	479	tgctacctcttgagacttcaac	tgctgaactttcaacactttct
(33)	1617	atggactctgcacgaaatatac	caactaggtcaatgtggaaaac	114	249	cgatttggtagaaagtgttgaa	caaaagcatctctcaaacacac
(34)	1175	agtctgcgtgaggttagtattg	taaccaactgaaatcattgacc	115	367	ctgtgatcttgctttctttgtc	ctgcatcttcgactgttacttt
26	319	gaagtttggtcaatgatttcag	ctgaattttcagcaccagatac	116	860	agtaacagtcgaagatgcagag	gacaaagaaagcaagatcacag
27	324	ggattcagctttatgtcctttt	ctgaattttcagcaccagatac	(24)	914	agtaacagtcgaagatgcagag	ggctaccaaaacttaaaacgat
28	548	ggtgctgaaaattcagtgttat	tcattaccgagttgatgaagat	(25)	1610	gatcgttttaagttttggtagc	gtatcaacgcaataatttggac
(10)	1097	caatgtatctggtgctgaaaat	ctgcgcataaacttaatcactt	117	743	agtttatcggtaaaggcaaagt	gtatcaacgcaataatttggac
29	1998	gtgattaagtttatgcgcagtc	gaatctttacgtttccaagtca	118	856	gtccaaattattgcgttgatac	gaacaaacatctttccattgag
30	750	cagcaaaggtaaaacctcagta	ggtagttccctctgaaatgaat	119	682	ctcaatggaaagatgtttgttc	atccgattcttgaaactcttct
31	503	attcatttcagagggaactacc	acaatccttgagctgtatgaaa	120	777	tgaacagaacacagttgctaga	attgctgctgaacgtaaataag
32	341	tactaaacggtaagcattggtc	ctcacacttgaatgatgagaca	121	477	ttcatagaaatgtcgaaagagc	gagttcaaaaagagtgctttga
33	879	aagcaaaaacgagtcaatgtag	ctaaagcagcttctacgacatc	122	176	ttcaaggcgttctattagtttc	gcttgaagatcttgagctactg
34	887	aagcacattaaattctgactgg	acactttctaccaatgcaacat	123	511	caagcttatccatcatgatctc	ttctacctcttctgcattgttc
(11)	1009	atgttgcattggtagaaagtgt	caatcgtgcttaatatctgtcc	124	622	gaacaatgcagaagaggtagaa	agcacgataatgatatcccata
35	913	ggtcagtttggtttcacaaata	caatcgtgcttaatatctgtcc	125	786	tatgggatatcattatcgtgct	acagtaaaatggcgtatgttgt
36	698	agtttatcggtaaaggcaaagt	ctgcatattcaggtccatagtc	126	1484	acaacatacgccattttactgt	tcaagacctatgggtttaatca
(12)	1197	ccactttagaaaaagtcgtgaa	gagttatcaccgtcaatcagag	127	585	aatcggtacttttagaggcagt	atttttagtgaagcgaactgtg
37	733	ttacttatgggctgaaagtctg	gagttatcaccgtcaatcagag	128	496	aacactatcggacgaaatgata	aactcataagcattcgtgaaac
38	390	acgaagtgtgactgtacaaggt	tcattaccgagttgatgaagat	129	878	gtttcacgaatgcttatgagtt	acacatcagacagcgatatttt
39	670	ctagagtttggtcaccagaaga	tgctgttgacactgatactgac	130	1874	aaaatatcgctgtctgatgtgt	gggtaaaacagtaaaaccacct
40	442	aaaggcactattcagctatgtg	ctgcgcataaacttaatcactt	131	512	tgacttagtcgttgtggaagat	ctttttgtgaaaccaagttgac
41	701	gaacgaagctgaagtgattaag	caaaatcagagcaattaggaga	132	704	gtcaacttggtttcacaaaaag	gctcaatgtagtaagcgtcaat
42	989	acaacatacgccattttactgt	cataaactccaaaactgaccac	133	933	gagcatattgaaaaaggtagca	ctcctgcaagtaaactagacca
43	484	gtggtcagttttggagtttatg	tgagttacaacatcccaagtct	134	1215	gctggcatagaactaaatgaaa	accactggtaacaaatctcaaa
44	545	agacttgggatgttgtaactca	atttgctgttgaacgactttag	135	767	agccaatatcatctggactgta	cgactccaagctgtagataaaa
45	817	ctaaagtcgttcaacagcaaat	ctctcaatgatcgtgagaaaat	136	238	ttttcgttgctggagtagttag	gaaatcgaaaaccacaaacac
(13)	1155	ctaaagtcgttcaacagcaaat	tgctaacattcgtacacacaac	137	274	gtgtttgtggttttcgatttct	aacgtacagcgataaacacatt
46	536	tgttgtgtgtacgaatgttagc	acggtctaactcaaacactacg	138	1540	tattgcctcattgagttgtagc	tgacattttgagctttaggttc
47	238	ttcatagaaatgtcgaaagagc	ttctctatttcccaaactcaca	(26)	1998	tattgcctcattgagttgtagc	ttgcttcatcacctctttatgt
48	848	tgtgagtttgggaaatagagaa	acactttctaccaatgcaacat	(27)	1163	aaagctcaaaatgtcagagttg	tacgatctgacgaataggttgt
(14)	536	atgttgcattggtagaaagtgt	aagcagctatctcaagttcgta	139	708	taaagaggtgatgaagcaaaag	tacgatctgacgaataggttgt
49	440	ggtcagtttggtttcacaaata	aagcagctatctcaagttcgta	140	436	gcatcagaaacacacacataac	caagaccaagatgcttcatatc
50	912	gcatcttgtgaaaacctgtatc	acgctttcacttactttcaatg	(37)	1493	tcagatcgtacgaatggtaaat	ccgtaaatggttagcaaataga
51	961	cacaatgaaaggtaagttggtc	actgagcaaatacatgaggtgt	141	525	tcaagcagagctaatgctaaac	ctggatgacgctgtactaactt
52	1259	tagcagctaatgttgtgtcaaa	gaaagccagaacacagagataa	142	500	gagttaagctcagctcttgttg	cgttttcagtagtagcccataa

(Continued)

Primer pair	Product Size*	Left Primer	Right Primer	Primer pair	Product Size*	Left Primer	Right Primer
53	914	ttatctctgtgttctggctttc	ttcttcataagagcatggatgt	143	296	gttagattggtacccgagagat	gctgaatacgcaggtactgata
54	514	ctcttatgaagaattcggatca	gagctcatggctaaaaatacaa	144	990	gtacctgcgtattcagctattc	tggaagatgcactctatgagat
55	702	cattgatgaagctatcaaggtc	tatatcttgtgaccctgctcat	145	656	acaccaatatgtgccttcatac	aaccagacgtcattaccaatta
56	453	tttacggatcttgtatgggtta	gcttatgattttcactgcactt	146	495	tcagagacttaatcgcatcttc	atccgattcttgaaactcttct
(15)	1221	tttacggatcttgtatgggtta	cgtattgggtgttaatttatgc	147	487	tgaacagaacacagttgctaga	cttcataaacgatggcattatc
57	399	gcataaattaacacccaatacg	cgtttcttctgctgttcactat	148	492	gataatgccatcgtttatgaag	atttgctgttgaacgactttag
(16)	1072	gcataaattaacacccaatacg	gtttctcttgtgatcaaacacc	149	804	ctaaagtcgttcaacagcaaat	gctcataatcatcgagcatatc
58	1213	ggtgtttgatcacaagagaaac	cgctctgtttcttgttctatgt	150	153	atcgatatgctcgatgattatg	tgaattgtcggagctaacttac
59	592	tacaaactccccattaactgaa	aggttagtgtcggtgtaatgaa	151	666	gtaagttagctccgacaattca	actcattatcttttccccaaac
60	1036	ttcattacaccgacactaacct	aagttggagctttacctctcat	152	573	gtttggggaaaagataatgagt	agttggtcatgtttagagtcca
61	283	aagttggtagcggtttttatct	aaaagtgacgcctacttaccttt	153	648	tggactctaaacatgaccaact	tgcttatatcatcggcttttac
(17)	1144	aaggtaagtaggcgtcactttt	gcgatgttacgtatatttggac	(28)	1046	tggactctaaacatgaccaact	cgctagagttgatgtcgatact
62	1144	aaggtaagtaggcgtcactttt	gcgatgttacgtatatttggac	154	1081	attgagttttgctctcataacg	acagtgctaacgcagtaatagg
(18)	906	cgattaccagtccaaatatacg	aagattttgacacaaggactca	155	302	gagaagttaacgtgacgaaaga	tagagccagaagaatacaccag
63	203	acgttttgtcataatgtcacct	aagattttgacacaaggactca	156	359	ctggtgtattcttctggctcta	tggctacttcaagcaatttatc
64	854	cttcatgagcttctttctggta	gcaaatctttcgttctgaagta	157	561	tcgaagatgaattaggtgatgt	catcttcgtccataagtctcac
65	584	ggatctttggatatgatgatga	gatacattgccacttcattagg	158	253	gtgagacttatggacgaagatg	aatgtgtttagctcttgctcac
(38)	1949	aatgtgtggtgagtgagtttgt	gtcgaatgggatcttttctaat	159	527	gttagctgtgcacaaagaattt	tcaatactggctctttcgtatc
66	788	ctgggactctttacatcacaac	gacaaagaaagcaagatcacag	160	371	gatacgaaagagccagtattga	ttcacgataattgaggtgttct
67	371	ctgtgatcttgctttctttgtc	ctttctcaacgattcgagatta	161	304	gaacacctcaattatcgtgaag	aacagtggcaatttgtgatact
68	1469	gtctgaggttttagggaatgtt	ctgaactttcaacgctttctac	162	511	ttactgcacaaatgttgttgtc	ttcgagtgcaatttcttagttc
(19)	1723	taatctcgaatcgttgagaaag	gttaagagtccgcacaaagaa	163	980	tccttgttggagtgacaactat	gcactcctacccttttaaaactc
(35)	1970	gtctgaggttttagggaatgtt	ggaaacacatggcttattacac	164	407	gagttttaaagggtaggagtgc	cgacaaataaaagttgagtgct
(36)	895	tttcctgatgttgatttggtag	tgtccattaaggactttgaaga	165	833	cttaagtgcatggttcaaaatc	ataagcataaactttgcgtcac
(20)	1022	gaagtttgcttctttggtgtaa	tagcttcttgtgccagataaac	166	1020	gtgacgcaaagtttatgcttat	tggagagctttcttgtaaattg
69	647	aaagtccttaatggacatgaca	tagcttcttgtgccagataaac	167	383	ggttttattattggtcggactt	aagacctaatagggcatcaatc
70	682	gtttatctggcacaagaagcta	aaatgagcatgtagatcggata	168	697	caagcagcataaattacctcat	tatctttcgctagttcgatgac
71	996	gtagtgtggcttctgctgtaat	gtgataactcaatgaggcaatg	169	257	gtcatcgaactagcgaaagata	ccaacttgtgaaatttggatac
72	431	aatgtgtctgtcgctttaagtt	cgacaaaactgaacgaaatg	(29)	426	gtcatcgaactagcgaaagata	gacaaagaaaaccagatcacag
73	1253	gtcggcaattcaatattgttc	atgtcgtttatcatcctcgtaa	170	462	ctgtgatctggttttctttgtc	ggaaaaaccctcagatacattt
74	489	ttacgaggatgataaacgacat	caactgcatgttgaatgttaag	171	335	aaggctagagctaaagaacagg	tgcgtgtactacatcagaaaca
75	434	cgagaaatcgttatcaagaaga	gacaaagaaagcaagatcacag	(30)	735	gtttttcgatgatgtttaccag	ctctcaatgatcgtgagaaaat
(21)	628	ctgtgatcttgctttctttgtc	tcaccacttcaccgtatttact	(31)	750	tttagaaccatcaacccaatac	ctctgcatcttcgactgttaat
76	588	ccaacttaaccttgatcgtttt	tcaccacttcaccgtatttact	172	229	tgttgtgtgtacgaatgttagc	ctctgcatcttcgactgttaat
77	1322	agtaaatacggtgaagtggtga	actggagttgcattactttcag	173	699	attaacagtcgaagatgcagag	atttgctgttgaacgactttag
78	1169	ctgagtatgaaaaagccatgtt	aaagagcaatctgtcccttagt	174	976	ctaaagtcgttcaacagcaaat	acgttttgtcataatgtcacct
79	379	actaagggacagattgctcttt	gtttggtgttttcacatctagg	(32)	1582	ctaaagtcgttcaacagcaaat	gttcttgtgaagcctttctgta
80	504	cctagatgtgaaaacaccaaac	gtgttggcatctcatttttatc	175	617	gcgatgttacgtatatttggac	gtagacattgggggattgatac
81	1043	ctcgatttgatattggtagtgg	actcattatcttttccccaaac	A	27055	cagtgggttgatggtgttgc	cgtctacttcgtgtaatcgtgc

Note. Product Size* is the length of the amplicon with the correlative primers in N16961.