SAPIENZA
UNIVERSITÀ DI ROMA

Faculty of Mathematical, Physical and Natural Sciences

Comparative genomics of *V. cholerae* 7th pandemic strains: analysis of Integrative Conjugative Elements, Genomic Islands and prophages

Matteo Spagnoletti

Dep. of Biology and Biotechnology "Charles Darwin"
Sapienza University of Rome

Directed by: Prof. Mauro M. Colombo

Tutor: Prof. Renato Fani

PhD Coordinator: Prof. Rodolfo Negri

A dissertation submitted in candidature for the degree of
Doctor of Philosophy in
Cellular and Developmental Biology (XXIV cycle)
Rome, December 2011

A Giulia e Leonardo

Acknowledgements

This work would not have been possible without the support of many people.

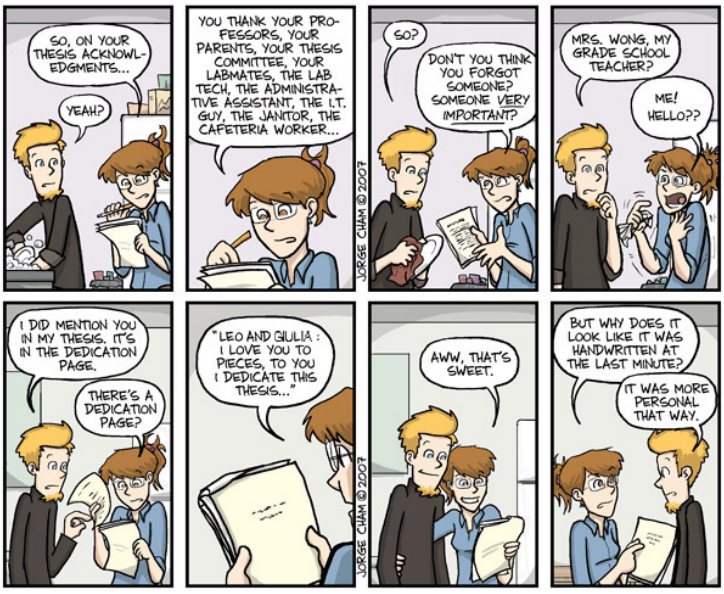
I am grateful to Prof. Mauro Colombo for his thoughtful guidance and help throughout my graduate studies and work towards the PhD. degree. He deeply stimulated my scientific interest with endless effort and devotion; much of the independence and confidence that I have acquired as a researcher over these years I owe to him.

My deepest gratitude also goes to Dr. Daniela Ceccarelli for her constant help and supervision. She has been a role model for me, and many of the ideas developed in my work arose during our discussions together. We spent a huge amount of time working in close collaboration and she has always shown willingness and endless tolerance towards my bad temper! I will never be able to thank her enough.

Several cooperators have contributed to the work presented in my dissertation, I'm sorry but there isn't enough space to mention everyone! I thank them all for their support and help. I am especially thankful to Elisa Taviani and Marco Fondi for the time they have dedicated to me and for the work done together. Special thanks also go to my PhD. tutor, Prof. Renato Fani.

I am grateful to many student colleagues for providing an inspiring and fun environment during these years: I want to thank all the lab members, Jusy, Nancy, Romy and Giulia, and all the Bernardini, Grossi and Colonna's laboratory members for their help and friendship.

Last but not least, I would like to thank my family and friends who have never lost faith in supporting me during my never-ending studies…



List of Publications

Most of the data presented in this thesis has been published on the following peer reviewed papers:

- Wozniak R.A.F., Fouts D.E., Spagnoletti M., Colombo M. M., Ceccarelli D., Garriss G.,. Déry C, Burrus V. and Waldor M. K. **Comparative ICE Genomics: Insights into the Evolution of the SXT/R391 Family of ICEs.** 2009. PLoS Genet 5(12): e1000786. doi:10.1371/journal.pgen.1000786.

- Ceccarelli D.*, Spagnoletti M. *, Bacciu D., Danin-Poleg Y., Kashi Y., Cappuccinelli P. and Colombo M. M. **ICE*Vch*Ind5 is prevalent in epidemic *Vibrio cholerae* O1 El Tor strains isolated in India**, Int J Med Microbiol. 2010. doi:10.1016/j.ijmm.2010.11.005. *equal contribution

- Ceccarelli D., Spagnoletti M., Cappuccinelli P., Burrus V. and Colombo M. M. **Origin of *Vibrio cholerae* in Haiti**, Lancet Infect. Dis. 2011. doi:10.1016/S1473-3099(11)70078-0

- Ceccarelli D., Spagnoletti M., Bacciu D., Cappuccinelli P., Colombo M. M. **New *V. cholerae* atypical El Tor variant emerged during the 2006 epidemic outbreak in Angola.** BMC Microbiol. 2011 Jun 13;11(1):130.

- Spagnoletti M., D. Ceccarelli and M. M. Colombo. **Rapid detection by multiplex PCR of Genomic Islands, prophages and Integrative Conjugative Elements in *V. cholerae* 7th pandemic variants.** J Microb Methods. 2011. Doi: doi:10.1016/j.mimet.2011.10.017.

Table of Contents

CHAPTER I

Introduction

Vibrio cholerae

Vibrio cholerae is a Gram-negative, rod-shaped, waterborne bacterium that causes the severe watery diarrheal disease known as cholera. It is a member of the genus *Vibrio,* of the γ-proteobacterium group which includes most of the enteric bacteria such as *E. coli* and *S. enterica*. The natural habitat of *V. cholerae* is the aquatic ecosystem where it can exist in free living form associated with phytoplankton and zooplankton, in biofilms, and also in a viable but non-culturable form (VBNC) [1-3].
Environmental *V. cholerae* and other *Vibrio* species seem to be pre-adapted to cause occasional infections in mammals when the opportunity arises but, in general, they lack the capacity of transmission between such hosts. However one or more strains of V. *cholerae* adapted to the human environment [4].

Pathogenicity

Pathogenicity of *V. cholerae* is mainly associated with the toxigenic effect of cholera toxin (CT), which disrupts ion transport of human intestinal epithelial cells [5]. When *V. cholerae* enters the small intestine, it colonizes the epithelium by means of flagella, toxin coregulated pili (TCP), and other virulence factors that aid in the attachment and colonization of the gut. After successful colonization, *V. cholerae* secretes CT, an AB_5 toxin responsible for the massive watery diarrhea characteristic of cholera. CT binds to a specific receptor (GM1) on host enterocytes and is internalized leading to elevated intracellular cAMP levels and resulting in a major loss of water and electrolytes [5]. The genes related to CT are part of the genome of a lysogenic phage named CTXΦ [6]. The genetics and ecology of this phage have been extensively studied [7-9], and the genes coding for CT and other phage-associated virulence factors were broadly used as epidemiological markers. CTXΦ will be introduced in the coming sessions and then further discussed in Chapter 4.

Subtype classification

Current serotyping of *V. cholerae* is based on the O-antigen of the lipopolysaccharide (LPS), and more than 200 serogroups have been described to date [10,11]. Most of the clinical strains belonging to serogroups O1 and O139 are toxigenic, and are responsible for all the major cholera epidemics and pandemics [12]. Other *V. cholerae* serogroups, collectively called non-O1, non-O139, are rarely toxigenic and have seldom caused epidemics, but can represent a genetic reservoir for the emergence of novel pathogenic clones by lateral gene transfer (LGT) [13-15]. Based on phenotypic differences, *V. cholerae* O1 strains are further classified into two biotypes: 'Classical' or 'El Tor'. Phenotypic tests based on haemolysis of sheep erythrocytes, agglutination of chicken erythrocytes, Voges–Proskauer reaction, and sensitivity to polymyxin B and specific phages were used to differentiate the two biotypes [12]. Additionally, genotypic analyses of genes *tcpA* (toxin co-regulated pilin A), *ctxB* (cholera toxin B), and *rstR* (repeat sequence transcriptional regulator), together with epitope analysis of one of the two subunits of the cholera toxin (CTB epityping) have also been used [12].

Due to the dissimilarity found among sub-types of *V. cholerae*, a clear classification is still challenging. It is now widely accepted that the O139 serogroup originated from an O1 El Tor strain with the acquisition of the O139 antigen by lateral gene transfer [16], rather than from non-O1, non-O139 strain(s) that acquired the virulence factors necessary for pandemic cholera [17]. Similarly, recent studies indicated that the gene clusters encoding the O1 antigen can be readily transferable between environmental and clinical clones [14,18] and that major pandemic clones in Asia and Africa showed contradicting biotype-related phenotypes [12].
Due to the extensive lateral gene transfer that takes place in *V. cholerae*, Cho and colleagues [17] suggested that classically important diagnostic markers such as serotype and biotype are not reliable anymore and new methods are required to identify emerging epidemic strains. This topic will be discussed in detail in Chapter 6.

Epidemiology

The first six cholera pandemics

It is currently accepted that seven cholera pandemics occurred since 1817. The sub-species associated with the first five pandemics remain unknown, as they occurred before the causative agent of cholera was first cultured [12]. The Classical biotype strains that caused the sixth pandemic (1899–1923) are therefore the first known pandemic biotypes. The 6th pandemic continued until 1923, but was largely confined to Asia thus not being pandemic anymore. All the strains isolated in this forty years period are defined as pre-seventh pandemic strains.
Although O1 El Tor strains existed and were associated with sporadic cholera during the sixth pandemic and then increasingly reported during the pre-seventh pandemic period [12], it was not until 1961 that this biotype caused the seventh pandemic and hence was recognized as an epidemic threat. El Tor strains have then gradually displaced the Classical strains as the cause of cholera [19]. In Bangladesh, the Classical biotype apparently disappeared in 1973, but re-emerged in 1982 and co-circulated with the El Tor biotype for at least a decade (the last isolation was reported in 1992) [20]. Curiously, the transient reappearance of O1 Classical strains was observed only in Bangladesh. Classical strains are now believed to be extinct, although a low occurrence in the environment cannot be completely excluded [12].

The seventh pandemic

The seventh pandemic started in 1961 in Indonesia and spread to cover most of Asia by 1966. After a 5 year break in 1971 an upsurge took it to Africa and Europe causing harsh epidemics in several African countries. Cholera remained relatively low in the 1980s and was mostly confined to Asia and Africa. In the 1990s there was a dramatic increase of cholera in South America and Asia as well as a sharp increase of cholera cases in Africa, where several war scenarios with large numbers of refugees created the favorable conditions to its spread. In 1991 cholera spread to South America for the first time since 1895: it started in Peru and spread rapidly in the neighboring countries. In 1992, a new form of the seventh pandemic strain characterized by the O139 O antigen appeared in India and rapidly spread all over Asia, prompting concerns that it could

represent the beginning of the eighth pandemic [21]. However the O139 serogroup has been confined in Asia to date, as new variants of *V. cholerae* O1 El Tor arose in the same years and rapidly spread outside Asia, confirming *V. cholerae* O1 as the only known pandemic serogroup. In conclusion, the 1990s are marked by a third wave of the O1 seventh pandemic strains as well as the introduction of a new El Tor-related epidemic serogroup, the *V. cholerae* O139 [4].

The seventh pandemic and *V. cholerae* Atypical El Tor

Since the late 1980s a number of new variants of *V. cholerae* O1 appeared, showing traits of both the Classical and the El Tor biotypes [30–33]. Safa *et al.* coined the term 'Atypical El Tor' for these clones [12], which are currently replacing previously circulating El Tor strains, retrospectively defined as 'prototypical' El Tor. Several Atypical El Tor strains have been recently identified and reported in the literature, which include Matlab types I, II, and III [22], altered El Tor [23], Mozambique El Tor [24] and hybrid El Tor strains [12,25]. These strains might have evolved from El Tor variants that acquired certain attributes from the Classical genome.

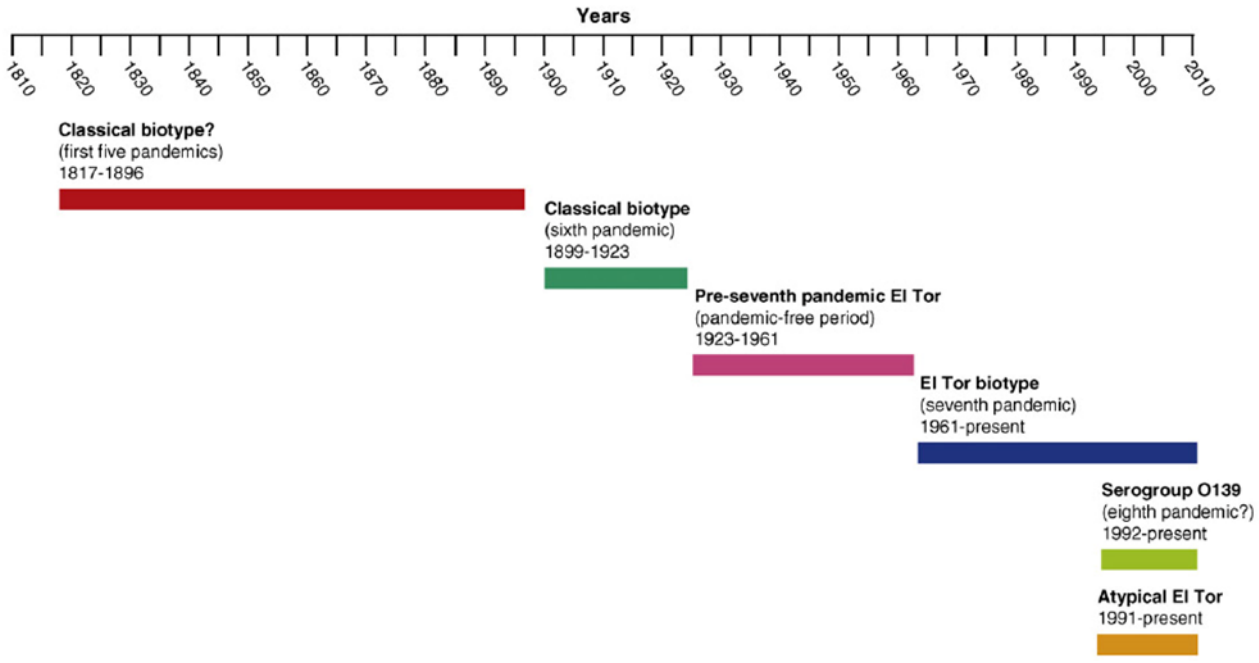


Figure 1. Key evolutionary events in cholera epidemiology since 1817 (adapted from Safa *et al*. [12]). The first five pandemics were caused by the Classical biotype, whereas the Classical strains appear to have caused the sixth pandemic. Between the sixth and seventh pandemics, there was an intervening period of 38 years during which cholera only occurred as local outbreaks caused by El Tor strains.

Notably, genes coding for cholera toxin derived from the Classical biotype. A particular kind of Atypical El Tor strains (the altered El Tor variant) has completely replaced the prototype seventh pandemic El Tor clone in Bangladesh [12,26]. These strains were recently isolated in other countries in Asia, Africa and Central America and they will be discussed in detail in Chapters 3, 4 and 5, respectively. An overview of *V. cholerae* epidemiology is summarized in Figure 3. It is

becoming increasingly clear that El Tor biotype and its derivatives (including the O139 serogroup) appear to be better adapted for global dissemination compared with their Classical predecessors. This idea is supported by the fact that the current seventh pandemic is the longest known cholera pandemic, and by the observation that in the last decade Atypical El Tor strains were the cause of all major cholera outbreaks worldwide [12].

V. cholerae in the Genomic Era

V. cholerae genomics

The first completely sequenced *V. cholerae* strain was the O1 El Tor strain N16961, isolated in Bangladesh in 1975. The complete sequence was published in 2000 by Heidelberg *et al.* [27].
The analysis of *V. cholerae* N16961 genome revealed that it is composed by two circular chromosomes of 2.97 Mb (chromosome 1) and 1.07 Mb (chromosome 2), respectively [27]. 12.8% of the genes show high similarity with genes already present on both chromosomes suggesting recent gene duplications [27]. Chromosome 1 bears most of the genes critical for the cell functions such as metabolism and pathogenesis. Genes encoding the O1 specific LPS are also on chromosome 1. On chromosome 2, the genes are mostly hypothetical proteins with unidentified functions, although there are also some components of the essential regulatory and metabolic pathways. Because of the presence of many genes of plasmid origin, chromosome 2 was suggested to be derived from a megaplasmid acquired by an ancestral *Vibrio* species [27].
Over the last decade, the advancement of sequencing technology has led to an exponentially increasing number of sequenced strains, allowing a great advancement understanding the evolution of this bacterial pathogen. More than 200 strains of *V. cholerae* have been sequenced to date, and this number is growing. This huge availability of data allowed the development of a number of comparative genomic studies to investigate the evolutionary relationships that exist between the different sub-types of this species [17,18,28]. Moreover, these studies have been directed in understanding the evolution of pathogenic strains, giving new insights on the Atypical El Tor variants responsible for the seventh pandemic.

Comparative Genomics of seventh pandemic strains

Comparative genomics is the study of the relationship of genome structure and function across different biological species or strains. Whole genome-based comparisons of representative *V. cholerae* strains belonging to various serogroups and genetically distinctive lineages indicated that the isolates responsible for the seventh pandemic (including the O139 serotype) are clonal and belong to a single phyletic line, the seventh pandemic clade (7P, see figure 2) [17,18].
All of the seventh pandemics *V. cholerae* strains sequenced so far share an almost identical genome backbone, diversified mainly by a large set of mobile (or potentially mobilizable) genetic elements. This large set includes Genomic Islands (GIs), prophages, and Integrative Conjugative Elements mainly acquired and lost by lateral gene transfer [17,18]. Detailed comparative genomics of currently predominant Atypical El Tor variants of the seventh pandemic suggested that insertion

of a hybrid type CTXΦ and an Integrative Conjugative Element may have led to their establishment as successful clinical clones over other seventh pandemic members [17].

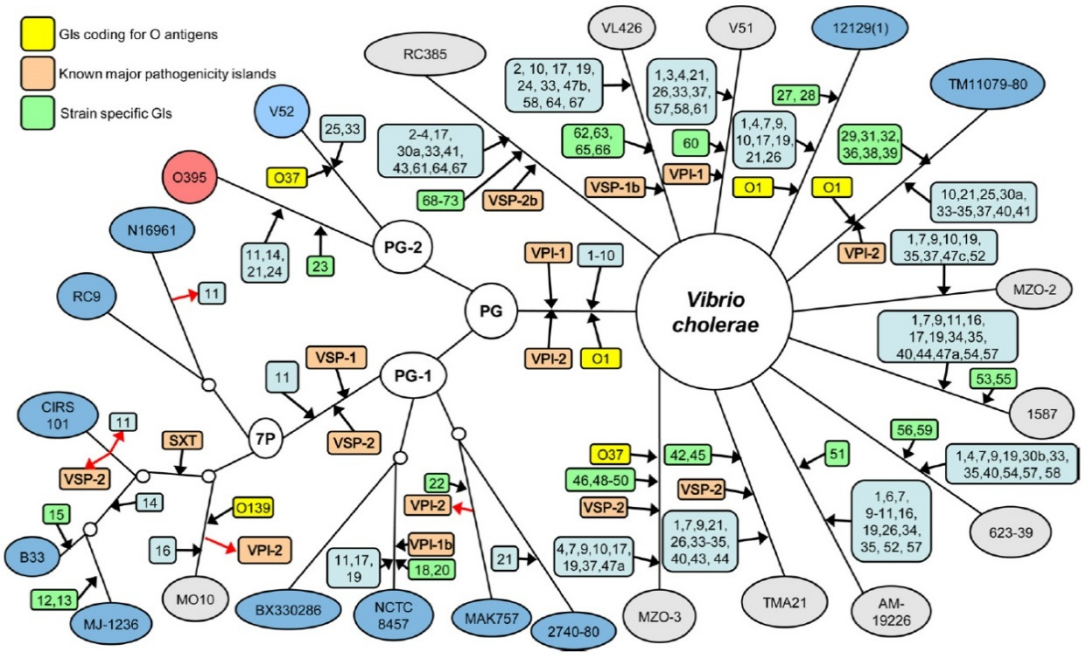


Figure 2. Proposed hypothetical evolutionary pathway of the *V. cholerae* species (adapted from Chun *et al.* [18]). Probable insertions and deletions of genomic islands found in 23 *V. cholerae* strains are indicated by black and red arrows, respectively, along the phylogenetic tree based on genome sequence data. Hypothetical ancestral strains are indicated by open circles.

V. cholerae mobilome

The mobilome comprises all the mobile genetic elements present in a genome, and most of the prokaryotic species are characterized by a large mobilome [17,29]. Chun *et al.* [15] developed a comparative genomic approach to define the first mobilome database of *V. cholerae*, comprising 73 newly identified genomic islands in addition to known mobile elements such as CTXΦ and Vibrio Phatogenicity Islands [4]. A total of 13 genomic regions (eight in the large and five in the small chromosome) were found to have a cassette-like property, whereby different GIs occupy the same or a similar region [18]. Information residing in the mobilome is perhaps the most important key to understand the evolution of *V. cholerae* [17,29].

As will be discussed more in detail in Chapter 6, the presence, absence or genetic rearrangements in a distinct subset of *V. cholerae* mobilome allows discrimination between different clones of the seventh pandemic clade. Such elements, together with the main virulence-associated mobile genetic elements of *V. cholerae* are introduced in the following sections.

Virulence-associated prophages and Genomic Islands

The major virulence factors in *V. cholerae* are contained within regions of the genome of foreign origin [7]. Genes *ctxAB* encoding for CT as well as accessory toxins *ace* and *zot* are part of the genome of the lysogenic phage CTXΦ [6], that can transduce non-toxigenic strains of *V. cholerae*

to harbor cholera toxin [30]. CTXΦ can be incorporated into both chromosomes of *V. cholerae* at specific positions [18]. CTXΦ found in Classical or El Tor biotypes differs in the sequence of the repressor gene, *rstR*, as well as in the sequence of CT subunit B, and was initially classified as $CTX\Phi^{Class}$ and $CTX\Phi^{ElTor}$, according to the biotype of the original hosts in which it was described [12]. CTXΦ is linked to another flanking genetic element, RS1, which is involved in the replication and site specific integration of the phage [7]. Several arrangements of these two elements are possible and have been described in O1 and O139 strains, including the Atypical El Tor variants [25,31,32] (see figure 3).

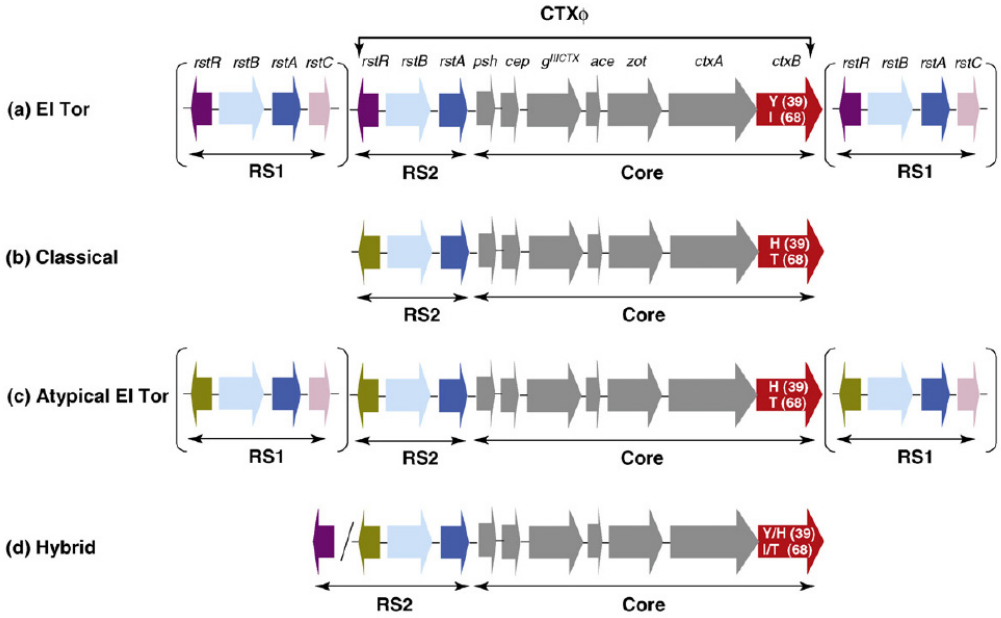


Figure 3. Structure of CTXΦ and the RS1 element in different *Vibrio cholerae* O1 variants (adapted from Safa *et al.* [12]). In the prototype El Tor (a) and Atypical El Tor (c) strains, parentheses delineate the RS1 elements that flank the integrated CTXΦ genome. (b) CTXΦ in Classical strains, which lack the RS1 element. (d) A hypothetical hybrid CTXΦ. Polymorphic positions within *ctxB* are indicated for H (histidine), T (threonine), Y (tyrosine) or I (isoleucine). The different $rstR^{El}$ and $rstR^{Cla}$ alleles are highlighted in different colors.

The receptor for CTXΦ is a type IV pilus, TCP, which is also involved in the colonization of the intestine of the human host. Genes encoding the synthesis of TCP are located on the *Vibrio* Pathogenicity Island I (VPI-I), a molecular tag of all seventh pandemic isolates [33]. The two genetic elements are linked in a complex pathway that controls the toxin expression under the regulative action of the ToxR transcriptional activator [34]. The close linkage of CTXΦ and VPI-I provides a fascinating example of co-evolution involving different genetic entities and their relationship with the host genome and biology, suggesting how different lateral events provide synergism in the construction of new genetic variants.

A second genomic island in the genome of pandemic *V. cholerae*, the Vibrio Pathogenicity Island 2 (VPI-2), has been recently described [33,35]. VPI-2 is a 57.3 kb GI including several gene clusters, such as a type 1 restriction modification system, a sialic acid metabolism gene cluster, and genes required for utilization of aminosugars, including the *nanH* neuramidase [35]. The neuraminidase protein is involved in *V. cholerae* pathogenicity in two ways: (i) it generates the GM1 ganglioside receptor for cholera toxin from higher order sialogangliosides, with release of sialic acid (an amino

sugar present in all mucous membranes); (ii) it appears to be part of the mucinase complex that is responsible for digestion of the intestinal mucus, enhancing *V. cholerae* colonization of the gut [33]. At its discovery, VPI-2 showed all of the characteristics of a horizontally transferred element [34,35]**.** This genomic island encodes for a phage-like integrase, which is responsible for insertion/excision of the island to and from the host genome [33]. It has been shown that VPI-2 forms a circular intermediate (CI) upon excision, which likely represents the mobile form of this genetic element, although the mobilization of VPI-2 has not yet been demonstrated. The absence of a gene cluster encoding a conjugative apparatus (*tra* gene operon), or type IV secretion system, suggests that the island is mobilized by other genetic elements or that it is an ancient mobile element which lost part of its genetic structure. To date, all toxigenic *V. cholerae* O1 serogroup isolates contain VPI-2, whereas non-toxigenic isolates miss it [33].

Two other gene clusters associated with the seventh pandemic strains were recently identified by comparative genomics and named *Vibrio* Seventh Pandemic I and II (VSP-I and VSP-II) [36]. These Genomic Islands were absent in Classical and pre-pandemic *V. cholerae* El Tor strains. VSP-I is a 16-kb region inserted in chromosome I and containing 11 genes [33]. VSP-II is 26.9-kb large, its integration site is a tRNA-methionine locus, and among others it encodes a type IV pilin, two methyl-accepting chemotaxis proteins, and a DNA repair protein [37]. Murphy *et al*. [33] found that both VSP-I and VSP-II are able to excise from the chromosome, forming an extra-chromosomal circular intermediate through site-specific recombination mediated by the P4-like integrase encoded by the island.

Other Genomic Islands and prophages associated with seventh pandemic strains

A number of genomic islands were identified in Atypical El Tor strains belonging to the seventh pandemic clade [38]. A novel rearrangement of VSP-II, showing a 14.4 kb deletion, was recently identified in an Atypical El Tor variant circulating in Bangladesh [39]. This variant, which has its representative in the completely sequenced strain CIRS-101, is now considered the prevalent seventh pandemic clone [17,38]. GI-12, GI-14 and GI-15 were recognized as novel GIs able to discriminate, with their presence or absence, closely related Atypical El Tor strains such as MJ1236 from Bangladesh and B33 from Mozambique [38]. These two strains are also characterized by the presence of another prophage (Kappa) that has been recently described by Kapfhammer et al [40].

Chromosomal attachment sites for CTXΦ are known to harbor other genetic elements, including toxin-linked cryptic (TLC) and VSK (pre-CTX) prophages [18,41]. TLC is typically found in the genomes of *V. cholerae* epidemic strains, and was recently recognized as an important player in *V. cholerae* toxigenic conversion [42].

As already mentioned, beside genomic islands and phages, also Integrative Conjugative Elements take part in the evolutionary process of *V. cholerae*; these mobile genetic elements are introduced in the next session.

Integrative Conjugative Elements

Integrative Conjugative Elements (ICEs) are a large family of self-transmissible mobile genetic elements. They are widespread in all bacteria and comprise self-transmissible integrative and conjugative mobile elements irrespective of their mechanisms of integration or conjugation, including elements previously classified as mobile genomic islands or conjugative transposons [43]. Initially these elements were observed in low G+C Gram positive Bacteria and Bacterioides, since the first elements described were *Tn916* from *Enterococcus faecalis*, and *CTnDOT* from *Bacteroides thetaiotaomicron,* originally classified as conjugative transposons [44,45]. However, with the discovery of numerous conjugative elements that integrate via different mechanisms of site-specific recombination, the nomenclature for such elements has been questioned and the term ICE was adopted by Burrus *et al*. in 2002 [46].

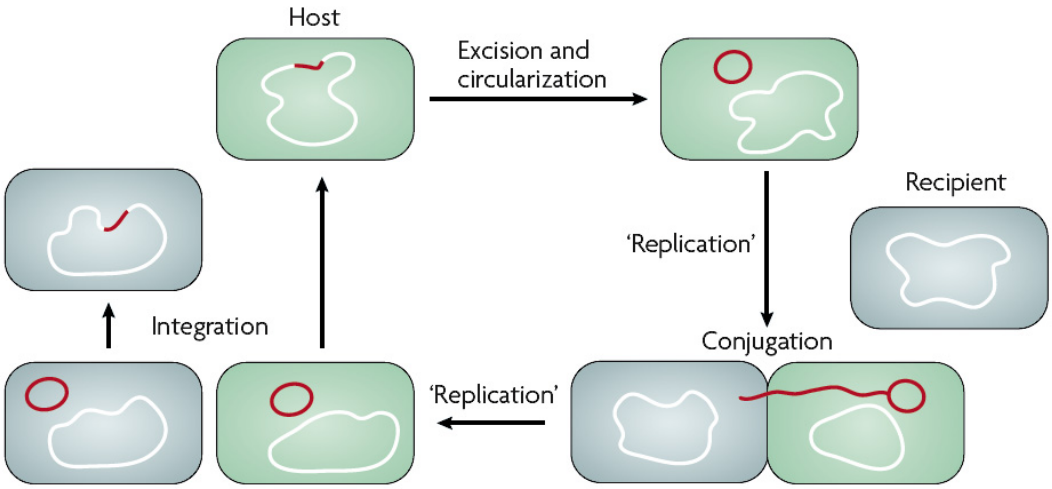


Figure 4 Schematic of a typical integrative and conjugative element life cycle (adapted from Wozniack *et al*.[43]). An integrative and conjugative element (ICE) is integrated into one site in the host chromosome and is bounded by specific sequences on the right (*attR*) and left (*attL*). Excision yields a covalently closed circular molecule as a result of recombination between *attL* and *attR* to yield *attP* (in the ICE) and *attB* (in the host chromosome). An ICE-free cell can serve as a potential recipient. During conjugation, the donor and recipient are brought in close contact, and a single DNA strand is transferred to the new host through the action of rolling circle replication. Following transfer, DNA polymerase in the recipient synthesizes the complementary strand to regenerate the double-stranded, circular form. A recombination event between *attP* and *attB* results in integration into the host chromosome.

ICEs integrate into and replicate as part of the host chromosome. Certain conditions induce the excision of ICEs from the chromosome after which they circularize and are replicated and then transferred to new hosts by the conjugation machinery encoded in the elements. The ICE in the recipient cell integrates into the chromosome, and the copy of the ICE that remains in the donor cell reintegrates into the donor cell chromosome. Thus, these elements combine features of other classes of mobile genetic elements, such as phages (which often integrate into and excise from the host chromosome but are not transmitted by conjugation), transposons (which integrate into and excise from the chromosome but are not transferred horizontally) and plasmids (which sometimes transfer from cell to cell by conjugation but replicate autonomously). ICEs, unlike plasmids, cannot be maintained in an extrachromosomal form, as they seem to be incapable of autonomous replication, although this is still under investigation [43].

ICEs usually show a modular organization, with genes clustered according to the process to which they contribute. They utilize a variety of genes to mediate the core ICE functions of chromosome integration, excision and conjugation. These genes are organized in three major functional modules which constitute the ICE backbone involved in maintenance, transfer and regulation [47]. Integration and excision of ICEs is mediated by site-specific recombinases (Integrase Int), mainly tyrosine recombinases, which usually show a broad host range and have a minimal requirement for specific host factors [48]. These enzymes, similarly to the role of Int in the lambda phage, promote the recombination between a specific sequence in the circular form of the ICE, *attP,* and a sequence on the chromosome, *attB*. Integration of the circular intermediate generates two sequences at the boundaries of the integrated element, called *attL* (left) and *attR* (right). Together with Int, many ICEs encode for a protein called RDF or Xis, which co-operates with Int in the excision of the element [49]. For a number of characterized ICEs no genes coding for RDFs have been identified to date, which may reflect the fact that they are difficult to detect using bioinformatics approaches. Thus, additional RDFs are likely to be discovered even in characterized elements [43]. Excision and transfer of ICEs can be activated by different factors, such as sub-inhibitory concentrations of antibiotics or by the host SOS response system activation, which can consequently induce the ICE regulation module [49-51].

Similarly to conjugative plasmids, different clusters of *tra* genes encode all factors necessary for the ICE transfer and the mating machinery. ICE transfer process is still under investigation as different ICEs show dissimilarities in some aspects of their transfer mechanism when compared to conjugative plasmids. Usually, conjugative DNA transfer takes place in two key steps: (a) biochemical processing of the DNA molecule for transfer and (b) assembly of a mating apparatus bridging the donor and recipient cells to allow DNA transfer [52]. Typically conjugative DNA transfer is initiated at a specific *cis*-acting site called the origin of transfer (*oriT*), required for efficient translocation of the DNA to the recipient cell [52,53]. A DNA relaxase encoded by the ICE recognizes the *oriT* locus and cleaves one strand to initiate the transfer through the mating bridge [54,55]. Within the recipient cell, host enzymes convert the transferred single-stranded DNA into double-stranded DNA that can be recircularized and/or recombined into the recipient chromosome [52].

Beside the backbone genes, ICEs often enclose in their nucleotide sequence other genetic elements, such as transposons and insertion sequences as well as genes encoding recombinases [56]. These genetic elements and specific genes mediate the acquisition of additional modules encoding functions such as resistance to antibiotics and several other traits, which offer a selective advantage to the bacterial host under certain environmental conditions [43,47]. Sequencing of several ICEs has shown that a large number of genes with different functions are acquired by these elements. Often they contain genes that allow the host to grow in hostile environments, such as in the presence of heavy metals, or they acquire additional functions involved in DNA repair, UV protection, recombination systems, and symbiotic growth with plants [47]. Furthermore, some features of the pathogenicity islands of *Legionella pneumophila*, *Agrobacterium tumefaciens,* and *Bartonella tribocorum*, such as the ability to encode type IV secretion systems (T4SS), suggest that these pathogenicity islands might be considered ICEs or defective ICEs [47]. With the availability of an increasing number of fully sequenced bacterial genomes, continued efforts to annotate and

understand the diversity and distribution of ICEs are necessary. As experimental characterization of ICEs has been carried out on only a few elements, it is unclear if these characterized ICEs are representative of all ICEs [43].
So far, several families of ICEs have been described (REF). The data reported in this thesis are focused on the SXT/R391 family that will is introduced in the following section and discussed in detail in the next chapters.

SXT/R391 family of ICEs

In 1996 Waldor *et al.* discovered in the chromosome of *V. cholerae* O139 MO10 from India a new Integrative Conjugative Element, that they named SXT [57]. The name is an acronym for sulfamethoxazole and trimethoprim resistances, the major antibiotic resistances provided by this element to the host strain. Since then SXT has been first considered as a conjugative transposon [57], then as a CONSTIN [58] and eventually as an ICE [46].
The molecular characterization of SXT revealed that this ICE bears, among others, a multiple-antibiotic resistance gene cluster conferring resistance to sulfamethoxazole (Su), trimethoprim (Tm), chloramphenicol (Cm), and streptomycin (Sm). Its discovery deeply changed the understanding of resistance circulation in *V. cholerae* epidemic strains and promoted several other studies to better investigate the genetics of this ICE [58-60]. SXT shares several common features with R391, an element originally isolated in South Africa in 1967, conferring resistance to kanamycin and mercury [61]. R391 was discovered in clinical strains of *Providencia rettgeri*, and was initially considered as a conjugative plasmid belonging to the incompatibility group IncJ [62]. Due to the impossibility of isolating an extrachromosomal replicative form of R391, the integration of this mobile element into the chromosome was assumed and confirmed [63]. Comparison of the complete DNA sequences of SXT (99.5 kb) and R391 (89 kb), revealed that both ICEs share a highly conserved set of genes that code for their regulation, excision-integration, and conjugative transfer functions [64]. Beside the conserved scaffold, these elements also harbor insertions, such as the antibiotic resistance genes, that confer element-specific properties. The enzyme involved in integration/excision of both SXT and R391 is an integrase (int_{SXT}), member of the tyrosine recombinase family and able to catalyze the site-specific insertion of the element into the host chromosome at the 5' end of *prfC* gene, restoring the gene integrity and preserving its functionality.
In the past 20 years several other ICEs were identified and recognized similar to SXT or R391 and thus grouped into the same family. All ICEs characterized by highly related int_{sxt} gene and by the ability to integrate into *prfC* are now grouped in the SXT/R391 family [65]. A schematic of the overall genetic organization of SXT–R391 family of ICEs is depicted in figure 5. Although the names of SXT and R391 are too rooted in the scientific terminology to be changed, it has been proposed to rename more recently described ICEs. The current nomenclature uses the prefix ICE followed by the abbreviation of the species and country of origin, then a number, in order to distinguish different elements isolated from the same species and country. For example, ICE*Vch*Ind5 is the fifth characterized ICE identified in a strain of *V. cholerae* isolated in India.
As it will be detailed in the next chapters, ICEs of SXT/R391 family are mostly widespread in *V. cholerae* clinical strains since the late 1980s and are virtually present in all the Atypical El Tor

strains. This observation, with the advance of sequencing techniques that have recently allowed the characterization of several identical ICEs, has created some complications in their nomenclature. Current nomenclature gave different names to elements isolated in different times and places irrespective from their sequence homology, which can be misleading. To solve this problem and maintain the current terminology, we introduced the term “siblings” to define two ICEs with the same nucleotide sequence but isolated in different regions and time (see Chapter 3). The SXT/R391 family of ICEs is the main topic of Chapter 2 and it will be further discussed in all the chapters of this thesis.

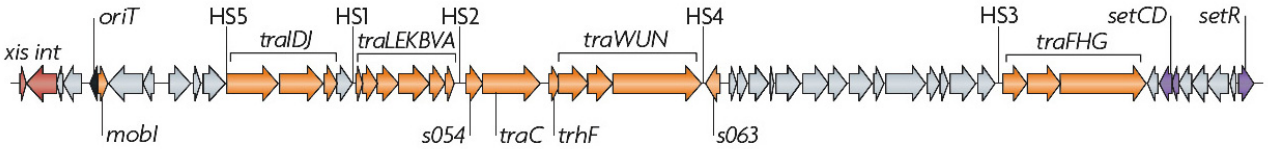


Figure 5. Schematic of the genetic organization of SXT–R391 family of integrative and conjugative elements (Adapted from Wozniak *et al.*[43]). All SXT–R391 family of ICEs have a set of nearly identical core genes (see chapter 2). These ICEs contain different DNA insertions at five hotspots (labelled HS1–HS5). Orange genes are conjugation related, purple genes are involved in regulation, red genes are involved in integration and excision, and grey genes have accessory or unknown functions. *int*, integrase; *oriT*, origin of transfer; *tra*, conjugal transfer; *xis*, excisionase.

Aim of the thesis

The aim of this thesis was to study the genomic evolution of *V. cholerae* focusing on the epidemic clones responsible for the ongoing seventh cholera pandemic. Particular attention has been dedicated to the study of a class of mobile genetic elements related to *V. cholerae*, the SXT/R391 family of ICEs, with a comparative genomics approach. These topics were analyzed from different points of view such as ICE genetics and evolution, the development of a rapid method to identify *V. cholerae* pandemic variants by molecular tracking of mobile genetic elements, and *V. cholerae* epidemiology in India, Angola and Haiti.

My study contributed to the understanding of:

- the genetics of SXT/R391 ICEs;
- the routes of dissemination of *V. cholerae* Atypical El Tor strains (carrying SXT/391 ICEs) in different epidemic events during the last 20 years;
- the development of a rapid molecular method to identify *V. cholerae* pandemic variants.

Chapter 2 illustrates a comparative genomic study on several ICEs of the SXT/R391 family. Comparative analyses of the DNA sequences of 13 ICEs revealed that they have an identical genetic structure consisting of syntenous, highly conserved core genes that are interrupted by clusters of diverse variable genes. Unexpectedly, many genes in the core backbone proved non-essential for ICE transfer. Comparisons of the variable gene content in the ICEs revealed that these elements are mosaics whose genomes have been shaped by inter–ICE recombination. Finally, this work suggests that ICEs contribute to a larger gene pool that connects all types of mobile elements.

Chapter 3 studies the occurrence and circulation of SXT/R391 ICEs in *V. cholerae* strains responsible of a cholera epidemic in India. We demonstrated that ICE*Vch*Ind5, described in Chapter 2, is prevailing in *V. cholerae* O1 clinical strains isolated in Wardha province (Maharashtra, India) from 1994 to 2005. Genetic characterization by ribotyping and multiple-locus SSR analysis proved the same clonal origin for *V. cholerae* O1 isolates in Wardha province over an 11-year period and was used to assess the correlation between strain and ICE content among ours and different Indian reference strains.

Chapter 4 shows that the epidemic *V. cholerae* O1 El Tor strain responsible for the 2006 outbreak in Angola is clonally and genetically different from El Tor strains circulating in the 1990s in the same area. Strains from 2006 carry ICE*Vch*Ang3 of the SXT/R391 family. This ICE is associated with a narrower multidrug resistance profile compared to the one conferred by plasmid p3iANG to strains of the 1990s. The CTX prophage carried by 2006 El Tor strains is characterized by $rstR^{ET}$ and $ctxB^{Cla}$ alleles organized in a RS1-RS2-Core array on chromosome I. Interestingly, the newly emerging Atypical El Tor strain belongs to a clade previously known to comprise only clinical isolates from the Indian Subcontinent that also contain the same ICE of the SXT/R391 family. Our

findings remark the appearance of a novel *V. cholerae* epidemic variant in Africa with a new CTXΦ arrangement previously described only in the Indian Subcontinent.

Chapter 5 focuses briefly on the recent outbreak of cholera in Haiti. We used a comparative genomics approach to compare the genomes of 3 sequenced Haitian strains with several *V. choleare* genomes available on public databases. Whole-genome alignment and comparative genomic analysis of the Haitian strains, with the representative *V. cholerae* O1 variants from Central America and the Indian Subcontinent, confirmed that the Haitian strain is strictly phylogenetically related to a *V. cholerae* altered El Tor stain from India.

Chapter 6 illustrates the development of a multiplex PCR assay for the rapid detection of the major 7th pandemic *V. cholerae* O1 and O139 variants. Three specific genomic islands (GI-12, GI-14 and GI-15), two phages (Kappa and TLC), *Vibrio* Seventh Pandemic Island 2 (VSP-II), and the ICEs of the SXT/R391 family were selected as targets of the multiplex PCR based on a comparative genomic approach. The optimization and specificity of the multiplex PCR was assessed on 5 *V. cholerae* 7th pandemic reference strains, and other 34 *V. cholera*e strains from various epidemic events were analysed to validate the reliability of our method.

Chapter 7 is about the genomic analysis of ICE*Vch*Ban8, an unusual ICE found in the genome of *V. cholerae* O37 strain MZO-3. ICE*Vch*Ban8 shares most of its genetic structure with SXT/R391 family of ICEs. However ICE*Vch*Ban8 codes for a different integration/excision module and it is located at a different insertion site, the same of Vibrio Pathogenicity Island II (VPI-2), a well described GI present in epidemic strains *V. cholerae*. Furthermore, part of its genetic cargo shows homology with genes coded by other pathogenicity islands.

In **Chapter 8** I summarized the obtained results and gave suggestions for scientific implications and future research.

References

1. Schoolnik GK, Yildiz FH (2000) The complete genome sequence of *Vibrio cholerae*: a tale of two chromosomes and of two lifestyles. Genome Biol 1: 1016.1011-1016.1013.
2. Chaiyanan S, Huq A, Maugel T, Colwell RR (2001) Viability of the nonculturable *Vibrio cholerae* O1 and O139. Syst Appl Microbiol 24: 331-341.
3. Broza M, Gancz H, Kashi Y (2008) The association between non-biting midges and *Vibrio cholerae*. Environ Microbiol 10: 3193-3200.
4. Reeves PR, Lan R (1998) Cholera in the 1990s. Br Med Bull 54: 611-623.
5. Vanden Broeck D, Horvath C, De Wolf MJ (2007) *Vibrio cholerae*: cholera toxin. Int J Biochem Cell Biol 39: 1771-1775.
6. Waldor M, Mekalanos JJ (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 272: 1910-1914.
7. Davis BM, Waldor MK (2000) CTXphi contains a hybrid genome derived from tandemly integrated elements. Proc Natl Acad Sci USA 97: 8572-8577.
8. McLeod SM, Kimsey HH, Davis BM, Waldor MK (2005) CTXphi and *Vibrio cholerae*: exploring a newly recognized type of phage-host cell relationship. Mol Microbiol 57: 347-356.
9. Waldor MK, Friedman DI (2005) Phage regulatory circuits and virulence gene expression. Curr Op Micr 8: 459-465.
10. Gardner AD, Venkatraman KV (1935) The antigens of the cholera group of *Vibrios*. J Hyg 35: 262-282.
11. Kaper JB, J Glenn Morris JR, Levine MM (1995) Cholera. Clin Microbiol Rev 8: 48–86.
12. Safa A, Nair GB, Kong RYC (2010) Evolution of new variants of *Vibrio cholerae* O1. Trends Microbiol 18: 46-54.
13. Maiti D, Das B, Saha A, Nandy RK, Nair GB, et al. (2006) Genetic organization of pre-CTX and CTX prophages in the genome of an environmental *Vibrio cholerae* non-O1, non-O139 strain. Microbiology 152: 3633-3641.
14. Blokesch M, Schoolnik GK (2007) Serogroup conversion of *Vibrio cholerae* in aquatic reservoirs. PLoS Pathog 3: e81. doi:10.1371/journal.ppat.0030081.
15. Dalsgaard A, Serichantalergs O, Forslund A, Lin W, Mekalanos J, et al. (2001) Clinical and environmental isolates of *Vibrio cholerae* serogroup O141 carry the CTX phage and the genes encoding the toxin-coregulated pili. J Clin Microbiol 39: 4086-4092.
16. Yamasaki S, Garg S, Nair GB, Takeda Y (1999) Distribution of *Vibrio cholerae* O1 antigen biosynthesis genes among O139 and other non-O1 serogroups of *Vibrio cholerae.* FEMS Microbiol Lett 179: 115-121.
17. Cho YJ, Yi H, Lee JH, Kim DW, Chun J (2010) Genomic evolution of *Vibrio cholerae*. Curr Opin Microbiol 13: 646-651.
18. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, et al. (2009) Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. Proc Natl Acad Sci USA 106: 15442-15447.
19. Siddique AK, Baqui AH, Eusof A, Haider K, Hossain MA, et al. (1991) Survival of classic cholera in Bangladesh. Lancet 337: 1125-1127.
20. Samadi AR, Huq MI, Shahid N, Khan MU, Eusof A, et al. (1983) Classical *Vibrio cholerae* biotype displaces EL tor in Bangladesh. Lancet 1: 805-807.
21. Nair GB, Bhattacharya SK, Deb BC (1994) *Vibrio cholerae* O139 Bengal: the eighth pandemic strain of cholera. Indian J Public Health 38: 33-36.
22. Nair GB, Faruque SM, Bhuiyan NA, Kamruzzaman M, Siddique AK, et al. (2002) New variants of *Vibrio cholerae* O1 biotype El Tor with attributes of the classical biotype from hospitalized patients with acute diarrhea in Bangladesh. J Clin Microbiol 40: 3296-3299.
23. Nair GB, Safa A, Bhuiyan NA, Nusrin S, Murphy D, et al. (2006) Isolation of *Vibrio cholerae* O1 strains similar to pre-seventh pandemic El Tor strains during an outbreak of gastrointestinal disease in an island resort in Fiji. J Med Microbiol 55: 1559-1562.

24. Ansaruzzaman M, Bhuiyan N, Nair B, Sack D, Lucas M, et al. (2004) Cholera in Mozambique, variant of *Vibrio cholerae*. Emerg Infect Dis 10: 2057-2059.
25. Safa A, Sultana J, Cam PD, Mwansa JC, Kong RYC (2008) *Vibrio cholerae* O1 Hybrid El Tor Strains, Asia and Africa. Emerg Infect Dis 14: 987-988.
26. Nair GB, Qadri F, Holmgren J, Svennerholm AM, Safa A, et al. (2006) Cholera due to altered El Tor strains of *Vibrio cholerae* O1 in Bangladesh. J Clin Microbiol 44: 4211-4213.
27. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, et al. (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406: 477-483.
28. Thompson CC, Vicente AC, Souza RC, Vasconcelos AT, Vesth T, et al. (2009) Genomic taxonomy of Vibrios. BMC Evol Biol 9: 258.
29. Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3: 722-732.
30. Faruque SM, Albert MJ, Mekalanos JJ (1998) Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*. Microbiol Mol Biol Rev 62: 1301-1314.
31. Faruque SM, Tam VC, Chowdhury N, Diraphat P, Dziejman M, et al. (2007) Genomic analysis of the Mozambique strain of *Vibrio cholerae* O1 reveals the origin of El Tor strains carrying classical CTX prophage. Proc Natl Acad Sci USA 104: 5151-5156.
32. Safa A, Bhuyian NA, Nusrin S, Ansaruzzaman M, Alam M, et al. (2006) Genetic characteristics of Matlab variants of *Vibrio cholerae* O1 that are hybrids between classical and El Tor biotypes. J Med Microbiol 55: 1563-1569.
33. Murphy RA, Boyd EF (2008) Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates. J Bacteriol 190: 636-647.
34. Karaolis DKR, Lan R, Kaper JB, Reeves PR (2001) Comparison of *Vibrio cholerae* pathogenicity islands in sixth and seventh pandemic strains. Infect Immun 69: 1947-1952.
35. Jermyn WS, Boyd EF (2002) Characterization of a novel Vibrio pathogenicity island (VPI-2) encoding neuraminidase (nanH) among toxigenic *Vibrio cholerae* isolates. Microbiology 148: 3681-3693.
36. Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, et al. (2002) Comparative genomic analysis of *Vibrio cholerae*: Genes that correlate with cholera endemic and pandemic disease. Proc Natl Acad Sci USA 99: 1556-1561.
37. O'Shea YA, Finnan S, Reen FJ, Morrissey JP, O'Gara F, et al. (2004) The *Vibrio* seventh pandemic island-II is a 26.9 kb genomic island present in *Vibrio cholerae* El Tor and O139 serogroup isolates that shows homology to a 43.4 kb genomic island in *V. vulnificus*. Microbiology 150: 4053-4063.
38. Grim CJ, Hasan NA, Taviani E, Haley B, Chun J, et al. (2010) Genome sequence of hybrid *V. cholerae* O1 MJ-1236, B-33 and CIRS101 and comparative genomics with *V. cholerae*. J Bacteriol 192: 3524-3533.
39. Taviani E, Grim CJ, Choi J, Chun J, Haley B, et al. (2010) Discovery of novel *Vibrio cholerae* VSP-II genomic islands using comparative genomic analysis. FEMS Microbiol Lett 308: 130-137.
40. Kapfhammer D, Blass J, Evers S, Reidl J (2002) *Vibrio cholerae* phage K139: complete genome sequence and comparative genomics of related phages. J Bacteriol 184: 6592-6601.
41. Rubin EJ, Lin W, Mekalanos JJ, Waldor MK (1998) Replication and integration of a *Vibrio cholerae* cryptic plasmid linked to the CTX prophage. Mol Microbiol 28: 1247-1254.
42. Hassan F, Kamruzzaman M, Mekalanos JJ, Faruque SM (2010) Satellite phage TLCphi enables toxigenic conversion by CTX phage through dif site alteration. Nature 467: 982-985.
43. Wozniak RAF, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol 8: 552-563.
44. Shoemaker NB, Barber RD, Salyers AA (1989) Cloning and characterization of a Bacteroides conjugal tetracycline-erythromycin resistance element by using a shuttle cosmid vector. J Bacteriol 171: 1294-1302.
45. Franke AE, Clewell DB (1981) Evidence for a chromosome-borne resistance transposon (Tn916) in Streptococcus faecalis that is capable of "conjugal" transfer in the absence of a conjugative plasmid. J Bacteriol 145: 494-502.
46. Burrus V, Pavlovic G, Decaris B, Guedon G (2002) Conjugative transposons: the tip of the iceberg. Mol Microbiol 46: 601-610.

47. Burrus V, Waldor MK (2004) Shaping bacterial genomes with integrative and conjugative elements. Res Microbiol 155: 376-386.
48. Rajeev L, Malanowska K, Gardner JF (2009) Challenging a Paradigm: the Role of DNA Homology in Tyrosine Recombinase Reactions. Microbiol Mol Biol Rev 73: 300-309.
49. Burrus V, Waldor MK (2003) Control of SXT integration and excision. J Bacteriol 185: 5045-5054.
50. McGrath BM, O'Halloran JA, Pembroke JT (2005) Pre-exposure to UV irradiation increases the transfer frequency of the IncJ conjugative transposon-like elements R391, R392, R705, R706, R997 and pMERPH and is recA+ dependent. FEMS Microbiol Lett 243: 461-465.
51. Salyers AA, Shoemaker NB, Stevens AM, Li LY (1995) Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. Microbiol Rev 59: 579-590.
52. Ceccarelli D, Daccord A, Rene M, Burrus V (2008) Identification of the origin of transfer (*oriT*) and a new gene required for mobilization of the SXT/R391 family of integrating conjugative elements. J Bacteriol 190: 5328-5338.
53. Grohmann E, Muth G, Espinosa M (2003) Conjugative plasmid transfer in Gram-positive bacteria. Microbiol Mol Biol Rev 67: 277-301.
54. Matson S, Morton B (1991) *Escherichia coli* DNA helicase I catalyzes a site- and strand-specific nicking reaction at the F plasmid oriT. J Biol Chem 266: 16232-16237.
55. Williams SL, Schildbach JF (2006) Examination of an inverted repeat within the F factor origin of transfer: context dependence of F TraI relaxase DNA specificity. Nucleic Acids Res 34: 426-435.
56. Toleman MA, Bennett PM, Walsh TR (2006) ISCR elements: novel gene-capturing systems of the 21st century? Microbiol Mol Biol Rev 70: 296-316.
57. Waldor MK, Tschape H, Mekalanos JJ (1996) A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. J Bacteriol 178: 4157-4165.
58. Hochhut B, Waldor MK (1999) Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. Mol Microbiol 32: 99-110.
59. Hochhut B, Marrero J, Waldor MK (2000) Mobilization of plasmids and chromosomal DNA mediated by the SXT element, a constin found in *Vibrio cholerae* O139. J Bacteriol 182: 2043-2047.
60. Beaber JW, Hochhut B, Waldor MK (2002) Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. J Bacteriol 184: 4259-4269.
61. Peters SE, Hobman JL, Strike P, Ritchie DA (1991) Novel mercury resistance determinants carried by IncJ plasmids pMERPH and R391. Mol Gen Genet 228: 294-299.
62. Coetzee JN, Datta N, Hedges RW (1972) R factors from *Proteus rettgeri*. J Gen Microbiol 72: 543-552.
63. Murphy D, Pembroke JT (1995) Transfer of the IncJ plasmid R391 to recombination deficient *Escherichia coli* K12: evidence that R391 behaves as a conjugal transposon. FEMS Microbiol Lett 134: 153-158.
64. Beaber JW, Burrus V, Hochhut B, KWaldor M (2002) Comparison of SXT and R391, two conjugative integrating elements: definition of a genetic backbone for the mobilization of resistance determinants. Cell Mol Life Sci 59: 2056-2070.
65. Burrus V, Marrero J, Waldor MK (2006) The current ICE age: biology and evolution of SXT-related integrating conjugative elements. Plasmid 55: 173-183.

CHAPTER II

Comparative ICE Genomics: Insights into the Evolution of the SXT/R391 Family of ICEs

Rachel A. F. Wozniak[1,2,3], Derrick E. Fouts[3], Matteo Spagnoletti[4], Mauro M. Colombo[4], Daniela Ceccarelli[5], Geneviève Garriss[5], Christine Dèry[5], Vincent Burrus[5], Matthew K. Waldor[1,2,6]

1 Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **2** Department of Genetics, Tufts Medical School, Boston, Massachusetts, United States of America, **3** J. Craig Venter Institute, Rockville, Maryland, United States of America, **4** Dipartimento di Biologia Cellulare e dello Sviluppo, Universita´ di Roma La Sapienza, Rome, Italy, **5** Centre d'E´tude et de Valorisation de la Diversite´ Microbienne, De´partement de Biologie, Universite´ de Sherbrooke, Sherbrooke, Que´bec, Canada, **6** Howard Hughes Medical Institute, Chevy Chase, Maryland, United States of America

Abstract

Integrating and conjugative elements (ICEs) are one of the three principal types of self-transmissible mobile genetic elements in bacteria. ICEs, like plasmids, transfer via conjugation, but unlike plasmids and similar to many phages, these elements integrate into and replicate along with the host chromosome. Members of the SXT/R391 family of ICEs have been isolated from several species of gram-negative bacteria, including *Vibrio cholerae*, the cause of cholera, where they have been important vectors for disseminating genes conferring resistance to antibiotics. Here we developed a plasmid-based system to capture and isolate SXT/R391 ICEs for sequencing. Comparative analyses of the genomes of 13 SXT/R391 ICEs derived from diverse hosts and locations revealed that they contain 52 perfectly syntenic and nearly identical core genes that serve as a scaffold capable of mobilizing an array of variable DNA. Furthermore, selection pressure to maintain ICE mobility appears to have restricted insertions of variable DNA into intergenic sites that do not interrupt core functions. The variable genes confer diverse element-specific phenotypes, such as resistance to antibiotics. Functional analysis of a set of deletion mutants revealed that less than half of the conserved core genes are required for ICE mobility; the functions of most of the dispensable core genes are unknown. Several lines of evidence suggest that there has been extensive recombination between SXT/R391 ICEs, resulting in re-assortment of their respective variable gene content. Furthermore, our analyses suggest that there may be a network of phylogenetic relationships between sequences found in all types of mobile genetic elements.

This chapter was published in Plos Genetics (2009)
doi:10.1371/journal.pgen.1000786

Introduction

There are three types of self-transmissible mobile genetic elements: plasmids, bacteriophages and integrative conjugative elements (ICEs). All three classes of elements enable horizontal transmission of genetic information and all have had major impacts on bacterial evolution [1-4]. ICEs, (aka conjugation transposons), like plasmids, are transmitted via conjugation; however, unlike plasmids, ICEs integrate into and replicate along with the chromosome. Following integration, ICEs can excise from the chromosome and form circular molecules that are intermediates in ICE transfer. Plasmids and phages have been the subject of more extensive study than ICEs and while there is growing understanding of the molecular aspects of several ICEs [5-10], to date there have been few reports of comparative ICE genomics [11,12] and consequently understanding of ICE evolution is only beginning to be unraveled.

Diverse ICEs have been identified in a variety of gram-positive and gram–negative organisms [13]. These elements utilize a variety of genes to mediate the core ICE functions of chromosome integration, excision and conjugation. In addition to a core gene set, ICEs routinely contain genes that confer specific phenotypes upon their hosts, such as resistance to antibiotics and heavy metals [14-18], aromatic compound degradation [19] or nitrogen fixation [20].

SXT is an ~100 Kb ICE that was originally discovered in *Vibrio cholerae* O139 [16], the first non-O1 serogroup to cause epidemic cholera [21]. SXT encodes resistances to several antibiotics, including sulfamethoxazole and trimethoprim (which together are often abbreviated as SXT) that had previously been useful in the treatment of cholera. Since the emergence of *V. cholerae* O139 on the Indian subcontinent in 1992, SXT or a similar ICE has been found in most clinical isolates of *V. cholerae*, including *V. cholerae* serogroup O1, from both Asia and Africa. Other *Vibrio* species besides *V. cholerae* have also been found to harbor SXT-related ICEs [22]. Furthermore, SXT-like ICEs are not restricted to vibrio species, as such ICEs have been detected in *Photobacterium damselae*, *Shewanella putrefaciens* and *Providencia alcalifaciens* [23-25]. Moreover, Hochhut *et al* [26] found that SXT is genetically and functionally related to the so-called 'Inc J' element R391, which was derived from a South African *Providencia rettgeri* strain isolated in 1967 [27]. It is now clear that Inc J elements are SXT-related ICEs that were originally misclassified as plasmids. In the laboratory, SXT has a fairly broad host range and can be transmitted between a variety of gram-negative organisms [16].

The SXT/R391 family of ICEs is now known to include more than 30 elements that have been detected in clinical and environmental isolates of several species of γ- proteobacteria from disparate locations around the globe [28]. SXT/R391 ICEs are grouped together as an ICE family because they all encode a nearly identical integrase, Int. Int, a tyrosine recombinase, is considered a defining feature of these elements because it enables their site-specific integration into the 5' end of *prfC*, a conserved chromosomal gene that encodes peptide chain release factor 3 [29]. Int mediates recombination between nearly identical element and chromosome sequences, *attP* and *attB* respectively [29]. When an SXT/R391 ICE excises from the chromosome, Int, aided by Xis, a recombination directionality factor, mediates the reverse reaction - recombination between the extreme right and left ends (*attR* and *attL*) of the integrated element - thereby reconstituting *attP*

and *attB* [6,29]. The excised circular SXT form is thought to be the principal substrate for its conjugative transfer. The genes that encode activities required for SXT transfer (*tra* genes) were originally found to be distantly related to certain plasmid *tra* genes [30-32]. The *tra* genes encode proteins important for processing DNA for transfer, mating pair formation and generating the conjugation machinery. Regulation of SXT excision and transfer is at least in part governed by a pathway that resembles the pathway governing the lytic development of the phage lambda. Agents that damage DNA and induce the bacterial SOS response are thought to stimulate the cleavage and inactivation of SetR, an SXT encoded λ cI-related repressor, which represses expression of *setD* and *setC*, transcription activators that promote expression of *int* and *tra* genes [5]. The complete nucleotide sequences of SXT (99.5kb) and R391 (89kb) were the first SXT/R391 ICE family genomes to be reported [14,32]. Comparative [33] and functional genomic analyses [5,32] revealed that these 2 ICEs share a set of conserved core genes that mediate their integration/excision (*int* and *xis*), conjugative transfer (various *tra* genes), and regulation (*setR*, *setCD*). In addition to the conserved genes, these 2 ICEs contain element specific genes that confer element specific properties such as resistance to antibiotics or heavy metals. Interestingly, many of these genes were found in identical locations in SXT and R391, leading Beaber *et al* [33] to propose that there are 'hotspots' where SXT/R391 ICEs can acquire new DNA. The genomes of two additional SXT/R391 ICEs, ICE*Pda*Spa1, isolated from *Photobacterium damselae* [23], and ICE*Spu*PO1, derived from an environmental isolate of *Shewanella putrefaciens* [24] are now also known. These two genomes also share most of the conserved set of core genes present in SXT and R391 and contain element specific DNA.

Determination of the sequences of SXT/R391 family ICE genomes was a fairly arduous task due to their size and predominantly chromosomal localization. Here, we developed a method to capture and then sequence complete SXT/R391 ICE genomes. In addition, we identified 3 as yet unannotated SXT/R391 ICE genomes in the database of completed bacterial genomes. Comparative analyses of the 13 SXT/R391 genomes now available allowed us to greatly refine our understanding of the organization and conservation of the core genes that are present in all members of this ICE family. Comparative and functional analyses also facilitated our proposal of the minimal functional SXT/R391 ICE genome. Furthermore, this work provides new knowledge of the considerable diversity of genes and potential accessory functions encoded by the variable DNA found in these mobile elements. Finally, this comparative genomics approach has allowed us to garner clues regarding the evolution of this class of mobile elements.

Results and Discussion

An ICE capture system

To date, ICE sequencing has been cumbersome because it has typically required construction of chromosome-derived cosmid libraries and screening for sequences that hybridize to ICE probes [23,32]. We constructed a vector (pIceCap) that enables capture of complete SXT/R391 ICE genomes on a low-copy plasmid to simplify the protocol for ICE sequencing. This plasmid is a derivative of the single-copy modified F plasmid pXX704 [34,35], which contains a minimal set of

genes for F replication and segregation but lacks genes enabling conjugation. We modified pXX704 to include an ~400bp fragment that encompasses the SXT/R391 attachment site (*attB*) and thereby enabled Int-catalyzed site-specific recombination between *attB* on pIceCap and *attP* on an excised and transferred ICE to drive ICE capture (Fig. 1). Conjugations between an SXT/R391 ICE-bearing donor strain and an *E. coli* recipient deleted for *prfC* (and thus chromosomal *attB*) and harboring pIceCap yielded exconjugants containing the transferred ICE integrated into pIceCap (Fig. 1). We used the Δ*prfC* recipient to bias integration of the transferred ICE into pIceCap rather than the chromosome. In these experiments, we selected for exconjugants containing the transferred ICE integrated into pIceCap, using an antibiotic marker present on the ICE as well as a marker present in pIceCap. The low copy IceCap::ICE plasmid was then isolated and used as a substrate for shotgun sequencing. We also found that the IceCap::ICE plasmids were transmissible. Thus, in principle this technique should facilitate capture of ICEs that do not harbor genes conferring resistance to antibiotics, by mating out the IceCap::ICE plasmid into a new recipient and selecting for the marker on pIceCap.

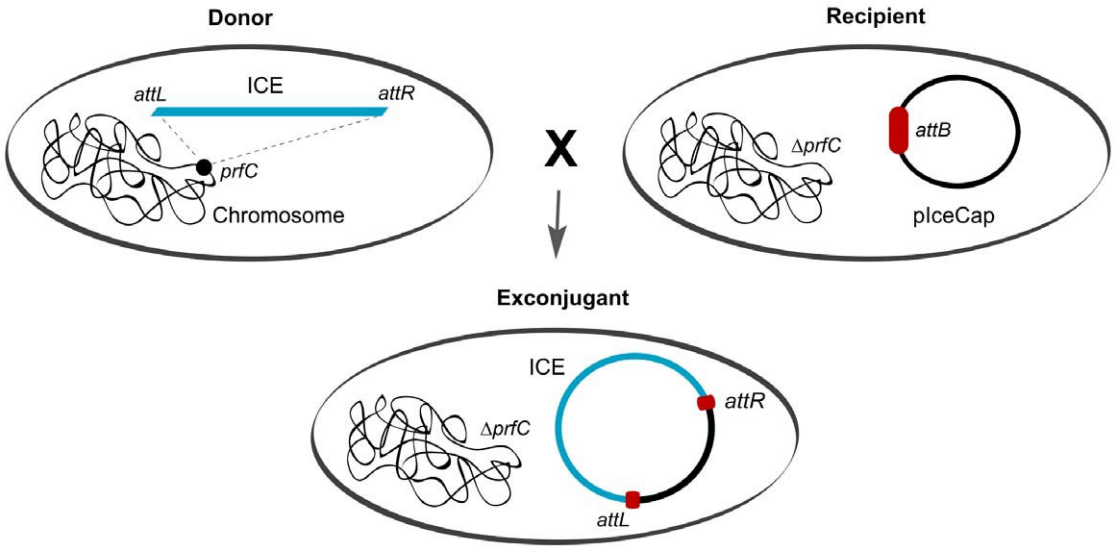


Figure 1. Schematic of the ICE capture system. Conjugation between a donor strain bearing a chromosomal ICE and a Δ*prfC* recipient strain harboring pIceCap, which contains *attB*, yields exconjugants that contain the transferred ICE integrated into pIceCap. Exconjugants were selected for using a marker on pIceCap and on the ICE. *attR* and *attL* represent the right and left ICE-chromosome junctions.

SXT/R391 ICEs included in this analysis

A list of the 13 SXT/R391 ICEs whose genomes were analyzed and compared in this study is shown in Table 1. All of the ICEs including in our analyses contain an *int* gene that was amplifiable using PCR primers for int_{SXT} [29]. They were isolated on 4 continents and from the Pacific Ocean during a span of more than 4 decades. They are derived from 7 different genera of γ-proteobacteria and the ICEs derived from *V. cholerae* strains are from both clinical and environmental isolates of 3 different *V. cholerae* serogroups.

Five of these ICE genome sequences were determined at the J. Craig Venter Institute (JCVI) using

the ICE capture system described above (Table 1, rows 1-5). In addition, we sequenced ICE*Vfl*Ind1, also at the JCVI, by isolating cosmids that encompassed this *V. fluvialis* derived ICE prior to developing the ICE capture technique (Table 1, row 6). Table 1 (rows 7-10) also includes 4 previously unannotated ICE genomes that we found in BLAST searches of the NCBI database of completed but as yet unannotated genomes; 3 of these ICEs are clearly members of SXT/R391 ICE family since they are integrated into their respective host's *prfC* locus and contain *int* genes that are predicted to encode Int proteins that are 99% identical to Int_{sxt}. The fourth element, ICE*Vch*Ban8 does not encode an Int_{sxt} orthologue; however, this element contains nearly identical homologues of most of the known conserved core SXT/R391 ICE family genes. ICE*Vch*Ban8 will be discussed in more detail below but since it does not contain an Int_{sxt} orthologue it is not considered a member of the SXT/R391 family of ICEs and thus not included in our comparative study. Finally, Table 1 also includes the 4 SXT/R391 ICEs that were previously sequenced (Table 1, rows 11-14).

Despite the diversity of our sources for SXT/R391 ICEs, the genomes of two pairs of ICEs that we analyzed proved to be very similar. SXT^{MO10} and ICE*Vch*Ind4 only differed by 13 SNPs in 7 genes and by the absence from ICE*Vch*Ind4 of *dfr18*, a gene conferring trimethoprim resistance. These ICEs were derived from *V. cholerae* O139 strains isolated in India from different cities at different times: SXT^{MO10} from Chennai in 1992 and ICE*Vch*Ind4 from Kolkata in 1997. The high degree of similarity of these two ICE genomes suggests that ICEs can be fairly stable over time. ICE*Vch*Ban9 and ICE*Vch*Moz10 were also extremely similar although ICE*Vch*Moz10 lacks *dfrA1*, another allele for trimethoprim resistance. These two ICEs were derived from *V. cholerae* O1 strains from Bangladesh (1994) and Mozambique (2004) respectively. The great similarity of these ICEs suggests that there has been spread of SXT-related ICEs between Asia and Africa in recent times. Studies of CTX prophage genomes have also suggested the spread of *V. cholerae* strains between these continents [36].

General structure and sizes of SXT/R391 genomes

The ICEs listed in Table 1 were initially compared using MAUVE [37] and LAGAN [38], programs that enable visualization of conserved and variable regions on a global scale. All of the SXT/R391 ICEs we analyzed share a common structure and have sizes ranging from 79,733 bp to 108,623 bp (Table 1 and Fig. 2). They contain syntenous sets of 52 conserved core genes (Fig. 2A) that total approximately 47kb and encode proteins with an average of 97% identity to those encoded by SXT. All of the individual ICEs also contain DNA that is relatively specific for individual elements (Fig. 2B); the differences in the sizes of the variable regions accounts for the range in ICE sizes.

Five sites within the conserved SXT/R391 ICE structure have variable DNA present in all of the ICEs in figure 2. Four of these sites were previously termed 'hotspots' for ICE acquisition of new DNA [33]. Due to similarities between SXT and R391, the fifth hotspot only became apparent through our comparison of the 13 ICEs examined here. Each of these hotspots (HS1 to HS5 in Fig. 2B) is found in an intergenic region (see below), suggesting that the acquisition of these variable DNA regions has not interrupted core ICE gene functions. In addition, some of the ICEs have variable DNA inserted in additional intergenic locations or in rumB (labeled I-IV in Fig. 2B). Previous

Table 1. SXT/R391 ICE family members analyzed in this study.

ICE	Host strain	Site and year of isolation	Size (bp)	% Identity to Int_{SXT}	Resistance profile	Notable Variable Genes	Genbank Accession Number	Strain or ICE References
ICE*Vch*Mex1	*Vibrio cholerae* non O1-O139	San Luis Potosi, Mexico 2001	82839	99% (410/413)	-	Fic family protein, diguanylate cyclase, restriction modification system	GQ463143	[66]
ICE*Vch*Ind4	*Vibrio cholerae* O139	Kolkata, India 1997	95491	100% (413/413)	*floR, strBA, sul2*	Toxin-antitoxin system	GQ463141	[54]
ICE*Vch*Ind5	*Vibrio cholerae* O1	Sevagram, India 1994	97847	99% (409/413)	*floR, strBA, sul2, dfrA1*	AraC family transcription regulator, glyoxoylase abx resistance	GQ463142	This study
ICE*Vch*Ban5	*Vibrio cholerae* O1	Bangladesh, 1998	102131	99% (409/413)	*floR, strBA, sul2, dfrA1*	AraC family transcription regulator, glyoxoylase abx resistance	GQ463140	[29]
ICE*Pal*Ban1	*Providencia alcalifaciens*	Bangladesh, 1999	96586	99% (409/413)	*floR, strBA, sul2, dfrA1*	Toxin-antitoxin system, phenazine biosynthesis protein, lysine exporter	GQ463139	[25]
ICE*Vfl*Ind1	*Vibrio fluvialis*	Kolkata, India 2002	91369[(a)]	99% (409/413)	*dfr18, floR, strBA, sul2*	Toxin-antitoxin system	GQ463144	[22]
ICE*Vch*Moz10	*Vibrio cholerae* O1	Beira, Mozambique 2004	104495	99% (409/413)	*floR, strBA, sul2, tetA'*	AraC family transcription regulator, glyoxoylase abx resistance, ATP-dependent Lon protease	ACHZ00000000	[67]
ICE*Pmi*Usa1	*Proteus mirabilis*	Maryland, United States 1986	79733	99% (409/413)	-	ATP-dependent helicase	AM942759	[68]
ICE*Vch*Ban9	*Vibrio cholerae* O1	Matlab, Bangladesh 1994	106124	99% (409/413)	*floR, strBA, sul2, dfrA1, tetA'*	AraC family transcription regulator, glyoxoylase abx resistance, ATP-dependent Lon protease	CP001485	[69]
ICE*Vch*Ban8	Vibrio cholerae non O1-O139	Bangladesh, 2001	105790[(a)]	25% (76/301)	-	Toxin-antitoxin system	NZ_AAUU00000000	This study
SXT[MO10]	*Vibrio cholerae* O139	Chennai, India 2002	99452	100%	*dfr18, floR, strBA, sul2*	Toxin-antitoxin system	AY055428	[16]
R391	*Providencia rettgeri*	Pretoria, South Africa 1967	88532	99% (410/413)	*kanR, merRTPCA*	Sulfate transporter, universal stress protein	AY090559	[27]
ICE*Pda*Spa1	*Photobacterium damselae*	Galicia, Spain 2003	102985	99% (412/413)	*tetAR*	ATP-dependent Lon protease, heat-shock protein	AJ870986	[23]
ICE*Spu*PO1	*Shewanella putrefaciens*	630m, Pacific Ocean 2000	108623	99% (409/413)	-	Zn/Co/Cd efflux system, restriction modification system	CP000503	[24]

[(a)]The sequence is not complete and therefore the true size is not known.

analyses [32] indicated that the insertion in rumB, did not impair SXT transmissibility. Overall, comparison of these 13 SXT/R391 ICE genomes suggests that: 1) these elements consist of the same perfectly syntenous and nearly identical 52 core genes that serve as a scaffold (see below) capable of mobilizing a large range of variable DNA; and 2) selection pressure to maintain ICE mobility has restricted insertions of variable DNA into sites that do not interrupt core functions.

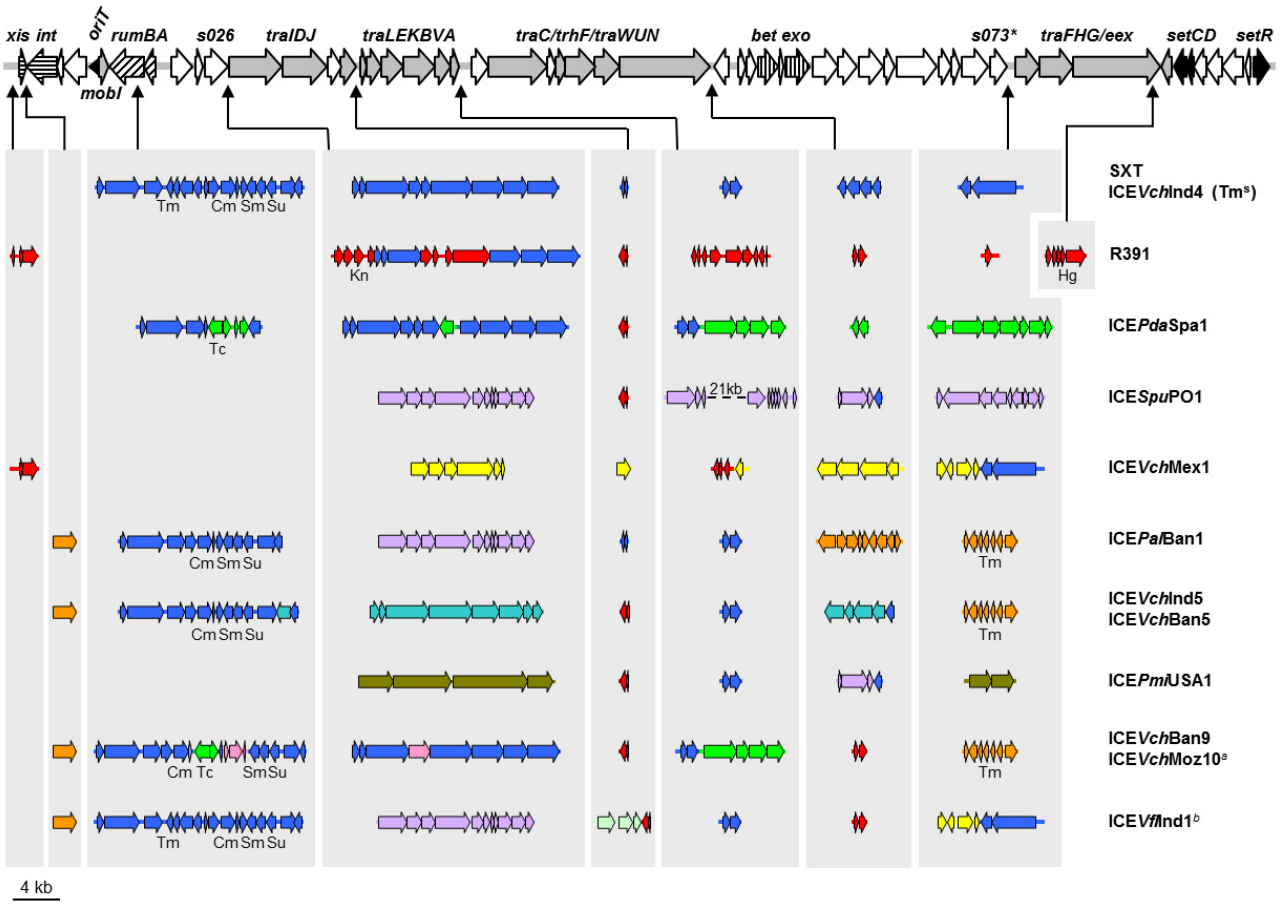


Figure 2 Structure of the genomes of 13 SXT/R391 ICEs. A) The upper line represents the set of core genes (thick arrows) and sequences common to all 13 SXT/R391 genomes analyzed. Hatched ORFs indicate genes involved in site-specific excision and integration (*xis* and *int*), error-prone DNA repair (*rumAB*), DNA recombination (*bet* and *exo*) or entry exclusion (*eex*). Dark gray ORFs correspond to genes involved in regulation (*setCDR*). Light gray ORFs represent genes encoding the conjugative transfer machinery, and white ORFs represent genes of unknown function. B) Variable ICE regions are shown with colors according to the elements in which they were originally described SXT (blue), R391 (red), ICE*Pda*Spa1 (green), ICE*Spu*PO1 (purple), ICE*Vch*Mex1 (yellow), ICE*Pal*Ban1 (orange), ICE*Vch*Ind5 (turquoise), ICE*Pmi*USA1 (olive), ICE*Vch*Ban9 (pink), ICE*Vfl*Ind1 (light green). Thin arrows indicate the sites of insertion for each variable region and HS1-HS5 represent hotspots 1-5. Roman numerals indicate variable regions not considered true hotspots. Cm, chloramphenicol; Hg, mercury; Kn, kanamycin; Sm, streptomycin; Su, sulfamethoxazole; Tc, tetracycline; Tm, trimethoprim. * indicates that *s073* is absent from ICE*Pda*Spa1. a ICE*Vch*Moz10, which lacks *dfrA1* in the integron structure, does not confer resistance to Tm. b The purple gene content of ICE*Vfl*Ind1 was deduced from partial sequencing, PCR analysis and comparison with ICE*Spu*PO1.

The SXT/R391 ICE core genes

The 52 core genes present in all the SXT/R391 ICEs analyzed include sets of genes that are known to be required for the key ICE functions of integration/excision, conjugative transfer and regulation [32] as well as many genes of unknown function. Most genes of known or putative

(based on homology) function (coded by gray shading or hatch marks in Fig. 2A) are clustered with genes that have related functions. For example, *int* and *xis*, genes required for integration and excision, are adjacent and *setR*, and *setC*/*D*, the key SXT regulators are near each other at the extreme 3' end of the elements, although separated by 4 conserved genes of unknown function. Each ICE also has four gene clusters implicated in conjugative DNA processing and transfer (shown in light gray in Fig. 2A). Finally, each of the ICEs has a nearly identical origin of transfer (oriT), a cis-acting DNA site that is thought to be nicked to initiate DNA processing events during conjugative transfer [39], in the same relative location. The conserved core genes include approximately as many genes of unknown function as genes of known function. Some of the genes of unknown function are found either interspersed amongst gene clusters that likely comprise functional modules (e.g *s091* between *traD* and *s043*) while others are grouped together (e.g. most genes between *traN* and *traF*). In several cases, the interspersed genes appear to be part of operons with genes of known function (e.g. *s086-s082* maybe in an operon with *setDC*).

Variable ICE DNA

In addition to sharing 52 core genes, all of the ICE genomes analyzed contain variable DNA regions, ranging in size from 676 to 29,210 bp. Most of the variable DNA sequences are found in 5 intergenic hotspots (Fig. 2B). However, some ICEs contain additional variable DNA inserts outside the 5 hotspots. For example, SXT and five other ICEs in figure 2 have variable DNA segments, corresponding to related ISCR2 elements, disrupting *rumB* (Fig. 2B, site III). ISCR2 elements are IS91-like transposable elements that tend to accumulate antibiotic resistance genes [40]. Interestingly, it is unusual for the contents of the hotspots and other variable regions to be found in only one ICE. Instead, the variable gene content of most of the ICEs shown in figure 2B is found in more than one ICE. For example, ICE*Spu*PO1, ICE*Pal*Ban1, and ICE*Vfl*Ind1, all have identical contents in hotspot 5 (lavender genes in hotspot 5 in Fig. 2B); however, the contents of the other hotspots in these 3 elements are almost entirely different. Thus, the variable gene content of the SXT/R391 ICEs reveals that these elements are mosaics. The overlapping distribution of variable DNA segments seen in the ICEs in figure 2B suggests that recombination among this family of mobile elements may be extensive. In addition, in some instances, the variable regions appear to subject to additional genetic modifications. For example, ICE*Pda*Spa1 and ICE*Vch*Ban9 contain ICE-specific DNA nested within the shared sequences inserted at hotspot 5 DNA (the green and pink genes in hotspot 5 in these elements, Fig. 2B).

The variable genes encode a large array of functions and only a few will be discussed here. A complete list of the diverse genes found in the hotspots is found in Table S1. Although we cannot predict functions for many genes found in the hotspots, since they lack homology to genes of known function, at least a subset of the known genes seem likely to confer an adaptive advantage upon their hosts. Most of the ICE antibiotic resistance genes are found within transposon-like structures (e.g., the ISCR2 elements noted above) but four ICEs contain a *dfrA1* cassette, which confers resistance to trimethoprim [25], in a class IV integron located in hotspot 3. A disproportionate number of variable genes are likely involved in DNA modification, recombination or repair, as they are predicted to encode diverse putative restriction-modification systems,

helicases and endonucleases. Such genes may provide the host with barriers to invasion by foreign DNA including phage infection and/or promote the integrity of the ICE genome during its transfer between hosts. Three ICEs contain genes that encode diguanylate cyclases (E Bordeleau, E Brouillette, N Robichaud, V Burrus, in press) in hotspot 3. These enzymes catalyze the formation of cyclic-diguanosine monophosphate (c-di-GMP), a second messenger molecule that regulates biofilm formation, motility and virulence in several organisms including *V. cholerae* [41,42]. Most SXT/R391 ICEs contain *mosA* and *mosT* in hotspot 2. These two genes encode a novel toxin-antitoxin pair that promotes SXT maintenance by killing or severely inhibiting the growth of cells that have lost this element [43]. Not all ICEs in the SXT/R391 family contain *mosAT*; however, those lacking these genes may encode similar systems to prevent ICE loss. For instance, R391 and ICE*Vch*Mex1 contain two genes (*orf2* and *orf3*) encoding a predicted HipA-like toxin and a predicted transcriptional repressor distantly related to the antitoxin HipB.

Locations of the ICE variable genes

The variable regions found in the 5 hotspots are found exclusively in intergenic regions, punctuating the conserved ICE backbone (Fig. 2). The boundaries between the conserved and variable sequences were mapped on the nucleotide level and compared (Fig. 3A-E). Each hotspot had a distinct boundary. Remarkably, even though the contents of the variable regions markedly differ, with few exceptions the left and right boundaries between conserved and variable DNA for each hotspot was identical among all the ICEs (Fig. 3). For example, the left junctions of the inserts in hotspot 2 immediately follow the stop codon of *traA* and the right junctions are exactly 79 bp upstream of the start of *s054* (Fig. 3B), despite the fact that the DNA contents within these borders greatly differ. In hotspot 2, the right junction appears to begin with a 15 bp sequence that has two variants (Fig. 3B, brown & light brown sequence). These sequences may reflect the presence of earlier insertions that have since been partially replaced. A similar pattern was found adjacent to the left boundary of hotspot 4 in several ICEs (Fig. 3D, lines 3-6). Once an insertion is acquired, the number of permissive sites for the addition of new variable DNA likely increases. There are two exceptions to the precise boundaries between variable and conserved DNA. Hotspot 1 and hotspot 3 in ICE*Vch*Mex1 and ICE*Pda*Spa1, respectively, contain variable DNA that extends beyond the boundary exhibited by all the other ICEs in these locations (Fig. 3A, line 3 and 3C, line 7). The only boundary that could not be identified was the left border of hotspot 5, the region containing genes between *s026* and *traI*. As discussed below, *s026* is the least conserved core gene and its variability obscured any consensus sequence abutting the variable DNA. Perhaps this border has eroded because s026 is not required for ICE mobility [32].

The relative precision of most boundaries between conserved and variable DNA sequences in all the ICEs analyzed suggests that a particular recombination mechanism, such as *bet*/*exo*-mediated recombination, may explain the acquisition of the variable regions. However, at this point, we cannot exclude the possibility that the precise location for variable DNA insertions simply reflects selection for optimal ICE fitness; i.e., ICEs can optimally accommodate variable DNA in these locations while preserving their essential functions.

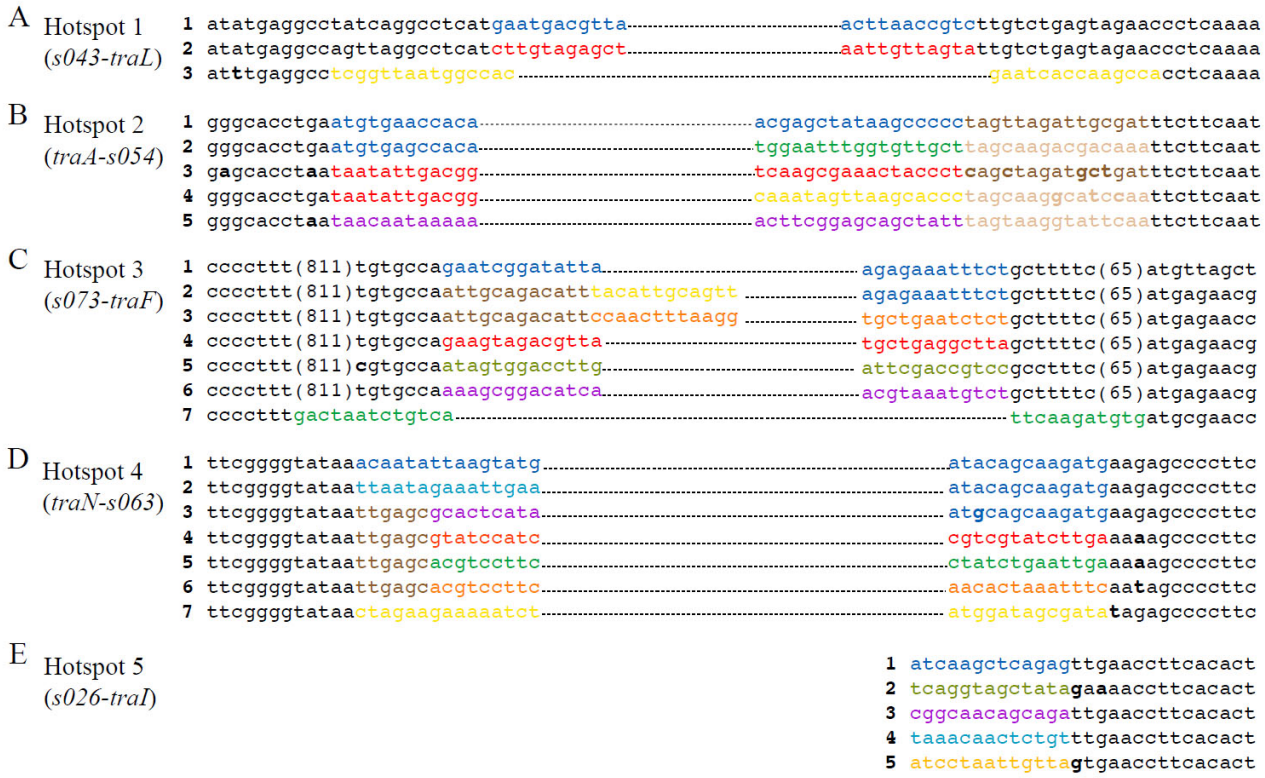


Figure 3. The boundaries of the 5 hotspots. The boundaries between conserved and hotspot variable regions are shown. Black typeface indicates conserved sequence, while color indicates variable sequence. Numbers in parentheses indicate the number of intervening nucleotides. The thin dotted lines indicate continuations of variable DNA. Bold letters indicate a non-conserved base. A) Hotspot 1, which is present between *traJ* and *traL*. Line 1: SXT, ICE*Vch*Ind4, ICE*Pal*Ban1; Line 2: R391, ICE*Pda*Spa1, ICE*Vch*Ban5, ICE*Vch*Ind5, ICE*Pmi*USA, ICE*Spu*PO1, ICE*Vfl*Ind1, ICE*Vch*Moz10, ICE*Vch*Ban9; Line 3: ICE*Vch*Mex1. B) Hotspot 2, which is present between *traA* and *s054*. Line 1: SXT, ICE*Vch*Ind4, ICE*Pmi*USA, ICE*Vfl*Ind1, ICE*Vch*Ind5, ICE*Pal*Ban1, ICE*Vch*Ban5; Line 2: ICE*Pd*aSpa1, ICE*Vch*Moz10, ICE*Vch*Ban9; Line 3: R391; Line 4: ICE*Vch*Mex1; Line 5: ICE*Spu*PO1. C) Hotspot 3, which is present between *s073* and *traF*. Line 1: SXT, ICE*Vch*Ind4; Line 2: ICE*Vch*Mex1, ICE*Vfl*Ind1; Line 3: ICE*Vch*Moz10, ICE*Vch*Ban9, ICE*Vch*Ind5, ICE*Pal*Ban1, ICE*Vch*Ban5. Line 4: R391; Line 5: ICE*Pmi*USA; Line 6: ICE*Spu*PO1; Line 7: ICE*Pda*Spa1. D) Hotspot 4, which is present between *traN* and *s063*. Line 1: SXT, ICE*Vch*Ind4. Line 2: ICE*Vch*Ind5, ICE*Vch*Ban5; Line 3: ICE*Spu*PO1, ICE*Pmi*USA; Line 4: R391, ICE*Vch*Moz10, ICE*Vch*Ban9, ICE*Vfl*Ind1. Line 5: ICE*Pda*Spa1; Line 6: ICE*Pal*Ban1; Line 7: ICE*Vch*Mex1. E) Hotspot 5, which is present between *s026* and *traI*. Line 1: SXT, ICE*Vch*Ind4, ICE*Pda*Spa1, R391, ICE*Vch*Moz10, ICE*Vch*Ban9; Line 2: ICE*Pmi*USA; Line 3: ICE*Spu*PO1, ICE*Pal*Ban1, ICE*Vfl*Ind1; Line 4: ICE*Vch*Ind5, ICE*Vch*Ban5; Line 5: ICE*Vch*Mex1.

Similarity of SXT/R391 ICE and IncA/C plasmid core genes

Unexpectedly, BLAST analyses revealed that most of the conserved core SXT/R391 genes are also present in IncA/C conjugative plasmids. These multidrug resistance plasmids are widely distributed among *Salmonella* and other enterobacterial isolates from agricultural sources [44,45]. Recently, members of this family of plasmids have also been identified in *Yersinia pestis*, including from a patient with bubonic plague [46], and in aquatic γ-proteobacteria [47], including *Vibrio cholerae* [48,49]. To date, the closest known relatives of the SXT/R391 transfer proteins are found in the IncA/C plasmids. Every predicted SXT transfer protein is encoded by the IncA/C plasmid pIP1202 isolated from *Y. pestis* [49] and the identities of these predicted protein sequences vary from 34 to 78% (Fig. 4A). Furthermore, there is perfect synteny between the four gene clusters

encoding the respective conjugative machineries of these two mobile elements (yellow and orange genes in Fig. 4A). Despite the extensive similarity of the SXT and IncA/C conjugative transfer systems, these plasmids lack homologues of *setR* and *setD*/*C* as well as *int*/*xis*, suggesting that regulation of conjugative transfer differs between these elements.

The similarity of IncA/C plasmids and SXT/R391 ICEs is not limited to genes important for conjugal DNA transfer. Ten genes of unknown function (shown in black in Fig. 4A), some of which are interspersed within likely tra gene operons and some of which are clustered together between *traN* and *traF*, are similar in the two elements. Furthermore, most of these ten genes are in identical locations in the two elements. Both elements also contain homologs of bet and *exo* (shown in green in Fig. 4A); these are the only known homologs of the λ Red recombination genes found outside of bacteriophages. Together, the similarity of DNA sequences and organization of SXT/R391 ICEs and IncA/C plasmids suggests that these elements have a common ancestor. The fact that the contents of the hotspots in the two classes of elements are entirely distinct suggests that their evolutionary paths diverged prior to acquisition of these variable DNA segments.

The minimal functional SXT/R391 ICE gene set

The conservation of the 52 core genes in all 13 SXT/R391 ICEs analyzed suggested that many or even all of these genes would be required for key ICE functions of excision/integration, conjugative transfer and regulation. The presence of ten ICE core genes of unknown function in IncA/C plasmids (black genes in Fig. 4A) is also consistent with the hypothesis that these genes might be required for ICE transfer. However, our previous work demonstrated that not all genes recognized here as part of the conserved core gene set are required for SXT transfer. Beaber *et al* showed that deletion of *rumB–s026* (which includes 5 cores genes) from SXT had no detectable influence on SXT excision or transfer [32]. Therefore, we systematically deleted all of the core ICE genes whose contributions had not previously been assessed, in order to explore the hypothesis that these genes (especially those also present in IncA/C plasmids) would be essential for ICE transfer and to define the minimum functional SXT/R391 gene set.

Surprisingly, deletion of most of the ICE core genes of unknown function, including genes with homologues in IncA/C plasmids, did not alter SXT transfer efficiency. Deletion of *s002* or *s003*, which are located downstream of *int* in all SXT/R391 ICEs, did not alter the frequency of SXT transfer; similarly, deletion of *s082*, *s083*, and *s084*, core genes of unknown function that are found near the opposite end of SXT/R391 ICEs but not in IncA/C plasmids, also did not influence SXT transfer frequency (Fig. 4B). Furthermore, deletion of *s091*, which is found between *traD* and *s043* in ICEs and IncA/C plasmids, did not reduce SXT transfer (Fig. 4B). In contrast, deletion of *s043*, which has weak homology to *traJ* in the F plasmid (a gene important in DNA processing) and is located in a transfer cluster containing *traI* and *traD*, abolished transfer (Fig. 4B, Δd), suggesting that *s043*, here re-named *traJ* is required for SXT transfer. It is unlikely that the transfer defect of SXTΔ*traJ* can be explained by polar effects of the deletion on downstream genes, since *traJ* appears to be the last gene of an operon found immediately upstream of hotspot 1. Similarly, deletion of *s054*, which is found immediately 5' of *traC* and is homologous to a disulfide-bond isomerase *dsbC*, also abolished transfer (Fig. 4B, Δe). Interestingly, disulfide bond-isomerases are

present in several other conjugative systems [50]. However, it is not clear at this point if the deletion of *s054* from SXT accounts for the transfer defect of SXTΔ*s054*, since we could not restore transfer by complementation.

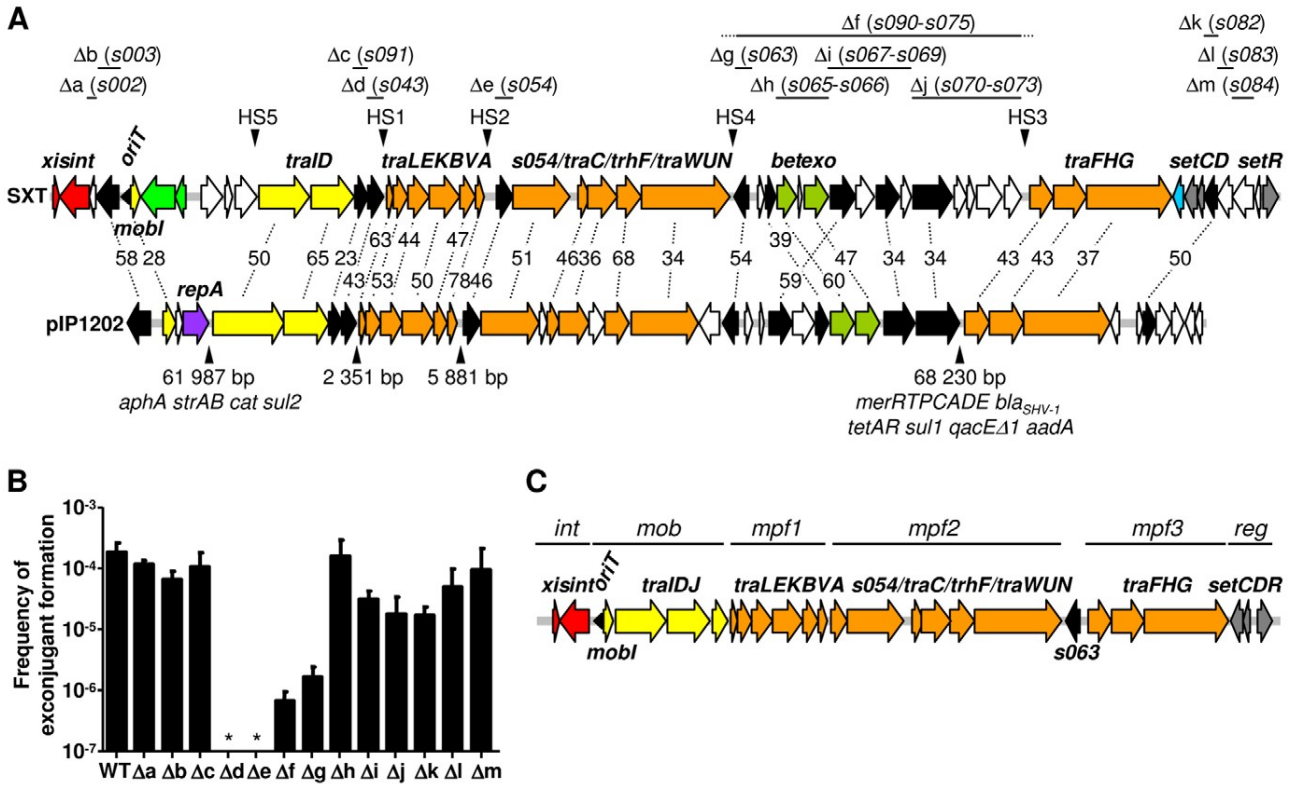


Figure 4 Comparison of the SXT/R391 core genome with the genome of pIP1202 and defining the minimal functional SXT/R391 gene set. A) Alignment of the conserved core genes of SXT/R391 ICEs with the genome of the IncA/C conjugative plasmid pIP1202 from *Yersinia pestis*. The top line shows the same core ICE genes shown in figure 2A. ORFs are color coded as follows: DNA processing, yellow; mating pair formation, orange; DNA recombination and repair, green; integration/excision, red; replication, purple; regulation, gray; entry exclusion, blue; homologous genes of unknown function, black; genes without corresponding counterparts in ICEs and pIP1202, white. Numbers shown in the middle represent % identity between the orthologous proteins encoded by SXT and pIP1202 [GenBank: NC_009141]. The positions of the hotspots in SXT/R391 ICEs are marked by downward pointing arrowheads. For pIP1202, the size of the sequences (which include IncA/C backbone DNA as well as variable DNA) found at these locations as well as resistance markers are indicated by upward pointing arrowheads. *aphA*, *aadA* and *strAB* confer resistance to aminoglycosides. *sul1* and *sul2* confer resistance to sulfonamides. *cat*, *blaSHV-1*, *tetAR*, *qacED1* and *merRTPCADE* confer resistance to chloramphenicol, β- lactams, tetracyclines, quaternary ammonium compounds and mercury ions, respectively. Detailed descriptions of the conserved backbone of the IncA/C conjugative plasmids have been published elsewhere [47,49]. Regions that were deleted from SXT to investigate the function of genes of unknown function (see panel B) are indicated with straight lines. Dotted lines indicate that the deletion included DNA in the adjacent hotspots. B) Influence of deletion of genes of unknown function on the frequency of SXT transfer. The mean values and standard deviations from three independent experiments are shown. * indicates that the frequency of transfer was below the detection level ($<10^{-8}$). Deletion mutants SXTΔa, SXTΔk and SXTΔl, transferred at frequencies that were not significantly different from that of wild-type SXT (data not shown). C) Proposed minimal set of genes necessary for a functional SXT/R391 ICE. int, integration/excision module; mob, DNA processing module; mpf, mating pair formation modules; reg, regulation module.

Additionally, Beaber *et al* found that deletion of *s060* through *s073* in SXT, which includes 7 genes that are also found in IncA/C plasmids reduced SXT transfer more than 100-fold [32]. We constructed several smaller deletions in this region and found that deletion of *s063*, which is also found in pIP1202, reduced the transfer frequency of SXT by ~100-fold, nearly the same amount as deleting the entire region (Fig. 4B). Complementation analyses revealed that the absence of *s063* accounted for the transfer defect of SXTΔ*s063* (data not shown). Even though SXTΔ*s063* was still capable of transfer, in our view, the drastic reduction in the transfer frequency of this mutant warrants inclusion of *s063* into the minimum functional SXT ICE genome (shown in Fig.4C). Other deletions in this region, including deletions of *bet*, *exo*, *s067*, *s068* and *s070*, which have orthologues in IncA/C plasmids, resulted in ≤ 10-fold reductions in transfer frequency. We therefore did not include these genes in the minimal functional core SXT/R391 genome (Fig. 4C).

The findings from our experiments testing the transfer frequencies of SXT derivatives harboring core gene deletions (shown in Fig. 4B), coupled with our previous work demonstrating the requirements for the predicted SXT *tra* genes in the element's transfer [32], suggest a minimal functional SXT/R391 ICE structure as shown in figure 4C. This minimum element is ~29.7 kb and consists of 25 genes. Genes with related functions, which in some cases encode proteins that likely form large functional complexes (such as the conjugation apparatus) are grouped together in the minimal genome. At the left end of the minimum ICE genomes are *xis* and *int*, the integration/excision module of SXT/R391 ICEs. In the minimal ICE genome, the ICE oriT and *mobI*, which encodes a protein required for SXT transfer [39], are no longer separated from the other genes (*traIDJ*) that are also thought to play roles in the DNA processing events required for conjugative DNA transfer. The genes required for formation of the conjugation machinery, including the pilus, and mating pair formation and stabilization [32,39] are divided between three clusters (denoted mpf1-3 in Fig. 4C). Finally, at the right end of the minimal functional genome are the genes that regulate ICE transfer (*setC*/*D* and *setR*). Thus, the minimal functional SXT/R391 ICE is relatively small and organized into 3 discrete functional modules that mediate excision/integration, conjugation, and regulation.

Even though deletion of 27 out of 52 SXT/R391 ICE core genes proved to have little or no effect on SXT transfer frequency, and hence these genes were not included in figure 4C, it is reasonable to presume that these genes encode functions that enhance ICE fitness given their conservation. For example, the presence of highly conserved bet and exo genes in all SXT/R391 ICEs suggests that there has been selection pressure to maintain this ICE-encoded recombination system that promotes ICE diversity by facilitating inter ICE recombination (G Garriss, MK Waldor, V Burrus, in press). A key challenge for future studies will be to determine how core genes of unknown function promote ICE fitness.

Variations in the similarity of core genes

To identify genes in the SXT/R391 core genome that may be subject to different selection pressures, we compared the percent identity of each ICE's core genes to the corresponding SXT gene (Fig. 5). Most of the ICEs' core genes exhibited 94% to 98% identity on the nucleotide level to SXT's core genes. There was no discernable difference in the degree of conservation of most core

genes that were or were not part of the minimal ICE, suggesting that there are equal selective pressures on essential and non- essential genes. However, we identified 8 genes (*s026*, *traI*, *orfZ*, *s073*, *traF*, *eex*, *s086*, and *setR*) that exhibit significantly different degrees of conservation (Fig. 5 and Fig. S1). Three of these showed unusually high conservation, while the other 5 had below average conservation. Two of the highly conserved genes, *setR* and *s086*, are found at the extreme 3' end of the elements. The conservation of *setR* may reflect the key role of this gene in controlling SXT gene expression. *s086* may also play a role in regulating SXT transfer [51]. The other highly conserved gene, *orfZ*, is found between *bet* and *exo* and has no known function.

s026 and *s073* are the most divergent of all the genes in the backbone. *s026* encodes a hypothetical protein with homologues in many gram negative organisms. Although S026 is predicted to contain a conserved domain, COG2378, which has a putative role in transcription regulation, this protein is not required for SXT transfer [32]. The significant divergence of *s026* along with its lack of essentiality suggests that this gene could become a pseudogene. A similar argument could be made for *s073*, which encodes a hypothetical protein that is also not required for ICE transfer. However, this argument does not hold for *traI* or *traF*, two genes which are essential for ICE transfer. Although the reasons which account for the different degrees of conservation of these 8 core genes are hard to ascertain at this point, the data in figure 5 suggests that individual core genes are subject to different evolutionary pressures.

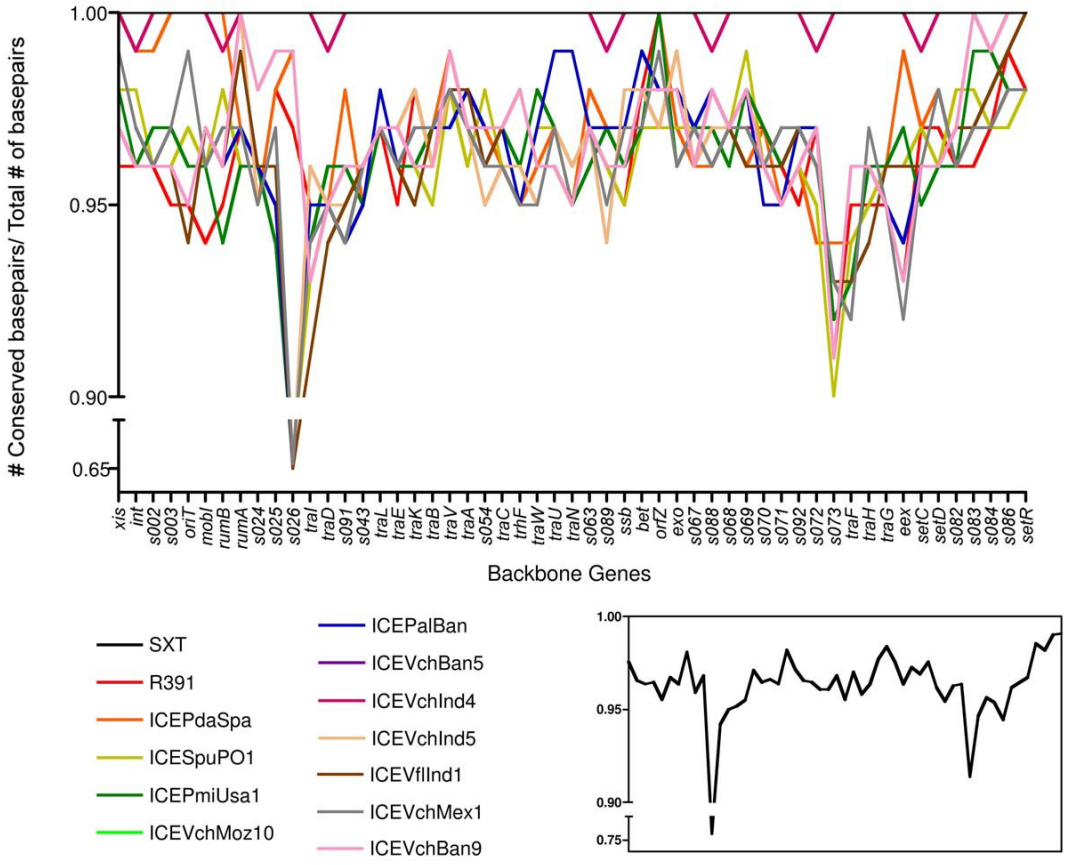


Figure 5. Variations in the nucleotide conservation of core ICE genes. The nucleotide sequence of each core gene from each ICE was compared to the corresponding sequence in SXT using pairwise BLASTn analyses to determine the percent identities. The average values for all of the ICEs, excluding SXT and ICEVchInd4, are shown in the inset.

Comparisons of core gene phylogenies

We created phylogenetic trees for each core gene based on their respective nucleotide sequences to further explore the evolution of the conserved backbone of SXT/R391 ICEs. Since we found such a high degree of conservation for most of the core genes, the bootstrap values for most of these trees were relatively low. Thus, we concentrated on the most polymorphic genes found in figure 5, *s026*, *s073*, *traI*, and *eex*, for phylogenetic analyses. As shown in figure 6A, the trees for *s026*, *traI* and *s073* exhibit 3 distinct branching patterns. The lack of similarity in these phylogenetic trees suggests that either individual core genes have evolved independently or that high degrees of recombination mask their common evolutionary history. The latter hypothesis seems more likely since experimental findings have revealed that SXT/R391 ICEs can co-exist in a host chromosome in tandem [26] and recombination between tandem elements can yield novel hybrid ICEs with considerable frequency [52] (G Garriss, MK Waldor, V Burrus, in press). Also, as noted above, the distributions of variable genes among the ICEs shown in figure 2 also supports the idea that inter-ICE recombination is commonplace.

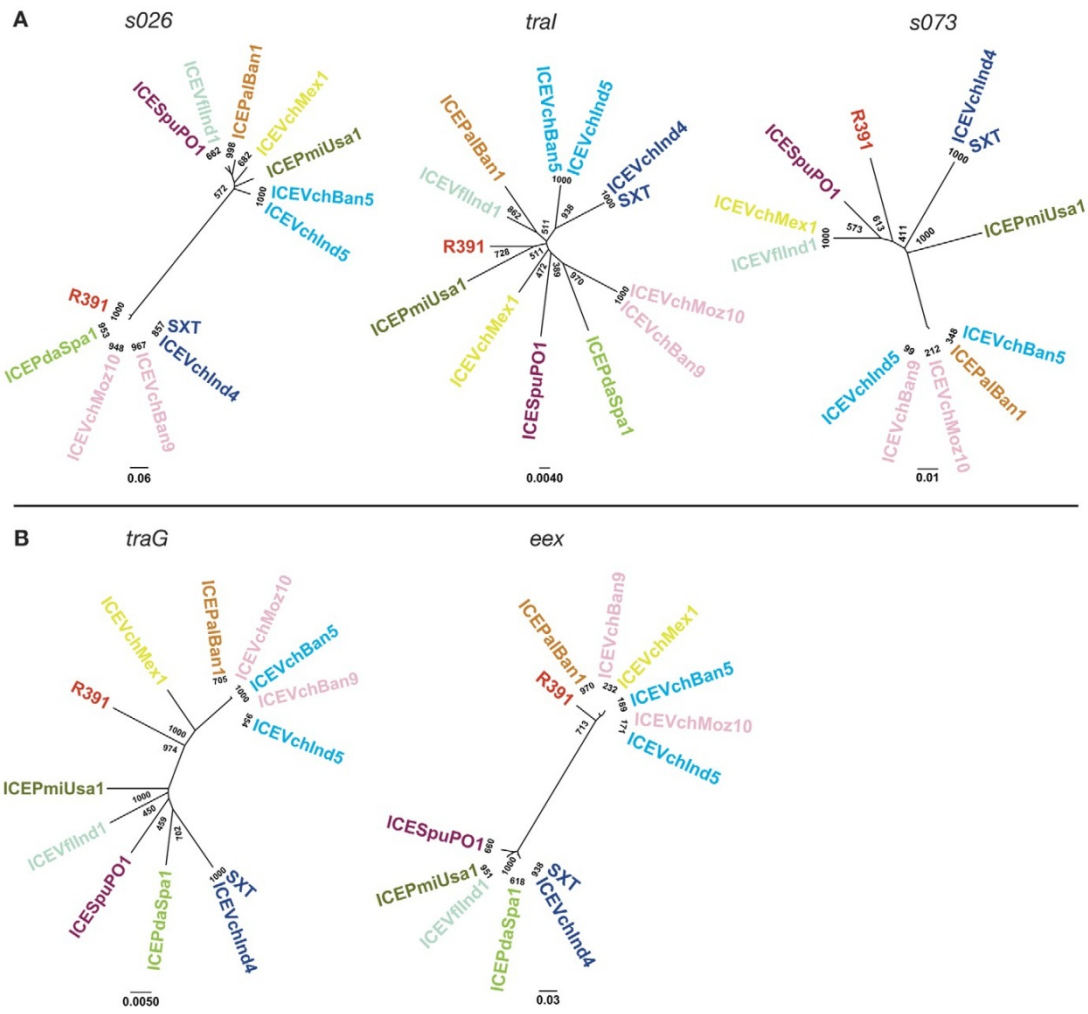


Figure 6. Phylogenetic analysis of several core ICE genes. Nucleotide sequences of the indicated core genes were used to generate the phylogenetic trees shown. Bootstrap values are indicated at branch points. The individual scale bars represent genetic distances and reflect the number of substitutions per residue.

Unlike most core genes, the trees for *traG* and *eex* were similar. In these two trees, the ICEs segregate into two evolutionarily distinct groups (Fig. 6B), confirming and extending previous observations that revealed that there are two groups of *eex* and *traG* sequences in SXT/R391 ICEs [53]. These two groups correspond to the two functional SXT/R391 ICE exclusion groups. Interactions between *traG* and *eex* of the same group mediate ICE exclusion [54]. Thus, the identical 2 clusters of *traG* and *eex* sequences observed in their respective trees reveals the co-evolution of the *traG*/*eex* functional unit. The two groups of *eex* sequences can also be observed in figure 5 where the bifurcating pattern reveals the 2 exclusion groups. This pattern is difficult to discern for *traG*, perhaps because of the large size of this multi-functional gene.

ICE*Vch*Ban8, an SXT-like ICE that lacks Int_{sxt}

The sequence of ICE*Vch*Ban8, which was derived from a non-O1, non-O139 *V. cholerae* strain, is incomplete but it appears to contain 49 out of 52 SXT/R391 core genes. However, since this strain lacks Int_{sxt} it was not included in our comparative analyses above. It is not known if ICE*Vch*Ban8 is capable of excision or transmission; however, it contains a P4-like integrase and a putative *xis*. It is tempting to speculate that the genome of ICE*Vch*Ban8 provides an illustration of how acquisition (presumably via recombination) of a new integration/excision module may generate a novel ICE family.

Perspectives

Comparative analysis of the genomes of the 13 SXT/R391 ICEs studied here has greatly refined our understanding of this group of mobile genetic elements. These elements, which have been isolated from 4 continents and the depths of the Pacific Ocean, all have an identical genetic structure, consisting of the same syntenous set of 52 conserved core genes that are interrupted by clusters of diverse variable genes. All the elements have insertions of variable DNA segments in the same five intergenic hotspots that interrupt the conserved backbone. Furthermore, some of the elements have additional insertions outside the hotspots; however, in all cases the acquistion of variable DNA has not compromised the integrity of the core genes required for ICE mobility. Functional analyses revealed that less than half of the conserved genes are necessary for ICE transmissibility and the contributions of the 27 core genes of unknown function to ICE fitness remains an open question. Finally, several observations presented here suggest that recombination between SXT/R391 ICEs has been a major force in shaping the genomes of this widespread family of mobile elements.

Although comparisons of the 13 ICE genomes analyzed here strongly suggest that these mobile elements have undergone extensive recombination during their evolutionary histories, there is a remarkable degree of similarity among the SXT/R391 ICEs. All of these ICEs consist of the same syntenous and nearly identical 52 genes. In contrast, other families of closely related mobile elements, such as lambdoid or T4-like phages for example, exhibit greater diversity [55,56]. Since the elements that we sequenced were isolated from several different host species and from diverse locations, the great degree of similarity of the SXT/R391 ICE family does not likely reflect

bias in the elements that we sequenced. It is possible that this family of mobile elements is a relatively recent creation of evolution and has yet to undergo significant diversification.

To date, relatively few formal comparative genomic analyses of other ICE families have been reported. Mohd-Zain *et al* [11] identified several diverse ICEs and genomic islands that shared a largely syntenous set of core genes with ICE*Hin*1056, an ICE originally identified in *Haemophilus influenzae*. However, even though these elements share a similar genomic organization, they exhibit far greater variability in the sites of insertion of variable DNA and in the degree of conservation in their core genes compared to SXT/R391 ICEs. Thus, although this group of elements appears to share a common ancestor, they seem to have diverged earlier in evolutionary history than the SXT/R391 ICEs. However, when comparative genomic analyses were restricted to ICEHin1056-related ICEs found in only two *Haemophilus* sp., Juhas *et al* found that, like the SXT/R391 family of ICEs, these 7 ICE*Hin*1056-related ICEs share greater than 90% similarity at the DNA level in their nearly syntenous set of core genes [12]. It will be interesting to learn the extent of conservation of genetic structure and DNA sequence in additional ICE families to obtain a wider perspective on ICE evolution.

Comparative genomic studies of bacteriophages have led to the idea that the full range of phage sequences are part of common but extremely diverse gene pool [57,58]. The SXT/R391 ICE genomes suggest that there may be an even larger network of phylogenetic relationships linking sequences found in all types of mobile genetic elements including phages, plasmids, ICEs and transposons. The genomes of SXT/R391 ICEs appear to be amalgams of genes commonly associated with other types of mobile elements. Many of the ICE core genes are usually associated with phages, such as *int*, *bet*, *exo* and *setR*, or with plasmids, such as the *tra* genes. Additionally, the SXT/R391 ICEs and IncA/C plasmids clearly have a common ancestor, as we found that the entire set of SXT/R391 *tra* genes are also present in IncA/C plasmids. Thus, the genes present in all types of mobile genetic elements appear to contribute to a common gene pool from which novel variants of particular elements (such as ICE*Vch*Ban8) or perhaps even novel types of mobile genetic elements can arise.

Materials and Methods

ICE Sequencing

ICE*Pal*Ban1, ICE*Vch*Mex1, ICE*Vch*Ind4, ICE*Vch*Ind5 and ICE*Vch*Ban5 were isolated using the plasmid capture system described in figure 1. The SXT chromosomal attachment sequence, *attB*, was introduced into the modified F plasmid pXX704 [34] to create pIceCap. This plasmid was then introduced into a Δ*prfC* derivative of the TcR *E. coli* strain CAG18439. Exconjugants derived from matings between this strain and those harboring the 5 ICEs listed above resulted in strains carrying a pIceCap::ICE plasmid. Once captured, the plasmids were isolated using the Qiagen plasmid midi kit for low-copy plasmids (Qiagen). Isolated pIceCap::ICE plasmids were then sequenced.

ICE*Vfl*Ind1 genome was determined by sequencing several overlapping cosmids that encompassed this ICE's genome. Briefly, genomic DNA from a *Vibrio fluvialis* strain carrying ICE*Vfl*Ind1 was prepared using the GNome DNA kit (QBIOgene). Sau3A1 restricted genomic DNA was used to

create a SuperCos1 (Stratagene)-based cosmid library according the manufacturer's instructions. The library was subsequently screened for cosmids containing ICE-specific sequences using PCR with primers to conserved core ICE sequences. Four cosmids containing overlapping ICE*Vfl*Ind1 sequences were identified and sequenced.

The genomes of 6 ICEs were sequenced by the Sanger random shotgun method [59]. Briefly, small insert plasmid libraries (2–3 kb) were constructed by random nebulization and cloning of pIceCap::ICE DNA or of cosmid DNA for ICE*Vfl*Ind1. In the initial random sequencing phase, 8-12 fold sequence coverage was achieved. The sequences of either pIceCap or pSuperCos were subtracted and the remaining sequences were assembled using the Celera Assembler [60]. An initial set of open reading frames (ORFs) that likely encode proteins was identified using GLIMMER [61], and those shorter than 90 base pairs (bp) as well as some of those with overlaps eliminated.

Bioinformatics

Nucleotide and amino acid conservation were assessed with the appropriate BLAST algorithms. ICEs were aligned using clustalW with default settings [62]. MAUVE [37] and LAGAN [38] were used to identify core genes in figure 2. To map the boundaries of the hotspots, sequence comparisons were made using MAUVE and then manually compared to find boundaries between conserved and variable DNA as shown in figure 3.

Phylogenetic trees were generated from alignments of nucleotide sequences using the neighbor-joining method as implemented by ClustalX software, version 2.011 [63]. The reliability of each tree was subjected to a bootstrap test with 1000 replications. Trees were edited using FigTree 1.22 (http://tree.bio.ed.ac.uk/software/figtree/).

Generation and testing of SXT deletion mutants

CAG81439 harboring SXT was used as the host strain to create the SXT deletion mutants shown in figure 3; the deletions were constructed using one-step gene inactivation as previously described [43,64]. The primers used to create the deletion mutants are available upon request. Matings were conducted as previously described [16,43] using deletion mutants and a KnR *E. coli* recipient, CAG18420. Exconjugants were selected on LB agar plates containing chloramphenicol, 20µg/ml (for SXT selection) and kanamycin, 50 µg/ml. The frequency of exconjugant formation was calculated by dividing the number of exconjugants by the number of donors.

Acknowledgements

We thank Brigid Davis and Frederique LeRoux for helpful comments on the manuscript. We thank Robert Hall for facilitating the work described here and Yoshiharu Yamaichi for suggesting the design of pIceCap.

References

1. Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3: 722-732.
2. Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol 3: 679-687.
3. Lawrence JG, Hendrickson H (2005) Genome evolution in bacteria: order beneath chaos. Curr Opin Microbiol 8: 572-578.
4. Ochman H, Lerat E, Daubin V (2005) Examining bacterial species under the specter of gene transfer and exchange. Proc Natl Acad Sci U S A 102 Suppl 1: 6595-6599.
5. Beaber JW, Hochhut B, Waldor MK (2004) SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature 427: 72-74.
6. Burrus V, Waldor MK (2003) Control of SXT integration and excision. J Bacteriol 185: 5045-5054.
7. Auchtung JM, Lee CA, Monson RE, Lehman AP, Grossman AD (2005) Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. Proc Natl Acad Sci U S A 102: 12554-12559.
8. Caparon MG, Scott JR (1989) Excision and insertion of the conjugative transposon Tn916 involves a novel recombination mechanism. Cell 59: 1027-1034.
9. Cheng Q, Paszkiet BJ, Shoemaker NB, Gardner JF, Salyers AA (2000) Integration and excision of a *Bacteroides* conjugative transposon, CTnDOT. J Bacteriol 182:4035-4043.
10. Hagege J, Pernodet JL, Friedmann A, Guerineau M (1993) Mode and origin of replication of pSAM2, a conjugative integrating element of *Streptomyces ambofaciens*. Mol Microbiol 10: 799-812.
11. Mohd-Zain Z, Turner SL, Cerdeno-Tarraga AM, Lilley AK, Inzana TJ, et al. (2004). Transferable antibiotic resistance elements in *Haemophilus influenzae* share a common evolutionary origin with a diverse family of syntenic genomic islands. J Bacteriol 186: 8114-8122.
12. Juhas M, Power PM, Harding RM, Ferguson DJ, Dimopoulou ID, et al. (2007) Sequence and functional analyses of *Haemophilus* spp. genomic islands. Genome Biol 8: R237.
13. Burrus V, Waldor MK (2004) Shaping bacterial genomes with integrative and conjugative elements. Res Microbiol 155: 376-386.
14. Boltner D, MacMahon C, Pembroke JT, Strike P, Osborn AM (2002) R391: a conjugative integrating mosaic comprised of phage, plasmid, and transposon elements. J Bacteriol 184: 5158-5169.
15. Whittle G, Shoemaker NB, Salyers AA (2002) The role of *Bacteroides* conjugative transposons in the dissemination of antibiotic resistance genes. Cell Mol Life Sci 59: 2044-2054.
16. Waldor MK, Tschape H, Mekalanos JJ (1996) A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. J Bacteriol 178: 4157-4165.
17. Shoemaker NB, Barber RD, Salyers AA (1989) Cloning and characterization of a *Bacteroides* conjugal tetracycline-erythromycin resistance element by using a shuttle cosmid vector. J Bacteriol 171: 1294-1302.
18. Franke AE, Clewell DB (1981) Evidence for a chromosome-borne resistance transposon (Tn916) in Streptococcus faecalis that is capable of "conjugal" transfer in the absence of a conjugative plasmid. J Bacteriol 145: 494-502.
19. Ravatn R, Studer S, Springael D, Zehnder AJ, van der Meer JR (1998) Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from Pseudomonas sp. Strain B13. J Bacteriol 180:

4360-4369.

20. Sullivan JT, Ronson CW (1998) Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc Natl Acad Sci U S A 95: 5145-5149.
21. Cholera Working Group (1993) Large epidemic of cholera-like disease in Bangladesh caused by *Vibrio cholerae* O139 synonym Bengal. Lancet 342: 387-390.
22. Ahmed AM, Shinoda S, Shimamoto T (2005) A variant type of *Vibrio cholerae* SXT element in a multidrug-resistant strain of *Vibrio fluvialis*. FEMS Microbiol Lett 242: 241-247.
23. Osorio CR, Marrero J, Wozniak RA, Lemos ML, Burrus V, et al. (2008) Genomic and functional analysis of ICE*Pda*Spa1, a fish-pathogen-derived SXT-related integrating conjugative element that can mobilize a virulence plasmid. J Bacteriol 190: 3353-3361.
24. Pembroke JT, Piterina AV (2006) A novel ICE in the genome of *Shewanella putrefaciens* W3-18-1: comparison with the SXT/R391 ICE-like elements. FEMS Microbiol Lett 264: 80-88.
25. Hochhut B, Lotfi Y, Mazel D, Faruque SM, Woodgate R, et al. (2001) Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins. Antimicrob Agents Chemother 45: 2991-3000.
26. Hochhut B, Beaber JW, Woodgate R, Waldor MK (2001) Formation of chromosomal tandem arrays of the SXT element and R391, two conjugative chromosomally integrating elements that share an attachment site. J Bacteriol 183: 1124-1132.
27. Coetzee JN, Datta N, Hedges RW (1972) R factors from *Proteus rettgeri*. J Gen Microbiol 72: 543-552.
28. Burrus V, Marrero J, Waldor MK (2006) The current ICE age: biology and evolution of SXT-related integrating conjugative elements. Plasmid 55: 173-183.
29. Hochhut B, Waldor MK (1999) Site-specific integration of the conjugal *Vibrio cholerae* SXT element into prfC. Mol Microbiol 32: 99-110.
30. Maeda K, Nojiri H, Shintani M, Yoshida T, Habe H, et al. (2003) Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain CA10 indicates its mosaicity and the presence of large catabolic transposon Tn4676. J Mol Biol 326: 21-33.
31. Murata T, Ohnishi M, Ara T, Kaneko J, Han CG, et al. (2002) Complete nucleotide sequence of plasmid Rts1: implications for evolution of large plasmid genomes. J Bacteriol 184: 3194-3202.
32. Beaber JW, Hochhut B, Waldor MK (2002) Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. J Bacteriol 184: 4259-4269.
33. Beaber JW, Burrus V, Hochhut B, Waldor MK (2002) Comparison of SXT and R391, two conjugative integrating elements: definition of a genetic backbone for the mobilization of resistance determinants. Cell Mol Life Sci 59: 2065-2070.
34. Niki H, Hiraga S (1997) Subcellular distribution of actively partitioning F plasmid during the cell division cycle in E. coli. Cell 90: 951-957.
35. Yamaichi Y, Niki H (2004) migS, a cis-acting site that affects bipolar positioning of oriC on the Escherichia coli chromosome. Embo J 23: 221-233.
36. Faruque SM, Tam VC, Chowdhury N, Diraphat P, Dziejman M, et al. (2007). Genomic analysis of the Mozambique strain of *Vibrio cholerae* O1 reveals the origin of El Tor strains carrying classical CTX prophage. Proc Natl Acad Sci U S A 104: 5151-5156.
37. Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14: 1394-1403.
38. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, et al. (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res 13: 721-731.

39. Ceccarelli D, Daccord A, Rene M, Burrus V (2008) Identification of the origin of transfer (oriT) and a new gene required for mobilization of the SXT/R391 family of integrating conjugative elements. J Bacteriol 190: 5328-5338.
40. Toleman MA, Bennett PM, Walsh TR (2006) ISCR elements: novel gene-capturing systems of the 21st century? Microbiol Mol Biol Rev 70: 296-316.
41. Jenal U, Malone J (2006) Mechanisms of cyclic-di-GMP signaling in bacteria. Annu Rev Genet 40: 385-407.
42. Romling U, Gomelsky M, Galperin MY (2005) C-di-GMP: the dawning of a novel bacterial signalling system. Mol Microbiol 57: 629-639.
43. Wozniak RA, Waldor MK (2009) A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. PLoS Genet 5: e1000439.
44. Lindsey RL, Fedorka-Cray PJ, Frye JG, Meinersmann RJ (2009) Inc A/C plasmids are prevalent in multidrug-resistant *Salmonella enterica* isolates. Appl Environ Microbiol 75: 1908-1915.
45. Llanes C, Gabant P, Couturier M, Bayer L, Plesiat P (1996) Molecular analysis of the replication elements of the broad-host-range RepA/C replicon. Plasmid 36: 26-35.
46. Galimand M, Guiyoule A, Gerbaud G, Rasoamanana B, Chanteau S, et al. (1997) Multidrug resistance in *Yersinia pestis* mediated by a transferable plasmid. N Engl J Med 337: 677-680.
47. Welch TJ, Fricke WF, McDermott PF, White DG, Rosso ML, et al. (2007) Multiple antimicrobial resistance in plague: an emerging public health risk. PLoS One 2:e309.
48. Pan JC, Ye R, Wang HQ, Xiang HQ, Zhang W, et al. (2008) Vibrio cholerae O139 multiple-drug resistance mediated by *Yersinia pestis* pIP1202-like conjugative plasmids. Antimicrob Agents Chemother 52: 3829-3836.
49. Fricke WF, Welch TJ, McDermott PF, Mammel MK, LeClerc JE, et al. (2009) Comparative genomics of the IncA/C multidrug resistance plasmid family. J Bacteriol 191: 4750-4757.
50. Elton TC, Holland SJ, Frost LS, Hazes B (2005) F-like type IV secretion systems encode proteins with thioredoxin folds that are putative DsbC homologues. J Bacteriol 187: 8267-8277.
51. Beaber JW, Waldor MK (2004) Identification of operators and promoters that control SXT conjugative transfer. J Bacteriol 186: 5945-5949.
52. Burrus V, Waldor MK (2004) Formation of SXT tandem arrays and SXT-R391 hybrids. J Bacteriol 186: 2636-2645.
53. Marrero J, Waldor MK (2007) The SXT/R391 family of integrative conjugative elements is composed of two exclusion groups. J Bacteriol 189: 3302-3305.
54. Marrero J, Waldor MK (2005) Interactions between inner membrane proteins in donor and recipient cells limit conjugal DNA transfer. Dev Cell 8: 963-970.
55. Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, et al. (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J Mol Biol 299: 27-51.
56. Filee J, Bapteste E, Susko E, Krisch HM (2006) A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. Mol Biol Evol 23: 1688-1696.
57. Hendrix RW, Hatfull GF, Smith MC (2003) Bacteriophages with tails: chasing their origins and evolution. Res Microbiol 154: 253-257.
58. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc Natl Acad Sci U S A 96: 2192-2197.
59. Fouts DE, Mongodin EF, Mandrell RE, Miller WG, Rasko DA, et al. (2005) Major structural differences

and novel potential virulence mechanisms from the genomes of multiple campylobacter species. PLoS Biol 3: e15.

60. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, et al. (2000) A whole- genome assembly of *Drosophila*. Science 287: 2196-2204.
61. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27: 4636-4641.
62. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.
63. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.
64. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97: 6640-6645.
65. Burrus V, Quezada-Calvillo R, Marrero J, Waldor MK (2006) SXT-related integrating conjugative element in New World *Vibrio cholerae*. Appl Environ Microbiol 72: 3054-3057.
66. Das B, Halder K, Pal P, Bhadra RK (2007) Small chromosomal integration site of classical CTX prophage in Mozambique *Vibrio cholerae* O1 biotype El Tor strain. Arch Microbiol 188: 677-683.
67. Pearson MM, Sebaihia M, Churcher C, Quail MA, Seshasayee AS, et al. (2008) Complete genome sequence of uropathogenic *Proteus mirabilis*, a master of both adherence and motility. J Bacteriol 190: 4027-4037.
68. Nair GB, Faruque SM, Bhuiyan NA, Kamruzzaman M, Siddique AK, et al. (2002) New variants of *Vibrio cholerae* O1 biotype El Tor with attributes of the classical biotype from hospitalized patients with acute diarrhea in Bangladesh. J Clin Microbiol 40: 3296-3299.

CHAPTER III

ICE*Vch*Ind5 is prevalent in epidemic *Vibrio cholerae* O1 El Tor strains isolated in India

Daniela Ceccarelli[a,*], Matteo Spagnoletti[a,*], Donatella Bacciu[b], Yael Danin-Poleg[c], Deepak K. Mendiratta[d], Yechezkel Kashi[c], Piero Cappuccinelli[b], Vincent Burrus[e], Mauro M. Colombo[a]

***These authors contributed equally to this work**

a Dipartimento di Biologia e Biotecnologie Charles Darwin, Sapienza Università di Roma, Via dei Sardi 70, 00185 Rome, Italy **b** Dipartimento di Scienze Biomediche, Università di Sassari, Viale San Pietro 43b, 07100 Sassari, Italy **c** Department of Biotechnology and Food Engineering, Technion-Israel Institute of Technology, Haifa 3200, Israel **d** Department of Microbiology, Mahatma Gandhi Institute of Medical Sciences, Sevagram, 442102 MS, India **e** Département de Biologie, Université de Sherbrooke, 2500 Boul. Université, J1K2R1 Sherbrooke, QC, Canada

Abstract

Integrating Conjugative Elements (ICEs) of the SXT/R391 family are self-transmissible mobile elements mainly involved in antibiotic resistance spread among γ-proteobacteria, including *V. cholerae*. We demonstrated that the recently described ICE*Vch*Ind5 is prevailing in *V. cholerae* O1 clinical strains isolated between 1994 and 2005 in Wardha province (Maharashtra, India). Genetic characterization by ribotyping and multiple-locus SSR analysis proved the same clonal origin for *V. cholerae* O1 isolates in Wardha province over an 11 year period, and was used to assess the correlation between strain and ICE content among ours and different Indian reference strains. In *silico* analysis showed the existence of at least three sibling ICEs of ICE*Vch*Ind5 in *V. cholerae* O1 El Tor reference strains, isolated after 1992 in the Indian Subcontinent.

This chapter was published in International Journal of Medical Microbiology (2010)
doi:10.1016/j.ijmm.2010.11.005

Introduction

ICEs are self-transmissible mobile elements, able to integrate into the host bacterial chromosome and to transfer by conjugation [1]. They are recognized for their important role in genome plasticity of many Gram-positive and Gram-negative bacteria [2,3]. In the last decade, the SXT/R391 family of ICEs has been the subject of a growing number of functional and epidemiological studies [4-6]. The name of the family originates from the elements SXT^{MO10} and R391 found in clinical strains of *Vibrio cholerae* and *Providencia rettgeri*, respectively. SXT^{MO10} was identified in India and confers resistance to chloramphenicol, streptomycin, sulfamethoxazole, and trimethoprim [7]. R391 was isolated in South Africa and confers resistance to kanamycin and mercury [8]. SXT^{MO10} and R391 share a highly conserved genetic backbone, coding for their integration/excision, conjugative transfer, and regulation [9]. Scattered in the conserved sequence and localized in five intergenic regions considered as hotspots for acquisition/integration of foreign DNA are ICE-specific genes conferring a number of properties upon the host bacteria, including resistance to antibiotics and heavy metals, new toxin/antitoxin systems, restriction/modification systems, and alternative metabolic pathways [9-11]. More than 50 ICEs, ranging in size from 82 to 108 kb [4] were identified and grouped within the SXT/R391 family to date, 30 of them in clinical and environmental *V. cholerae* strains [4,9].

Cholera has been endemic in the Ganges and Brahmaputra belt of eastern India and Bangladesh from very early times [12]. In a condition of dynamic equilibrium, the two epidemic serogroups O1 and O139 coexist in the cholera-prone areas of the Indian Subcontinent since 1992 [13], when O139 emerged as new epidemic serogroup [14]. One of the main differences between the two serogroups was the presence of SXT^{MO10} in O139 strains [7]. This distinction was rapidly leveled by the emergence of a new O1 clone [15] characterized by ICE*Vch*Ind1, the first ICE of the SXT/R391 family described in serogroup O1 [16]. To date 15 SXT-related ICEs isolated in India or Bangladesh between 1992 and 2001 have been identified, and six of them completely sequenced and annotated [4]: SXT^{MO10} (Chennai, Tamil Nadu), ICE*Vch*Ind4 (Kolkata, West Bengal), ICE*Vch*Ban5 (Bangladesh), ICE*Vch*Ban10 (Dhaka, Bangladesh), ICE*Vch*Ban9 (Matlab, Bangladesh), and ICE*Vch*Ind5 (Wardha, Maharashtra).

The genome of ICE*Vch*Ind5, an ICE isolated from a *V. cholerae* O1 strain belonging to our collection comprises elements shared with SXT^{MO10}, R391, ICE*Vch*Ban9 and ICE*Pda*Spa1, alongside some unique insertions: (i) a gene coding for a protein similar to the *E. coli dam*-directed mismatch repair protein MutL (Variable Region 2); (ii) a possible transposon of the IS*21* family (Hotspot 4); and (iii) a 14.8 kb hypothetical operon of unknown function (Hotspot 5) [4]. ICE*Vch*Ind5 belongs to the R1 exclusion group, which is consistent with previous data showing that this group includes only *V. cholerae* O1 clinical strains from India and Bangladesh isolated between 1994 and 2005 [17]. In this study we analysed a collection of *V. cholerae* O1 clinical strains isolated from 1994 to 2005 in the Indian province of Wardha (State of Maharashtra) where ICE*Vch*Ind5 was originally isolated. This region is periodically subjected to recurrent cholera epidemics especially during summer, due to heavy rains associated with monsoons. Like the rest of India, Wardha province was confronted with O139 cholera epidemics in the first half of the 90s [18,19]. Yet, by the end of

the decade the O1 El Tor serogroup prevailed over O139, re-establishing itself as the main cause of cholera in this area [20,21]. In order to combine the diffusion of ICE*Vch*Ind5 with *V. cholerae* epidemic circulation we tracked the strains from Wardha and various Indian O1 reference strains by ribotyping and SSR typing. Ribotyping of 16S/23S has been extensively applied for molecular epidemiology [22], and in the past few years simple sequence repeat analysis (SSR, VNTR- variable number of tandem repeats) has provided a highly discriminative detailed typing that was efficiently used for genotyping a number of pathogens including *V. cholerae* [23].

Strain	Serogroup	Year of isolation	Resistances in addition to common profile[a]	Ribotype	ICE
VC7452	O1	1994	STR, NAL	R1	ICE*Vch*Ind5[b]
VC8909	O1	1997	STR, NAL	R1	ICE*Vch*Ind5
VC8636	O1	1997	STR, NAL	ND	ICE*Vch*Ind5
VC9223	O1	1997	STR, NAL	ND	ICE*Vch*Ind5
VC9906	O1	1998	NAL	R1	ICE*Vch*Ind5
VC6884	O1	1998	STR, NAL	ND	ICE*Vch*Ind5
VC9227	O1	1998	STR, NAL	ND	ICE*Vch*Ind5
VC8677	O1	1998	STR, NAL	ND	ICE*Vch*Ind5
VC10693	O1	1998	STR, NAL	R1	ICE*Vch*Ind5
VC7070	O1	1998	STR, NAL	ND	ICE*Vch*Ind5
VC15699	O1	1999	STR, NAL	R1	ICE*Vch*Ind5
VC9258	O1	1999	STR, NAL	R1	ICE*Vch*Ind5
VC13672	O1	2000	STR, NAL	ND	ICE*Vch*Ind5
VC13571	O1	2000	STR, NAL	R1	ICE*Vch*Ind5
VC11675	O1	2001	STR, NAL	R1	ICE*Vch*Ind5
VC12516	O1	2002	STR, NAL	R1	ICE*Vch*Ind5
VC14100	O1	2002	STR, NAL	R1	ICE*Vch*Ind5
VC16770	O1	2003	STR, NAL	R1	ICE*Vch*Ind5
VC16102	O1	2003	STR, NAL	ND	ICE*Vch*Ind5
VC18245	O1	2004	STR, NAL	R3	ICE*Vch*Ind5
VC24531	O1	2004	STR, NAL	ND	ICE*Vch*Ind5
VC25249	O1	2005	STR, NAL	R1	ICE*Vch*Ind5
VC23701	O1	2005	STR, NAL	ND	ICE*Vch*Ind5
VC25959	O1	2005	STR, NAL	ND	ICE*Vch*Ind5
VC24585	O1	2005	STR, NAL	ND	ICE*Vch*Ind5
VC11020	non-O1, non-O139	1997	–	R2	–
VC11454	non-O1, non-O139	1998	–	R2	–
VC11529	non-O1, non-O139	1998	STR, NAL	R2	–
VC22417	non-O1, non-O139	2005	–	R2	–
VC21294	non-O1, non-O139	2005	NAL	R2	–
Reference strains used in ribotype analysis					
CO840	O1	1995 (Kolkata)	SPT[c]	R1	ICE*Vch*Ind1
VC20	O1	1989 (Kolkata)	SPT[c]	R5	–
MO10	O139	1992 (Chennai)	CHL, STR, SXT[c]	R6	SXT^{MO10}
SG24	O139	1992 (Kolkata)	CHL, STR, SXT[c]	R4	SXT^{MO10}

Table 1. Indian *V. cholerae* strains used in this study. [a]Common profile: AMP, PEN, SPT, SXT. [b]ICE fully sequenced [4]. [c]Common profile absent. AMP, ampicillin; CHL, chloramphenicol; NAL, nalidixic acid; PEN, penicillin; STR, streptomycin; SPT, spectinomycin; SXT, sulfamethoxazole-trimethoprim; ND, not determined

Materials and Methods

Bacterial strains and drug susceptibility

We analyzed 25 *V. cholerae* O1 and 5 non-O1, non-O139 clinical strains isolated between 1994 and 2005 at the Kasturba Hospital-Mahatma Gandhi Institute of Medical Sciences (MGIMS), situated in the rural area of Sevagram (Wardha province, Maharashtra - India) (Table 1). Strains were selected from an already-existing collection at the Department of Microbiology of the MGIMS [24]. Strains were isolated from stool samples and/or rectal swabs from patients, and after isolation on thiosulfate citrate bile sucrose agar and biochemical identification, bacterial strains were routinely grown in Luria-Bertani (LB) or agar plates at 37°C and maintained at -80° in LB broth containing 30% (vol/vol) glycerol. Antimicrobial susceptibility was tested at the following concentrations:

ampicillin (AMP), 100 µg/ml; chloramphenicol (CHL), 20 µg/ml; kanamycin (KAN), 50 µg/ml; nalidixic acid (NAL), 40 µg/ml; rifampin (RIF), 100 µg/ml; spectinomycin (SPT), 50 µg/ml; streptomycin (STR), 50 µg/ml; sulfamethoxazole (SUL), 160 µg/ml; tetracycline (TET), 12 µg/ml; and trimethoprim (TMP), 32 µg/ml. Strains *V. cholerae* O139 MO10 [7], *V. cholerae* O139 SG24 ([25]), *V. cholerae* O1 VC20 [26], *V. cholerae* O1 CO840 [26], *V. cholerae* O1 N16961 [27] and *E. coli* AB1157:R391 [28] were appropriately used as negative or positive controls in PCR experiments for ICE detection and ribotype analysis. Strain *V. cholerae* E4: ICE*Vch*Ind1 [16] was used for the comparative analysis of ICE*Vch*Ind5 and ICE*Vch*Ind1. Thirty-one previously studied clinical and environmental *V. cholerae* isolates [29] (hereafter "reference set") together with *V. cholerae* O139 MO10 [7] were used as reference strains in the genomic SSR analysis.

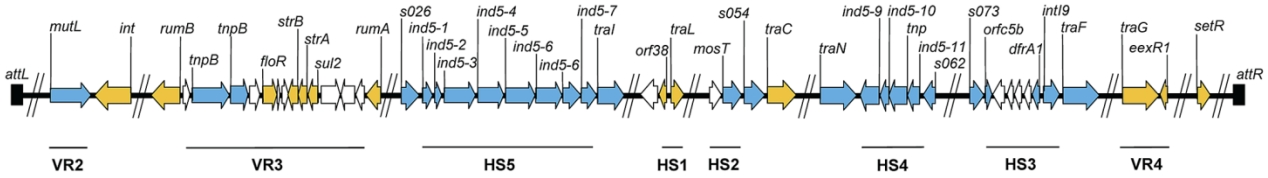


Figure 1. Schematic linear representation (not to scale) of the 37 out of a total of 94 genes amplified by PCR to check ICE*Vch*Ind5 molecular structure. The thick black line represents the conserved backbone of the SXT/R391 family. Genes in orange were tested with primer set A. Genes in blue were tested with primer set B. Genes not tested inside the hotspots are shown in white. The thin black lines below the genes correspond to Hotspots and Variable Regions. GenBank accession no. of the full sequence of ICE*Vch*Ind5 is GQ463142 [4].

Molecular biology procedures

Bacterial DNA for PCR analysis was prepared with a DNeasy Blood & Tissue Kit (Qiagen), as described in the manufacturer's instructions. The amplicons to be sequenced were directly purified from PCR by Nucleospin extract kit (Macherey-Nagel), or extracted from agarose gel by Rapid Gel extraction system (Marligen BioSciences) according to the manufacturer's instructions. DNA sequences were determined by BMR Genomics (Padova, Italy). PCR reactions were set in a 50-µl volume of reaction buffer containing 1 U of *Taq* polymerase as directed by the manufacturer (Promega).

Detection of ICEs of the SXT/R391 family was performed by PCR analysis, using 17 specific primer pairs (set A, orange genes Fig. 1) previously described by our group [22]. We screened for the presence of int_{SXT}, *prfC*/SXT right junction, *floR*, *strA*, *strB*, *sul2*, *dfrA18*, *dfrA1*, *rumAB* operon, *traI*, *traC*, *setR*, and Hotspots or Variable Regions *s026/traI*, *s043/traL*, *traA/s054*, *s073/traF* and *traG/s079*. A second set of 15 new primer pairs (set B) designed on the specific sequence of ICE*Vch*Ind5 (blue genes Fig. 1, and Table 2) were used to detect ICE*Vch*Ind5 specific Hotspots and Variable Regions in all *V. cholerae* O1 strains.

Ribotyping and SSR analysis

Ribotyping of *V. cholerae* strains was performed by *Bgl*I restriction of chromosomal DNA as already described [22] with fluorescent-labeled 16S and 23S rDNA (Gene Images 3540 RPn3510, by Amersham), generated by reverse transcriptase polymerase chain reaction of ribosomal RNAs.

SSR typing was performed using unique primers for six SSR loci (VC0147-$(6)_9$; VC0437-$(7)_7$; VC1457-$(7)_4$; VC1650-$(9)_7$; VCA0171-$(6)_{23}$ and VCA0283-$(6)_{14}$) as previously described [29]. Size variation was performed by Capillary electrophoresis in 3130 Genetic Analyzer and analyzed with GeneMapper-v4.0 (Applied-Biosystems Inc., Foster City, CA, USA). A nonparametric analysis of variations in SSR-size alleles was used for SSR loci. An additional allele (a null allele) was counted when there was no amplification product. Phylogenetic relationships among *V. cholerae* isolates were inferred from SSR data by using the Nei coefficient of association and generating the corresponding matrix with SAS 9.2 (2002-2008 SAS Institute, Inc., Cary, NC). A genetic-distance matrix was generated based on 55 polymorphic points (the sum of alleles across 6 SSR loci) to create a dendrogram based on the UPGMA method. Genetic relationships are based on unweighted pair group method with arithmetic mean cluster analysis of VNTR variation using MEGA 4 software [30].

Hotspot or variable region	Target	Primer	Nucleotide sequence (5′ to 3′)	Positions (GenBank accession no. GQ463142)
Hotspot 5	Intergenic region *s026/vchInd5-1*	s026-F	AGTCACCGAGGCCATACACAA	26456–26476
		Ind51-R	AAAGACTTCCCCCGCCCTAC	27509–27528
	vchInd5-1/vchInd5-2/vchInd5-3	Ind51-F	GCACAGATTTCACAGGACGACA	27362–27383
		Ind53-R	AATTCGTAGAGGTGCCGCC	28453–28471
	Intergenic region *vchInd5-3/vchInd5-4*	Ind53-F	CACCGAAACAACCCCGCAA	31817–31870
		Ind54-R	AGCCGCAAGCTGGATCAA	32864–32881
	Intergenic region *vchInd5-4/vcIind5-5*	Ind54-F	TGGCCTGCCGATTACCCTT	34839–34857
		Ind55-R	TCCGCCTCACCGCTTTCAT	35840–35858
	Intergenic region *vchInd5-5/vchInd5-6*	Ind55-F	ACAACTATCTTTACTTCGCCCGCA	37613–37636
		Ind56-R	CCCAAATACCACCCATCAACATCT	38500–38523
	vchInd5-6/vchInd5-7/vchInd5-8	Ind56-F	GTTGTCTGAGAATGGCGCGA	40013–40032
		Ind58-R	CGTTTAGCCTTTGAGCCCCC	41148–41167
	Intergenic region *vchInd5-8/traI*	Ind58-F	ACAAACCCGAAGACCCGCA	41248–41266
		TraI-R	ACGTAATAAACCACCAGCACCAC	42200–42222
Hotspot 2	Intergenic region *mosT/s054*	most-F	CTGTATGGTGGGAAATTGTGTGCT	53595–53618
		s054-R	TCGGTGGCCGCTTTCTTGT	54714–54732
Hotspot 4	Intergenic region *traN/vchInd5-9*	traN-F	AAATGAACGCCAAGCGGGT	63545–63563
		Ind59-R	GGAAGAAAGGCAGCCAGAGAG	64582–64602
	vchInd5-9/vchInd5-10/tnp	Ind59-F	AAGTCAAGCACAGCGCCAT	65642–65660
		tnp-R	CCCCCTTCCAAACCTTGCCA	66822–66841
	tnp/vchInd5-11	tnp-F	CGCTGTTTGCGTGGCTTTTT	67801–67820
		Ind511-R	TTTGGTTTGCTGGGGGCTGT	68787–68806
	Intergenic region *vchInd5-11/s062*	Ind511-F	ACAGCCCCCAGCAAACCAA	68787–68805
		s062-R	AACCAAGGAAGCAACCGAAACGA	69850–69872
Hotspot 3	Intergenic region *s062/orfc5b*	s073-F	CGGTAGAGAAACGGGAAGATTGG	81831–81853
		orfc5b-R	GGGCTTGGGGCAATAGGATGA	82826–82846
	Intergenic region *intI9/traF*	intI9-F	AAGAACATGGGGGAACAAGAGG	86139–86160
		traF-R	GGCGGGCAGGATGAATAGAAAGA	87284–87306
Variable Region II	*vchInd5-13*	Orf1-F	CAGAGGGCGGTATTCATTT	1183–1201
		Orf1-R	GTTTTGTTGCGCAGTTCTT	1800–1818

Table 2. DNA sequences and positions of primer set B on ICE*vch*Ind5.

Results

V. cholerae O1 and non-O1, non-O139 strains isolated in Wardha province share a common antibiotic resistance profile

We analyzed a collection of 30 *V. cholerae* clinical strains isolated in a rural hospital situated in Sevagram (Wardha province) from 1994 to 2005. All but five strains were serogroup O1 (see Table 1 for details). All the strains, irrespective of the serogroup, showed sensitivity to rifampin, kanamycin, chloramphenicol and tetracycline. A common multi-resistant profile to ampicillin/penicillin, spectinomycin, sulfamethoxazole and trimethoprim was observed in almost all the strains. Most strains also showed resistance to streptomycin and nalidixic acid (Table 1). Such a complex resistance profile is quite common in Indian *V. cholerae* strains, known to be resistant to cotrimoxazole, beta-lactams, fluoroquinolones and aminoglycosides as previously described ([31,32]).

Same resistance profiles for O1 and non-O1, non-O139 strains could be anticipated, considering the recent emergence of multidrug resistant non-O1, non-O139 isolates ([33]) and taking into consideration that the two groups of strains likely share the same environmental and biological reservoir. Resistances to beta-lactams, fluoroquinolones and aminoglycosides such as spectinomycin are not linked to ICE*Vch*Ind5 [4], therefore they must be related to other gene cassettes either integrated in the chromosome or located in other mobile elements.

ICE*Vch*Ind5 is present in *V. cholerae* O1 clinical strains isolated between 1994 and 2005

As ICEs of the SXT/R391 family are known to be widespread in *V. cholerae* strains we decided to investigate for their presence in our strain collection. Using the highly conserved integrase gene int_{SXT} as a specific ICE marker, all of the *V. cholerae* O1 strains (25 strains) were positive for the presence of this family of ICEs, whereas all of the non-O1, non-O139 (5 strains) were negative. Further testing showed that all int_{SXT}^{+} strains contained the conserved genes *traI*, *traC* and *setR* coding for a putative relaxase, a putative conjugation coupling protein and a transcriptional repressor, respectively [9].

A defining feature of the SXT/R391 group of ICEs is their chromosomal integration into the *prfC* gene [9]. All int_{SXT}^{+} strains were positive for the right junction SXT^{MO10}/*prfC*. Based on these results we included the ICE under study in the SXT/R391 family.

Members of the SXT/R391 family exhibit significant genetic polymorphisms in hotspots interspersed in their common backbone [4]. We used specific primer pairs (set A) designed to discriminate between SXT^{MO10} and R391 [22] on all the twenty-five int^{+} strains. All the strains exhibited the same SXT^{MO10}/R391 hybrid ICE pattern. Intergenic regions *traG/eex* (Variable Region 4) and *traA/s054* (Hotspot 2) showed the same molecular rearrangement as in SXT^{MO10}, whereas region *s043/traL* (Hotspot 1) was organized as in R391. This preliminary approach was insufficient to completely characterize the ICE integrated in the *V. cholerae* O1 strains of our collection, and named ICE*Vch*Ind5. Therefore, this ICE was included in a larger comparative ICE genomic study and it was fully sequenced (GenBank accession no. GQ463142, [4]).

In order to confirm the identity of the ICE present in our strain collection we designed a new set of primers based on the specific sequence of ICE*Vch*Ind5 (set B). Twenty-six different genes were analysed to identify the specific insertions of ICE*Vch*Ind5 in Hotspots 2, 3, 4 and 5 and Variable Regions 2 and 3 (Table 2 and Fig. 1). All the strains tested showed the same ICE profile, providing evidence that ICE*Vch*Ind5 is present in the epidemic area of Wardha in a period of 11 years.

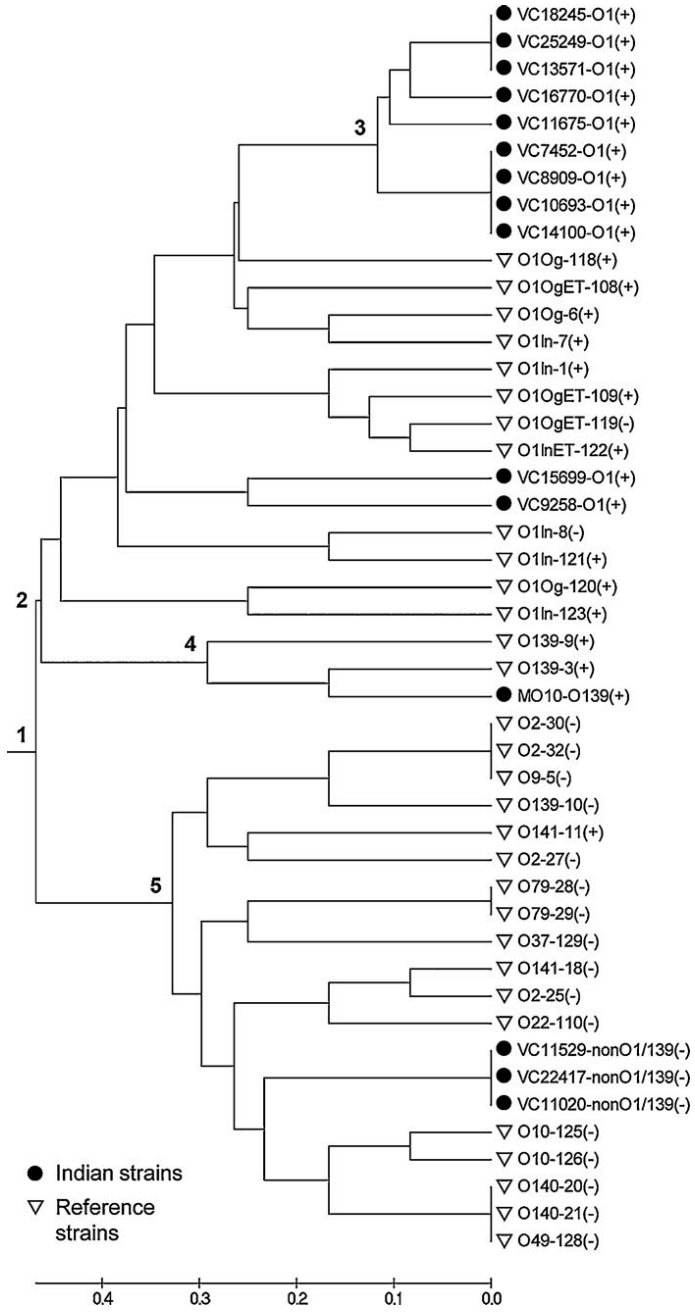


Figure 2 Phylogenetic relationships among 46 *V. cholerae* isolates based on simple-sequences repeat (SSR) variation data at 6 loci. A genetic-distance matrix was generated based on 55 polymorphic points (the sum of alleles across 6 SSR loci). Genetic relationships are based on unweighted pair group method with arithmetic mean cluster analysis of VNTR variation using MEGA 4 software. Scale bar represents genetic distance. (+) and (-) indicate presence or absence of *ctxA.*

ICE*Vch*Ind5 siblings and nomenclature

We included in our strain collection the laboratory strain *V. cholerae* E4:ICE*Vch*Ind1. This strain contains ICE*Vch*Ind1, a reference ICE for the SXT/R391 family, being the first one isolated in *V. cholerae* O1 in India in 1994 [16]. ICE*Vch*Ind1 was never entirely sequenced, but the presence of *intI9* integron in Hotspot 3 was known.

Our previous work of comparative ICE genomics [4] showed that this Hotspot is quite common in ICEs of the SXT/R391 family (i.e. ICE*Vch*Ban5, ICE*Pal*Ban1, ICE*Vch*Ban9, ICE*Vch*Moz10, and ICE*Vch*Ind5). Given this common feature between ICE*Vch*Ind1 and ICE*Vch*Ind5, and their same epidemiological origin, we decided to use primer set B on ICE*Vch*Ind1 to determine its hotspot content. All ICE*Vch*Ind5-specific hotspot insertions were present in ICE*Vch*Ind1.

Fifty-two core genes are present in all the SXT/R391 ICEs analyzed so far [4], spanning ~ 65% of the ICE whole sequence, and our analysis proved ICE*Vch*Ind1 hotspot content to be identical to ICE*Vch*Ind5's. Therefore, although we cannot claim that the two ICEs are identical because we lack ICE*Vch*Ind1 complete sequence, we are confident that the two ICEs share the same gene content.

ICE*Vch*Ind5 was known to be identical to ICE*Vch*Ban5 (GenBank accession no. GQ63140) integrated in a *V. cholerae* O1 strain isolated in Bangladesh in 1998 [4]. By GenBank Blast analysis we also proved ICE*Vch*Ind5 to be identical, with few single nucleotide polymorphisms, to ICE*Vch*Ban10, a SXT-related ICE fully sequenced with the entire genome of *V. cholerae* O1 CIRS101 (GenBank accession no. NZ_ACVW0000000) isolated in Bangladesh in 2002.

Genomic comparison of the four ICEs shows that they are genetically identical but epidemiologically different. The current ICE nomenclature names any newly described ICE according to the species and country of origin and not to its genetic content [9]. We introduce here the term "sibling" for ICEs that share the same genetic content but were isolated in different countries or even species, to help track the genetic content of ICEs without losing the epidemiological information resulting from the accepted nomenclature.

ICE*Vch*Ind5 is associated with a clonal strain of *V. cholerae* O1 from 1994 to 2005

Given the persistence of ICE*Vch*Ind5 in different cholera outbreaks in Wardha province over an 11-year period (1994 to 2005), we decided to verify the clonality of the strains of our collection. We applied two molecular methods able to distinguish among strains within the same species: ribotyping and SSR (VNTR) loci analysis.

Ribotyping was performed on 18 representative strains from various years and serogroups (Table 1). Three *Bgl*I restriction patterns were observed with DNA fragments ranging in size from 2.3 to 22.0 kb (Table 1). Ribotype profile R1 (2.3, 4.2, 5.8, 6.1, 6.3, 8.5, 9.4, 10.8, 22.0 kb) included all *V. cholerae* O1 strains but VC18245 that showed a different profile R3 (0.5, 4.1, 4.2, 4.6 kb). This discrepancy was leveled by the higher discriminative power of the SSR (VNTR) analysis (see below, Fig. 2). Strain CO840 (epidemic *V. cholerae* O1 containing ICE*Vch*Ind1) [26] shared the same ribotype profile R1 with the O1 isolates of this study. This corresponded to pattern RIII which was first described for O1 strains that re-emerged in 1993 after the spread of the O139 serogroup ([34]). Ribotype profile R2 (1.0, 2.3, 4.2, 5.8, 6.0, 9.3, 22.0 kb) included the 5 ICE*Vch*Ind5-free non-

O1, non-O139 *V. cholerae* strains and was quite distinct from those obtained for non-toxigenic strains previously described [35,36], confirming the high variability among these serogroups. Thus, partitioning of *V. cholerae* strains into two main ribotype patterns (R1 and R2) corresponds with the presence or absence of ICE*Vch*Ind5 and with the serogroup of the strains: *V. cholerae* O1 strains containing ICE*Vch*Ind5 belonged to profile R1, and *V. cholerae* non-O1, non-O139 strains lacking ICE*Vch*Ind5 belonged to profile R2.

Beside *V. cholerae* strain CO840, we included in this study three other Indian reference strains: two containing SXT^{MO10} and one devoid of any ICE (Table 1). They gave three different but closely related ribotypes: strain SG24 (*V. cholerae* O139, SXT^{MO10}) gave profile R4 (2.3, 4.2, 4.6, 5.8, 6.0, 6.5 kb); strain VC20 (*V. cholerae* O1, contains class 1 integrons but no ICE) gave profile R5 (2.3, 4.2, 4.6, 5.8, 6.0, 6.5, 9.4, 9.6, 22.0 kb); and strain MO10 (*V. cholerae* O139, SXT^{MO10}) gave profile R6 (2.3, 4.2, 4.6, 5.8, 6.0, 6.5, 9.4, 22.0 kb) differing from profile R4 by a single fragment of 9.4 kb. The discrepancies between profiles R4 and R6 of the two O139 strains was likely to be, given the high variability shown in O139 ribotype isolated in India and Bangladesh between 1992 and 1998 ([37]).

V. cholerae strains were also screened by SSR analysis in order to investigate their fine phylogenetic relationship. Six highly polymorphic L-SSR (VNTR) loci scattered in the genome were selected based on our previous results [29]: VC0147-$(6)_9$, VC0437-$(7)_7$, VC1457-$(7)_4$, VC1650-$(9)_7$, VC0171-$(6)_{23}$, VC0283-$(6)_{14}$. A total of 14 *V. cholerae* isolates from Wardha province were studied and compared to strain *V. cholerae* O139 MO10 and a reference set of 31 clinical or environmental *V. cholerae* strains previously studied and well characterized [29]. High polymorphism was detected among the 46 *V. cholerae* isolates at the level of the tested SSR loci, showing number of alleles ranging from 5 to 14 (Fig. 2). The 46 isolates were discriminated into 34 different SSR types, while the 23 *V. cholerae* O1 isolates collected from various locations were discriminated into 18 different SSR types, reflecting the high variation among O1 isolates.

SSR variation data were used to calculate genetic relationships among the isolates and a genetic distance matrix was generated followed by cluster analysis. The resulting dendrogram showed clear separation between the 3 main serogroups analysed. Indeed, *V. cholerae* O139 isolates from various locations were separated from *V. cholerae* O1 strains and formed a separate cluster (Fig. 2, branch 4). Among the pathogenic *V. cholerae* isolates, both O1 and O139 strains (Fig. 2, branch 2) formed a separate cluster compared to the group of *V. cholerae* non-O1, non-O139 strains (Fig. 2, branch 5), matching our previous data [29].

Most *V. cholerae* O1 El Tor strains containing ICE*Vch*Ind5, presented a distinct cluster (Fig. 2, branch 3) except VC15699 and VC9258, both isolated in 1999. Genetic distances among branch 3 (*V. cholerae* O1 El Tor isolates containing ICE*Vch*Ind5) were rather low (average 0.167 ± 0.119) compared to the high genetic distances found among all O1 isolates (average 0.627 ± 0.259), indicating a clonal origin of these closely related strains. Due to the high mutation rate, L-SSR revealed variations between the 9 strains containing ICE*Vch*Ind5 that clustered into 4 SSR types in branch 3 of the dendrogram (Fig. 2). Our results suggested that *V. cholerae* O1 El Tor isolates containing ICE*Vch*Ind5 belong to a common clonal origin and this finding is in accordance with ribotype analysis showing the same profile R1 for most of these isolates. Although containing ICE*Vch*Ind5 and sharing the same R1 ribotype profile, strains VC15699 and VC9258 were separated

from the main *V. cholerae* O1 El Tor cluster circulating in Wardha, suggesting a different genomic origin or a different genomic background (single point mutations or alteration in the length of the tandem repeats in one of the 6 SSR loci analysed) where ICE*Vch*Ind5 was integrated. Taken together, our results suggest that most *V. cholerae* O1 El Tor isolates containing ICE*Vch*Ind5 originate from a single clone.

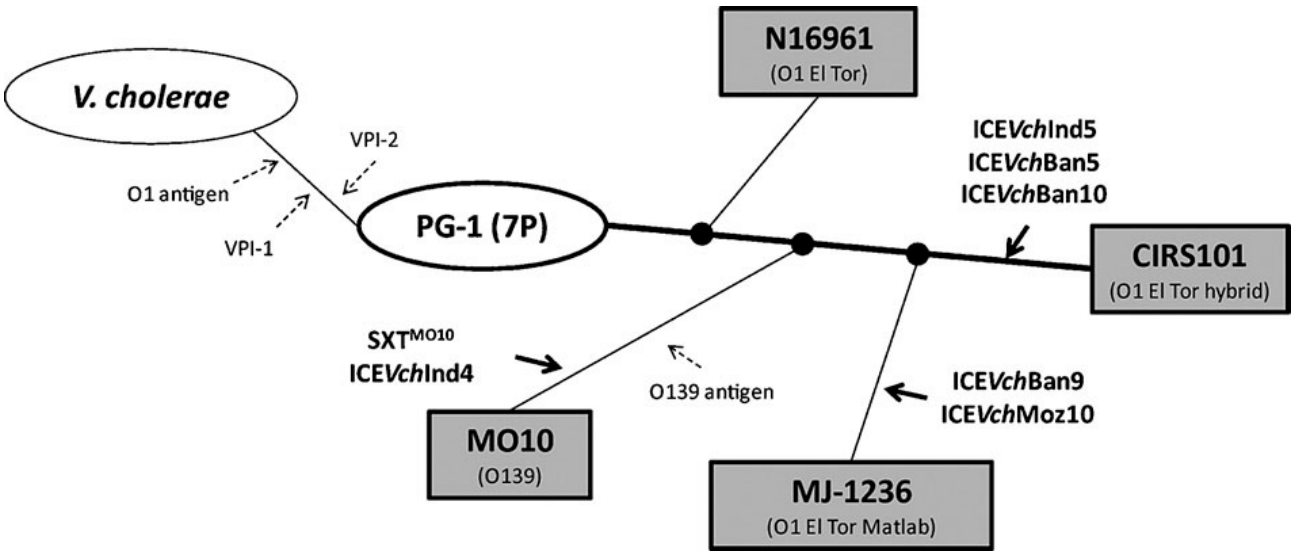


Figure 3. ICE distribution along the hypothetical evolutionary pathway of *V. cholerae* epidemic strains in the Indian Subcontinent (adapted from [38]). 7th cholera pandemic strains belong to a single phyletic line (thick black line) and grey boxes represent completely sequenced *V. cholerae* reference strains isolated in the Indian Subcontinent (GenBank accession no.: N16961, AE003852/AE003853; CIRS101, ACVW00000000; MJ-1236, CP001485/CP001486; MO10, AAKF03000000). Strain serogroup is indicated in brackets and hypothetical ancestral strains by thin open circles. PG-1 (7P): hypothetical 7th pandemic ancestor. Vibrio Pathogenicity Islands and serogroup acquisition are indicated by hatched arrows. ICE acquisition is indicated by solid arrows.

Discussion

The current 7th cholera pandemic is caused by *V. cholerae* strains belonging to a single lineage [38], distinguished either in O139 or O1 El Tor prototype and Matlab variants (Fig. 3) [39]. We analysed a collection of *V. cholerae* O1 strains isolated from 1994 to 2005 in a centrally located province in the state of Maharashtra. Province of Wardha represents a peculiar epidemic area, far from the Ganges-Brahmaputra Delta, the main epidemic source of *V. cholerae* strains in India. Since the beginning of the 90s, most *V. cholerae* epidemic strains such as the emerging O139 serogroup, were associated to ICEs of the SXT/R391 family conferring multiple drug resistance [5,7,22]. We found that ICE*Vch*Ind5 is the only ICE of the SXT/R391 family circulating during an 11-year period in *V. cholerae* O1 strains in the cholera endemic area of Wardha. Complete sequencing of ICE*Vch*Ind5 revealed its composite structure, a new example of the mosaic nature of these elements known for their high recombinogenic potential [4]. ICE*Vch*Ind5 shares genome portions with several other ICEs of the SXT/R391 family, but is also characterized by specific insertions mostly coding for genes of unknown function.

Through the comparison of ICE*Vch*Ind5 with all the other known ICEs circulating in the Indian

Subcontinent, we were able to integrate diffusion and molecular organization of ICEs of the SXT/R391 family with strain epidemiological data. Analysis of these ICEs pointed out the presence of three groups of SXT/R391 ICEs in the genome of epidemic *V. cholerae* strains circulating to date in the Indian Subcontinent. The first group is composed by SXT^{MO10} and ICE*Vch*Ind4, isolated in Chennai and Kolkata (India), respectively. The second group includes three ICEs sharing the same hotspot content with ICE*Vch*Ind5: ICE*Vch*Ind1 (India), ICE*Vch*Ban5 and ICE*Vch*Ban10 (Bangladesh). ICE*Vch*Ban9 suggests the presence of a third group, isolated in the new *V. cholerae* O1 El Tor Matlab variant in Bangladesh [40]. This ICE has no close relative in the Indian Subcontinent, but a sibling ICE*Vch*Moz10 isolated in Mozambique [4].

A closer look at this ICE distribution in the Indian Subcontinent pointed out that different ICEs correspond to different serogroups. Indeed, *V. cholerae* O139 is associated to SXT^{MO10} or siblings; *V. cholerae* O1 El Tor Matlab variant is associated to ICE*Vch*Ban9 or siblings, and *V. cholerae* O1 El Tor hybrid variant is associated to ICE*Vch*Ind5 or siblings. This association was never investigated before and it is worth noting that it reflects a recently proposed classification to describe homologous intraspecific groups of *V. cholerae* [38]. Clinical isolates associated with the current seventh cholera pandemic were grouped on the basis of a full genome alignment. A monophyletic group, designated as the 7th pandemic clade (PG-1) (Fig. 3, a simplified adaptation of the cholera evolutionary pathway described by Chun and colleagues), comprised all *V. cholerae* O1 and O139 epidemic strains analyzed. The proposed hypothetical evolutionary pathway for *V. cholerae* was mainly based on the loss or acquisition of Genomic Islands from a common ancestor to the rising of the contemporary *V. cholerae* pandemic strains [38]. The role of an SXT-like element in the evolutionary process of *V. cholerae* strains was also suggested and our data can confirm this hypothesis. Indeed, the seventh pandemic clade subgroups match the three groups of ICEs circulating in the Indian Subcontinent we observed (Fig. 3): strain MO10, representing the O139 serogroup, contains SXT^{MO10}; strain MJ-1236, representing the O1 El Tor Matlab variant, contains ICE*Vch*Ban9; and strain CIRS101, representing the hybrid O1 El Tor variant, contains ICE*Vch*Ban10. Finally, strain N16961 does not contain any ICE and represents the prototype El Tor variant, a forth group in the main PG-1 clade.

Given all these considerations, the ICE/biotype association we noted in the Indian Subcontinent represents what happens on a larger scale in *V. cholerae* evolutionary scenario. As a consequence, finding ICE*Vch*Ind5 or siblings as the prevalent ICEs associated to O1 El Tor strains circulating in India after 1992, can be reinterpreted. Indeed, *V. cholerae* strains under study share the same clonal origin (ribotype profile R1) with several O1 El Tor strains isolated from Kolkata after 1992 [26]. This clone, which proved to be different from *V. cholerae* O1 strains isolated in Kolkata in 1989 [15], emerged and replaced O139 strains in 1993, spread all over India, including Wardha province, and eventually reached Africa in Guinea-Bissau [26]. The clonality of these O1 El Tor strains and the circulation of the same ICE among them, led us to propose ICE*Vch*Ind5 and siblings as the prevalent group of SXT-related ICEs circulating in the main clone of *V. cholerae* O1 El Tor strains circulating in the Indian Subcontinent after 1992.

Acknowledgments

This research was funded by a grant from Ministero dell'Istruzione, dell'Università e della Ricerca - Italy (PRIN 2007), and by Ateneo Federato, Università di Roma Sapienza (AST 2008). DC was supported by a fellowship from Cenci Bolognetti - Institute Pasteur Foundation, Italy. MS is the recipient of a PhD fellowship from Scuola di Dottorato in Biologia Cellulare e dello Sviluppo, Sapienza Università di Roma. YK acknowledges support from The Grand Water Research Institute, Technion and The Israeli Water Commission. VB holds a Canada Research Chair in molecular biology, impact and evolution of bacterial mobile elements and is grateful for support from the Natural Sciences and Engineering Research Council of Canada (Discovery Grant Program). We thank R. Pino, E. Suffredini and K. Bhutada for technical assistance in Italy and India. We are also grateful to Tony Pembroke, Didier Mazel, and Mathew K. Waldor for the kind gift of some strains.

References

1. Wozniak RAF, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol 8: 552-563.
2. Seth-Smith H, Croucher NJ (2009) Genome watch: breaking the ICE. Nat Rev Microbiol 7: 328-329.
3. Churchward G (2008) Back to the future: the new ICE age. Mol Microbiol 70: 554-556.
4. Wozniak RAF, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, et al. (2009) Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. PLoS Genet 5: e1000786. doi: 1000710.1001371/journal.pgen.1000786.
5. Ceccarelli D, Bani S, Cappuccinelli P, Colombo MM (2006) Prevalence of *aada1* and *dfra15* class 1 integron cassettes and SXT circulation in *V. cholerae* O1 isolates from Africa. J Antimicrob Chemother 58: 1095-1097.
6. Ceccarelli D, Daccord A, Rene M, Burrus V (2008) Identification of the origin of transfer (*oriT*) and a new gene required for mobilization of the SXT/R391 family of integrating conjugative elements. J Bacteriol 190: 5328-5338.
7. Waldor MK, Tschape H, Mekalanos JJ (1996) A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. J Bacteriol 178: 4157-4165.
8. Peters SE, Hobman JL, Strike P, Ritchie DA (1991) Novel mercury resistance determinants carried by IncJ plasmids pMERPH and R391. Mol Gen Genet 228: 294-299.
9. Burrus V, Marrero J, Waldor MK (2006) The current ICE age: biology and evolution of SXT-related integrating conjugative elements. Plasmid 55: 173-183.
10. Bordeleau E, Brouillette E, Robichaud N, Burrus V (2010) Beyond antibiotic resistance: integrating conjugative elements of the SXT/R391 family that encode novel diguanylate cyclases participate to c-di-GMP signalling in *Vibrio cholerae*. Environ Microbiol 12: 510–523.
11. Wozniak RAF, Waldor MK (2009) A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. PLoS Genet 5: e1000439. doi:1000410.1001371/journal.pgen.1000439.
12. Kaper JB, J Glenn Morris JR, Levine MM (1995) Cholera. Clin Microbiol Rev 8: 48–86.
13. Alam M, Hasan NA, Sadique A, Bhuiyan NA, Ahmed KU, et al. (2006) Seasonal cholera caused by *Vibrio cholerae* serogroups O1 and O139 in the coastal aquatic environment of Bangladesh. Appl Environ Microbiol 72: 4096-4104.

14. Faruque SM, Sack DA, Sack RB, Colwell RR, Takeda Y, et al. (2003) Emergence and evolution of *Vibrio cholerae* O139. Proc Natl Acad Sci USA 100: 1304-1309.
15. Garg P, Nandy RK, Chaudhury P, Chowdhury NR, De K, et al. (2000) Emergence of *Vibrio cholerae* O1 biotype El Tor serotype Inaba from the prevailing O1 Ogawa serotype strains in India. J Clin Microbiol 38: 4249-4253.
16. Hochhut B, Lotfi Y, Mazel D, Faruque SM, Woodgate R, et al. (2001) Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins. Antimicrob Agents Chemother 45: 2991-3000.
17. Marrero J, Waldor MK (2007) The SXT/R391 family of integrative conjugative elements is composed of two exclusion groups. J Bacteriol 189: 3302-3305.
18. Ingole K, Jalgaonkar S, Fule C, Fule R (1997) Changing pattern of *Vibrio cholerae* serotype El Tor and O139 in Yavatmal (Maharashtra, India) during 1992 to 1994. Indian J Pathol Bacteriol 40: 369-371.
19. Narang P, Mendiratta D, Kannathe J, Ramnani V, Wasekar V, et al. (1994) Characteristics of *Vibrio cholerae* O139 strains isolated in Sevagram (Maharashtra) during April-August 1993. Indian J Med Res 99: 103-104.
20. Batra P, Saha A, Vilhekar KY, Chaturvedi P, Mendiratta DK (2006) *Vibrio cholerae* O1 ogawa (Eltor) diarrhoea at Sevagram. Indian J Pediatr 73: 543.
21. Mishra M, Mohammed F, Akulwar SL, Katkar VJ, Tankhiwale NS, et al. (2004) Re-emergence of El Tor *vibrio* in outbreak of cholera in and around Nagpur. Indian J Med Res 120: 478-480.
22. Bani S, Mastromarino PN, Ceccarelli D, Van AL, Salvia AM, et al. (2007) Molecular characterization of ICE*Vch*Vie0 and its disappearance in *Vibrio cholerae* O1 strains isolated in 2003 in Vietnam. FEMS Microbiol Lett 266: 42-48.
23. Raz N, Danin-Poleg Y, Broza YY, Arakawa E, Ramakrishna BS, et al. (2010) Environmental monitoring of *Vibrio cholerae* using chironomids in India. Environ Microbiol Rep, in press.
24. Narang P, Mendiratta D, Deotale V, Narang R (2008) Changing patterns of *Vibrio cholerae* in Sevagram between 1990 and 2005. Indian J Med Microbiol 26: 40-44.
25. Khetawat G, Bhadra RK, Kar S, Das J (1998) *Vibrio cholerae* O139 Bengal: combined physical and genetic map and comparative analysis with the genome of *V. cholerae* O1. J Bacteriol 180: 4516-4522.
26. Sharma C, Ghosh A, Dalsgaard A, Forslund A, Ghosh RK, et al. (1998) Molecular evidence that a distinct *Vibrio cholerae* O1 biotype El Tor strain in Calcutta may have spread to the African continent. J Clin Microbiol 36: 843-844.
27. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, et al. (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406: 477-483.
28. McGrath BM, O'Halloran JA, Piterina AV, Pembroke JT (2006) Molecular tools to detect the IncJ elements: a family of integrating, antibiotic resistant mobile genetic elements. J Microbiol Methods 66: 32-42.
29. Danin-Poleg Y, Cohen LA, Gancz H, Broza YY, Goldshmidt H, et al. (2007) *Vibrio cholerae* strain typing and phylogeny study based on simple sequence repeats. J Clin Microbiol 45: 736-746.
30. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599.
31. Krishna BVS, Patil AB, Chandrasekhar MR (2006) Fluoroquinolone-resistant *Vibrio cholerae* isolated during a cholera outbreak in India. Trans R Soc Trop Med Hyg 100: 224-226.
32. Thungapathra M, Amita, Sinha KK, Chaudhuri SR, Garg P, et al. (2002) Occurrence of antibiotic resistance gene cassettes *aac(6')*-Ib, *dfrA5*, *dfrA12*, and *ereA2* in class 1 integrons in non-O1, non-O139 *Vibrio cholerae* strains in India. Antimicrob Agents Chemother 46: 2948-2955.

33. Chandrasekhar MR, Krishna BV, Patil AB (2008) Changing characteristics of *Vibrio cholerae*: emergence of multidrug resistance and non-O1, non-O139 serogroups. South Asian J Trop Med Public Health 39: 1092-1097.
34. Sharma C, Nair GB, Mukhopadhyay AK, Bhattacharya SK, Ghosh RK, et al. (1997) Molecular characterization of *Vibrio cholerae* O1 biotype E1 Tor strains isolated between 1992 and 1995 in Calcutta, India: evidence for the emergence of a new clone of the El Tor biotype. J Infect Dis 175: 1134-1141.
35. Mohapatra SS, Ramachandran D, Mantri CK, Colwell RR, Singh DV (2009) Determination of relationships among non-toxigenic *Vibrio cholerae* O1 biotype El Tor strains from housekeeping gene sequences and ribotype patterns. Res Microbiol 160: 57-62.
36. Popovic T, Bopp C, Olsvik O, Wachsmuth K (1993) Epidemiologic application of a standardized ribotype scheme for *Vibrio cholerae* O1. J Clin Microbiol 31: 2474-2482.
37. Faruque SM, Saha MN, Asadulghani, Bag PK, Bhadra RK, et al. (2000) Genomic diversity among *Vibrio cholerae* O139 strains isolated in Bangladesh and India between 1992 and 1998. FEMS Microbiol Lett 184: 279-284.
38. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, et al. (2009) Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. Proc Natl Acad Sci USA 106: 15442-15447.
39. Safa A, Bhuyian NA, Nusrin S, Ansaruzzaman M, Alam M, et al. (2006) Genetic characteristics of Matlab variants of *Vibrio cholerae* O1 that are hybrids between classical and El Tor biotypes. J Med Microbiol 55: 1563-1569.
40. Nair GB, Faruque SM, Bhuiyan NA, Kamruzzaman M, Siddique AK, et al. (2002) New variants of *Vibrio cholerae* O1 biotype El Tor with attributes of the classical biotype from hospitalized patients with acute diarrhea in Bangladesh. J Clin Microbiol 40: 3296-3299.
41. Safa A, Nair GB, Kong RYC (2010) Evolution of new variants of *Vibrio cholerae* O1. Trends Microbiol 18: 46-54.
42. Lee JH, Choi SY, Jeon YS, Lee HR, Kim EJ, et al. (2009) Classification of hybrid and altered *Vibrio cholerae* strains by CTX prophage and RS1 element structure. J Microbiol 47: 783-788.
43. Ang GY, Yu CY, Balqis K, Elina HT, Azura H, et al. (2010) Molecular evidence of cholera outbreak caused by a toxigenic *Vibrio cholerae* O1 El Tor variant strain in Kelantan, Malaysia. J Clin Microbiol: JCM.01086-01010.
44. Safa A, Sultana J, Cam PD, Mwansa JC, Kong RYC (2008) *Vibrio cholerae* O1 Hybrid El Tor Strains, Asia and Africa. Emerg Infect Dis 14: 987-988.
45. Nair G, Qadri F, Holmgren J, Svennerholm A, Safa A, et al. (2006) Cholera due to altered El Tor strains of *Vibrio cholerae* O1 in Bangladesh. J Clin Microbiol 44: 4211-4213.
46. Nguyen BM, Lee JH, Cuong NT, Choi SY, Hien NT, et al. (2009) Cholera outbreaks caused by an altered *Vibrio cholerae* O1 El Tor biotype strain producing classical cholera toxin B in Vietnam in 2007 to 2008. J Clin Microbiol 47: 1568-1571.
47. Ansaruzzaman M, Bhuiyan N, Nair B, Sack D, Lucas M, et al. (2004) Cholera in Mozambique, variant of *Vibrio cholerae*. Emerg Infect Dis 10: 2057-2059.
48. Quilici M-L, Massenet D, Gake B, Bwalki B, Olson DM (2010) *Vibrio cholerae* O1 variant with reduced susceptibility to ciprofloxacin, Western Africa. Emerg Infect Dis 16: 1804-1805.
49. Ceccarelli D, Salvia AM, Sami J, Cappuccinelli P, Colombo MM (2006) New cluster of plasmid-located class 1 integrons in *Vibrio cholerae* O1 and a *dfrA15* cassette-containing integron in *Vibrio parahaemolyticus* isolated in Angola. Antimicrob Agents Chemother 50: 2493-2499.

50. Ceccarelli D, Spagnoletti M, Bacciu D, Danin-Poleg Y, Mendiratta D, et al. (2011) ICE*Vch*Ind5 is prevalent in epidemic *Vibrio cholerae* O1 El Tor strains isolated in India. Int J Med Microbiol 301: 318–324.
51. Colombo MM, Mastrandrea S, Leite F, Santona A, Uzzau S, et al. (1997) Tracking of clinical and environmental *Vibrio cholerae* O1 strains by combined analysis of the presence of toxin cassette, plasmid content and ERIC PCR. FEMS Immunol Med Microbiol 19: 33-45.
52. WHO (2007) Cholera 2006. Wkly Epidemiol Rec 31: 273-284.
53. Taviani E, Ceccarelli D, Lazaro N, Bani S, Cappuccinelli P, et al. (2008) Environmental *Vibrio spp*., isolated in Mozambique, contain a polymorphic group of integrative conjugative elements and class 1 integrons. FEMS Microbiol Ecol 64: 45-54.
54. Bhattacharya T, Chatterjee S, Maiti D, Bhadra RK, Takeda Y, et al. (2006) Molecular analysis of the *rstR* and *orfU* genes of the CTX prophages integrated in the small chromosomes of environmental *Vibrio cholerae* non-O1, non-O139 strains. Environ Microbiol 8: 526-634.
55. Keasler SP, Hall RH (1993) Detecting and biotyping *Vibrio cholerae* O1 with multiplex polymerase chain reaction. Lancet 341: 1661.
56. Olsvik O, Wahlberg J, Petterson B, Uhlen M, Popovic T, et al. (1993) Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. J Clin Microbiol 31: 22-25.
57. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, et al. (1990) Current protocols in molecular biology; Sons JWa, editor. New York: Green Publishing Associates and Wiley.
58. Beaber JW, Hochhut B, Waldor MK (2002) Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. J Bacteriol 184: 4259-4269.
59. Grim CJ, Hasan NA, Taviani E, Haley B, Chun J, et al. (2010) Genome sequence of hybrid *V. cholerae* O1 MJ-1236, B-33 and CIRS101 and comparative genomics with *V. cholerae*. J Bacteriol 192: 3524-3533.
60. Cho YJ, Yi H, Lee JH, Kim DW, Chun J (2010) Genomic evolution of *Vibrio cholerae*. Curr Opin Microbiol 13: 646-651.
61. Chatterjee S, Patra T, Ghosh K, Raychoudhuri A, Pazhani GP, et al. (2009) *Vibrio cholerae* O1 clinical strains isolated in 1992 in Kolkata with progenitor traits of the 2004 Mozambique variant. J Med Microbiol 58: 239-247.
62. Cappuccinelli P, Colombo MM, Morciano C (1995) Epidemiologia del colera in Africa. In: IILA, editor. Il colera, una malattia da debellare Il colera oggi in America Latina.
63. Dalsgaard A, Mortensen HF, Molbak K, Dias F, Serichantalergs O, et al. (1996) Molecular characterization of *Vibrio cholerae* O1 strains isolated during cholera outbreaks in Guinea-Bissau. J Clin Microbiol 34: 1189-1192.
64. Ceccarelli D, Spagnoletti M, Cappuccinelli P, Burrus V, Colombo MM (2011) Origin of *Vibrio cholerae* in Haiti. Lancet Infect Dis 11: 260.
65. Ghosh-Banerjee J, Senoh M, Takahashi T, Hamabata T, Barman S, et al. (2010) Cholera toxin production by the El Tor variant of *Vibrio cholerae* O1 compared to prototype El Tor and Classical biotypes. J Clin Microbiol 48: 4283-4286.

CHAPTER IV

New *V. cholerae* atypical El Tor variant emerged during the 2006 epidemic outbreak in Angola

Daniela Ceccarelli[1], Matteo Spagnoletti[1], Donatella Bacciu[2], Piero Cappuccinelli[2] and Mauro M Colombo[1]

1 Dipartimento di Biologia e Biotecnologie Charles Darwin, Sapienza Università di Roma, Rome, 00185, Italy. **2** Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, 07100, Italy.

Abstract

V. cholerae is the etiological agent of cholera, a major public health concern in most developing countries. Virulence of *V. cholerae* relies on the powerful cholera toxin, encoded by the CTX prophage. The emergence of new pathogenic variants in the recent years has been mostly associated with new CTX prophage rearrangements. In this retrospective study, we show that the epidemic *V. cholerae* O1 El Tor strain responsible for the 2006 outbreak in Angola is clonally and genetically different from El Tor strains circulating in the 1990s in the same area. Strains from 2006 carry ICE*Vch*Ang3 of the SXT/R391 family. This ICE is associated with a narrower multidrug resistance profile compared to the one conferred by plasmid p3iANG to strains of the 1990s. The CTX prophage carried by 2006 El Tor strains is characterized by $rstR^{ET}$ and $ctxB^{Cla}$ alleles organized in a RS1-RS2-Core array on chromosome I. Interestingly, the newly emerging atypical strain belongs to a clade previously known to comprise only clinical isolates from the Indian subcontinent that also contain the same ICE of the SXT/R391 family. Our findings remark the appearance of a novel *V. cholerae* epidemic variant in Africa with a new CTXΦ arrangement previously described only in the Indian Subcontinent.

This chapter was published in BMC Microbiology (2011)
doi:10.1186/1471-2180-11-130

Background

Vibrio cholerae is the etiological agent of the severe watery diarrhoeal disease known as cholera, a major public health concern in most developing countries. More than 200 serogroups have been described on the basis of different somatic O antigens [12], but only serogroups O1 and O139 have the ability to cause harsh epidemics. Serogroup O1 is further divided into two main biotypes, Classical and the 7th pandemic El Tor. Beside their phenotypic characteristics, differences in specific genetic markers, such as toxin structure, confer distinct features to these biotypes.

Pathogenic *V. cholerae* strains carry the genes encoding the cholera toxin (CT) on the CTXΦ prophage. Different CTXΦ arrangements have been described within the O1 serogroup [41]. These arrangements depend on the genotype of the CT gene *ctxB* and on the organization and chromosomal location of several gene clusters of phage origin, namely the core, RS2, and RS1 [41]. Although the Classical biotype is considered extinct, new El Tor strains holding the Classical *ctxB* allele, generically labeled as atypical El Tor (including hybrid El Tor, altered El Tor and Mozambique variants) [41], were identified from 1993 to date mostly in Asia [40,42-46] with few cases in Africa [44,47,48]. The atypical variants are characterized by a new CTXΦ arrangement, holding El Tor and/or Classical alleles of *rstR* and *ctxB* genes [41]. As a consequence of these genetic arrangements in CTX prophage, toxigenic *V. cholerae* O1 El Tor strains have changed in the last 20 years. Initially, atypical variants were only sporadically identified in the Indian Subcontinent along with prototype El Tor. However they are now in the process of replacing it worldwide [41].

Prototype El Tor strains often contain multi-resistant conjugative plasmids [49] whereas O139 and atypical O1 El Tor *V. cholerae* epidemic strains usually harbor Integrative Conjugative Elements (ICEs) of the SXT/R391 family [4].

SXT/R391 ICEs are self-transmissible mobile elements, ranging in size from 79 to 108 kb, able to integrate into the host bacterial chromosome and to transfer by conjugation. They are recognized for their important role in bacterial genome plasticity [2] and as vectors of antibiotic resistance and alternative metabolic pathways [4]. The name of the SXT/R391 family originates from elements SXT^{MO10} and R391, respectively discovered in clinical strains of *Vibrio cholerae* in India [7] and *Providencia rettgeri* in South Africa [8]. The two elements are associated with different multi-resistance profiles: chloramphenicol, streptomycin, sulfamethoxazole, and trimethoprim for SXT^{MO10}, and kanamycin, and mercury for R391 [4]. They share a highly conserved genetic backbone encoding their integration/excision, conjugative transfer, and regulation, but also contain variable DNA found in five insertion sites of the backbone [4]. Each ICE of the family holds specific genes scattered in the conserved sequence that code for resistance to antibiotics and heavy metals, new toxin/antitoxin systems, restriction/modification systems, and alternative metabolic pathways [4]. To date more than 50 ICEs have been identified and grouped within the SXT/R391 family, most of them discovered in *V. cholerae* strains.

To date, only a few SXT-related ICEs were identified in Africa, most of them through the characterization of the integrase int_{SXT}. Only ICE*Vch*Moz10 from Mozambique (2004) has been completely sequenced and annotated [4]. This ICE has no close relative in Africa except its sibling

ICE*Vch*Ban9 isolated in Bangladesh (1994), suggesting the possible spread of SXT-related ICEs between the two continents in recent times. Although the use of horizontally-transferred elements as genetic markers for strain discrimination might appear risky, we recently showed the existence of an ICE/strain association in epidemic *V. cholerae* strains circulating in the Indian Subcontinent [50]. The association between ICE and *V. cholerae* reflects the classification proposed by Chun and colleagues to describe homologous intraspecific groups of *V. cholerae* based on the whole genome alignment of 23 strains isolated over the past 100 years [38].

In this retrospective study, we analysed *V. cholerae* O1 clinical strains isolated in Luanda (Angola) in 2006. Angola is an endemic area for cholera and was subjected to two major epidemic events in the past three decades. The first outbreak (1987-1993) [51] was followed by a thirteen year remission phase until cholera reemerged in 2006 in one of the most severe epidemic outbreaks of the last decade, counting about 240.000 cases [52].

Here we demonstrate that the *V. cholerae* O1 El Tor strain responsible for the 2006 Angolan outbreak is an atypical O1 El Tor variant previously detected only in Asia [42]. This variant is significantly different from those isolated during previous cholera outbreaks in the 1990s in the same geographic area. Indeed, it holds a peculiar CTXΦ array and the SXT-like element ICE*Vch*Ang3. Ribotype analysis suggests that this strain might have spread to West Africa from the Indian Subcontinent.

Methods

Bacterial strains, susceptibility tests and transfer of drug resistances

We analyzed *V. cholerae* strains isolated in Angola or India between 1992 and 2006 (Table 1). All strains were isolated from stool samples and/or rectal swabs from patients, and after isolation on thiosulfate citrate bile sucrose agar and biochemical identification, bacterial strains were routinely grown in Luria-Bertani (LB) or agar plates at 37°C and maintained at -80° in LB broth containing 30% (vol/vol) glycerol. Antimicrobial susceptibility was tested according to their MIC breakpoints at the following concentrations: ampicillin (Ap), 100 µg/ml; chloramphenicol (Cm), 20 µg/ml; kanamycin (Km), 50 µg/ml; nalidixic acid (Nx), 40 µg/ml; penicillin (Pn), 20 µg/ml; rifampin (Rf), 100 µg/ml; spectinomycin (Sp), 50 µg/ml; streptomycin (Sm), 50 µg/ml; sulfamethoxazole (Su), 160 µg/ml; tetracycline (Tc), 12 µg/ml; and trimethoprim (Tm), 32 µg/ml. Antibiotics were included in ISO sensitest (Oxoid) agar plates and bacterial strains were spotted onto the plates as previously described [49].

Conjugation assays were used to transfer ICE*Vch*Ang3 from *V. cholerae* into rifampin-resistant derivatives of *E. coli* 803 strain. Mating assays were performed by mixing equal volumes of overnight cultures of donor and recipient strains. Briefly, the cells were harvested by centrifugation and resuspended in a 1/20 volume of LB broth. Cell suspensions were poured onto LB agar plates and incubated at 37°C for 6 h. The cells were then resuspended in 1 ml of LB medium, and serial dilutions were plated onto appropriate selective media to determine the numbers of donors, recipients, and exconjugants. Frequency of transfer was expressed as the

number of exconjugant cells per donor cells in the mating mixture at the time of plating. *V. cholerae* O139 MO10 [7], *V. cholerae* E4:ICE*Vch*Ind1 [16], *V. cholerae* O1 VC20 [26], *V. cholerae* N16961 [27], *V. cholerae* O1 CO840 [26], *V. cholerae* O1 VC7452, VC15699, and VC9258 isolated in India (Maharashtra) [50], and *E. coli* AB1157:R391 [28] were appropriately used as negative or positive controls for class 1 integrons, ICE, *tcpA*, and *rstR* detection, CTXΦ array and ribotype analysis.

Strain	Isolation		Antibiotic resistance profile	Antibiotic resistance genes	ICE content	CTXΦ array	Ribotype	Reference
	Place	Year						
VC175	Angola (Luanda)	2006	Ap, Pn, Sm, Su, Tp	*floR, strA, strB, dfrA1, sulII*[b]	ICE*Vch*Ang3	B	R1	This study
VC189	Angola (Luanda)	2006	Ap, Pn, Sm, Su, Tp	*floR, strA, strB, dfrA1, sulII*[b]	ICE*Vch*Ang3	B	R1	This study
VC582	Angola (Luanda)	1992	Ap, Cm, Kn, Pn, Sm, Sp, Su, Tc, Tp[a]	*aph, tetG, cat1, blaP1, dfrA15, aadA8, sul2*[c]	-	A	R2	[11]
VC1383	Angola (Benguela)	1994	Ap, Cm, Kn, Pn, Sm, Sp, Su, Tc, Tp[a]	*aph, tetG, cat1, blaP1, dfrA15, aadA8, sul2*[c]	-	A	R3	[11]
VC547	Angola (Bengo river)	1994	Ap, Cm, Kn, Pn, Sm, Sp, Su, Tc, Tp[a]	*aph, tetG, cat1, blaP1, dfrA15, aadA8, sul2*[c]	-	A	R4	[11]
VC7452	India (Sevagram)	1995	Ap, Nx, Pn, Sm, Sp, Su, Tp	*floR, strA, strB, dfrA1, sulII*[b]	ICE*Vch*Ind5[d]	B	R1	[16]
VC15699	India (Sevagram)	1999	Ap, Nx, Pn, Sm, Sp, Su, Tp	*floR, strA, strB, dfrA1, sulII*[b]	ICE*Vch*Ind5	B	R1	[16]
VC9258	India (Sevagram)	1999	Ap, Nx, Pn, Sm, Sp, Su, Tp	*floR, strA, strB, dfrA1, sulII*[b]	ICE*Vch*Ind5	B	R1	[16]

Table 1. *V. cholerae* O1 strains analyzed in this study. [a]Resistance profile conferred by conjugative plasmid p3iANG [11]; [b]located on the ICE; [c]located on p3iANG; [d]ICE fully sequenced. Abbreviations: Ap, ampicillin; Cm, chloramphenicol; Kn, kanamycin; Nx, nalidixic acid; Pn, penicillin; Sm, streptomycin; Sp, spectinomycin; Su, sulfamethoxazole; Tc, tetracycline; Tp,trimethoprim.

Molecular biology procedures

Bacterial DNA for PCR analysis was prepared with a Wizard Genomic DNA Purification kit (Promega). Amplicons to be sequenced were directly purified from PCR or extracted from agarose gel by Wizard SV Gel and PCR Clean-up System (Promega) according to the manufacturer's instructions. DNA sequences were determined by BMR Genomics (Padova, Italy).

Class 1 integron detection was performed by PCR amplification with specific primer pairs as previously described [49]. ICEs of the SXT/R391 family were screened by PCR analysis, using 17 specific primer pairs previously described by our group [22,53]. int_{SXT}, *prfC*/SXT^{MO10} right junction, *floR*, *strA*, *strB*, *sul2*, *dfrA18*, *dfrA1*, *rumAB* operon, *traI*, *traC*, *setR*, and Hotspots or Variable Regions *s026/traI*, *s043/traL*, *traA/s054*, *s073/traF* and *traG/eex* were screened. A second set of 15 primer pairs designed on the specific sequences of ICE*Vch*Ind5 [50] were used to detect ICE*Vch*Ind5 and ICE*Vch*Ang3 specific Hotspots and Variable Regions.

All PCR reactions were set in a 50-µl volume of reaction buffer containing 1 U of *Taq* polymerase as directed by the manufacturer (Promega).

Ribotype analysis

Ribotyping of *V. cholerae* strains was performed by *Bgl*I restriction of chromosomal DNA with fluorescent-labeled 16S and 23S DNA (Gene Images 3540 RPn3510, Amersham) generated by reverse transcriptase polymerase chain reaction of ribosomal RNAs, as already described [22].

CTX array analysis and *ctxB*, *tcpA*, *rstR* biotype characterization

The structure of CTX array was determined by multiple PCR analysis (see table 2) and by Southern Blot hybridization. The genetic structure of the two CTX prophage arrays described in figure 1 was determined using the primers described in table 2. Briefly, combination of primers tlcF/rstAR, tlcF/rstCR, rstCF/rstAR, ctxAF/rstAR, rstCF/rtxR and ctxAF/rtxR were used to detect the presence of CTXΦ on chromosome 1 and to determine the position of the RS1 element (see table S1 for complete amplicon profiles). The absence of CTXΦ or RS1 on chromosome 2 was established using primers chr2F/chr2R. Primers ctxAF/cepR were used to determine the presence of CTX tandem arrays.

Primer	Nucleotide sequence (5′ to 3′)	Position (GenBank Accession no. NC_002505-6)
tlcF	CCAAAACAACAGAAGCAACAGAGCAACG	1574460-1574487
rstCF	GGCGCTTATACAGACGAAATCGCTC	1564180-1564201
rstCR	AGCGCCTGAACGCAGATATAAA	1564290-1564311
rstAR	CGACAAAAACAAACGGAGAAGCGT	1572748-1572771
ctxAF	CTCAGACGGGATTTGTTAGGCACG	1567895-1567918
rtxR	CAAGCTGCGATCAGCATGGCGTGGTC	1563652-1563671
cepR	CAGTGTTTTGGTGACTTCCGT	1571101-1571121
chr2F	CTCACGCTGAACAGCAAGTC	507564-507583
chr2R	AAACCGGGAGAAGTGATTGC	509487-509506

Table 2. Primers used to determine CTX prophage array structure

Three previously described primer sets were used to detect: (i) Classical, El Tor, or Kolkata type *rstR* gene [54], (ii) *ctxB* genotype sequencing [55], (iii) and Classical or El Tor biotypes for *tcpA* [56]. PCR results on organization and location of CTXΦ on chromosome 1 were further confirmed by Southern Blot hybridization assays. DNA probes were produced by PCR using the chromosomal DNA of *V. cholerae* strain N16961 as template: *ctxA* gene (564 bp) with primers CTX-2 (CGGGCAGATTCTAGACCTCCTG) and CTX-3 (CGATGATCTTGGAGCATTCCCAC); *rstA* gene (789 bp) with primers rstA1F (AAACCTGCAAAATACCCCT) and rstA1R (ACAACTCGATACAAACGCT). Probes for hybridization were labeled with alkaline phosphatase with AlkPhos Direct™ Labelling and Detection System with CDP-*Star™* kit (Ge Healthcare), according to manufacturer's instructions.
Strains were cultured in Luria-Bertani medium and 1 ml of culture was used to extract and purify the genomic DNA using the DNeasy Blood & Tissue Kit (Qiagen). Aliquots of the extracted DNA (1,5 µg) were digested with *EcoR*V for CTXΦ element restriction fragment length polymorphism analysis. The digested fragments were separated by agarose gel electrophoresis (1 % gel) and were blotted on nitrocellulose membranes using standard methods [57]. Southern blots were hybridized O/N with *ctxA* or *rstA* labeled probes, and washed under stringent conditions, according to manufacturer's instructions. Addition of CDP-Star Detection Reagent was followed by 10 min incubation, and autoradiography (20 min to 1 h) was performed to generate a signal.

Nucleotide sequence accession numbers

Nucleotide sequences of *ctxB* genes were deposited in GenBank under accession no. HQ599507 (*V. cholerae* 1383), HQ599508 (*V. cholerae* 7452), HQ599509 (*V. cholerae* 547), HQ599510 (*V. cholerae* 582), and HQ599511 (*V. cholerae* 175).

Results

V. cholerae strains from 2006 show reduced resistance profile compared to previous epidemic strains

We analyzed two *V. cholerae* O1 El Tor clinical strains, VC175 and VC189 (Table 1), isolated at the Luanda Central Hospital (Angola). These strains were collected during the peak (May) of the cholera outbreak reported in Angola in 2006.

The two strains were sensitive to tetracycline, chloramphenicol, and kanamycin but showed a multiresistant profile to ampicillin, penicillin, streptomycin, trimethoprim, and sulfamethoxazole. Despite this significant multidrug resistance, these strains showed a narrower resistance profile compared to those isolated in the previous 1987-1993 cholera epidemic, which were also resistant to tetracycline, chloramphenicol, spectinomycin and kanamycin [49]. We found no evidence for the presence of conjugative plasmids or class 1 integrons in the 2006 strains analyzed (data not shown), which might explain their reduced drug resistance profile. Indeed, strains from 1987-1993 were associated with the conjugative plasmid p3iANG that holds genes encoding the resistance to tetracycline, chloramphenicol, kanamycin, and spectinomycin [49].

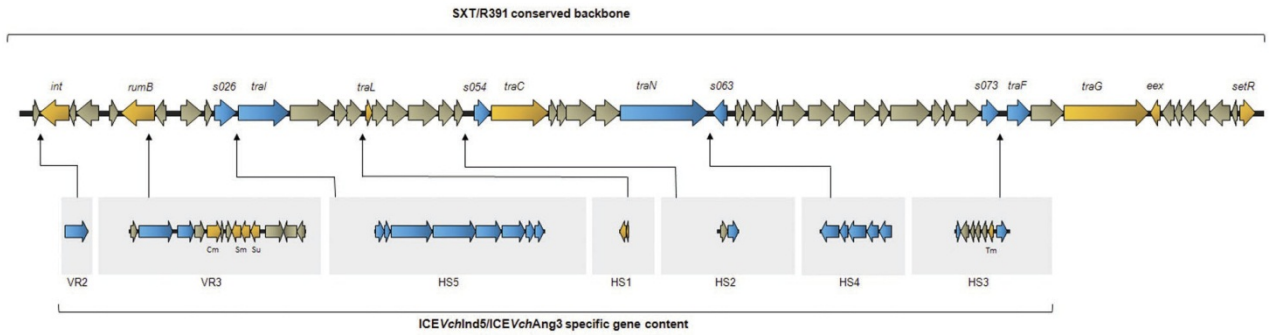


Figure 1. ICEVchAng3 genetic structure. Schematic linear representation (adapted from Wozniak et al., 2009) of the genes amplified by PCR to define the molecular structure of ICEVchAng3. The upper line represents the conserved backbone of the SXT/R391 family members. The black arrows indicate insertion sites for ICEVchInd5/ICEVchAng3 specific gene content. Genes in orange were tested with primer set A. Genes in blue were tested with primer set B. Genes not tested are shown in grey. VR: Variable Region; HS: Hotspot. GenBank accession no. of the full sequence of ICEVchInd5 is GQ463142 [12].

ICE*Vch*Ang3 is a sibling of ICE*Vch*Ind5

We assessed the presence of SXT/R391 family ICEs since they are a major cause of antibiotic resistance spread among *V. cholerae* strains.. Both strains were int_{SXT}^{+}, were shown to contain an ICE integrated into the *prfC* gene, and contained the conserved genes *traI*, *traC* and *setR*, respectively encoding a putative relaxase, a putative conjugation coupling protein, and a transcriptional repressor found in all SXT/R391 family members [58]. Based on these results we included this ICE in the SXT/R391 family and named it ICE*Vch*Ang3 according to the accepted nomenclature [9].

SXT/R391 ICEs exhibit significant genetic polymorphisms in hotspot content [4]. We used a first set of primers (primer set A), designed to discriminate between SXT^{MO10} and R391 specific sequences [22], in order to prove the identity of the ICE circulating in the 2006 Angolan strains. Genes *floR*, *strA*, *strB*, *sul2*, *dfrA18*, *dfrA1*, the *rumAB* operon, and Hotspots or Variable Regions *s026/traI*, *s043/traL*, *traA/s054*, *s073/traF* and *traG/eex* were screened.

The 2006 strains exhibited the same SXT^{MO10}/R391 hybrid ICE pattern. Intergenic regions *traG*/*eex* (Variable Region 4) and *traA/s054* (Hotspot 2) showed the molecular arrangement described in SXT^{MO10}, whereas region *s043/traL* (Hotspot 1) was organized as in R391. Variable Region 3, inserted into the *rumB* locus, contained genes that mediate resistance to chloramphenicol, streptomycin and sulfamethoxazole: *floR*, *strA*, *strB*, *sul2*. Interestingly, ICE*Vch*Ang3 lacks *dfr18*, the gene conferring resistance to trimethoprim found in SXT^{MO10}, and carries instead *dfrA1* in Hotspot 3. This preliminary analysis revealed that ICE*Vch*Ang3 exhibits a hybrid genetic content similar to that of the completely sequenced ICE*Vch*Ind5, the most widespread ICE circulating in *V. cholerae* El Tor O1 strains in the Indian Subcontinent [50].

Given these similarities we analyzed ICE*Vch*Ang3 using a second set of primers (primer set B) previously designed to assess the hotspot content of ICE*Vch*Ind5 [50]. This analysis confirmed that all the peculiar insertions found in ICE*Vch*Ind5 were also present in ICE*Vch*Ang3: (i) a gene encoding a protein similar to the *E. coli* dam-directed mismatch repair protein MutL (Variable Region 2); (ii) intI9 integron (Hotspot 3); (iii) a possible transposon of the IS21 family (Hotspot 4); and (iv) a 14.8-kb hypothetical operon of unknown function (Hotspot 5). On account of our results and of the common backbone shared by SXT/R391 ICEs (~65% of the ICE), we are confident that ICE*Vch*Ang3 is a sibling of ICE*Vch*Ind5 [50]. A map (not to scale) of ICE*Vch*Ang3 is shown in figure 1.

We performed mating experiments to assess the ability of ICE*Vch*Ang3 to transfer by conjugation between *V. cholerae* strain VC 175 or VC 189 and *E. coli* 803Rif. The frequency of transfer of ICE*Vch*Ang3 was $4{,}4 \times 10^{-5}$, a frequency of transfer similar to that of most of the ICEs of this family. Ten *E. coli* exconjugant colonies were tested and proved to be positive for the presence of int_{SXT}, confirming the mobilization of ICE*Vch*Ang3.

A new CTXΦ array in Africa

The variability of CTXΦ and the emergence of atypical El Tor variants in the ongoing 7th pandemic [41] les us to analyze the organization of CTXΦ arrays and the presence of different alleles of *ctxB*, *rstR* and *tcpA* genes. The genetic structure of CTX prophage in the genome of the Angolan isolates

from both epidemic events was determined by multiple PCR analysis, hybridization, and sequencing, when required.

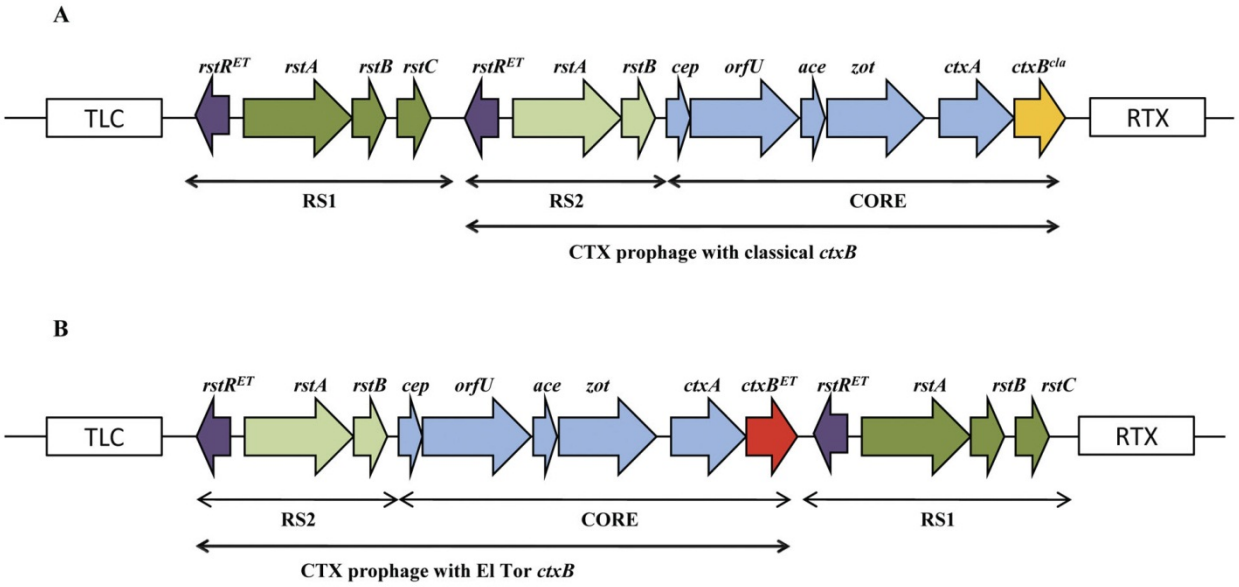


Figure 2. Comparison of the genetic structures of the two CTX prophage arrays identified in the *V. cholerae* strains under study. Both prophages are integrated into the large chromosome. Arrows indicate the transcription direction of each gene. (A) CTX prophage array profile A: RS1-RS2-CORE; (B) CTX prophage array profile B: RS2-CORE-RS1. Map is not to scale. rstR^{ET} (purple arrow): El Tor type rtsR; ctxB^{ET} (red arrow): El Tor type ctxB; ctxB^{cla} (yellow arrow): Classical type ctxB; TLC: toxin-linked cryptic plasmid; RTX: RTX (repeat in toxin) gene cluster.

Combining the results obtained by multiple PCR analysis and hybridization we were able to show that the strains analyzed contained two distinct CTXΦ arrays (A and B), both of which were found integrated in the large chromosome (figure 2, table S1). These strains also proved to be negative for any CTXΦ integration on the small chromosome and devoid of CTX tandem arrays as detected by primer pairs chr2F/chr2R and ctxAF/cepR, respectively. The Angolan strains isolated in 2006 (VC 175 and VC 189) belonged to profile A, in which the RS1 element is followed by CTXΦ, both being located between the toxin-linked cryptic (TLC) element and the chromosomal RTX (repeat in toxin) gene cluster (figure 2a). In contrast, strains from the first outbreak (1987-1993) contained CTXΦ followed by the RS1 element (profile B) (figure 2b). Both CTXΦ arrays were characterized by El Tor type *rstR* genes (both in RS1 and RS2) but showed a noteworthy difference in their *ctxB* genotype (table 3). CTXΦ arrays belonging to profile A contained a histidine and a threonine at the 39th and 68th amino acid positions, respectively, which are representative of Classical genotype 1 CtxB. The CTXΦ arrays belonging to profile B held a tyrosine, a phenylalanine and an isoleucine at positions 39th, 46th and 68th, respectively, typical of an El Tor genotype 3 CtxB.

Angolan and Indian strains share the same clonal origin

In order to verify their clonal relationship, we analysed by ribotyping the strains from the two Angolan epidemics of the 1990s and of 2006, as well as the Indian strains collected from 1993 to

2005 (Table 1) [50]. Strains from 1987-1993 outbreak (VC582, VC1383 and VC547) were chosen according to their epidemiological role (clinical or environmental isolate) and the presence of plasmid p3iANG [49].

Angolan strains isolated between 1992 and 1994 showed an assorted ribotype profile: clinical strains VC582 and VC1383 were characterized by profiles R2 (2.3,4.2, 4.6, 5.7, 6.0 kb) and R3 (2.3,4.2, 4.6, 5.7, 6.0, 9.6, 18.0 kb), respectively, and environmental isolate VC547 by a third completely different profile R4 (1.0, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 3.8, 5.5 kb). This heterogeneity is not surprising if we consider the Angolan clinical strains on a larger sample scale. Indeed, our data showed that there was a clonal shift in Angola from 1992 to 1993/1994 with consequent change of ribotype (D.C personal communication) that can explain the discrepancies observed here. Strains VC175 and VC189 isolated in 2006 were characterized by the same ribotype profile R1 (2.3, 4.2, 5.8, 6.1, 6.3, 8.5, 9.4, 10.8, 22.0 kb) which corresponds to the ribotype profile of the Indian strains carrying ICE*Vch*Ind5 [50], suggesting a common clonal origin.

Discussion

2006 was a crucial year for cholera worldwide. The number of reported cases was higher than ever and exceeded the levels of the late 1990s. Major outbreaks affected some of the largest African countries, including Angola, which reported to WHO one of the most exceptional epidemics experienced in Africa in the last decade [52].

This is the first study on the causative agent of this dramatic outbreak and our analysis revealed significant differences between the Angolan strains of 2006 and those isolated in the previous 1987-1993 cholera epidemic. The 1987-1993 epidemic was the longest in Angolan history and the *V. cholerae* epidemic strains were characterized by the presence of the conjugative plasmid p3iANG that carries three class 1 integrons [49]. Interestingly, the strains from the 2006 outbreak lack p3iANG but harbor an SXT-like ICE sibling of ICE*Vch*Ind5, previously described only in Asian *V. cholerae* strains [50]. The gene content of ICE*Vch*Ang3 comprises elements shared with SXT^{MO10}, R391, ICE*Vch*Ban9, and ICE*Pda*Spa1, alongside some unique insertions of unknown function that might provide the strain with increased fitness. In light of its genetic content we included ICE*Vch*Ang3 in the subgroup of SXT/R391 ICEs that characterizes *V. cholerae* O1 El Tor strains circulating in several epidemic areas of the Indian Subcontinent, of which ICE*Vch*Ind5 is the reference strain [4,50].

Beside the analysis of the Mozambican variant, extensive studies of CTXΦ arrangements in *V. cholerae* strains isolated in Africa lack so far.

Our analysis reports that the strains of the 2006 outbreak contain an RS1-CTX array on the large chromosome with a classical *ctxB* allele, which classifies them as *V. cholerae* O1 altered El Tor. This variant was responsible for major epidemics in India in 2004-2006 [42] and in Vietnam in 2007 [46]. It is considered as prevalent in Asia nowadays [59,60] and forms a monophyletic group with other variants of the 7^{th} pandemic clade [38]. This variant arose in the Indian Subcontinent at the beginning of the 90s and slowly diffused to Asian countries [40,45]. The possible spread to Africa

was only suggested [42,59] and some authors gave partial evidences supporting this hypothesis by strain ribotyping [26] or *ctxB* genotyping [44]. With this work we ascertain the presence of this atypical El Tor variant in Africa and demonstrate it holds the responsibility for the 2006 cholera epidemic in Angola.

Strain	*rstR*	*tcpA*	*ctxB*	
			Genotype[a]	Amino acid position[b]
VC582	ET	ET	3 (ET)	20 (His); 24 (Gln); 28 (Asp); 34 (His); 39 (Tyr); 46 (Phe); 55 (Lys); 68 (Ile)
VC547	ET	ET	3 (ET)	20 (His); 24 (Gln); 28 (Asp); 34 (His); 39 (Tyr); 46 (Phe); 55 (Lys); 68 (Ile)
VC1383	ET	ET	3 (ET)	20 (His); 24 (Gln); 28 (Asp); 34 (His); 39 (Tyr); 46 (Phe); 55 (Lys); 68 (Ile)
VC175	ET	ET	1 (Cla)	20 (His); 24 (Gln); 28 (Asp); 34 (His); 39 (His); 46 (Phe); 55 (Lys); 68 (Thr)
VC7452	ET	ET	1 (Cla)	20 (His); 24 (Gln); 28 (Asp); 34 (His); 39 (His); 46 (Phe); 55 (Lys); 68 (Thr)

Table 3. Biotype characterization and *ctxB* genotype comparison of *V. cholerae* O1 isolates from Angola and India. Cla, Classical type; ET, El Tor type; [a]According to *ctxB* genotyping by Safa et al., 2010 [2]; [b]Nucleotide position +1 corresponds to the A of the ATG start codon in *ctxB*.

The Angolan variant is the second example of atypical El Tor variant described in Austral Africa, the first being the Mozambican strain B33 [47]. However, this variant is different from the Angolan one, since it holds a tandem CTXΦ array on the small chromosome [59], contains a different ICE (ICE*Vch*Moz10) [4], and is closely related to the Bangladeshi strain MJ-1236 [38,40].

Unlike B33 whose progenitor was identified as a Kolkata hybrid strain from 1992 [61], we have no clear information on how the variant we found in Angola penetrated Austral Africa. We can speculate that it arrived from the Indian Subcontinent through the same Sub-Saharan corridor used by cholera to enter Africa at the beginning of the 7th pandemic [62]. During the '70s it spread from the Horn of Africa to Senegal, Guinea Bissau and eventually arrived in Angola: the new atypical variant might have disseminated by a similar route. This supposition might find some confirmation in the analysis performed by Sharma and colleagues who proposed the spread of a distinct *V. cholerae* O1 strain from India to Guinea Bissau, where it was associated with an epidemic of cholera in 1994 [26]. This hypothesis was based on the ribotype analysis of pre- and post- O139 *V. cholerae* O1 strains circulating in both countries. Our ribotype analysis confirmed these data since the Angolan strain from 2006, the clinical strains isolated in Guinea Bissau in 1994/1995 [63], and clinical post-O139 *V. cholerae* O1 strains from India [26] share the same profile, suggesting a common clonal origin. Unfortunately, the genetic content of the strains isolated in Guinea Bissau, in terms of ICE structure and CTXΦ array, was never investigated and our speculations cannot go any further.

Whichever route of dissemination used by the new variant to spread from the Indian Subcontinent to Africa, many evidences indicate that atypical *V. cholerae* strains are in the process of globally replacing the prototype El Tor strains, as observed in Angola.

Conclusions

Cholera remains a global threat to public health and the recent outbreak in Haiti is a distressing example of this situation [64]. In 2006, Angola, which had reported no cholera cases since 1998, was affected by a major outbreak due to an atypical *V. cholerae* O1 El Tor strain that was analyzed for the first time in our study. This altered El Tor strain holds an RS1-CTX array on the large chromosome and a Classical *ctxB* allele and likely replaced the previous prototype O1 El Tor strain reported till 1994. The success of the new variant might depend on the combination of the respective predominant features of the El Tor and Classical biotypes: a better survival in the environment [41] and the expression of a more virulent toxin [65].

Acknowledgements

We are grateful to Dr. M. Francisco (Dept. of Microbiology, Faculty of Medicine, University A. Neto, Luanda - Angola) for providing us with Angolan *V. cholerae* strains from 2006, and to A. Crupi for technical assistance. We are grateful to G. Garriss for manuscript revision.
This work was supported by Ministero Istruzione Università e Ricerca (MIUR) (Grant n. 2007W52X9M to MMC and PC), and Ministero Affari Esteri – Direzione Generale Cooperazione Sviluppo (MAE-DGCS) (Grant n. AID89491 to MMC), Italy. DC was supported by a fellowship from Institute Pasteur - Fondazione Cenci Bolognetti, Italy. MS is the recipient of a PhD fellowship from the Doctorate School in Cellular and Developmental Biology, Sapienza Università di Roma.

References

1. Kaper JB, J Glenn Morris JR, Levine MM (1995) Cholera. Clin Microbiol Rev 8: 48–86.
2. Safa A, Nair GB, Kong RYC (2010) Evolution of new variants of *Vibrio cholerae* O1. Trends Microbiol 18: 46-54.
3. Lee JH, Choi SY, Jeon YS, Lee HR, Kim EJ, et al. (2009) Classification of hybrid and altered *Vibrio cholerae* strains by CTX prophage and RS1 element structure. J Microbiol 47: 783-788.
4. Ang GY, Yu CY, Balqis K, Elina HT, Azura H, et al. (2010) Molecular evidence of cholera outbreak caused by a toxigenic *Vibrio cholerae* O1 El Tor variant strain in Kelantan, Malaysia. J Clin Microbiol: JCM.01086-01010.
5. Safa A, Sultana J, Cam PD, Mwansa JC, Kong RYC (2008) *Vibrio cholerae* O1 Hybrid El Tor Strains, Asia and Africa. Emerg Infect Dis 14: 987-988.
6. Nair G, Qadri F, Holmgren J, Svennerholm A, Safa A, et al. (2006) Cholera due to altered El Tor strains of *Vibrio cholerae* O1 in Bangladesh. J Clin Microbiol 44: 4211-4213.
7. Nair GB, Faruque SM, Bhuiyan NA, Kamruzzaman M, Siddique AK, et al. (2002) New variants of *Vibrio cholerae* O1 biotype El Tor with attributes of the classical biotype from hospitalized patients with acute diarrhea in Bangladesh. J Clin Microbiol 40: 3296-3299.
8. Nguyen BM, Lee JH, Cuong NT, Choi SY, Hien NT, et al. (2009) Cholera outbreaks caused by an altered *Vibrio cholerae* O1 El Tor biotype strain producing classical cholera toxin B in Vietnam in 2007 to

2008. J Clin Microbiol 47: 1568-1571.
9. Ansaruzzaman M, Bhuiyan N, Nair B, Sack D, Lucas M, et al. (2004) Cholera in Mozambique, variant of *Vibrio cholerae*. Emerg Infect Dis 10: 2057-2059.
10. Quilici M-L, Massenet D, Gake B, Bwalki B, Olson DM (2010) *Vibrio cholerae* O1 variant with reduced susceptibility to ciprofloxacin, Western Africa. Emerg Infect Dis 16: 1804-1805.
11. Ceccarelli D, Salvia AM, Sami J, Cappuccinelli P, Colombo MM (2006) New cluster of plasmid-located class 1 integrons in *Vibrio cholerae* O1 and a *dfrA15* cassette-containing integron in *Vibrio parahaemolyticus* isolated in Angola. Antimicrob Agents Chemother 50: 2493-2499.
12. Wozniak RAF, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, et al. (2009) Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. PLoS Genet 5: e1000786. doi: 1000710.1001371/journal.pgen.1000786.
13. Seth-Smith H, Croucher NJ (2009) Genome watch: breaking the ICE. Nat Rev Microbiol 7: 328-329.
14. Waldor MK, Tschape H, Mekalanos JJ (1996) A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. J Bacteriol 178: 4157-4165.
15. Peters SE, Hobman JL, Strike P, Ritchie DA (1991) Novel mercury resistance determinants carried by IncJ plasmids pMERPH and R391. Mol Gen Genet 228: 294-299.
16. Ceccarelli D, Spagnoletti M, Bacciu D, Danin-Poleg Y, Mendiratta D, et al. (2011) ICE*Vch*Ind5 is prevalent in epidemic *Vibrio cholerae* O1 El Tor strains isolated in India. Int J Med Microbiol 301: 318–324.
17. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, et al. (2009) Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. Proc Natl Acad Sci USA 106: 15442-15447.
18. Colombo MM, Mastrandrea S, Leite F, Santona A, Uzzau S, et al. (1997) Tracking of clinical and environmental *Vibrio cholerae* O1 strains by combined analysis of the presence of toxin cassette, plasmid content and ERIC PCR. FEMS Immunol Med Microbiol 19: 33-45.
19. WHO (2007) Cholera 2006. Wkly Epidemiol Rec 31: 273-284.
20. Hochhut B, Lotfi Y, Mazel D, Faruque SM, Woodgate R, et al. (2001) Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins. Antimicrob Agents Chemother 45: 2991-3000.
21. Sharma C, Ghosh A, Dalsgaard A, Forslund A, Ghosh RK, et al. (1998) Molecular evidence that a distinct *Vibrio cholerae* O1 biotype El Tor strain in Calcutta may have spread to the African continent. J Clin Microbiol 36: 843-844.
22. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, et al. (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406: 477-483.
23. McGrath BM, O'Halloran JA, Piterina AV, Pembroke JT (2006) Molecular tools to detect the IncJ elements: a family of integrating, antibiotic resistant mobile genetic elements. J Microbiol Methods 66: 32-42.
24. Bani S, Mastromarino PN, Ceccarelli D, Van AL, Salvia AM, et al. (2007) Molecular characterization of ICE*Vch*Vie0 and its disappearance in *Vibrio cholerae* O1 strains isolated in 2003 in Vietnam. FEMS Microbiol Lett 266: 42-48.
25. Taviani E, Ceccarelli D, Lazaro N, Bani S, Cappuccinelli P, et al. (2008) Environmental *Vibrio spp*., isolated in Mozambique, contain a polymorphic group of integrative conjugative elements and class 1 integrons. FEMS Microbiol Ecol 64: 45-54.
26. Bhattacharya T, Chatterjee S, Maiti D, Bhadra RK, Takeda Y, et al. (2006) Molecular analysis of the *rstR*

and *orfU* genes of the CTX prophages integrated in the small chromosomes of environmental *Vibrio cholerae* non-O1, non-O139 strains. Environ Microbiol 8: 526-634.

27. Keasler SP, Hall RH (1993) Detecting and biotyping *Vibrio cholerae* O1 with multiplex polymerase chain reaction. Lancet 341: 1661.
28. Olsvik O, Wahlberg J, Petterson B, Uhlen M, Popovic T, et al. (1993) Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. J Clin Microbiol 31: 22-25.
29. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, et al. (1990) Current protocols in molecular biology; Sons JWa, editor. New York: Green Publishing Associates and Wiley.
30. Beaber JW, Hochhut B, Waldor MK (2002) Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. J Bacteriol 184: 4259-4269.
31. Burrus V, Marrero J, Waldor MK (2006) The current ICE age: biology and evolution of SXT-related integrating conjugative elements. Plasmid 55: 173-183.
32. Grim CJ, Hasan NA, Taviani E, Haley B, Chun J, et al. (2010) Genome sequence of hybrid *V. cholerae* O1 MJ-1236, B-33 and CIRS101 and comparative genomics with *V. cholerae*. J Bacteriol 192: 3524-3533.
33. Cho YJ, Yi H, Lee JH, Kim DW, Chun J (2010) Genomic evolution of *Vibrio cholerae*. Curr Opin Microbiol 13: 646-651.
34. Chatterjee S, Patra T, Ghosh K, Raychoudhuri A, Pazhani GP, et al. (2009) *Vibrio cholerae* O1 clinical strains isolated in 1992 in Kolkata with progenitor traits of the 2004 Mozambique variant. J Med Microbiol 58: 239-247.
35. Cappuccinelli P, Colombo MM, Morciano C (1995) Epidemiologia del colera in Africa. In: IILA, editor. Il colera, una malattia da debellare Il colera oggi in America Latina.
36. Dalsgaard A, Mortensen HF, Molbak K, Dias F, Serichantalergs O, et al. (1996) Molecular characterization of *Vibrio cholerae* O1 strains isolated during cholera outbreaks in Guinea-Bissau. J Clin Microbiol 34: 1189-1192.
37. Ceccarelli D, Spagnoletti M, Cappuccinelli P, Burrus V, Colombo MM (2011) Origin of *Vibrio cholerae* in Haiti. Lancet Infect Dis 11: 260.
38. Ghosh-Banerjee J, Senoh M, Takahashi T, Hamabata T, Barman S, et al. (2010) Cholera toxin production by the El Tor variant of *Vibrio cholerae* O1 compared to prototype El Tor and Classical biotypes. J Clin Microbiol 48: 4283-4286.

CHAPTER V

Origin of *Vibrio cholerae* in Haiti

Daniela Ceccarelli, Matteo Spagnoletti, Piero Cappuccinelli, Vincent Burrus, Mauro M Colombo

Dipartimento di Biologia e Biotecnologie Charles Darwin, Sapienza Università di Roma, Rome, Italy (DC,MS, MMC); Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, Italy (PC); Centre d'Étude et de Valorisation de la Diversité Microbienne, Département de Biologie, Université de Sherbrooke, Sherbrooke, QC, Canada (VB)

This chapter was published in The Lancet Infectious Diseases (2011)
doi:10.1016/S1473-3099(11)70078-0

Origin of *Vibrio cholerae* in Haiti

As pointed out in the December 2010 editorial,[1] the ongoing cholera outbreak in Haiti placed this diarrhoeal infectious disease at the forefront of the global public health agenda. As of Dec 3 2010 WHO reported 121518 cases, and 2591 deaths associated with cholera infection. Since Haiti was not previously affected by cholera during the current seventh pandemic its population is more susceptible to *Vibrio cholerae* infection. The epidemic strain responsible for the outbreak was identified as *V cholerae* O1 biotype El Tor, resistant to co-trimoxazole (trimethoprim-sulfamethoxazole), furazolidone, sulfafurazole, streptomycin, and nalidixic acid.[2]

We analysed the genome of three clonal isolates sequenced recently by the US Centers for Disease Control and Prevention (AELH00000000, AELI00000000, and AELJ00000000). The Haitian strains contain an integrative conjugative element (ICE) of the SXT/R391 family, a major drug resistance-spreading vector in bacteria, which is 99% identical to ICE*Vch*Ind5. This ICE, which confers resistance to co-trimoxazole, sulfafurazole, and streptomycin, was originally identified in strains of *V cholerae* isolated in India, which are also resistant to nalidixic acid, and clonally belong to the most prevalent epidemic clade in the Indian subcontinent, represented by the reference strain CIRS101.[3] The Haitian clone carries a genotype 7 *ctxB* gene coding for the cholera toxin subunit B.[4] This genotype was described only in an altered El Tor *V cholerae* variant isolated during the harsh cholera epidemic in Orissa, India, in 2007.[5]

Whole-genome alignment and comparative genomic analysis of the Haitian strains, with the representative *V cholerae* O1 variants from Central America and Indian subcontinent, confirmed that the Haitian strain is strictly phylogenetically related to CIRS101 from India. This strain is one of the highly virulent Indian *V cholerae* O1[6] that are gradually spreading all over the world; it is not surprising that this strain easily took advantage of the susceptibility of the Haitian people to the disease, and the poor sanitation caused by the earthquake in Haiti.

References

[1] The Lancet Infectious Diseases. As cholera returns to Haiti, blame is unhelpful. *Lancet Infect Dis* 2010; **10:** 813.

[2] CDC. Update: Cholera outbreak—Haiti, 2010. *MMWR Morb Mortal Wkly Rep* 2010; **59:** 1637–41.

[3] Ceccarelli D, Spagnoletti M, Bacciu D, et al. ICE*Vch*Ind5 is prevalent in epidemic *Vibrio cholerae* O1 El Tor strains isolated in India. *Int J Med Microbiol* 2011; published online Jan 26. DOI:10.1016/j.ijmm.2010.11.005.

[4] Safa A, Nair GB, Kong RYC. Evolution of new variants of *Vibrio cholerae* O1. *Trends Microbiol* 2010; **18:** 46–54.

[5] Goel AK, Jain M, Kumar P, Bhadauria S, Kmboj DV, Singh L. A new variant of *Vibrio cholerae* O1 El Tor causing cholera in India. *J Infect* 2008; **57:** 280–81.

[6] Ghosh-Banerjee J, Senoh M, Takahashi T, et al. Cholera toxin production by the El Tor variant of *Vibrio cholerae* O1 compared to prototype El Tor and classical biotypes. *J Clin Microbiol* 2010; **48:** 4283–86.

CHAPTER VI

Rapid detection by multiplex PCR of Genomic Islands, prophages and Integrative Conjugative Elements in *V. cholerae* 7th pandemic variants

Matteo Spagnoletti*, Daniela Ceccarelli, Mauro M. Colombo

Dipartimento di Biologia e Biotecnologie Charles Darwin, Sapienza Università di Roma, Via dei Sardi 70, 00185, Rome, Italy *Corresponding author.

Abstract

Vibrio cholerae poses a threat to human health, and new epidemic variants have been reported so far. Seventh pandemic *V. cholerae* strains are characterized by highly related genomic sequences but can be discriminated by a large set of Genomic Islands, phages and Integrative Conjugative Elements. Classical serotyping and biotyping methods do not easily discriminate among new variants arising worldwide, therefore the establishment of new methods for their identification is required. We developed a multiplex PCR assay for the rapid detection of the major 7th pandemic variants of *V. cholerae* O1 and O139. Three specific genomic islands (GI-12, GI-14 and GI-15), two phages (Kappa and TLC), *Vibrio* Seventh Pandemic Island 2 (VSP-II), and the ICEs of the SXT/R391 family were selected as targets of our multiplex PCR based on a comparative genomic approach. The optimization and specificity of the multiplex PCR was assessed on 5 *V. cholerae* 7th pandemic reference strains, and other 34 *V. cholera*e strains from various epidemic events were analysed to validate the reliability of our method. This assay had sufficient specificity to identify twelve different *V. cholerae* genetic profiles, and therefore has the potential to be used as a rapid screening method.

This chapter was published in Journal of Microbiological Methods (2011)
doi:10.1016/j.mimet.2011.10.017

Introduction

Cholera is still an emerging problem in developing countries and the recent harsh cholera epidemic in Haiti is a prominent example of this scenario [1]. It is well documented that the current 7th cholera pandemic is caused either by *V. cholerae* O1 biotype El Tor or *V. cholerae* O139 strains [2,3]. Recent studies demonstrated that the main *V. cholerae* O1 El Tor variants and the O139 strains emerged in India in 1992 and belong to a single phyletic line [4]. These variants are characterized by highly related genome sequences that differ by assorted combinations of genomic islands (GI), prophages, and Integrative Conjugative Elements (ICEs) [5]. Lately, this composed "mobilome" has been used to analyze recent *V. cholerae* evolution and distinguish among the different phyletic lines comprised in the *V. cholerae* 7th pandemic clade [6]. Indeed, the acquisition, loss, and rearrangement of six genomic elements such as *Vibrio* Seventh Pandemic Island 2 (VSP-II), prophage Kappa (also known as GI-11), GI-12, GI-14, GI-15, and prophage TLC were investigated.

VSP-II is a polymorphic GI, integrated in a methionine tRNA locus [7] and it is considered a molecular tag for 7th pandemic *V. cholerae* strains. Indeed, a novel rearrangement of this GI, showing a 14.4 kb deletion, was found in *V. cholerae* O1 El Tor strain CIRS101, which is dominant in recent isolates from Bangladesh [8]. GI-12, GI-14 and GI-15 were recognized as novel GIs able to discriminate, with their presence or absence, closely related atypical El Tor strains such as MJ1236 from Bangladesh and B33 from Mozambique [2,6]. Prophages Kappa and TLC are typically found in the genomes of *V. cholerae* epidemic strains, and the latter was recently recognized as an important player in *V. cholerae* toxigenic conversion [9].

Beside genomic islands, also the ICEs of the SXT/R391 family take part in the evolutionary process of *V. cholerae* [4]. SXT/R391 ICEs are self-transmissible mobile elements able to integrate into the host bacterial chromosome and to transfer by conjugation [10]. They are acknowledged for their important role in bacterial genome plasticity and as vectors of antibiotic resistance and alternative metabolic pathways [11]. The name of the SXT/R391 family originates from elements SXT and R391 that share a highly conserved genetic backbone encoding their integration/excision, conjugative transfer, and regulation. Each element contains specific genes scattered in five insertion sites (Hotspots) on the backbone [11]. We recently showed the existence of at least three conserved genetic ICE profiles circulating in 7th pandemic strains from the Indian Subcontinent [12]. The three profiles have their representatives in SXT, ICE*Vch*Ind5, and ICE*Vch*Ban9, that can be discriminated by the specific content of Hotspot IV (see Section 2.1 for further details) [11].

Due to the extensive lateral gene transfer that takes place in *V. cholerae*, Cho and colleagues [5] suggested that classically important diagnostic markers such as serotype and biotype are not reliable anymore and new methods are required to identify emerging epidemic strains. To address these shortcomings we developed a rapid screening method based on two multiplex PCRs aimed at characterizing the mobilome of *V. cholerae* O1 and O139. This method identifies GIs, phages and ICEs established as discriminative targets by comparative genomics [6], and represents a quick approach for 7th pandemic *V. cholerae* variants tracking.

Material and methods

Primer design

PCR primers were specifically designed to be used in multiplex PCR assays to avoid cross-dimer formation using Java web tools for PCR [13]. Information about primer sequences, PCR product sizes and their molecular targets are shown in Table 1. Bioinformatic analysis of the genes amplified in the two multiplex PCR assays is shown in supplementary figure S1. Oligonucleotides were obtained from PRIMM srl (Milano, Italy).

The first multiplex assay (six primer pairs) detects i) the presence and genetic rearrangements of *Vibrio* 7th Pandemic Island II (VSP-II), and ii) the presence and genetic rearrangements of SXT/R391 ICEs. Detection of VSP-II was obtained by primers VSPIIintF/VSPIIintR targeting VSP-II integrase. The specific VSP-II deletion peculiar of the Altered El Tor strain CIRS101 was detected by primer VSPIIcutF/VSPIIcutR. These primers target a specific gene (*vc0512* in the genome of N16961) present only in the prototypical VSP-II, giving an amplicon of 245bp. When the region is deleted, primers give no amplification as in the VSP-II arrangement of strain CIRS101. The identification of SXT/R391 ICEs was obtained by primers ICEdetF/ICEdetR targeting integrase int_{SXT}. ICE identity was then established by three additional primer pairs, targeting Hotspot IV of the three main ICE variants isolated in *V. cholerae*: i) SXTdetF/SXTdetR targeting gene *s060*, specific for Hotspot IV in SXT; ii) Ind5detF/Ind5detR targeting *vchind5-10*, specific for Hotspot IV in ICE*Vch*Ind5; and iii) Moz10detF/Moz10detR targeting genes *vchB33-4* and *orf65,* specific for Hotspot IV in ICE*Vch*Moz10 (sibling of ICE*Vch*Ban9) [11,12].

Target	Primer	Nucleotide sequence (5′ to 3′)	GenBank acc. num.	Nucleotide position	Fragment length (bp)
Multiplex 1					
VSP-II integrase	VSPIIintF	CCGACAAAGAATACACTCTCTCTGATGG	NC_002505	548835–548862	170
	VSPIIintR	ACGTCTTTTCCTTGCCTCGGCAAGAG		548979–549004	
Prototypical VSP-II	VSPIIcutF	TTATCTACGACCACACCAGACAGC	NC_002505	541391–541414	245
	VSPIIcutR	ATGGGCATAGCAAAGGCACTTACCCA		541610–541635	
SXT/R391 ICEs integrase	ICEdetF	TCAGTTAGCTGGCTCGATGCCAGG	GQ463142	4256–4279	505
	ICEdetR	GCAGTACAGACACTAGGCGCTCTG		3775–3798	
SXT Hotspot IV	SXTdetF	ACTTGTCGAATACAACCGATCATGAGG	AY055428	68416–68442	357
	SXTdetR	CAGCATCGGAAAATTGAGCTTCAAACTCG		68088–68116	
ICE*Vch*Ind5 Hotspot IV	Ind5detF	TGCACATTGAGGCCCTGCAAGCAC	GQ463142	66514–66537	423
	Ind5detR	GTGCATTCACCAGCTCTAACGTCG		66115–66138	
ICE*Vch*Moz10 Hotspot IV	Moz10detF	CGGAAGATGACGAAGACCGCCTAAGC	NZ_ACHZ01000011	15980–16005	712
	Moz10detR	ATTTGCCTTCGAACAAAAGGGGCA		15294–15317	
Multiplex 2					
TLC phage	TLCdetF	AATCAACTCACGGGTGCAGACCTC	NC_002505	1575253–1575276	449
	TLCdetR	TCCGCCAAGAAGTGACGTTGTAGC		541610–541635	
Kappa phage	KdetF	CGTCCGTAACCTTAAAGATGGCAGC	NZ_DS990136	1560575–1560599	230
	KdetR	TCGTATGTCCGTGAACTTGCCACC		1560781–1560804	
GI 12	GI12detF	CTACGGTTGAGCCGCTCCATTTGTC	NC_012668	1629546–1629570	571
	GI12detR	GTGCCTTCTAAATTGACCAAACGCGGCA		1630072–1630099	
GI 14	GI14detF	AGACGAGTATCTAGTAAACGCCAAACC	NC_012667	940275–940301	142
	GI14detR	CTTTGCTTGCACTGGCAACCTCAG		940393–940416	
GI 15	GI15detF	CAGACCGCGAAGGAAAACGCTCTTTGC	NZ_ACHZ01000008	53780–54937	348
	GI15detR	AGCGTCTCAGATGATGTCCGGCTG		54132–54155	

Table 1. DNA sequence and genome position of the primers designed for multiplex 1 and multiplex 2.

The second multiplex assay (five primer pairs) was developed to detect the presence of phage TLC, phage Kappa (GI-11), GI-12, GI-14, and GI-15. Amplification of the integrase genes of phage Kappa and GI15 (primers KdetF/KdetR and GI15detF/GI15detR, respectively), and amplification of unique genes of GI-12 (primers GI12detF/GI12detR) and GI-14 (primers GI14detF/GI14detR) were used as

discriminative targets. Although the rearrangement of the CTXΦ genetic cluster is not the aim of this analysis, detection of the CTXΦ associated cryptic phage TLC (primers TLCdetF/TLCdetR) can give preliminary information on the chromosomal position of CTXΦ. Indeed, it was reported that some 7th pandemic reference strains, such as B33 and MJ-1236 are devoid of the TLC element on chromosome 1, and carry multiple copies of CTXΦ on chromosome 2 [4,14].
Single reactions of the two multiplex PCR reactions were also performed to purify the amplification products and confirm their identity by sequencing.

DNA extraction and PCR conditions

Genomic DNA was prepared with a DNeasy Blood & Tissue Kit (Qiagen), according to manufacturer's instructions. After genomic extraction DNA concentration of each sample was measured using a spectrophotometer and adjusted to 50 ng/µL. The amplification was performed in an automated thermocycler (BioRad MJ-Mini), and PCR conditions were optimized as follows: a total reaction mix of 50µl contained 1.3U of DreamTaq DNA polymerase (Fermentas), 1X PCR Magnesium buffer, dNTP 250µM each, 0.3µM concentrations of each primer and 1 µl of genomic DNA template.
The optimized thermal cycling conditions for both multiplex PCRs were 35 cycles of denaturation at 94°C for 30 sec (2 min and 30 sec for the first cycle), annealing at 59°C (*multiplex* 1) or 64.5°C (*multiplex* 2) for 30 sec, and polymerization at 72°C for 20 sec (5 min for the last cycle). Amplified products (5 µl) were resolved by 1.5% agarose gel electrophoresis at 110 V for 1h. The gel was then stained with ethidium bromide, and the bands were visualized under UV illumination at 254 nm. Amplicons to be sequenced were directly purified from PCR by Nucleospin extract kit (Macherey-Nagel) or extracted from agarose gel by Rapid Gel extraction system (Marligen BioSciences) according to the manufacturer's instructions. DNA sequences were determined by Bio-Fab Research (Rome, Italy).

***V. cholerae* strains and media**

All *V. cholerae* strains analysed in this work are listed in Table 2. All bacterial strains were routinely grown on LB Agar at 37°C for 16-18h with appropriate antibiotic selection and were maintained at -80°C in LB broth containing 15% (vol/vol) glycerol. Five completely sequenced *V. cholerae* 7th pandemic reference strain were used to develop the multiplex PCR: *V. cholerae* O1 N16961 (acc. num. NC_002505), *V. cholerae* O1 CIRS101 (acc. num. NZ_ACVW00000000), *V. cholerae* O1 B33 (acc. num. NZ_ACHZ00000000), *V. cholerae* O1 MJ-1236 (acc. num. NC_012668), and *V. cholerae* O139 MO10 (acc. num. NZ_AAKF00000000). Thirty-four additional *V. cholerae* O1 El Tor strains belonging to our strain collection and isolated from different epidemic outbreaks were then screened to verify the specificity of our method and its ability to distinguish further profiles compared to the reference strains.

Results and discussion

Multiplex PCR optimization on 7th pandemic reference strains

To optimize and test the efficiency of our double multiplex PCR analysis we applied this approach to five completely sequenced reference strains (Table 2): *V. cholerae* O1 N16961 (prototype El Tor), *V. cholerae* O1 CIRS101 (Altered El Tor), *V. cholerae* O1 B33 (Mozambique variant), *V. cholerae* O1 MJ1236 (Matlab type I), and *V. cholerae* O139 MO10. Strains CIRS101, B33 and MJ-1236 are enclosed in the term "atypical El Tor" [2], and correspond to the main pandemic strains currently circulating in Asia and Africa [4].

As expected, combining the information obtained with the two multiplex assays, the five strains showed unique amplicon profiles corresponding to their specific mobilome, confirming the reliability of our method (Fig. 1). The five profiles belonging to the reference strains were named from A to E (Table 2). As an example, we here describe profile D of strain MJ-1236, that gave 7 specific amplicons (Fig. 1 lanes 4 and 10). Multiplex 1 gave the following amplicons: 170 bp (VSP-II integrase), 245 bp (prototypical VSP-II), 505 bp (int_{SXT}), and 712 bp (ICE*Vch*Ban9 Hotspot IV); multiplex 2 gave the amplicons: 142 bp (GI-14), 230 bp (prophage Kappa), and 571 bp (GI-12). As a proof of the susceptibility of our method, we were able to discriminate between the two closely related variants B33 and MJ-1236. They show the same amplicon profile in multiplex 1 (Fig. 1, lanes 3 and 4) corresponding to the presence of ICE*Vch*Moz10 and the prototypical VSP-II, but are distinguished in multiplex 2 by a different set of genomic islands: GI15 in strain B33 (Fig. 1 lane 9, 348bp amplicon) and GI12 in strain MJ-1236 (Fig. 1, lane 10, 571bp amplicon). Once the multiplex PCR method was optimized on the five reference strains we applied it to the screening of two distant epidemic events in India and Mozambique.

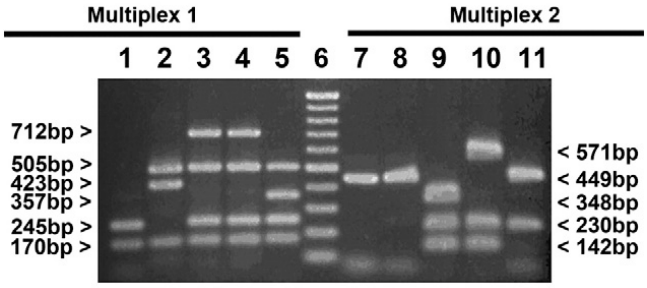


Figure 1. Amplification patterns showing PCR-amplified products in multiplex 1 and 2 for five reference strains. Banding profiles of multiplex 1 (lanes 1 to 5) and multiplex 2 (lanes 7 to 11) for each reference strain are shown (see Table 2 for complete profiles). Lanes 1 and 7, profile A (strain N16961, India 1975); lanes 2 and 8, profile B (strain CIRS101, Bangladesh 2002); lanes 3 and 9, profile C (strain B33, Mozambique 2004); lanes 4 and 10, profile D (strain MJ-1236, Bangladesh 1994); lane 5 and 11, profile E (strain MO10, India 1992). Lane 6, 100 bp molecular size marker.

Analysis of a *V. cholerae* Indian epidemic outbreak (1994-2005)

We analyzed 15 *V. cholerae* O1 clinical strains isolated from 1994 to 2005 at the Mahatma Gandhi Institute of Medical Science, situated in the rural area of Sevagram (Wardha province, Maharashtra, India). These strains were previously characterized as *V. cholerae* O1 El Tor all bearing ICE*Vch*Ind5 [12].

Combining the results from multiplex 1 and multiplex 2 we obtained two amplicon profiles: 14 out of 15 strains showed a new profile F (Fig. 2, lanes 1 and 11, representative strain 7452 is shown), and only one strain (25959) showed profile B as reference strain CIRS101 (Fig. 2, lanes 2 and 12). Profile F was associated to strains holding ICE*Vch*Ind5 and the prototypical VSP-II. No phage K or genomic islands GI-12, GI-14 and GI-15 were detected. Profile B of strain 25959 differed from profile F only for the presence of a truncated version of VSP-II (Fig. 2, lane 2, 245 bp amplicon missing) which is typical of CIRS101 genome (Table 2).

The study of the Indian epidemic revealed the appearance of the deleted version of VSP-II in one strain isolated in 2005, while all previous strains showed the same prototypical VSP-II profile as for *V. cholerae* strain N16961. Taviani and colleagues were able to detect the shift between 'old' and 'new' pandemic clones of *V. cholerae* O1 by tracking VSP-II variants in Bangladesh [8]. The role of this deletion is still unclear, but seems to be stable in recent *V. cholerae* O1 Altered El Tor strains like the recently isolated ones in Haiti [1,15].

			Multiplex 1		Multiplex 2					Final profile
Strain	Serotype	Isolation	ICE	VSP-II	Prophage Kappa	GI-12	GI-14	GI-15	Prophage TLC	
Reference										
N16961	O1	India – 1975	–	+	–	–	–	–	+	A
CIRS101	O1	Bangladesh – 2002	ICE*Vch*Ind5	+ (truncated)	–	–	–	–	+	B
B33	O1	Mozambique – 2004	ICE*Vch*Moz10[a]	+	+	–	+	+	–	C
MJ-1236	O1	Bangladesh – 1994	ICE*Vch*Ban9[a]	+	+	+	+	–	–	D
MO10	O139	India – 1992	SXT	+	+	–	–	–	+	E
India										
7452	O1	1994	ICE*Vch*Ind5	+	–	–	–	–	+	F
8909	O1	1997	ICE*Vch*Ind5	+	–	–	–	–	+	F
8636	O1	1997	ICE*Vch*Ind5	+	–	–	–	–	+	F
9906	O1	1998	ICE*Vch*Ind5	+	–	–	–	–	+	F
6884	O1	1998	ICE*Vch*Ind5	+	–	–	–	–	+	F
10693	O1	1998	ICE*Vch*Ind5	+	–	–	–	–	+	F
13571	O1	2000	ICE*Vch*Ind5	+	–	–	–	–	+	F
11075	O1	2001	ICE*Vch*Ind5	+	–	–	–	–	+	F
14100	O1	2002	ICE*Vch*Ind5	+	–	–	–	–	+	F
16102	O1	2003	ICE*Vch*Ind5	+	–	–	–	–	+	F
16770	O1	2003	ICE*Vch*Ind5	+	–	–	–	–	+	F
18245	O1	2004	ICE*Vch*Ind5	+	–	–	–	–	+	F
23701	O1	2005	ICE*Vch*Ind5	+	–	–	–	–	+	F
24585	O1	2005	ICE*Vch*Ind5	+	–	–	–	–	+	F
25959	O1	2005	ICE*Vch*Ind5	+ (truncated)	–	–	–	–	+	B
Mozambique										
7698	O1	Tete – 1998	ICE*Vch*Moz10	+	+	–	+	+	+	G
7798	O1	Tete – 1998	ICE*Vch*Moz10	+	+	–	+	+	+	G
7898	O1	Tete – 1998	ICE*Vch*Moz10	+	+	–	+	+	+	G
71c2	O1	Zambezia – 1998	ICE*Vch*Moz10	+	+	–	+	+	+	G
9398	O1	Zambezia – 1998	ICE*Vch*Moz10	+	+	–	+	+	+	G
9298	O1	Zambezia – 1998	ICE*Vch*Moz10	+	+	–	+	+	+	G
9198	O1	Zambezia – 1998	ICE*Vch*Moz10	+	+	–	+	+	+	G
9598	O1	Zambezia – 1998	ICE*Vch*Moz10	+	+	–	+	+	+	G
101	O1	Zambezia – 1998	ICE*Vch*Moz10	+	+	–	+	+	+	G
5556	O1	Maputo – 1997	ICE*Vch*Moz10	+	+	–	+	+	–	C
5806	O1	Maputo – 1997	ICE*Vch*Moz10	+	+	–	+	+	–	C
6127	O1	Maputo – 1997	ICE*Vch*Moz10	+	+	–	+	+	–	C
7036	O1	Maputo – 1997	ICE*Vch*Moz10	+	+	–	+	+	–	C
5594	O1	Maputo – 1997	ICE*Vch*Moz10	+	+	–	+	+	–	C
Other profiles										
VC20	O1	India – 1992	–	+	–	–	–	+	+	H
CO840	O1	India – 1995	ICE*Vch*Ind5	+	–	–	–	–	–	I
VC1	O1	Vietnam – 2003	–	+	–	–	–	–	–	J
VC90	O1	Vietnam – 1990	ICE*Vch*Vie0	+	+	–	–	–	–	K
806	O1	Italy – 1994	ICE*Vch*Moz10	+	+	–	–	+	–	L

Table 2. General characteristics, specific genetic element content and final profile of the *V. cholerae* strains used in this study.

Analysis of a *V. cholerae* Mozambique epidemic outbreak (1997-1998)

The second analysis was performed on 14 *V. cholerae* O1 strains isolated from the epidemic outbreak occurred in 1997-1998 in three different regions of Mozambique: Tete (North-West), Zambezia (Centre), and Maputo (South).

We were able to recover two main profiles. Nine strains showed a new profile named G (Fig. 2, lanes 3 and 13), and five strains showed profile C (Fig. 2, lanes 4 and 14). Both profiles are consistent with the presence of a prototypical VSP-II and a sibling ICE of ICE*Vch*Moz10, the presence of phage Kappa, GI-14 and GI-15, and with the absence of GI-12 (Table 2). Profile G holds an additional 449 bp amplicon (Fig. 2, lane 13) indicating the presence of prophage TLC.

Interestingly, only the five strains isolated in the region of Maputo showed the same profile C as reference strain B33 indicating the absence of prophage TLC (Table 2). This finding is of particular importance because it revealed that the atypical variant B33, isolated for the first time in the central province of Beira in 2004 [16], was already present in the area of Maputo in 1997.

On the other hand, strains from the regions of Tete and Zambezia showed profile G, characterized by the presence of prophage TLC on chromosome 1. Since TLC is invariably associated to CTXΦ prophage [17], we further investigated the presence of CTXΦ and associated genetic clusters, finding an RS1-RS1 array between TLC prophage and RTX chromosomal gene (data not shown). Even if the analysis of the CTXΦ region is not the aim of this study, this arrangement was never described in atypical El Tor strains isolated in Africa. Indeed, a strain with the same genetic organization was found only in the atypical El Tor strain MG116926 isolated in Bangladesh in 1991 [18].

Unexpected *V. cholerae* 7th pandemic profiles

Since the profiles we obtained with our approach were mostly similar to known *V. cholerae* 7th pandemic variants, to test the sensitivity of our method on unpredicted genomic profiles, we further investigated our strain collection, identifying five *V. cholerae* O1 El Tor strains with new GI, ICE, and prophage profiles (Table 2, profiles H to L). Strain VC20, isolated in India in 1992 [19], showed the presence of prototypical VSP-II, GI15, and TLC, but no ICE of the SXT/R391 family was identified (profile H, Fig. 2, lanes 5 and 15). Strain CO840, an atoxigenic strain from India isolated in 1995 [20], was characterized by ICE*Vch*Ind5 and prototypical VSP-II, but no GI12, GI14, GI15, TLC or phage Kappa were detected (profile I, Fig. 2, lanes 6 and 16). Strain VC1 isolated in Vietnam in 2003 [21] showed the presence of the prototypical VSP-II but no ICE of the SXT/R391 family, prophages or other GIs were identified (profile J, Fig. 2, lanes 7 and 17). Strain VC90 from Vietnam, isolated in a local epidemic outbreak in 1990 [21], showed the presence of the prototypical VSP-II and int_{SXT} in multiplex 1 and prophage Kappa in multiplex 2 but had no amplification for any of the other GIs and phages (profile K, Fig. 2, lanes 8 and 18). As for Hotspot IV content, the absence of any of the three expected amplicons suggests a possible genetic rearrangement of this region. Finally, strain 806 isolated in the South of Italy in 1994 showed another peculiar profile with the prototypical VSP-II and an ICE sibling of ICE*Vch*Moz10, genomic island GI-15, and phage Kappa (profile L, Fig. 2, lanes 9 and 19). Profiles H to L significantly differed in their mobile genetic

content from the five 7th pandemic reference strains. The reliability of our method was validated by the identification of new profiles associated to a variety of isolates from Asia, Africa and Europe.

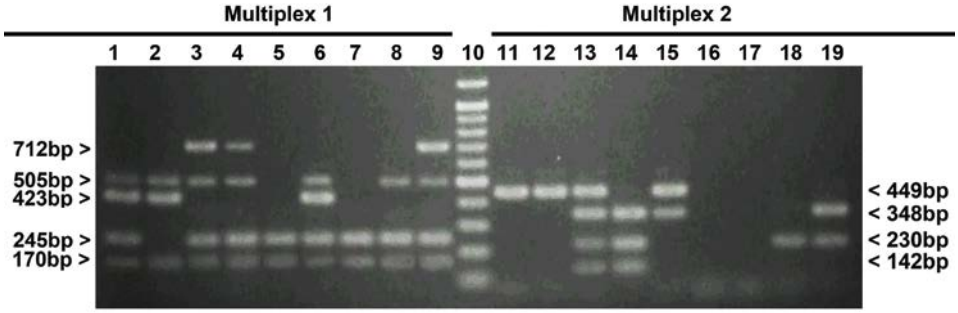
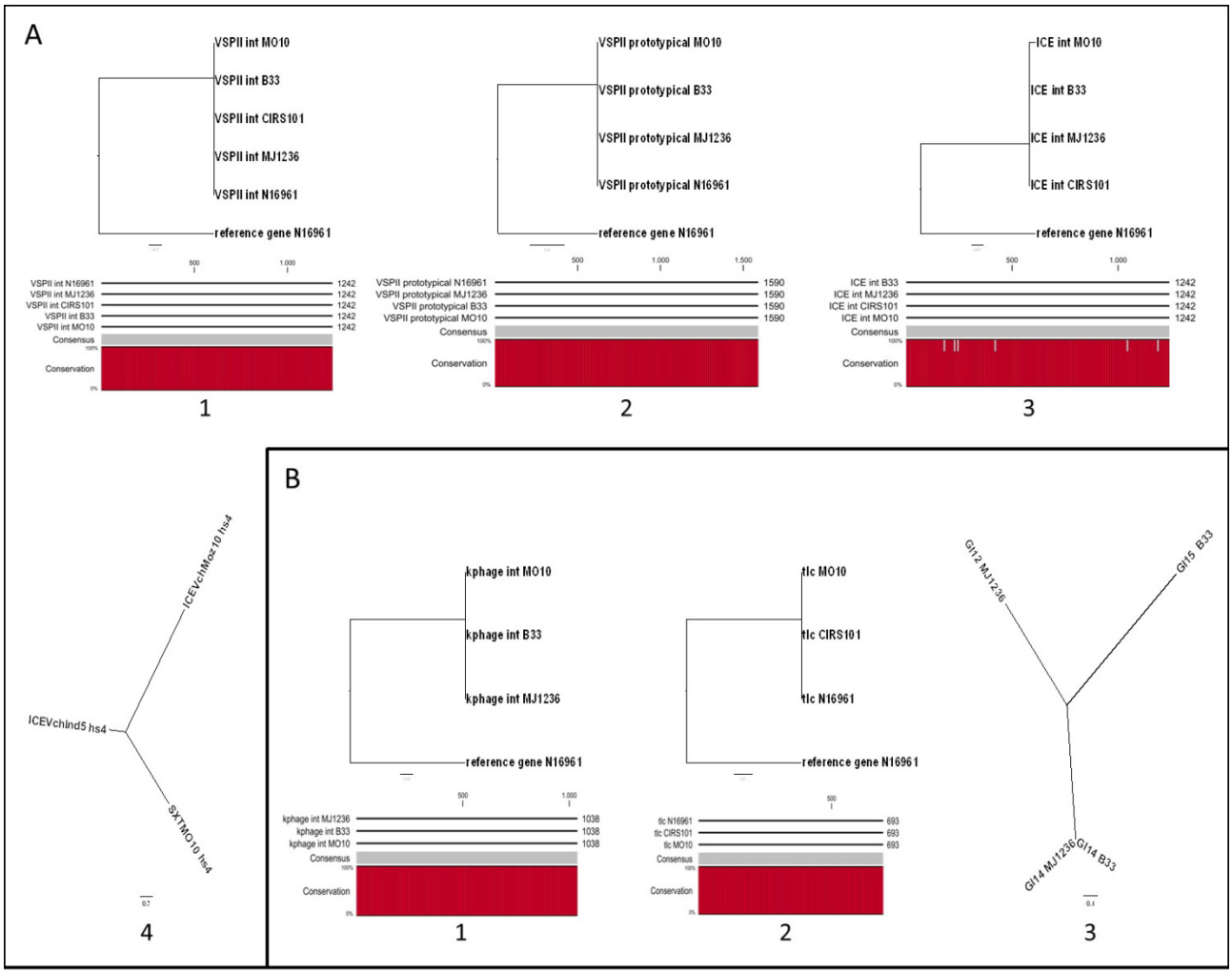


Figure 2. Amplification patterns showing PCR-amplified products in multiplex 1 and 2 for representative Indian, Mozambican and other epidemic *V. cholerae* strains. Banding profiles of multiplex 1 (lanes 1 to 9) and multiplex 2 (lanes 11 to 19) for each strain are shown (see Table 2 for complete banding profiles). Lanes 1 and 11, profile F (strain 7452, India 1994); lanes 2 and 12, profile B (strain 25959, India 2005); lanes 3 and 13 profile G (strain 7698, Mozambique 1998); lanes 4 and 14, profile C (strain 5556, Mozambique 1997); lanes 5 and 15, profile H (strain VC20, India 1992); lanes 6 and 16, profile I (strain CO840, India 1995); lanes 7 and 17, profile J (strain VC1, Vietnam 2003); lanes 8 and 18, profile K (strain VC90, Vietnam 1990); lanes 9 and 19, profile L (strain 806, Italy 1994). Lane 10, 100bp molecular size marker.

Conclusions

The double multiplex PCR assay we developed is a rapid molecular method able to detect genetic elements that characterize the main *V. cholerae* seventh pandemic variants. This approach is advantageous for quick source tracking, showing sensitivity and specificity. It can be used to analyze a large number of samples in a short period of time, it can identify new genomic arrangements, and it could become an useful tool for cholera epidemic surveillance. Therefore, from an epidemiological and public health point of view, our approach can be very functional in giving direct information on the genetic background of the new epidemic variants.

To date, there are no other options to better investigate *V. cholerae* 7th pandemic mobilome but total genome sequencing, which remains an expensive and time consuming procedure. The effectiveness of this method was demonstrated by its application to two different outbreaks in India and in Mozambique, which gave evidence of the rearrangement of mobile genetic content of these strains. Atypical El Tor variants are increasingly reported worldwide [22,23] and may contribute to the increasing number of cholera cases reported to WHO in the last 10 years (http://www.who.int/wer/en/).



Supplementary figure S1. Bioinformatics analysis of the genes amplified in the two multiplex PCR assays. The comparison was performed using the genomic sequences of five selected V. cholerae 7th pandemic reference strains (see Table 2). The analysis of the genes amplified in multiplex 1 and multiplex 2 are shown in panel A and B, respectively. **Panel A:** 1, VSPII (GeneID: 2615808); 2, VSPII prototypical (GeneID: 2615804); 3, SXT/R391 ICEs integrase (GeneID: 21885342); 4, SXT/ICEVchInd5/ICEVchMz10 hotspot IV, (GeneID: 21885308; 259156420; 229509096).

Panel B: 1, prophage Kappa (GeneID: 7855560); 2, prophage TLC (GeneID: 2613973); 3, GI-12/GI-14/GI-15 (GeneID: 7856472; 7853980; 229508872). Phylogenetic trees were generated from the sequence alignments of the selected homologous genes using ClustalW algorithm. An unrelated DNA sequence (rRNA 16S gene of V. cholerae strain N16961) was used as outgroup. ClustalW sequence alignments of the same genes (without the 16S reference sequence) are show below every tree. A red scale-bar depicts the sequence conservation level. For the discriminative genes that identify GI-12/GI-14/GI-15 (panel A, figure 4) and the three different ICE profiles (panel B, figure 3) phylogenetic trees are represented using a radial tree. Since each gene is unique among the selected reference strains, sequence alignments for these trees are not shown (GI-14 integrase is an exception because is present in both B33 and MJ-1236 reference strains). Phylogenetic trees were generated using the neighbor-joining method as implemented by MEGA 4.1 software, and were edited using FigTree 1.31 (http://tree.bio.ed.ac.uk/software/figtree/).

Acknowledgments

This research was funded by Ministero dell' Istruzione, dell' Università e della Ricerca - Italy (PRIN 2007) and Ministero Affari Esteri, DGCS AID 9397 support to Biotechnology Center in Mozambique. We are grateful to R. Valia for manuscript revision. MS is the recipient of a PhD fellowship from the Doctorate School in Cellular and Developmental Biology, Sapienza Università di Roma. DC was supported by a fellowship from Institute Pasteur - Fondazione Cenci Bolognetti, Italy.

References

1. Ceccarelli D, Spagnoletti M, Cappuccinelli P, Burrus V, Colombo MM (2011) Origin of *Vibrio cholerae* in Haiti. Lancet Infect Dis 11: 260.
2. Safa A, Nair GB, Kong RYC (2010) Evolution of new variants of *Vibrio cholerae* O1. Trends Microbiol 18: 46-54.
3. Ramamurthy T, Yamasaki S, Takeda Y, Nair GB (2003) *Vibrio cholerae* O139 Bengal: odyssey of a fortuitous variant. Microbes Infect 5: 329-344.
4. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, et al. (2009) Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. Proc Natl Acad Sci USA 106: 15442-15447.
5. Cho YJ, Yi H, Lee JH, Kim DW, Chun J (2010) Genomic evolution of *Vibrio cholerae*. Curr Opin Microbiol 13: 646-651.
6. Grim CJ, Hasan NA, Taviani E, Haley B, Chun J, et al. (2010) Genome sequence of hybrid *V. cholerae* O1 MJ-1236, B-33 and CIRS101 and comparative genomics with *V. cholerae*. J Bacteriol 192: 3524-3533.
7. O'Shea YA, Finnan S, Reen FJ, Morrissey JP, O'Gara F, et al. (2004) The *Vibrio* seventh pandemic island-II is a 26.9 kb genomic island present in *Vibrio cholerae* El Tor and O139 serogroup isolates that shows homology to a 43.4 kb genomic island in *V. vulnificus*. Microbiology 150: 4053-4063.
8. Taviani E, Grim CJ, Choi J, Chun J, Haley B, et al. (2010) Discovery of novel *Vibrio cholerae* VSP-II genomic islands using comparative genomic analysis. FEMS Microbiol Lett 308: 130-137.
9. Hassan F, Kamruzzaman M, Mekalanos JJ, Faruque SM (2010) Satellite phage TLCphi enables toxigenic conversion by CTX phage through dif site alteration. Nature 467: 982-985.
10. Wozniak RAF, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol 8: 552-563.
11. Wozniak RAF, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, et al. (2009) Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. PLoS Genet 5: e1000786. doi: 1000710.1001371/journal.pgen.1000786.
12. Ceccarelli D, Spagnoletti M, Bacciu D, Danin-Poleg Y, Mendiratta D, et al. (2011) ICE*Vch*Ind5 is prevalent in epidemic *Vibrio cholerae* O1 El Tor strains isolated in India. Int J Med Microbiol 301: 318–324.
13. Kalendar R (2010) Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analyses. [http://primerdigital.com/tools/]
14. Faruque SM, Tam VC, Chowdhury N, Diraphat P, Dziejman M, et al. (2007) Genomic analysis of the Mozambique strain of *Vibrio cholerae* O1 reveals the origin of El Tor strains carrying classical CTX prophage. Proc Natl Acad Sci USA 104: 5151-5156.

15. Chin C-S, Sorenson J, Harris JB, Robins WP, Charles RC, et al. (2011) The origin of the haitian cholera outbreak strain. New Eng J Med 364: 33-42.
16. Ansaruzzaman M, Bhuiyan N, Nair B, Sack D, Lucas M, et al. (2004) Cholera in Mozambique, variant of *Vibrio cholerae*. Emerg Infect Dis 10: 2057-2059.
17. Rubin EJ, Lin W, Mekalanos JJ, Waldor MK (1998) Replication and integration of a *Vibrio cholerae* cryptic plasmid linked to the CTX prophage. Mol Microbiol 28: 1247-1254.
18. Lee JH, Choi SY, Jeon YS, Lee HR, Kim EJ, et al. (2009) Classification of hybrid and altered *Vibrio cholerae* strains by CTX prophage and RS1 element structure. J Microbiol 47: 783-788.
19. Chakraborty S, Mukhopadhyay AK, Bhadra RK, Ghosh AN, Mitra R, et al. (2000) Virulence genes in environmental strains of *Vibrio cholerae*. Appl Environ Microbiol 66: 4022-4028.
20. Garg P, Nandy RK, Chaudhury P, Chowdhury NR, De K, et al. (2000) Emergence of *Vibrio cholerae* O1 biotype El Tor serotype Inaba from the prevailing O1 Ogawa serotype strains in India. J Clin Microbiol 38: 4249-4253.
21. Bani S, Mastromarino PN, Ceccarelli D, Van AL, Salvia AM, et al. (2007) Molecular characterization of ICE*Vch*Vie0 and its disappearance in *Vibrio cholerae* O1 strains isolated in 2003 in Vietnam. FEMS Microbiol Lett 266: 42-48.
22. Ceccarelli D, Spagnoletti M, Bacciu D, Cappuccinelli P, Colombo M (2011) New *V. cholerae* atypical El Tor variant emerged during the 2006 epidemic outbreak in Angola. BMC Microbiol 11: 130.
23. Taneja N, Mishra A, Sangar G, Singh G, Sharma M (2009) Outbreaks caused by new variants of *Vibrio cholerae* O1 El Tor, India. Emerg Infect Dis 15: 352-354.

CHAPTER VII

Genomic analysis of an atypical ICE in *Vibrio cholerae* strain MZO-3

Elisa Taviani[1]*, Matteo Spagnoletti[2]*, Daniela Ceccarelli[2], Bradd J. Haley[1], Nur A. Hasan[1], Arlene Chen[1], Mauro M Colombo[2], Anwar Huq[1], and Rita R. Colwell[3,4*]

***These authors contributed equally to this work**

1 Maryland Pathogen Research Institute, University of Maryland College Park, MD 20742, USA **2** Dipartimento Biologia e Biotecnologie Charles Darwin, Sapienza Università di Roma, Rome, Italy. **3** John Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA **4** Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD 20742, USA

Abstract

Genomic islands (GIs) and integrative conjugative elements (ICEs) are major players in bacterial evolution since they encode genes involved in adaptive functions of medical or environmental importance. Here we performed the genomic analysis of ICE*Vch*Ban8, an unusual ICE found in the genome of a clinical non toxigenic *Vibrio cholerae* O37 isolate. ICE*Vch*Ban8 shares most of its genetic structure with SXT/R391 ICEs. However, this ICE codes for a different integration/excision module, is located at a different insertion site, and part of its genetic cargo shows homology to other pathogenicity islands of *V. cholerae*.

Manuscript in preparation

Introduction

Recent comparative genomics studies revealed that prokaryotic genomes are dynamic entities capable of acquiring and discarding large amounts of genetic material via lateral gene transfer (LGT) and recombination [1]. The role of LGT in the evolution of *Vibrio cholerae*, an autochthonous inhabitant of riverine and marine environments as well as a human pathogen, is extensively documented [1-4]. In this species, major virulence genes and several important adaptive functions are known to be clustered in regions of the chromosome laterally acquired from conspecific or distantly related organisms [5,6]. The extent of LGT was confirmed by the presence of Integrative Conjugative Elements (ICEs) and a large set of Genomic Islands (GIs) in *V. cholerae* sequenced genomes, contributing to genomic plasticity and generating a significantly heterogeneous group of strains in this species [1,4].

GIs are not self-transmitting gene clusters, since they are devoid of transfer genes, but they can excise and form circular intermediates [7-10]. They are of significant interest since they often encode genes of clinical importance, such as virulence factors otherwise known as pathogenicity islands (PAI) [9]. Among all the *V. cholerae* associated PAIs, *Vibrio* pathogenicity island 2 (VPI-2) is present in toxigenic O1 strains, and occasionally in non-O1 non-O139 *V. cholerae* [11,12]. VPI-2 is a 57.3 kb island integrated at a *tRNA-Ser*, encoding 52 ORFs including proteins for release, transport, and catabolism of sialic acid [12].

Integrative conjugative elements (ICEs) are a class of mobile elements integrated into the prokaryotic host chromosome. They comprise mosaic genetic element with plasmid-, phage-, and transposon-like features organized in functional modules and are capable of self-transfer from a donor to a recipient cell via conjugation [13]. These elements were first described in *V. cholerae* with the discovery of the SXT element [14], and have since been found in the majority of *V. cholerae* and related *Vibrio* spp. [15,16]. All ICEs that have been described in *V. cholerae* to date belong to the SXT/R391 family [17], which share a conserved genetic scaffold of 52 genes that incorporates unique sequences coding for resistance to antibiotics and heavy metals, new toxin/antitoxin systems, restriction/modification systems, and alternative metabolic pathways [16,17].

In this study, we report the genomic analysis of ICE*Vch*Ban8, an unusual ICE present in the genome of the non-toxigenic clinical isolate *V. cholerae* O37 MZO-3, collected in Bangladesh in 2001 [1,16]. Unlike most *V. cholerae* non-O1 non-O139 isolates, strains of serogroup O37 have been shown to have epidemic potential [18,19]. *V. cholerae* MZO-3 lacks CTXΦ [1] but shows the presence and/or major rearrangements of two of the main pathogenicity islands found in *V. cholerae* 7th pandemic isolates: a novel version of *Vibrio* seventh pandemic island 2 (VSP-2) [20] and the replacement of VPI-II with ICE*Vch*Ban8, which is the subject of the present work.

Materials and methods

ICE assembly, annotation and comparative genomics

Nucleotide sequence of ICE*Vch*Ban8, was obtained from the genome sequence of *V. cholerae* O37 strain MZO-3 (AAUU00000000) [1]. Gaps between two contigs were closed by manual editing using Consed and subjected to custom primer walk and PCR amplification. Complete nucleotide sequence of ICE*Vch*Ban8 was deposited in GenBank under accession no. JQ345361.

ICE genetic analysis was accomplished in four steps: (i) ICE*Vch*Ban8 nucleotide sequence was aligned with SXT using NUCmer [21]; (ii) ORFs were identified and annotated using the RAST annotation pipeline (http://rast.nmpdr.org); (iii) ICE sequence and genetic organization was compared with published ICEs using the Artemis Comparative Tool (ACT) (www.sanger.ac.uk/Software/ACT); and (iv) similarity in nucleotide and protein sequences for ICE*Vch*Ban8 was determined as % nucleotide or amino acid identity with other ICEs in GenBank employing BLASTN and BLAST-PSI [22].

PCR assays

To confirm accurate assembly of the two contigs and to close gaps between them we designed different primers: pMZO_1F/pMZO3_1R for contig 1 assembly; pMZO3_1F2 /pMZO3_54R for contigs 1 and 54 junction. EnterotoxinB-F/s063R primer pair was designed to localize the 3'-end of Hotspot 4 insertion and AcriF/AcriR to confirm the presence of the acriflavine resistance gene (Supplementary Table 1). Selected amplicons were sequenced using Applied Biosystems DNA sequencer 3730.

Results

Genomic organization of ICE*Vch*Ban8

Comparative analysis of the genome of *V. cholerae* O37 MZO-3 revealed sequences belonging to the genetic backbone of ICE*Vch*Ban8 on two contigs: nt. 27439–105212 of contig 1 (accession no. NZ_AAUU01000001.1), and nt. 1–25409 of contig 54 (accession no. NZ_AAUU01000001.54). The overall genetic organization of this element was already identified as highly similar to SXT/R391 ICEs, and named ICE*Vch*Ban8 [16]. We chose to retain the proposed nomenclature although it does not belong to the SXT/R391 family of ICEs. Assembly of the ICE was accomplished using NUCmer, with SXTM010 as reference, revealed one gap at the junction between the two contigs that was resolved by PCR and confirmed by sequencing (Fig. 1). Our analysis revealed a 103.381 bp sequence. The complete sequence of ICE*Vch*Ban8 was submitted to RAST pipeline for annotation and resulted in the identification of 106 ORFs. Annotation was then refined manually comparing it to SXT^{M010} and other ICEs of the SXT/R391 family found in *V. cholerae* [16].

ICE*Vch*Ban8 holds the same highly conserved core set of genes responsible for transfer and regulation of integration/excision in the SXT/R391 ICEs backbone (Fig. 1). Four operons of *tra* genes encoding the conjugative apparatus (*traIDJ*, *traLEKBA*, *traCFWUN*, and *traFHG*) share at least 97% similarity at the nucleotide level with the *tra* clusters in the conserved SXT/R391 ICEs backbone. As noted for all SXT related ICEs, specific inserted genes were identified in five hotspots and four variable regions within the core backbone (Fig. 1) [16]. In ICE*Vch*Ban8, Hotspot 1 and

Hotspot 2 have the same molecular arrangement as in SXT, with an additional gene coding for a hypothetical protein in the latter. Hotspot 3, and Variable Region I, II and III do not have any insertions. At the 3' end of the element, the variable region IV is disrupted by the rearrangement of *setR/eex* backbone genes. Overall, ICE*Vch*Ban8 holds some major rearrangements in its genetic structure, as the substitution of the integration/excision module, the localization of the regulation module and the presence of a large cluster of genes in Hotspot 4, never described before.

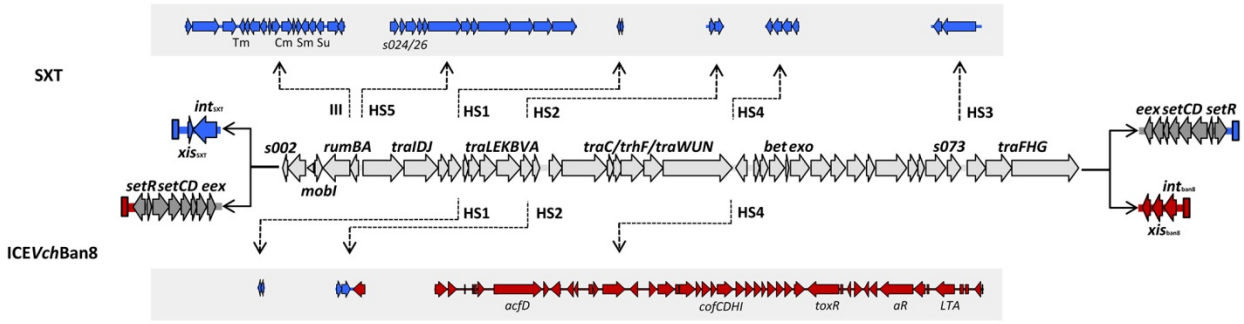


Figure 1. Structural comparison between SXT and ICE*Vch*Ban8. In the middle is a genetic map of core genes shared between the two ICEs (light gray). Specific regions for SXT and ICE*Vch*Ban8 are depicted in blue and red, respectively. In dark gray: inverted genes *setR*/*eex*. Hotspot insertions are indicated by dotted arrows and shown above (SXT) and below (ICE*Vch*Ban8) the shared backbone.

ICE*Vch*Ban8 integration/excision and regulation modules

To date, all ICEs described in *V. cholerae* belong to the SXT/R391 family, the distinguishing traits of which are the presence of a conserved integrase (int_{SXT}) and the insertion at the *prfC* locus [17]. Analysis of ICE*Vch*Ban8 insertion site into MZO-3 genome showed its insertion in an alternative locus, a *tRNA-Ser*, leaving the *prfC* gene intact. In epidemic *V. cholerae* strains the *tRNA-Ser* locus is the insertion site of the VPI-2 [23]. *V. cholerae* O37 strain MZO-3 lacks VPI-2 and is replaced by the ICE under study. The two 23bp inverted sequences (*attL* and *attR*) flanking the element were almost identical to the flanking sequences of VPI-2. Furthermore, alignment of the nucleotide sequence of the flanking genes showed that insertion of new DNA occurs at the same base pair for both VPI-2 and ICE*Vch*Ban8, suggesting the two elements insert via a similar recombination event and by similar integrase activity.

Located at the 3'-end of ICE*Vch*Ban8 we found a P4-like integrase gene, int_{BAN8}, different from the int_{SXT}. Int_{BAN8} showed low similarity with the integrase genes of *Vibrio parahaemolyticus* RIMD 2210633 and *V. cholerae* VPI-2 integrase. At the amino acid level, IntBAN8 was 75% and 74% similar to the phage integrases of *Shewanella sp*. MR-7 and *V. cholerae* VPI-2, respectively. Notably, there was no significant similarity at the amino acid level with any of the ICE integrases. In a recent study, Boyd and colleagues [8] reported results of a phylogenetic analysis of integrases distributed among different bacterial groups. Their finding showed that, with only a few exceptions, GI's integrases cluster together and separate from phage, plasmid, integron, and ICE integrases. They had included int_{BAN8} and showed that it did not cluster with SXT-like *int*, but instead within the GI group, yet on a separate branch from the *V. cholerae* VPI-2 integrases [8].

Downstream from the integrase gene, there is a small ORF of 213nt showing no similarity at the nucleotide level with other ORFs in the databases. Considering its position and length, we interpreted the role of this gene as a putative recombination directionality factor (RDF). Low conservation at the nucleotide level is considered a general feature of RDF as described for the *xis* gene encoded by SXT/R391 ICEs, which shares only general biochemical characteristics with other RDF [24,25]. At the amino acid level, Blast-PSI revealed an *alpA* superfamily domain and significant similarity with several proteins annotated as *alpA* phage transcriptional regulators. Notably, the putative RDF showed 47% (28/60) similarity with *vefA* of *V. cholerae* N16961 VPI-2, recently recognized as the main RDF involved in the excision of VPI-2 [26]. This ORF was also 55% (30/55) similar to the putative RDF of MGI*Vch*Usa1, a newly described genetic element in *V. cholerae* mobilized by SXT/R391 ICEs [27]. This suggests the gene functions as RDF during ICE*Vch*Ban8 excision and was annotated as xis_{BAN8}, and belongs to a large family of proteins functionally related to the RDF of *E. coli* prophage *alpA*, potentially acquired and exchanged by lateral gene transfer by phages, GI and ICEs.

The regulation module (*setR/eex*) of ICE*Vch*Ban8 showed a major rearrangement since is unusually located at the 5'-end of the element, with *setR* being the first ORF of the element. This peculiar genetic arrangement differs from the conserved backbone of SXT related ICEs, in which the *setCDR* operon is located at the 3'-end of the element (Fig. 1).

Figure 2. Muscle alignment of the predicted amino acid sequences of xis_{BAN8}, *vefA* of VPI2 from *V. cholerae* N16961 and putative RDF of MGI*Vch*USA1 from *V. cholerae* RC385. Amino acids conserved in all sequences are shown in red; amino acids conserved in two of three sequences are shown in pink. Gaps and amino acids divergent in all three sequences are shown in blue. Consensus sequence is depicted below the alignment.

Hotspot 4 organization

Hotspot 4, located between genes *traN* and *s063*, is characterized by a unique 49.5 kb region that includes 45 ORFs (Supplementary Table 2). Overall this hotspot merited more detailed analysis. Bioinformatics analysis of its genetic content revealed several ORFs related to known virulence determinants and pathogenic cellular properties.

Cellular transport. HS4 encodes for a replication protein and 3 membrane permeases containing ABC-transporter domain common in OmpA/B protein found in several *Vibrio* species. Furthermore, three putative accessory colonization factors with similarities with Acf in other Vibrios are interspersed in the hotspot.

Type IV pilus assembly operon. It contains 12 ORFs (*tcpA* to *toxT*) with similarities to proteins involved in the formation of pili such as the Longus, the CFA III of ETEC, the toxin-coregulated pilin (TCP) of *V. cholerae*, and the bundle-forming pilin (BFP) found in enteropathogenic *E. coli* (EPEC). These structures are known to be involved in essential features for pathogenicity such as

twitching, motility/adherence to host cells, biofilm formation, cell signaling, invasion, and evasion of the immune system.

Metallopeptidase. A putative ORF of 2922 bp was annotated as a *tagA* gene, encoding a ToxR-activated A protein with low similarity to gene *stcE* found in plasmids of *E. coli* O157:H7 and *Shigella boydii.* StcE belongs to the family of metallopeptidases involved in the pathogenesis of these organisms.

Acriflavine resistance family protein. This ORF is 71-78 % similar to AcrB/AcrD/AcrF family proteins involved in acriflavine resistance. In *V. cholerae* MZO3, the transcriptional regulator usually associated with this gene is truncated.

CtxA. This 1806 bp ORF contains a functional domain found in enterotoxins as the heat labile (LT) toxin in *E. coli* and CTX subunit A in *V. cholerae*. The second part of the protein has a weak similarity with PspC and CbpA surface proteins involved in immune response activation and host-pathogens relationship of *Streptococcus pneumoniae* in Pneumococcal Bacteremia.

Putative hemolysin. Hypothetical protein with 72% similarity with the *Vibrio mimicus* thermostable direct hemolysin gene.

Discussion

Based on the results reported here, ICE*Vch*Ban8 displayed unique genetic organization yet conserving 47 of the 52 core genes of SXT/R391 ICEs. However, the ICE is integrated at a different locus of the *V. cholerae* genome than the *prfC* locus and has an integrase gene not related to int_{SXT}. Therefore, ICE*Vch*Ban8 could not be classified as a member of the SXT/R391 family that includes all of the ICEs that have been described in *V. cholerae* to date. Moreover the presence of a new integration/excision module (*xis*/*int*) and the complete inversion and rearrangement of the regulation cluster (*eex*/*setR*), are sufficient to stimulate questions about its evolution and classification.

We propose a hypothetical model to elucidate the mechanism that might have lead to the formation of this hybrid element. The model is organized into four subsequent steps (see Fig. 3 for the detailed mechanism), with an initial recombination event involving acquisition of new genetic material into the variable region IV of a prototype SXT/R391 ICE. This initial recombination event is followed by the substitution of the integration/excision module, an anomalous rearrangement and excision of the ICE and the reintegration in a different site. Even though this model is speculative, it is supported by the well-documented ability of the SXT/R391 ICEs to incorporate new genetic material into conserved hotspots and recombine during the integration/excision process [28]. It is also known that SXT/R391 ICEs encode an additional recombination machinery involved in hybrid formation [28] which is conserved in the ICE*Vch*Ban8 backbone. In contrast to ICE*Vch*Ban8, recombination events in other SXT/R391 ICEs maintain synteny of the backbone without loss of core genes [16].

ICE*Vch*Ban8 holds a GI integration module and is inserted in a*t-RNA-Ser.* This peculiar organization suggests that recombination events may have generated this 'hybrid' element compromising the boundaries between the different classes of mobile elements. Previous studies investigating the

correlation between GIs and ICEs have shown the pathogenicity islands of *Legionella pneumophila*, *Agrobacterium tumefaciens,* and *Bartonella tribocorum*, to be either ICEs or defective ICEs [29]. The hypothesis that some GIs are derived from ICEs that have lost their transfer ability is also supported by the discovering of ICE*Ec*1, an ICE found in a strain of *E. coli* [30]. ICE*Ec*1 includes the high pathogenicity island HPI, previously described in *Yersinia*, but with an additional transfer module able to self-mobilize by conjugation with the entire element [30].
It is concluded that mobile elements previously believed to be divergent in evolution may have instead followed the same path in selected genomes, and that recombination can occur between elements otherwise assigned to different families.

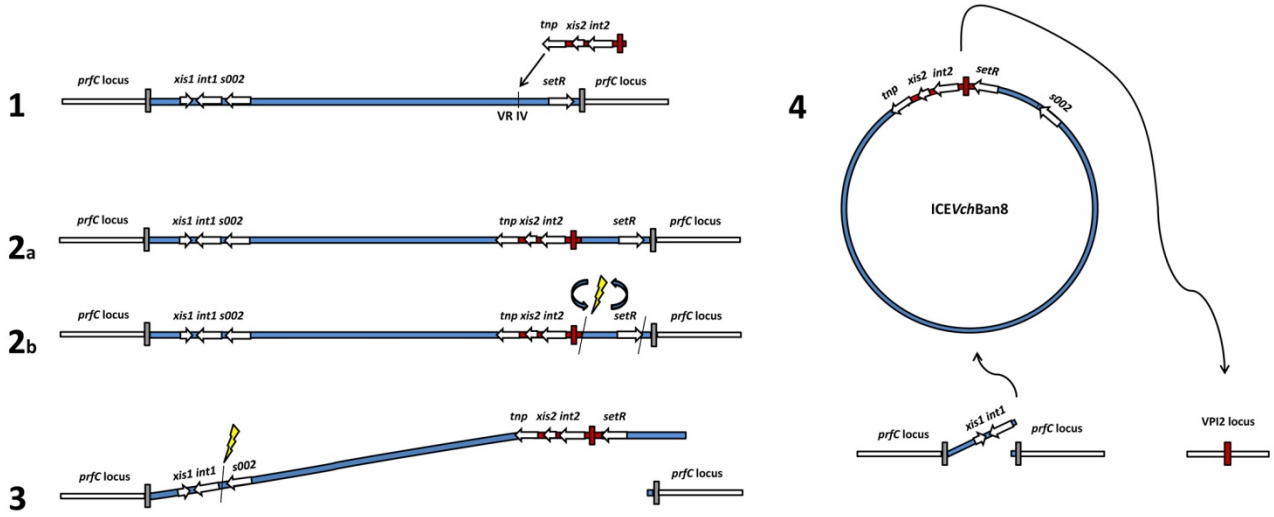


Figure 3. Hypothetical model of ICE*Vch*Ban8 formation. 1) A fragment of an unknown mobile genetic element (shown in red) transposes into VR IV of the SXT/R391 precursor of ICE*Vch*Ban8 (shown in blue), with the same mechanism responsible for acquisition of the new genetic material in the hotspots. **2a or 2b**) During transposition or in an unknown subsequent event, the 3' region of ICE*Vch*Ban8 precursor undergoes inversion rearrangement. This inversion causes loss of the old *attR* site, triggering excision of the rearranged element from the *prfC*locus. **3)** During abnormal excision, the ICE loses the 5' region containing the original integration module (int_{SXT}). This event could also result from competition between two integration modules in the linear or circular form. **4)** Reintegration of the newborn ICE*Vch*Ban8 in VPI2 locus is catalyzed by the new integrase (int_{Ban8}).

Acknowledgments

Support for the research was provided by National Institutes of Health grant no. 1 R01 A139129-01 (RRC) and by Ministero dell' Istruzione, dell' Università e della Ricerca – Italy (MMC).
ET received a PhD graduate assistentship from the Graduate School of University of Maryland, USA. MS is the recipient of a PhD fellowship from the Doctorate School in Cellular and Developmental Biology, Sapienza Università di Roma. DC was supported by a fellowship from Institute Pasteur - Fondazione Cenci Bolognetti, Italy.

References

1. Chun, J. et al. (2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. Proc. Natl. Acad. Sci. U.S.A. 106, 15442.
2. Colwell, R. and Spira, W. (1992) The ecology of *Vibrio cholerae.* In Cholera (Barua, D. and Greenough, W.I., ed.^eds), pp. 107-127. Plenum Medical Book Co., New York.
3.Faruque, S., Albert, M. and Mekalanos, J. (1998). Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae.* Microbiol Mol Biol Rev 62, 1301-1314.
4. Kaper, J.B., Morris, J.G.J. and Levine, M.M. (1995). Cholera. Clin. Microbiol. Rev. 8, 48-86.
5. Davis, B.M. and Waldor, M.K. (2000). CTXphi contains a hybrid genome derived from tandemly integrated elements. Proc Natl Acad Sci U S A 97, 8572-7.
6. Cho, Y.-J., Yi, H., Lee, J.H., Kim, D.W. and Chun, J. Genomic evolution of *Vibrio cholerae*. Current opinion in microbiology 13, 646-651.
7. Kaper, J.B. and Hacker, J. (1999) Pathogenicity Islands and Other Mobile Virulence Elements. pp. 366. Amer Soc for Microbiol, Washington, D.C.
8. Boyd, E.F., Almagro-Moreno, S. and Parent, M.A. (2009). Genomic islands are dynamic, ancient integrative elements in bacterial evolution. Trends in microbiology 17, 47-53.
9. Dobrint, U., Hochhut, B., Hentschel, U. and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol 2, 414-424.
10. Hentschel, U. and Hacker, J.r. (2001). Pathogenicity islands: the tip of the iceberg. Microbes and Infection 3, 545-548.
11. Jermyn, W.S. and Boyd, E.F. (2002). Characterization of a novel Vibrio pathogenicity island (VPI-2) encoding neuraminidase (*nanH*) among toxigenic *Vibrio cholerae* isolates. Microbiol 148, 3681-93.
12. Murphy, R.A. and Boyd, E.F. (2008). Three Pathogenicity Islands of *Vibrio cholerae* Can Excise from the Chromosome and Form Circular Intermediates. Journal of Bacteriology 190, 636-647.
13. Wozniak, R.A.F. and Waldor, M.K. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Micro 8, 552-563.
14. Waldor, M.K., Tschape, H. and Mekalanos, J.J. (1996). A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. J Bacteriol 178, 4157-65.
15. Taviani, E., Ceccarelli, D., Lazaro, N., Bani, S., Cappuccinelli, P., Colwell, R.R. and Colombo, M.M. (2008). Environmental Vibrio spp., isolated in Mozambique, contain a polymorphic group of integrative conjugative elements and class 1 integrons. FEMS Microbiol Ecol 64, 45-54.
16. Wozniak, R.A.F. et al. (2009). Comparative ICE Genomics: Insights into the Evolution of the SXT/R391 Family of ICEs. PLoS Genet 5, e1000786.
17. Burrus, V., Marrero, J. and Waldor, M.K. (2006). The current ICE age: biology and evolution of SXT-related integrating conjugative elements. Plasmid 55, 173-83.
18. Kamal, A.M. (1971). Outbreak of gastroenteritits by non-agglutinable (NAG) vibrios in the Republic of Sudan. J Egypt Public Health Assoc 46, 125-173.
19. Aldova, E., Lazickova, K., Stepankova, E. and Lietava, J. (1968). Isolation of noagglutinable vibrios from an enteritis outbreak in Czechoslovakia. J. Infect. Dis. 118, 25-31.
20. Taviani, E., Grim, C.J., Choi, J., Chun, J., Haley, B., Hasan, N.A., Huq, A. and Colwell, R.R. Discovery of novel *Vibrio cholerae* VSP-II genomic islands using comparative genomic analysis. FEMS Microbiology Letters 308, 130-137.

21. Delcher, A.L., Salzberg, S.L. and Phillippy, A.M. (2002) Using MUMmer to Identify Similar Regions in Large Sequence Sets. In Current Protocols in Bioinformatics ed.^eds). John Wiley & Sons, Inc.
22. Altschul, S.F., Madden, T.L., SchÃ¤ffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25, 3389-3402.
23. Murphy, R.A. and Boyd, E.F. (2008). Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates. J Bacteriol 190, 636-47.
24. Burrus, V. and Waldor, M.K. (2003). Control of SXT integration and excision. J Bacteriol 185, 5045-54.
25. O'Halloran, J.A., McGrath, B.M. and Pembroke, J.T. (2007). The *orf4* gene of the enterobacterial ICE, R391, encodes a novel UV-inducible recombination directionality factor, Jef, involved in excision and transfer of the ICE. FEMS Microbiology Letters 272, 99-105.
26. Almagro-Moreno, S., Napolitano, M. and Boyd, E. Excision dynamics of Vibrio pathogenicity island-2 from *Vibrio cholerae*: role of a recombination directionality factor *VefA*. BMC microbiology 10, 306.
27. Daccord, A., Ceccarelli, D. and Burrus, V. Integrating conjugative elements of the SXT/R391 family trigger the excision and drive the mobilization of a new class of Vibrio genomic islands. Molecular Microbiology 78, 576-588.
28. Garriss, G.v., Waldor, M.K. and Burrus, V. (2009). Mobile Antibiotic Resistance Encoding Elements Promote Their Own Diversity. PLoS Genet 5, e1000775.
29. Burrus, V. and Waldor, M. (2004). Shaping bacterial genomes with integrative and conjugative elements. Res Microbiol 155, 376-386.
30. Schubert, S., Dufke, S., Sorsa, J. and Heesemann, J. (2004). A novel integrative and conjugative element (ICE) of *Escherichia coli*: the putative progenitor of the Yersinia high-pathogenicity island. Molecular Microbiology 51, 837-848.

CHAPTER VIII

Concluding remarks

Concluding remarks

In recent years, the incredible step forward made by whole-genome sequencing technologies and the availability of new bioinformatic tools allowed major insights into the genomic evolution of *V. cholerae* [1]. It has become increasingly evident that several important adaptive functions of *V. choleare* are clustered in genomic regions that have been acquired from other conspecific or distantly related organisms, and that lateral gene transfer plays a crucial role in the emergence of new pathogenic strains [2,3]. Comparative genomic studies demonstrated that all seventh pandemic strains share a conserved genomic backbone diversified mostly by Lateral Gene Transfer (LGT) [2]. Indeed, the transition between different strains is characterized by the acquisition of genomic islands, prophages and Integrative Conjugative Elements [2]. The so-called Atypical El Tor variants, which appeared and spread from the late 1980s, are mostly characterized by several rearrangements in the CTXΦ region, as well as by the presence of novel genomic islands missing in the genome of previous El Tor strains [4]. Another remarkable feature of both *V. cholerae* O1 Atypical El Tor and the derivative *V. cholerae* O139 is the presence of ICEs of the SXT-R391 family in their genomes.

During my PhD I contributed to the understanding of the recent evolution of *V. cholerae* by studying the genetics of the SXT-R391 family of ICEs and by investigating the epidemiology of *V. cholerae* in different epidemic events occurred in the last 20 years. Furthermore, I used a comparative genomic approach to develop a rapid tool for the identification of mobile genetic elements which characterize the major *V. cholerae* seventh pandemic variants.

In Chapter 1, comparative analyses of 13 ICE sequences revealed that they have an identical genetic structure, consisting of syntenous highly conserved core genes that are interrupted by clusters of diverse variable genes. Unexpectedly, many genes in the core backbone proved non-essential for ICE transfer. Comparisons of the variable gene content in the ICEs under study, revealed that these elements are mosaics shaped by inter-ICE recombination.

An extreme result of this mechanism is likely the reason of ICE*Vch*Ban8 formation, discussed in Chapter 7. This element cannot be classified as an SXT-like ICE but is closely related to the SXT/R391 family, as proved by the presence of 47 out of 52 SXT/R391 backbone genes. Interestingly, ICE*Vch*Ban8 holds putative virulence factors that could be associated with potential pathogenic traits of *V. cholerae* non-O1, non-O139 strains.

One of the most intriguing aspects of SXT-R391 ICE biology is to elucidate their role in promoting the host fitness. Despite the antibiotic resistance cluster is one of the most noticeable features of the analyzed ICEs, its mere presence cannot explain by itself the spread and persistence of ICEs in the bacterial population. Moreover, certain ICEs do not encode for drug resistance genes, while others seem to have lost these genes during their recent evolution. Bioinformatic analysis revealed that several predicted ICE-encoded functions are related to cellular mechanisms responsible for an improved environmental fitness. Restriction modification systems appear to be involved in resistance to phages, which is one of the main factors influencing bacterial population trends. Several cellular messengers are involved in different pathways regulating motility and

biofilm formation and can likely give a selective advantage to the microorganisms in the environment [5,6].

Taken together, a number of ICE-encoded genes suggest a possible ecological role associated to SXT-R391 ICEs. This might confer a positive selection to the *V. cholerae* strain and to the maintenance of the ICE as a part of the genetic asset. Conversely, for a number of ICE-encoded genes, it is still not clear whether and how, they can be beneficial to the host strain.

New studies will be required to better understand the processes that led to the current diversity in this family of ICEs and the effects that these elements may have on the host fitness. Beside SXT/R391 ICEs genetic structure, investigating the epidemiology of *V. cholerae* in different epidemic events was also a main aim of my PhD thesis.

The presence of 3 different groups of SXT/R391 ICEs in the genome of *V. cholerae* epidemic strains circulating in the Indian Subcontinent was pointed out in Chapter 3. This association reflected the different lineages in the seventh pandemic clade described by Chun *et al* based on a full genome sequencing approach [2]. It was also demonstrated that ICE*Vch*Ind5 is associated with the prevalent Atypical El tor variant circulating in the Indian Subcontinent, which has its reference in the completely sequenced strain CIRS-101 [3].

Chapter 4 focused on the dreadful 2006 cholera outbreak in Angola where it was demonstrated that the responsible *V. cholerae* strain was clonally and genetically different from the El Tor strains circulating in the 1990s in the same area. This new strain can be classified as Atypical El Tor variant sharing the same genomic features as the prevalent clone circulating in the Indian Subcontinent (described in Chapter 3), demonstrating the recent spread of this pandemic clone from India to Africa. Finally, in chapter 5 a comparative genomics approach was used to analyze the complete genomes of three *V. cholerae* strains responsible of the harsh cholera outbreak occurred in Haiti in 2010. Whole-genome phylogenetic analysis revealed that these strains were closely related to the same pandemic clone circulating in India and Angola, i.e. CIRS-101 [3].

The genomic organization of *V. cholerae* 7th pandemic strains was the subject of the last part of my work, reported in Chapter 6. Indeed, I developed a rapid molecular tool to identify the discriminative mobile genetic elements associated to the 7th pandemic clade by multiplex PCR. From an epidemiological and public health point of view, this approach can be very functional in giving direct information on the genetic background of the new epidemic variants. To date, there are no other options to better investigate *V. cholerae* 7th pandemic mobilome but total genome sequencing, which remains an expensive and time consuming procedure.

The major outcome of these works was to notice that one of the Atypical El Tor variants is replacing all the other epidemic *V. cholerae* strains worldwide, as already suggested by other scientists [3]. A key question is why this pandemic clone is so successful? The success of Atypical El Tor variants seems to depend mostly by the combination of the respective predominant features of the El Tor and Classical biotypes: a better survival in the environment [4] and the expression of a more virulent toxin [7]. A noticeable feature of the prevalent pandemic clone is the presence of a truncated VSP-II pathogenicity island, together with the absence of specific phages and Genomic Islands that may have an ecological significance for other Atypical El Tor variants, perhaps in the form of reduced fitness [3]. The presence of ICE*Vch*Ind5 in its genome is another significant point

of distinction, although also *V. cholerae* O139 and other Atypical El Tor variants hold ICEs of the SXT-R391 family. As discussed above, the functional role of ICEs in *V. cholerae* is still far from being fully understood. Further studies are needed to better elucidate the processes that led to the current diversity in the SXT-R391 family, and the effects that ICEs and other mobile genetic elements may have on the fitness of *V. cholerae*.

References

1. Cho YJ, Yi H, Lee JH, Kim DW, Chun J (2010) Genomic evolution of *Vibrio cholerae*. Curr Opin Microbiol 13: 646-651.
2. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, et al. (2009) Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. Proc Natl Acad Sci USA 106: 15442-15447.
3. Grim CJ, Hasan NA, Taviani E, Haley B, Chun J, et al. (2010) Genome sequence of hybrid *V. cholerae* O1 MJ-1236, B-33 and CIRS101 and comparative genomics with *V. cholerae*. J Bacteriol 192: 3524-3533.
4. Safa A, Nair GB, Kong RYC (2010) Evolution of new variants of *Vibrio cholerae* O1. Trends Microbiol 18: 46-54.
5. Bordeleau E, Brouillette E, Robichaud N, Burrus V (2010) Beyond antibiotic resistance: integrating conjugative elements of the SXT/R391 family that encode novel diguanylate cyclases participate to c-di-GMP signalling in *Vibrio cholerae*. Environ Microbiol 12: 510–523.
6. Wozniak RAF, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol 8: 552-563.
7. Ghosh-Banerjee J, Senoh M, Takahashi T, Hamabata T, Barman S, et al. (2010) Cholera toxin production by the El Tor variant of *Vibrio cholerae* O1 compared to prototype El Tor and Classical biotypes. J Clin Microbiol 48: 4283-4286.