ABSTRACT

Title of Document: GENOMICS AND ECOLOGY OF INTEGRATIVE CONJUGATIVE ELEMENTS (ICE) IN *VIBRIO CHOLERAE*

Elisa Taviani, PhD, 2011

Directed By: Prof. Rita Colwell, UMIACS

Lateral gene transfer (LGT) is recognized as a major contributor in bacterial evolution. In this study, the importance of LGT was assessed by analyzing a class of mobile elements, Integrative Conjugative Elements (ICE), and two genomic islands, Vibrio Seventh Pandemic islands (VSP-I and VSP-II) in the bacterial species, *Vibrio cholerae.* The objective was to understand how acquisition of heterologous genetic material by *V. cholerae* has influenced its differentiation and adaptation to different niches in the aquatic environment, as well as its role as a human pathogen. ICEs are a class of self-transmissible mobile elements that mediate LGT via conjugation. Members of this class of mobile elements have been isolated from several species of

Gram-negative and Gram-positive bacteria, and have been observed to serve as vectors of drug resistance, virulence factors and genetic tools modulating ecological adaptation in *V. cholerae*.

In this study, two new ICEs discovered in seventh pandemic *V. cholerae* O1 El Tor strains associated with cholera epidemics in Asia and Africa, were analyzed. Overall, comparative analysis of the ICEs revealed an epidemiological relationship between new and old pandemic clones in cholera endemic areas. Furthermore, analysis of a novel ICE found in a clinical isolate of *V. cholerae* O37 revealed an unusual genetic organization and also association with a major pathogenicity island in *V. cholerae,* offering evidence of ICEs as progenitors and/or potential dissemination tools for PAIs, likely associated with the pathogenic potential of non-O1/O139 *V. cholerae*. It is concluded from an extensive survey of the geographical distribution of ICEs that they are present in environmental *V. cholerae* populations worldwide.

In summary, through analysis of ICEs and VSP-I and II it was possible to demonstrate the extent to which environmental *V. cholerae,* autochthonous to the aquatic environment worldwide, comprise a reservoir of mobile genetic elements. It is clear that mobile genetic elements have undergone significant molecular rearrangement, with a significant range of profiles, confirming their genetic plasticity. Finally, their variable content most likely contributes significantly to genomic evolution responsible for adaptation of this bacterial host to new ecological niches.

GENOMICS AND ECOLOGY OF INTEGRATIVE CONJUGATIVE ELEMENTS (ICE) IN *VIBRIO CHOLERAE*

By

Elisa Taviani

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor Rita R. Colwell, Chair
Professor Anwar Huq
Professor Sam Joseph
Professor James Kaper
Dr. Ivor Knight

Table of Contents

List of figures.

List of tables

Chapter 1. Lateral gene transfer and bacterial evolution.

"Wherever there is life there are bacteria" (Cohan, 2002). Bacteria are found in every environment, with or without eukaryotes, and no animal or plant in the natural environment is known to be free of bacterial commensals or pathogens. Therefore, microorganisms are the largest reservoir of biodiversity on earth (Konstantinidis, 2006) and are also the most ancient organisms on the planet (Staley, 2006). Although of great scientific interest and the object of intense research, microorganisms are not easy to study and many aspects of their biology are still poorly understood.

For taxonomy purposes, bacterial diversity is clustered as for higher organisms: individuals are grouped into discontinuous clusters on the basis of phenotypic, ecological and genomic characteristics (Cohan, 2007). However, "The bacterial species definition remains a very difficult issue to advance" (Konstantinidis, 2006). With increasing amounts of genetic information being obtained from whole genome sequencing, several examples can be offered that illustrate the intra-specific diversity that often reflects physiological or ecological differences of strains belonging to a single species (Lan and Reeves, 2000; Cohan, 2006; Staley, 2006; Kostantinidis *et al*., 2006). In almost all known species, the individual organisms

have different metabolic capabilities that indicate different ecological niches and coexist in the same geographic region (Mayr, 1982; Cohan, 2001). Traditional taxonomy based on phenotypic trait differences provides assemblages (species) in which is encountered significant divergence at the level of phenotype, genotype and ecology (Lan and Reeves, 2000; Cohan, 2001, 2006; Staley, 2006; Kostantinidis, 2006). The species *Vibrio cholerae*, for example, encompasses more than 200 serogroups, pathogenic and non-pathogenic strains, and significant genomic variability that accounts, at leas in part, for the extensive phenotypic and ecological diversity of this species (Thompson, 2004; Chun *et al.*, 2009). Functional analysis of the different serovars of *Salmonella enterica* has also revealed that most of the gene content differences involve metabolic pathways, not only virulence (Lan and Reeves, 2000). In contradiction to the notion of prokaryotic evolution being driven by clonality and periodic selection is the idea of evolution through "recombinational generation of gene combinations" (Gogarten *et al.*, 2002). By employing multilocus sequence typing (MLST), the role of homologous recombination and lateral gene transfer (LGT) in shaping genomic diversity among isolates has gained recognition sufficient to stimulate some investigators to consider networks, rather than trees, to represent lineages present in many bacterial species (Gogarten *et al.*, 2002). Darwinian evolution is traditionally viewed as cycles of gradual changes shaping organisms from a common ancestor to individuals evolved through accumulation of mutations. In contrast, changes introduced by lateral transfer are quicker and of bigger consequences, like the evolution of a pathogen from harmless organisms after the acquisition of virulence gene clusters via LGT. It is important to note that the

study of microbial genomes has validated the highly significant influence of LGT in the evolution of microorganisms and introduced the idea of a gene pool present in environmental reservoirs that assorts stochastically in individual genomes (Lawrence and Hendrickson, 2008). Therefore, the molecular variation observed in a species is the expression of a combination of genes belonging to the species gene pool (Gogarten *et al*., 2002; Chun *et al.*, 2009).

This model raises an important issue in the task of classification. Theoretically, LGT can remove the barriers between bacterial taxa, permitting genetic information and its ecological and functional counterparts to move freely between such barriers. In other words, if LGT is an important mechanism underlying genome evolution, it is necessary to assess the degree to which homologous recombination and gene transfer can or cannot abolish clonal history in each species. Interestingly, progress obtained from comparative genomic studies has revealed that, although genetic transfer occurs at a high level, such genetic flow does not lead to phylogenetic chaos (Gogarten *et al*., 2002). There are molecular and genetic constraints that affect both donor and recipient and rules, not completely understood yet, govern LGT, resulting in the phylogenetic hierarchy that is observed in the microbial world (Lawrence *et al.*, 2005).

Role of recombination in bacteria.

The main feature that accounts for the difference between evolution of prokaryotes and higher organisms is the mode of reproduction. Multicellular organisms, with rare exception, reproduce sexually and recombination is the fundamental means by which parental genomes rearrange their genomic assets as are observed in the new generation. This is the basis of genetic variability and evolution in these organisms and the limiting force that constrains diversity within a species (Cohan, 2002). Bacteria, in contrast, reproduce by binary fission in which recombination does not take place (Cohan, 2002). Recombination in bacteria occurs in two different ways. Homologous recombination is strictly related to sequence similarity. In contrast, site specific recombination occurs between short signature sequences that are present on mobile genetic elements and bacterial chromosomes and recognized by a specific class of recombinases. Plasmids, phages, transposons, integrons and ICEs are the mechanisms by which genetic information is moved among organisms and can be viewed as the molecular tools for HTG in bacteria (Gogarten *et al.*, 2002).

Recombination in bacteria is promiscuous because microorganisms can exchange genetic material within and between distantly related taxa, while plants and animals lose the ability to exchange genetic material when divergence in their mitochondrial DNA is >3% (Cohan, 2002). Bacteria can undergo homologous recombination with organisms that differ by 25% in DNA sequence (Cohan, 2002).

However, in bacteria the recombination rate decreases with donor-recipient genetic divergence (Cohan, 2002).

Although several studies have shown that recombination is responsible for prokaryotic genetic diversity (Gogarten *et al.*, 2002), other authors assign more or less importance to it as a driving force in bacterial speciation. In particular, one model considers speciation to be driven primarily by the extent of recombination, thus describing species as a "group of ecologically heterogeneous individuals" (Gogarten *et al.*, 2005). In contrast, another model follows the pattern established in population genetics and ecology, whereby species are described as a group of ecologically identical organisms that is irreversibly separated (Cohan, 2002: Gevers *et al.*, 2005).

Clonality and periodic selection: the ecotype model

The importance of the ecotype model resides in the broad theoretical speculation that its supporters created by infusing systematics and evolutionary biology with sequence based approaches that comprise the ecotype model (Cohan *et al.*, 2006; Gevers *et al.*, 2005). The underlying idea is that species "are more than only clusters of closely related organisms" (Cohan, 2002) and deserve dynamic properties that make them fundamental units of ecology and evolution (Cohan 2002, 2006; Gevers *et al.*, 2005). At the same time the recombination rate is considered to be "extremely rare", compared to the mutation rate, and is highly localized, as well as affecting only discrete part of the chromosome (Cohan, 2002, 2006; Gevers *et al.*, 2005). The evolution of organisms is thus represented as a cycle of alternating events of mutation

followed by selection. In the absence of recombination, an adaptive mutation will cause a round of natural selection that "purges" diversity at all loci (Cohan, 2002). This is because adaptive mutation will give rise to a mutant able to over-compete within the population, eventually replacing other individuals and eliminating variation otherwise present the population. At the end of this process, the mutant will give rise to a clonal descendent, thereby reducing diversity of the original population. The adaptive mutant is the ecotype and the purging force is called periodic selection (Cohan, 2002). The definition of an ecotype is a set of strains using the same or very similar ecological niches, such that an adaptive mutant from within the ecotype outcompetes to extinction all other strains of the same ecotype; an adaptive mutant does not however, drive the extinction strains from other ecotypes" (Cohan, 2002).

Impact of LGT on genome evolution

The multiple mechanisms of DNA acquisition, rearrangement, loss, and mutation that are associated with LGT are considered to be significant driving forces that operate at the base of prokaryotic genome plasticity. Thus chromosomes are more than a simple collection of genes. They are physically and structurally significant components of the bacterial cell, requiring both complex molecular mechanisms for maintenance and organization and accurate replication and segregation to daughter cells. The exchange of genetic material between genomes is not random, but there are genes that are transferred more efficiently than others.

Usually genes are transferred in clusters, known as genomic islands (GI) (Hacker and Kaper, 2002). First discovered as large DNA fragments conferring virulence to known pathogens and referred to as pathogenicity islands (PAI), these large portions of foreign DNA have been later described in several taxa and independently from pathogenicity. In fact, GI are associated with metabolic and adaptive functions in many bacteria. Despite their extensive presence in many known bacteria, the evolutionary history of GIs and PAIs is not yet completely understood. The origin and acquisition of an island is often masked by a long history of co-evolution within the host genome. In contrast to other genetic elements, such as plasmids, phages, transposons, and ICEs, genomic islands have lost their mobility and usually the functions associated with them (Hacker and Kaper, 2002). Although the lack of a complete understanding of the history of GIs, logical reasoning might help to speculate the potential fate of transferred genes. Regulatory factors, as well as amino acid biosynthesis and DNA maintenance and replication, are considered to be more stable, refractory to lateral transfer, while genes encoding surface proteins, signal transduction factors, and virulence more often are associated with mobile elements (Nakamura *et al.*, 2004). The frequency of transfer of one class of genes with respect to others is related to a selective advantage conferred to the recipient organism. Many genes that were acquired horizontally cause effects harmful for the receiving cell and consequently, they are similar to deleterious mutations. Other acquisitions can have neutral effects and survival of the bacteria carrying the additional genetic material will depend on random events, while genes conferring an advantage will impart the ability to persist in the cell and to proliferate within the bacterial population (Frost *et*

al., 2005). Consequently, despite dramatic changes in a bacterial genome, LGT does not affect all genes in the same manner.

The most obvious impact of LGT is acquisition of new genes by the host genome. Functional and phenotypic changes introduced by a LGT event largely differ from evolution of new genetic traits via mutation of existing genes. Transferred genes have already been under selection in a genome and can provide fully functional products for colonization of new niches upon integration in the new host. A LGT event can also introduce new protein families not found in the host genome and initiate complex functions when a cluster of genes and operons is mobilized (Lawrence and Hendrickson, 2004). However, the impact of gene transfer does not reside only in direct benefits through acquisition of new functions, but also, and probably mostly, in long term, indirect effects on the bacterial genome (Lawrence and Hendrickson, 2004). Insertion of large regions of DNA provides new sites for insertion of other incoming genes, thus facilitating more transfer events. Furthermore, genomic rearrangements, such as translocations, deletions, and inversions are often observed in regions of the genome affected by a transfer event (Lawrence and Hendrickson, 2008).

Identification of accessory genetic elements, such as plasmids, phages, transposons, and pathogenicity islands, help understanding the evolution of pathogenic microorganisms. Studies of the genome, along with pathogenesis, of *V. cholerae* represents a paradigm employed in this research. *V. cholerae* is both a human pathogen and a natural inhabitant of the aquatic environment. Pathogenesis is

effected through acquisition of virulence genes by phages and genomic islands via lateral transfer events (Faruque *et al.*, 1998). Consequently, to trace down the evolution of pathogens is of prime importance in the characterization of genetic elements that mediate transfer of virulence determinants, resistance to antibiotics, and advantageous metabolic pathways, leading to the emergence of new pathogenic organisms.

Lateral gene transfer in Vibrio cholerae.

Vibrio cholerae, an autochthonous aquatic bacterium, is the causative agent of cholera, a severe, watery, life-threatening diarrheal disease (Colwell, 1996). Strains of this species are serogrouped based on the variable somatic O antigen, with more than 200 serogroups identified (Chatterjee *et al.*, 2003). Although strains of most serogroups of *V. cholerae* are capable of causing a mild gastroenteritis and sporadic local outbreaks of cholera, only toxigenic strains of *V. cholerae* O1 and O139 have been linked to epidemics and pandemics. In particular *V. cholerae* serogroup O1 is responsible for the last two cholera pandemics, in which classical and El Tor biotypes were dominant in the sixth and the current seventh pandemics, respectively (Chun *et al.*, 2009). Genes encoding for cholera toxin, *ctxAB*, and other pathogenic factors have been shown to reside in various mobile genetic elements.

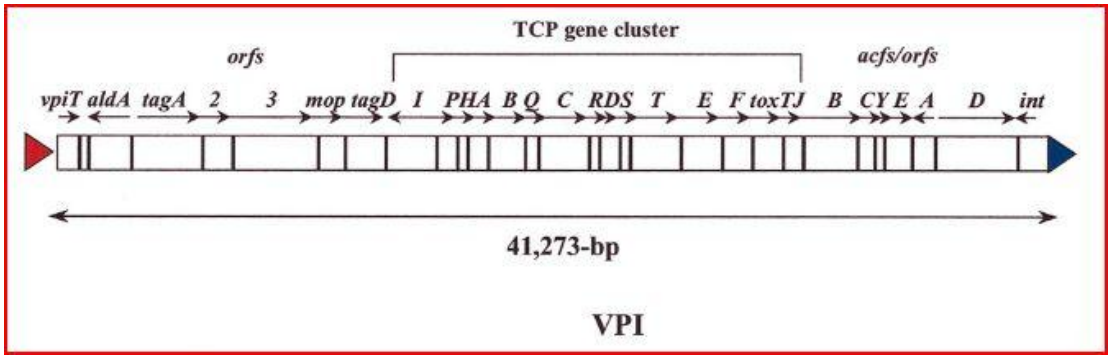
The role of LGT in the evolution of the *V. cholerae* genome remains the focus of ongoing studies, but it is well known that major virulence genes and several

important adaptative functions are clustered in regions of the chromosome that have been acquired by lateral transfer from other con-specific or distantly related organisms. This extreme genome plasticity has generated a significant heterogeneous group of strains capable of inhabiting different environments such as the water and marine environment worldwide and the human host (Colwell, 1996; Faruque *et al.*, 1998; Chun *et al.,* 2009). Due to the variation found among sub-clusters of this species, classification presents challenges. The serogroups-based classification system subdivides the species on the basis of more than 200 subsets, of which only two are associated with major epidemic potential, namely *V. cholerae* O1 and O139 (Kaper *et al.,* 1995). It is now widely accepted that the O139 serogroup originated by the replacement of the O1 antigen synthesis genes by those encoding the O139, acquired via lateral gene transfer. The LGT event caused a substitution of the gene clusters responsible for LPS synthesis of a *V. cholerae* O1 El Tor (Waldor *et al.*, 1994). The evolution of a new serogroup capable of causing epidemic cholera provides a fascinating example of how *V. cholerae* can improve its fitness by acquisition of new genes. Changes observed in *V. cholerae* O139 likely reflect a selective advantage by evading the evolving immunity of an endemic infected population (Faruque and Mekalanos, 2003). Furthermore, many studies have revealed that O1 classical and El Tor biotypes present a significant genetic variance, leading to the hypothesis that acquisition of the O1 antigen might have occurred independently.

Major virulence factors in *V. cholerae* are contained within regions of the genome of foreign origin (Davis and Waldor, 2000). Virulent strains carry the

genome of a temperate prophage, CTXΦ, that encompasses the genes *ctxAB*, encoding the cholera toxin CT, as well as accessory toxins *ace* and *zot* (Waldor and Mekalanos, 1996) (Fig. 1.1a). The CTXΦ is associated with another genetic element, RS1, that is involved in the replication and site specific integration of the phage (Davis and Waldor, 2000). The receptor for the CTX phage is a type IV pilus, the toxin coregulated pilus TCP, which is the bacterial appendix involved in the colonization of the intestine of the human host. Interestingly, genes encoding the synthesis of the pilus are located on the *Vibrio* Pathogenicity Island VPI-I (Fig. 1.1b). Furthermore, the two elements are linked in a complex pathway that controls expression of the toxin under the regulative action of the ToxR transcriptional activator (Karaolis *et al.*, 2001). It has been suggested that the VPI-I island represents the genome of a prophage integrated into the *V. cholerae* chromosome, but repeated attempts to induce and isolate the phage have failed and the hypothesis was discarded (Faruque *et al.*, 2003b). 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.



B.

Figure.1.1: Genetic organization of A) CTX; B) VPI-I (Rajanna et al., 2003).

A second pathogenicity island in the genome of pandemic *V. cholerae*, *Vibrio* Pathogenicity Island 2 (VPI-2), has been described (Jermyn and Boyd, 2002; Murphy and Boyd, 2008). VPI-2 is a 57.3 kb PAI encoding 52 ORFs and includes several gene clusters, such as a type 1 restriction modification (RM) system (*hsdR* and *hsdM*), a sialic acid metabolism gene cluster, genes required for utilization of aminosugars, including the *nanH* neuramidase, a gene cluster that shows homology to Mu phage. The neuramidase protein is involved in *V. cholerae* pathogenicity in two ways. On one hand, 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). On the other hand, 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 (Murphy and Boyd, 2008). At its discovery, VPI-2 showed all of the characteristics of an horizontally transferred element**.** The genomic island encodes for a phage-like integrase, which was found to be responsible for insertion/excision of the island to and from the host genome (Murphy and Boyd, 2008). It has been shown that VPI-2 forms a circular intermediate (CI) upon excision, which likely represents the mobile form of this genetic element. 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 has lost part of its genetic structure. To date, all toxigenic *V. cholerae* O1 serogroup isolates contained VPI-2, whereas non-toxigenicisolates lack the island (Murphy and Boyd, 2008).

Two other gene clusters associated with the seventh pandemic strains were identified by comparative genomics, using microarray analysis, and named *Vibrio* Seventh Pandemic (VSP) I and II (Dziejman *et al*., 2002). These clusters were absent in Classical and pre-pandemic *V. cholerae* El Tor strains and showed an unusual G+C content (40%), compared with the entire *V. cholerae* genome (47%) (Dziejman *et al.*, 2002). VSP-I is a 16-kb region inserted in chromosome I between ORFs VC0174 to VC0186 and comprising 11 genes, VC0175–0185. Recently, VSP-I has been reported in *V. cholerae* non-O1/non-O139 (O'Shea *et al*., 2004). VSP-II was originally identified as a 7.5-kb island, spanning genes VC0490 to VC0497 in *V. cholerae* O1 El Tor N16961 (Dziejman *et al.*, 2002) and, subsequently, found to include a larger 26.9-kb region, spanning from VC0490 to VC0516 (O'Shea *et al.*, 2004). Its site of

integration is a tRNA-methionine locus, VC0516.1. As described in *V. cholerae* O1 El Tor N16961, VSP-II encodes type IV pilin, two methyl-accepting chemotaxis proteins, an AraC-like transcriptional regulator, a DNA repair protein, and a P4-like integrase (VC0516) at the 3' end of the island. Murphy and Boyd (2008) 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 integrase encoded in the island (Murphy and Boyd, 2008).

Due to its extreme genome plasticity *V. cholerae* continually presents newly emerging pathogenic clones carrying diverse combinations of phenotypic and genotypic features and thereby significantly affecting public health efforts to control disease. In a recent study, we compared the genome sequences of 23 *V. cholerae* strains isolated from a variety of sources over the past 98 years in order to reveal evolutionary mechanisms underlying the genetic diversity of members of this species, with particular regard to its pandemic clones (Chun *et al.,* 2009). We concluded that *V. cholerae* undergoes extensive genetic recombination via lateral gene transfer. Specifically, transition between clones in the current seventh pandemic is characterized by different assortments of laterally transferred genomic islands, resulting in emergence of new pathogenic clones, exemplified by *V. cholerae* O139 and *V. cholerae* O1 El Tor hybrid clones. This suggests that genome assets are more informative in defining the different pandemic strains than serogroup classification (Chun *et al.,* 2009). Furthermore, in addition to the known genomic islands, our analysis revealed the presence of new regions in the genome of *V. cholerae* considered genomic islands that have likely been acquired by LGT (Fig. 1.2).

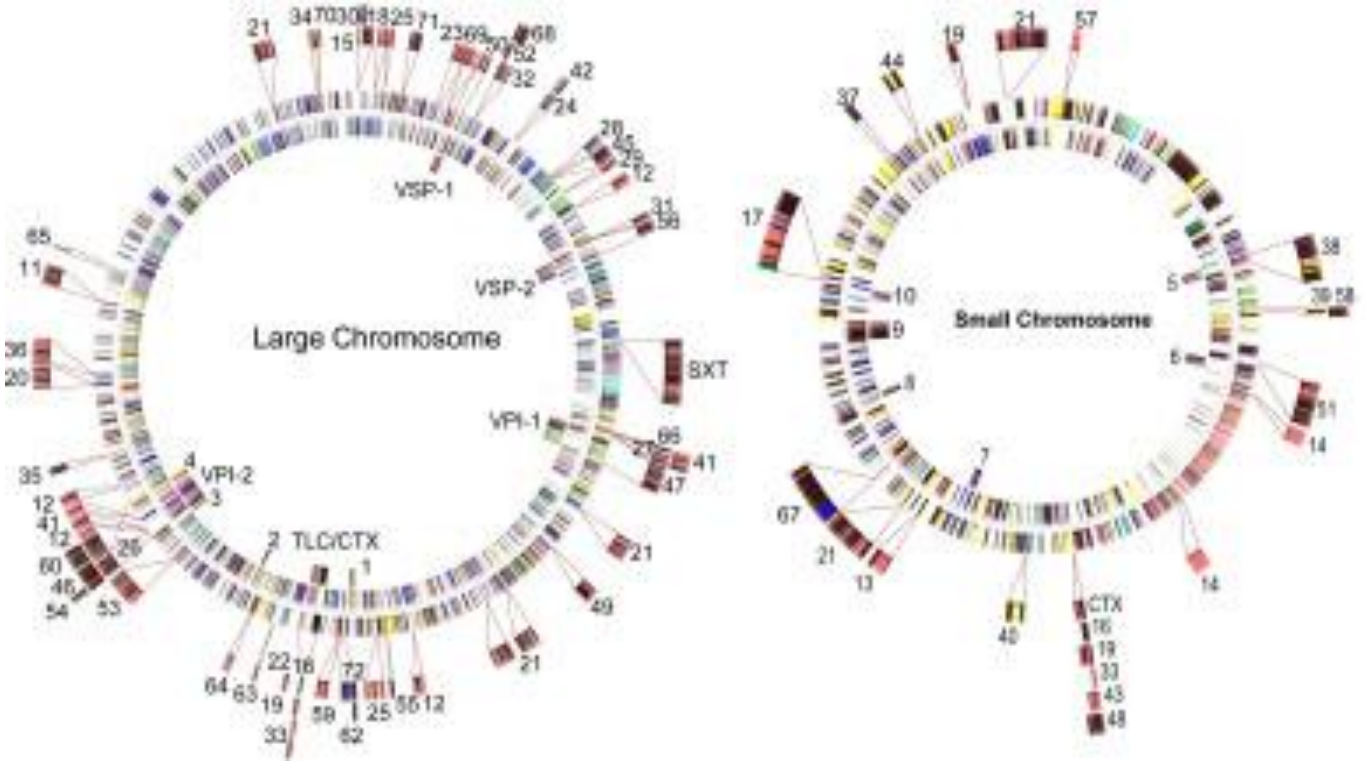


Figure 1.0.2: Genomic representation of genomic islands of both V. cholerae chromosomes. The two circles in the middle represent the genes in V. cholerae O1 El Tor N16961. The inner circle indicates genomic islands found in strain N16961, whereas the outer circles are those absent in strain N16961 (Chun et al., 2009).

Mechanisms of lateral gene transfer.

Lateral gene transfer, defined as the transfer of genetic material between two genetically distinct organisms, was first revealed with discovery of virulence determinants that could be transferred between Pneumococci in infected rats, a phenomenon that later has been attributed to the ability to acquire genetic material through a mechanism called transformation (Griffith, 1928). The successive identification of gene transfer via plasmids and viruses and the observation of transposable elements places the basis for today's emerging strains as being ‘genetic flow’ and the importance of mobile genetic elements (Thomas, 2000).

Transport of chromosomal and extrachromosomal material occurs through three main mechanisms: transduction, transformation, and conjugation. Today, these processes are recognized as mechanisms for gene transfer and are, moreover, responsible for the rapid expansion of resistance to drugs by bacteria (Burrus *et al.*, 2004).

Transduction

The process of transduction is closely linked to bacteriophages. During this process, some bacterial genes can be incorporated in the viral genome due to errors of the phage life-cycle. The phage can then transfer this information, within its genome, to a different organism during its successive cycle of infection. Two types of transduction exist: generalized transduction, which occurs when chromosomal DNA

becomes erroneously incorporated into phage particles. In this case, any gene can potentially be transferred, with the only limitation being the amount of DNA that can be contained in the phage capsid. Secondly, specialized transduction depends on errors during excission of the lysogenic phage. In this case, only genes adjacent to the integration site can be included in the capsid and transferred (Brussow, 2008). Bacteriophages have a narrow host spectrum, often limited to a single bacterial species. Furthermore, bacteria rapidly mutate, becoming resistant to phage infection. For these reasons, transduction and co-evolution of phages and their bacterial host are seen to be an important mechanism for genomic modification through horizontal gene transfer in the environment (Brussow, 2008). The most significant and well studied example of the influence of transduction is found in bacteriophages that mediate acquisition of virulence factors and in the evolution of the bacterium host. Particularly in *V. cholerae*, acquisition of virulence determinants associated with evolution of this species from an environmental microrganism to one of the most dreaded pathogens is tied to the presence of lysogenic phages in its genome.

Transformation

Transformation is a genetic mechanism by which a piece of naked DNA is taken up by a bacterial cell from its surrounding environment and incorporated in its genome or stabilized in the cytoplasmic matrix. A cell in a position to acquire DNA for transformation is called competent. Many bacterial species, among which *Haemophilus influenzae*, *Neisseria gonorrhoeae*, *Bacillus subtilus*, are naturally

competent (Lorenz and Wackernagel, 1994). For other bacterial species, like *Escherichia coli*, artificial systems exist that induce competence. It was recently demonstrated that *V. cholerae* can acquire natural competence in the presence of chitin, thus suggesting this mechanism to be favored when the bacterium is attached to its planktonic host in the environment (Meibom *et al.*, 2005).

Conjugation.

Conjugation refers to the transmission of genetic material from a cell (donor) to another (recipient) through a process that requires a direct cellular contact. Genetic material transferred is usually part of an extrachromosomal circular DNA molecule called plasmid. The plasmid, originally termed a fertility factor, or F factor, of *E. coli*, is the first plasmid to be described (Hayes, 1953). The F factor is approximately 94.5 kb and carries the genes necessary for the contact of two cells through the conjugative pilus and transfer of the plasmid to the recipient cell during the conjugation process. Most of the necessary information for transfer of the plasmid is localized inside of the *tra* operons. Approximately 40 genes are encoded within these operons, 15 of which encode proteins involved in synthesis and assembly of the conjugative pilus (Frost *et al.*, 1994). The entire process can be divided in two parts: formation of the conjugative apparatus and transfer of the DNA. The contact between cells begins through the interaction between the tip of the sexual pilus and the receiving cell. Once contact is established, one filament of the plasmid is cut at a specific locus, called *oriT* (origin of transfer), and transferred through the pilus to the receiving cell. In the

meanwhile, the complementary filament is synthesized *de novo* in both donor and recipient cells. This molecular mechanism assures that a copy of the plasmid is transferred to a recipient cell without being lost by the donor (Frost *et al.*, 1994). It is thought that the majority of genetic exchanges among bacteria in the environment involve conjugative mechanisms through different classes of mobile genetic element.

Mechanisms of genetic recombination

Clearly, from the above discussion, it is widely recognized that one of the major processes underlying bacterial species evolution is horizontal gene transfer (Davison, 1999; De la Cruz and Davies, 2000). Since the vehicles of this genetic flow, i.e. the mobile genetic elements, can exist integrated in the chromosome, it is necessary to understand the molecular processes that permit their integration in the bacterial genome. These are collectively grouped under the name of "genetic recombination", and can be subdivided in three processes: homologous recombination, site-specific recombination and transposition.

Homologous recombination.

Homologous recombination involves genetic exchange between any two DNA molecules (or segments of the same molecule) that share an extended region of sequence homology. The classical model of a recombination mechanism has been proposed by Holliday (1964) and comprises four points: (1) alignment of homologous DNA; (2) nick of one strand of both molecules and formation of the 'Holliday

intermediate; (3) branch migration of the Holliday structure, that is the sliding of the hinge of the heteroduplex region along the two aligned DNA molecules. The result is a physical transfer of a part of one molecule with that of the other; (4) the Holliday intermediate is resolved in different ways according to the orientation of the two filaments, leading to the formation of two recombinant molecules.

In *E. coli*, an efficient recombination requires several enzymes: the RecA protein, which promotes the pairing of the homologous filaments and their exchange; RecBCD enzyme, which acts as nuclease and elicase; RuvAB, involved in the branch migration; and RuvC, the resolvase in charge of the resolution of the Holliday structure. Other necessary factors are enzymes involved in the normal metabolism of the DNA, among which are the proteins SSB, DNA polymerase I, DNA gyrases, and DNA ligases (Kowalczykowski *et al.*, 1994). Homologous proteins have been identified in numerous bacterial species and proteins similar to RecA have been discovered in yeast and eukaryotes, suggesting that similar mechanisms are common to all organisms (Cox, 2001). Furthermore, homologous recombination plays an important role in the repair of damages caused by DNA replication, in a process called post-replicative repair (Kowalczykowski *et al.*, 1994) and it generates chromosomal rearrangements like inversions, deletions and duplications. Moreover, this process is involved in the integration of horizontally acquired foreign DNA into the bacterial chromosome. An example of this mechanism is integration of *E. coli* F plasmid inside the chromosome of Hfr strains, which typically occurs through homologous recombination between insertion elements, IS, on the plasmid and that on the chromosome (Deonier and Mirels, 1977).

Site-specific recombination

Site-specific recombination occurs between two DNA sequences (not necessarily homologous) in a reaction catalyzed by specialized enzymes, recombinases, together with several accessory factors. The site-specific recombinases physically interact with short DNA target sequences, joining the two strands in a process called synapse, and later catalyzing cut and exchange of filaments, formation of the Holliday intermediate, and its successive resolution (Gopaul and Van Duyne, 1999). One of the most studied models of site-specific recombination is the integration of the λ phage in the chromosome of *E. coli*. The infection occurs when the phage genome is injected in the host, in which it circularizes. Infection of the host can lead to activation of a lytic cycle, in which the phage replicates and kills the bacterium, or in a lysogenic cycle, in which the phage remains silently integrated in the genome of *E. coli* in the form of prophage, until reactivation by appropriate stimuli. Integration and excision of the phage from the chromosome occurs via recombination between specific sequences of the phage and the bacterial chromosome, called *att* (attachment) sites. The *att* site on the phage (*attP*) is approximately 240 bp long and is organized in a module, POP', in which O is a key sequence of 7 bp situated within 15 conserved bp, while P and P' are adjacent sequences containing the binding sites for the enzymes that catalyze the reaction. The bacterial *att* site (*attB*) is localized in *E. coli* in an intergenic region between the genes *gal* and *bio*. This site is only 21 bp long and is organized in the BOB' module, where O is a sequence of 7 bp identical to the one found on the phage, within 15 conserved bp, and B and B' are the adjacent sequences, containing the enzymes binding sites.

Integration (*attP* x *attB*) is catalyzed by the product of the gene *int*, the integrase (Int) and by a bacterial protein called IHF (integration host factor). Excision of the prophage from the chromosome (*attL* x *attR*) involves the product of the gene *xis* (excisionase) together with Int and IHF. This process will be analyzed in more detail in the discussion of the integration and excision mechanisms of ICEs of the SXT/R391 family, in the *prfC* gene.

Site-specific recombinases.

Site-specific recombinases are subdivided into two main families, based on evolutionary and functional parameters: the tyrosine-recombinase, the protein family of the λ integrase, and the serine-recombinase family of the resolvase. Members of the first family, of which Int protein of the λ phage is the prototype, have a rather variable amino acid structure, however they are characterized by four conserved amino acids located in two recurrent domain present at the C-terminal of the protein (Blakely and Sherratt, 1996; Esposito *et al.*, 1997; Smith, 2002). The first domain of 34-40 amino acids contains an arginine (R), followed by a histidine and arginine (HR), separated by two amino acids not conserved and the residual catalytic product of tyrosine (Y). This tetramer, RHRY, constitutes the catalytic site of these enzymes. Furthermore, highly conserved residues are found also in the inner domain that contributes to folding of the protein (Nunes-Düby *et. al.*, 1998). All of the four necessary cuts and exchanges in the process of recombination catalyzed by the λ integrase are followed by two reactions of trans-esterification via a mechanism

similar to the one of topo-isomerase I. The tyrosine-recombinase can be grouped into four classes, based on the complexity of the reactions they catalyze (Azaro and Landy, 2002). The first group is represented by the human topo-isomerase of group 1B and it is the only one with a monomer structure, the other three groups all operate as tetramers. The most studied members of the third group are the XerCD proteins. Many of these proteins catalyze integration and excision of mobile elements into and from the host chromosome. Enzymes belonging to this class include the integrase of λ phage, the Tn916 and, as discussed in the following chapter, the integrase of the SXT/R391 family of ICE (Azaro and Landy, 2002).

Transposition is a recombination reaction that mediates mobilization of discreet segments of DNA between sites dispersed within the host genome. Transposons are widely dispersed in the living world and carry out a fundamental role as motors of genomic plasticity in Prokaryotes as well as Eukaryotes (Mahillon and Chandler, 1998; Haren *et al.*, 1999). Transposition is mediated by enzymes called transposases and occurs through the endonucleolitic cuts of the phosphodiester bond and the transfer of the extremities to a new DNA molecule target. Transposons can be duplicate (replicative transposition) or simply transferred (conservative or non-replicative transposition). Examples of replicative transposition are found in the Tn3 family, the insertion sequences IS6 and the bacteriophage Mu (Haren *et al.*, 1999). Examples of conservative transposition are found in the members of families of Tn7, IS10 and IS50 (Haren *et al.*, 1999). The specificity for target sites in transposition is rather wide, allowing for the insertion of transposons at practically any locus of the chromosome, with some exceptions, like the Tn7 or IS91 (Craig, 2002).

Objective of this study

The importance of LGT in bacterial evolution is the focus of this research and was studied by analyzing a class of mobile elements, ICE, in the bacterial species *Vibrio cholerae,* an extensively researched and well known bacterial pathogen for humans. The objective was to understand how acquisition of heterologous genetic material by *V. cholerae* has influenced its differentiation and adaptation to different niches in the aquatic environment, as well as its role as a human pathogen.

The initial task was to find new ICEs circulating among pandemic *V. cholerae* O1 El Tor and, through comparative analysis of these elements, extrapolate the epidemiological relationship between new and old pandemic clones in cholera endemic areas. Specifically, the goal was to gain an understanding of the nature of ICEs, their importance as genetic machinery for dissemination of resistance, virulence, and a broad range of metabolic functions, and correlation with other genetic elements, like genomic islands. The goal subsequently proceeded from molecular analysis to the ecology and geographical distribution of ICEs, employing a large collection of strains isolated from several countries where they had not yet been described. The aim was to demonstrate the extent to which environmental *V. cholerae,* autochthonous to different geographical areas of the world, represents a reservoir for mobile genetic elements providing selective ecological advantage.

The general characteristics of ICE and their distribution in bacteria are presented in Chapter 2. A specific class of ICE present in *V. cholerae* and in both

related and unrelated species, namely the SXT/R391 family of ICE, is reviewed. The focus is on the unique genetic asset of these elements that confers upon them significant plasticity and the capacity to disseminate genetic material within and among bacterial populations. Current knowledge of the SXT/R391 family of ICE is reviewed, notably the work on genetics, functionality, ecology, and genomics of this class of mobile elements. The chapter concludes with the description of SXT/R391 ICEs reported to date.

Chapter 3 provides a genomic analysis of three ICEs found in three *V. cholerae* strains that had been included in a study of the complete genomic sequences of 23 strains of *V. cholerae*. The amount of information residing in whole genomic sequences of both the ICE and the host genome provided an incomparable source upon which to draw insight concerning the function of these elements. The genetic structure and content of the three elements are discussed with respect to their host and its genome. The 23 *V. cholerae* genomes that had been sequenced also provided data on a interesting genetic element. Characteristics of the insertion in a clinical isolate, *V. cholerae* O37 MZO3, discussed in Chapter 4, lay between ICEs and genomic islands, and, to date, represent the only one of its kind. A complete nucleotide sequence analysis was done and the functions encoded in new significant genetic cluster carried by the element were disussed. In particular, the hypothesis of the contribution of these genes to the virulence of this *V. cholerae* O37 strain is presented.

In Chapter 5, a geographic survey was done of ICEs in *V. cholerae* strains isolated from different parts of the world, focusing especially on elements that had

not yet been described. The concluding study, reported in Chapter 6, described two other genomic islands, VSP-1 and 2, their genetic polymorphisms, and distribution of different variants in clinical and environmental *V. cholerae* strains originating in a cholera endemic area.

Chapter 2. Integrative conjugative elements (ICE).

Conjugation is one of the mechanisms by which horizontal gene transfer takes place, usually mediated by plasmids. However, other elements have recently been described that can mobilize genetic information by a conjugative process involving integrative and conjugative elements (ICE). These elements are linear sequences integrated into the bacterial chromosome, behaving like plasmids, able to move through a conjugative process and, like temperate phages, able to integrate and replicate in the genome of the host. Initially the ICE family was observed only in low G+C Gram positive Bacteria and Bacterioides, since the first elements described were *Tn916*, isolated from *Enterococcus faecalis* and *CTnDOT* from *Bacteroides thetaiotaomicron*. These were classified as "conjugative transposons" (Scott and Churchward, 1995; Whittle *et al*., 2002), due to their ability to integrate and translocate between different loci in the host chromosome and to conjugate from a donor to a recipient. However, with the discovery in the past decade, of various conjugative elements that integrate via site-specific recombination, similar to bacteriophage λ, the nomenclature for such elements has been questioned and a more appropriate designation, 'integrating and conjugative elements' (ICE), was adopted (Burrus *et al*., 2002). To date, ICEs have been described in isolates belonging to many taxonomic groups, including *Actinomycetales*, *Rhizobiaceae* (α-proteobacteria),

Burkholderiaceae (β-protebacteria) and widely in the γ-proteobacteria (Burrus *et al.*, 2006).

General characteristics

The molecular architecture of the ICE encompasses three distinguishable functional modules found in all self-transmissible elements, which assure the maintenance, dissemination, and regulation of the element (Toussaint and Merlin, 2002). These modules are structured in clusters of genes that can be acquired and/or excised, and, therefore, exchanged between different elements or with the chromosome.

The first cluster, responsible for maintenance, promotes the integration of the ICE in a replicon of the host (a plasmid or a chromosome), which assures the replication and, therefore, transmission of the element, that is otherwise unable to replicate, to the next generation. This function is encoded by an integrase of the family of the site-specific recombinases (*int*). The enzyme, as in the integration of the λ phage, promotes 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 extremities of the integrated element, called *attL* (left) and *attR* (right). Different recombinases have been described to date on the basis of the specific reaction catalyzed, the majority belonging to the tyrosine-recombinase family, but also serine-recombinases as in *Clostridium difficile* (Wang *et al*., 2000). Together with Int and *attP*, many ICEs encode for a protein called Xis

(excisionase), which co-operates with Int in the excision of the element (Burrus and Waldor, 2003).

Like conjugative plasmids, ICEs are transmitted through conjugation. Different clusters of *tra* genes encode all factors necessary for the mating machinery, which allow establishing physical contact and the transfer of genetic material between donor and recipient cell. These clusters, like the maintenance cluster, may differ between different elements. ICEs are mobilized in a single strand form, with the exception of elements like *pSAM2*, isolated from *Streptomyces ambofaciens*, which transfers in a double stranded form (Burrus *et al*., 2004b). Moreover, the transfer machinery of various ICEs is able to mobilize other genetic elements (Hochhut *et al*., 2000).

The genes and mechanisms that regulate transfer of ICE are under study, but there is evidence that different regulation pathways exist among known ICEs. The regulation module comprises activators and repressors, different for each ICE, that regulate excision and transfer of the element and, especially, timing in response to environmental conditions perceived by the host.

Together, the three modules constitute the basic structure of the ICE, which can incorporate other genes that consequently gain a 'free ride' by the elements to other organisms. This hypothesis is exemplified by the genomic comparison of SXT and R391, two related ICEs, derived from *Vibrio cholerae* and *Providencia rettgeri*, respectively. These elements share almost identical modules which mediate regulation, conjugative transfer, and integration/excision. In addition to the 65 kb of conserved DNA, each of the two ICE carries specific insertions, which confer

peculiar properties to the element, like resistance to certain antibiotics and/or resistance to mercury (Beaber *et al*., 2002a). Preferential sites for insertion of these sequences have been identified (hotspot), but the mechanism by which these are integrated has not yet been clarified. Some of the insertions appear to be mediated by transposons or by homologous recombination. Moreover, different recombination mechanisms, not yet identified, might be involved in the insertions, as well as the one mediated by recombinases Red, a class of proteins encoded by both elements (Beaber *et al*., 2002b; Boltner *et al*., 2002).

Initially, ICEs were identified because they conferred specific phenotypes, like resistance to some antibiotics (Waldor *et al*., 1996). Sequencing of many 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, and acquire additional functions involved in DNA repair, UV protection, recombination systems, and symbiotic growth with plants (Burrus *et al*., 2004b). Furthermore, some characteristics of the pathogenicity islands of *Legionella pneumophila*, *Agrobacterium tumefaciens*, and *Bartonella tribocorum*, such as ability to encode type IV secretion systems (T4SS) and tyrosin-recombinases similar to Int involved in DNA processing, suggest that these pathogenicity islands might be considered ICEs, or defective ICEs (Burrus *et al*., 2004b).

The hypothesis that ICEs are involved in bacterial virulence, or that they can mobilize virulence determinants, is strongly supported by the discovery of a new

element in a strain of *E. coli*, ICE*Ec*1, which carries the high pathogenicity island HPI, encoding for the siderophore system "yersiniabactin", previously described in *Yersinia*. Although its conjugative potential has not yet been demonstrated, ICE*Ec*1 integrates at the *asnV* locus, is able to form a circular intermediate after excision, and encodes a pore forming system. (Schubert *et al*., 2004).

Another interesting example of correlation between pathogenicity islands, virulence, and ICEs comes from the discovery of ICE*Kp*1 in a *Klebsiella pneumoniae* strain isolated from a primary liver abscess (Lin *et al.*, 2007). Lin and colleagues demonstrated the presence of a putative ICE containing the HPI and its self-transmission and chromosomal integration. These two examples suggest ICEs are genetic tools for diffusion of the HPI among different bacterial species, a mechanism possibly shared by other genomic islands (GI) or pathogenicity islands (PAI).

IncJ elements and R391

A group of conjugative elements exhibiting structural similarity with enteric pathogenicity islands has been described in members of γ-Proteobacteria isolated from different locations and identified as carriers of genes for drug resistance (Boltner *et al*., 2002). These elements were initially identified in 1967, when exogenous DNA sequence was described in the genome of *Providencia rettgeri* isolates from South Africa that encoded resistance determinants for Kanamycin (Km) and mercury (Hg) (Coetzee *et al.,* 1972). Twelve related elements, all associated with resistance to

drugs, were successfully transferred to *E. coli* K12 via conjugation. Among these, three resistance factors for kanamycin were identified, R391, R392, and R397, all able to coexist with all known incompatibility groups. The elements were classified as conjugative plasmids mediating drug resistance, that is, R factors, and assigned to a new incompatibility group, IncJ, of which R391 was selected as the prototype.

Other IncJ elements have been described in *P. rettgeri,* namely R392 and R397. Furthermore, R705 and R706 from *Proteus vulgaris* and R748 and R749 from other *Providencia spp*., were isolated in the early 70's and are characterized by an identical restriction profile and several phenotypic features, such as Km and Hg resistance, low frequency of transfer, sensitivity to UV, and complementation with *umuCD* (Pembroke and Stevens, 1984; Pembroke and Murphy, 2000).

Only three other members of the IncJ elements outside Africa have been described. These are R997 in *Proteus mirabilis* isolated in 1997 in India and resistant to ampicillin, pMERPH, in *Shewanella putrefaciensis* in 1990 in England and resistant to Hg (Boltner *et al.*, 2004), and pJY1 from a clinical *V. cholerae* O1 isolated in the Philippines and resistant to streptomycin, chloramphenicol, and sulfamethoxazol (trimethoprim resistance was not tested) (Yokota and Kuwahara, 1977). Interestingly, pJY1 was described twenty years before the discovery of SXT in *V. cholerae* (discussed below), to which it is related by resistance determinants. pJY1 is no longer available for study, thus genetic comparison with more recent SXT-related ICEs is not possible.

The complete nucleotide sequence of R391 covers 88.5 kb, with 96 ORFs of a composite nature, evidenced also in its C+G content. This element has the

characteristics of ICEs, including modules of transfer, integration, and resistance (Boltner *et al*., 2002). Given the repeated failure to isolate a replicative extra-chromosomal form of IncJ elements from the host and demonstrated insertion of R391 in the bacterial chromosome, it was proposed to include these elements into the conjugative transposon group (Murphy and Pembroke, 1995). The site of integration of R391 is a 17bp sequence located at the 5' end of *prfC*, encoding a release factor involved in protein synthesis (RF-3). This gene is found at 99.3 min in the chromosome of *E. coli*.

Furthermore, the IncJ group encodes sensitivity to UV, shown to promote the transfer of the elements (McGrath and Pembroke, 2004). It is suggested, therefore, that there is an additional function of these elements, namely protection from UV, probably linked to the SOS response pathways that, by activating the *recA* dependent mechanism, induce transfer of the element and increase mutagenesis rate. In R391, this function is associated with the presence of two genes, *rumA* and *rumB*, which encode for homologous functions of the UmuDC proteins of *E. coli*. These genes encode the DNA polymerase V involved in DNA repair in both organisms. Downstream of *rumBA*, in R391, is a cluster of genes with homology to the ε subunit of DNA polymerase III, which has proofreading activity.

This complex genetic structure makes the classification of these elements difficult: from plasmids to conjugative transposons to CONSTIN (conjugative, self-transmitting, integrating elements) (Pembroke *et al*., 2002). The IncJ elements are now included in the most recent group of ICE (Burrus *et al*., 2006a).

ICEs in Vibrio cholerae: SXT elements

The SXT mobile element was originally discovered in *V. cholerae* O139 MO10, one of the first clinical strains of *V. cholerae* O139 to be isolated in Madras, India (Waldor *et al*., 1996). This new serogroup of *V. cholerae* emerged at the end of 1992 as the first *V. cholerae* non-O1 serogroup causing epidemic cholera (Cholera Working group, 1993). *V. cholerae* O139 soon appeared throughout Asia, initially replacing *V. cholerae* O1 in India and Bangladesh by the end of 1993. Besides encoding for a new O antigen, *V. cholerae* O139 differs from *V. cholerae* O1 El Tor, by a distinguishing pattern of antibiotic resistance. The first *V. cholerae* O139 strains to be isolated in India and Bangladesh were resistant to four antibiotics: sulfametoxazol; chloramphenicol; trimethoprim; and streptomycin. In *V. cholerae* MO10, the resistance genes were detected in a mobile element, initially called SXT and subsequently renamed SXT^{MO10}, that was classified as a conjugative transposon (Waldor *et al*., 1996; Hochhut *et al*., 2001b). SXT^{MO10} was detected in other strains of *V. cholerae* O139, but not in *V. cholerae* O1 (Amita *et al*., 2003).

In 1994, *V. cholerae* O1 El Tor re-emerged as the predominant causative agent of cholera on the Indian subcontinent. The new isolates of *V. cholerae* O1 El Tor were resistant to the same four antibiotics that were characteristic of *V. cholerae* O139 (Yamamoto *et al.*, 1995) and the encoding resistance genes were found in a mobile element related to SXT^{MO10}, called SXT^{ET} (Hochhut *et al*., 1999). Since 1994, different variants of SXT elements have been described in *V. cholerae* O139 and O1 from India (Amita *et al.*, 2003), Bangladesh, South Africa (Dalsgaard *et al.*, 2001, Ceccarelli *et al.,* 2006), Vietnam (Dalsgaard *et al.*, 1999; Ehara *et al.*, 2004; Bani *et*

al., 2007), Laos (Iwanaga *et al.*, 2004; Toma *et al.*, 2005) and Mexico (Burrus *et al.*, 2006b). These elements share a common molecular organization differentiated by peculiar insertions (discussed below).

As for the R391-like ICEs, the common structure of SXT-like elements includes genes derived from plasmids, phages, transposons, and various other genetic sources still unknown. The main property of the R391-like ICEs is the ability to transfer by conjugation between Gram-negative bacteria. Furthermore, SXT-like elements integrate in *prfC*, the same integration site as R391, via site-specific recombination.

The SXT/R391 family of ICEs

Because of the similarity between the two elements, SXT and R391, they are now classified as two subgroups of the same family of ICEs (Beaber *et al*., 2002a). The latest nomenclature for these elements comes from Burrus *et al*. (2006a), that is, mobile genetic elements encoding for an integrase closely related to that present in SXT (int_{SXT}) and integrating at the *prfC* locus are considered to be part of the SXT/R391 family of ICE (Burrus *et al*., 2006a). Although the nomenclature of SXT (which refers to resistance to sulfametoxazol and trimethoprim) and R391 (where R stands for resistance factor) is too rooted in the scientific terminology to be renamed, it has been proposed to rename more recently described ICEs. The current nomenclature uses the prefix ICE following from 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, the recently described ICE, ICE*Vch*Mex1, was identified in a strain of *V. cholerae* isolated in Mexico (Burrus *et al*., 2006b).

The complete nucleotide sequences of R391 (89 kb) and SXT^{MO10} (99,5 kb) revealed that these elements share 65 kb of conserved genes with 95% identity (Boltner *et al*., 2002; Beaber *et al.*, 2002b). The conserved regions include genes for integration (*attP, int)* andexcision (*xis*), for conjugative transfer (*traI-s043*, *traL-traA*, *s054-traN*, *traF-traG*), and for regulation of transfer (*setC-setD-setR*) (Böltner and Osborn, 2004; Burrus *et al*., 2006).

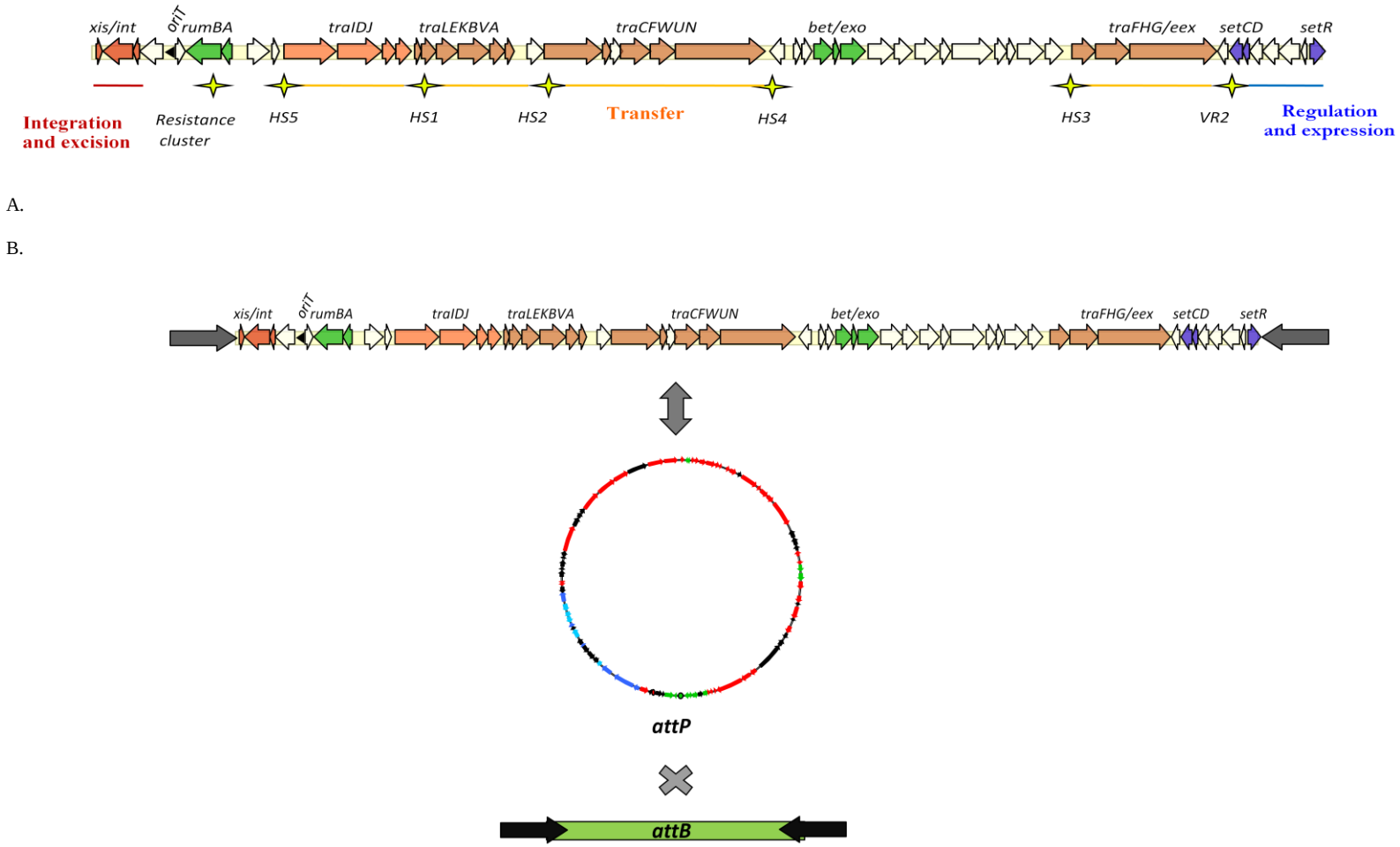


Figure 2.1. A. ICE combine features of transposons, plasmids and phages (Burrus et al., 2006). B. Mechanism of integration of ICE in the host chromosome.

Conserved regions

Integration and excision cluster

As does R391, SXT elements also integrate at the 5'-end of a nonessential chromosomal gene (*prfC*) encoding the peptide release factor 3 (RF3). Upon integration, the element provides a functional promoter and 5'-end, resulting in the correct expression of the RF3 (Hochhut and Waldor, 1999). The integrative mechanisms resemble the *rec-A* independent site-specific recmbination described for lambdoid phages. The recombination occurs between the almost identical 17 bp host encoded chromosomal *attB* and element encoded *attP* sequences and creates two inverted repeats flanking the element, *attL* and *attR*. Integration requires the element encoded integrase, Int (Hochhut and Waldor 1999), which contains in its 3'-end the tyrosine recombinase domain common to all site specific recombinases, enzymes involved in DNA breaking-rejoining, with a C-terminal catalytic domain. Members of this diverse protein superfamily include enzymes like the human topoisomerase I, the bacteriophage lambda integrase, the bacteriophage P1 Cre recombinase, the yeast Flp recombinase and the bacterial XerD/C recombinases which include the integrase found in ICEs.

The excision occurs via recombination between the *attL* and *attR* flanking sequences and results in a circular extrachromosomal form that is the substrate for the conjugative transfer of the ICE. It was demonstrated that Int is involved in the

excision with a second protein, Xis, a recombination directionality factor that promotes excision and prevents re-integration of the element (Burrus and Waldor 2003). The excisionases in SXT (Xis) and R391 (Jef) are almost identical, small (<100 amino acid) basic proteins and are orientated in opposite fashion with respect to the integrase, suggesting that the two are not coregulated (Burrus and Waldor, 2003; O' Halloran *et al.*, 2007). However, integrase and excisionase are under the regulation of two element-encoded transcriptional activators and a repressor which are linked to the cellular SOS-response (see below) (O' Halloran *et al.*, 2007).

Conjugative transfer clusters.

Genes involved in the conjugative transfer of SXT/R391 ICEs are grouped in four regions about 25 kb long: *traI-s043*, *traL-traA*, *s054-traN*, and *traF-traG*. The conjugative apparatus encoded by the SXT family of ICE is structurally and genetically related to conjugative plasmids, in particular the F- factor, and it is considered a variant of the type IV secretion system family (T4SS) (Lawley *et al.*, 2003). In Gram-negative bacteria, the T4SS, also known as mating pair formation (Mpf) apparatus, is one of the most complex cellular apparatus by which DNA and proteins travel across the cellular envelope, in and out of the cytoplasm via a protein structure that creates a pore or a channel between two cells. T4SS is involved in secretion of virulence factors, exchange of genetic material between a donor and a recipient cell, and uptake of DNA from the environment in natural transformation (Lawley *et al.*, 2003).

The first set of *tra* genes, including *traI, traD, traJ* and *s043*, encodes for the proteins involved in the relaxosome, namely all proteins involved in DNA processing and transfer initiation. The nickase, TraI, cuts one strand of the ICE circular intermediate (CI) to allow only a single-strand DNA molecule to pass to the host cell. The coupling protein, TraD, is then required to lead the ssDNA through the conjugative apparatus (Burrus *et al.,* 2006). The *traL-traA, s054-traN*, and *traF-traG* genes encode proteins for the mating pair formation, stabilization of the mating pair, formation of the pilus, and entry exclusion (see below).

The *tra*-gene clusters have the same genetic organization in all SXT/R391 ICEs described to date and all share >94% identity at the amino acid level. Interestingly, the core set of genes are present in synteny in the *tra* clusters of multidrug conjugative plasmids, IncA/C, found in *Salmonella spp*., *Yersinia pestis*, and aquatic γ-Proteobacteria including *V. cholerae* (Wozniak *et al.*, 2009).

Host-encoded proteins are required to regulate key functions of F plasmids and λ phage. In particular, the integration host factor (IHF) is a heterodimeric regulator involved in the expression of *tra* genes of F-plasmid and λ and promote the TraI nick at the origin of transfer. It was demonstrated that similar to other conjugative elements, ICEs were also impaired in their transfer in host strains defective of IHF. Interestingly, the effect of the mutation was host-specific and SXT transfer was not affected in a IHF$^-$*E. coli* (McLeod *et al.*, 2006).

Entry exclusion mechanism

Interestingly, despite their high genetic and structural homology, SXT and R391 belong to two different exclusion groups. The entry exclusion system is a cellular mechanism to prevent the acquisition of redundant genetic material common to conjugative plasmids (Marrero and Waldor, 2007). Beside IncF plasmids, which showed a surface interaction, the exclusion is usually mediated by two inner membrane proteins, encoded by the donor and recipient cells, whose interaction determines if the incoming genetic element is already present in the recipient cell, and inhibits the redundant conjugative transfer between cells carrying the same element. As in conjugative plasmids, in the SXT/R391 family, the entry exclusion is operated by the TraG and Eex inner membrane proteins, expressed in the donor and recipient cell, respectively. Interestingly, despite the high similarity in the transfer apparatus, SXT and R391 form two exclusion groups, S (SXT) and R (R391), thus, these two elements do not exclude each other (Marrero and Waldor, 2007).

Molecular details of the exclusion process are not yet known. It appears that, during the conjugation process, ICE exclusion is mediated by the interaction of only three amino acids of TraG (position 606, 607, and 608) and the C-terminal portion of Eex protein (Marrero and Waldor, 2007). The portions of these proteins involved in exclusion appear to be located in the cytoplasm, making it difficult to understand how they interact during and after the transfer. TraG-Eex interaction selectively inhibits incoming DNA avoiding redundant cell to cell transfer of ICEs of the same group. Exclusion is observed when EexS expressed in a cell carrying SXT recognizes the $TraG_S$, expressed in a donor also carrying a SXT-like element, thus

inhibiting the transfer of other SXT-like ICEs, but allowing acquisition of an R391-like element.

Analysis of exclusion determinants of other ICEs of this family showed different *eex* alleles, but all belonged to one of the two groups (S or R). Their protein sequences clustered into the two groups, with more than 93% identity between alleles and significant sequence discrepancies at the C-terminus. Five *eexS* alleles and four *eexR* alleles have been isolated worldwide to date from SXT/R391 ICEs (Marrero and Waldor, 2007).

In *E. coli*, a recipient cell bearing an ICE of a different exclusion group acts as an empty cell in terms of frequency of transfer. However, it appears that exclusion can be only partial and that transfer can occur with acquisition of the same group element, resulting in tandem arrays and possible formation of hybrid ICEs (Marrero and Waldor, 2007).

Regulative cluster

The regulation of integration/excision and transfer of the circular intermediate of ICEs resemble the pathway that regulates the lytic cycle of lambdoid phages (Beaber *et al.,* 2004). ICE regulation is governed by three ICE encoded genes: *setC, setD* and the repressor *setR*. The regulatory cascade is triggered by events that activate the host SOS response, including reactions that damage the DNA, by cleavage and inactivation of the repressor SetR. The cellular SOS response consist of a series of regulatory processes all connected in a RecA-dependent manner. Activated

RecA, stimulates autoproteolysis of SetR that is a phage lambda cI-like repressor, whose inactivation permits expression of SetC and SetD proteins, which are transcriptional activators of *int* and *tra* genes (Beaber *et al.*, 2004).

All characterized ICEs to date bear an almost identical regulatory cluster with *setC*, *setD* and *setR* genes, sharing >97% sequence homology and suggesting the same mechanism serves in the induction of SXT/R391 ICEs (Burrus *et al.*, 2006).

Other conserved genes: *s065* and *s066*

These two genes are highly conserved (≥96% identity) of all the known SXT/R391 ICEs that encode proteins similar to recombinase Bet and exonuclease Exo (71% and 38% similarity, respectively) found in the temperate bacteriophage λ and several other phages (Garris *et al*., 2009). In λ, Bet and Exo, along with Gam protein constitute an efficient *recA*-independent homologous recombination system dependent on DNA replication, known as λ Red. These proteins work on DNA with double-stranded breaks generated during the phage replication. Interestingly, λ Red-like recombination systems have not been described outside of phages (Garris *et al*., 2009). This recombination machinery, together with RecA, was shown to play a role in shaping the hybrid genomes commonly found in SXT/R391 ICEs, suggesting that both host- and element-encoded recombination systems promote formation of the mosaic genomes of these elements (Garris *et al*., 2009).

Variable regions and insertions

Comparative genomics of several ICEs led to the identification of a pool of genes specific for each element of the SXT/R391 family and interspersed in the conserved sequence conferring genetic specificity and a broad size range of the ICEs (Burrus *et al*., 2006; Wozniak *et al*., 2009). Some of the insertions are shared by different elements, some are unique and unrelated, but most encode adaptative functions and appears to have lost the genetic signature for mobilization. The five target loci of variable insertions were named 'hotspots' (HS1 to HS5). Furthermore, some ICEs have insertions in four additional loci and the *rumB* gene (Figs 2.2b) (Beaber *et al*., 2002; Boltner *et al*., 2002, Wozniak *et al*., 2009).

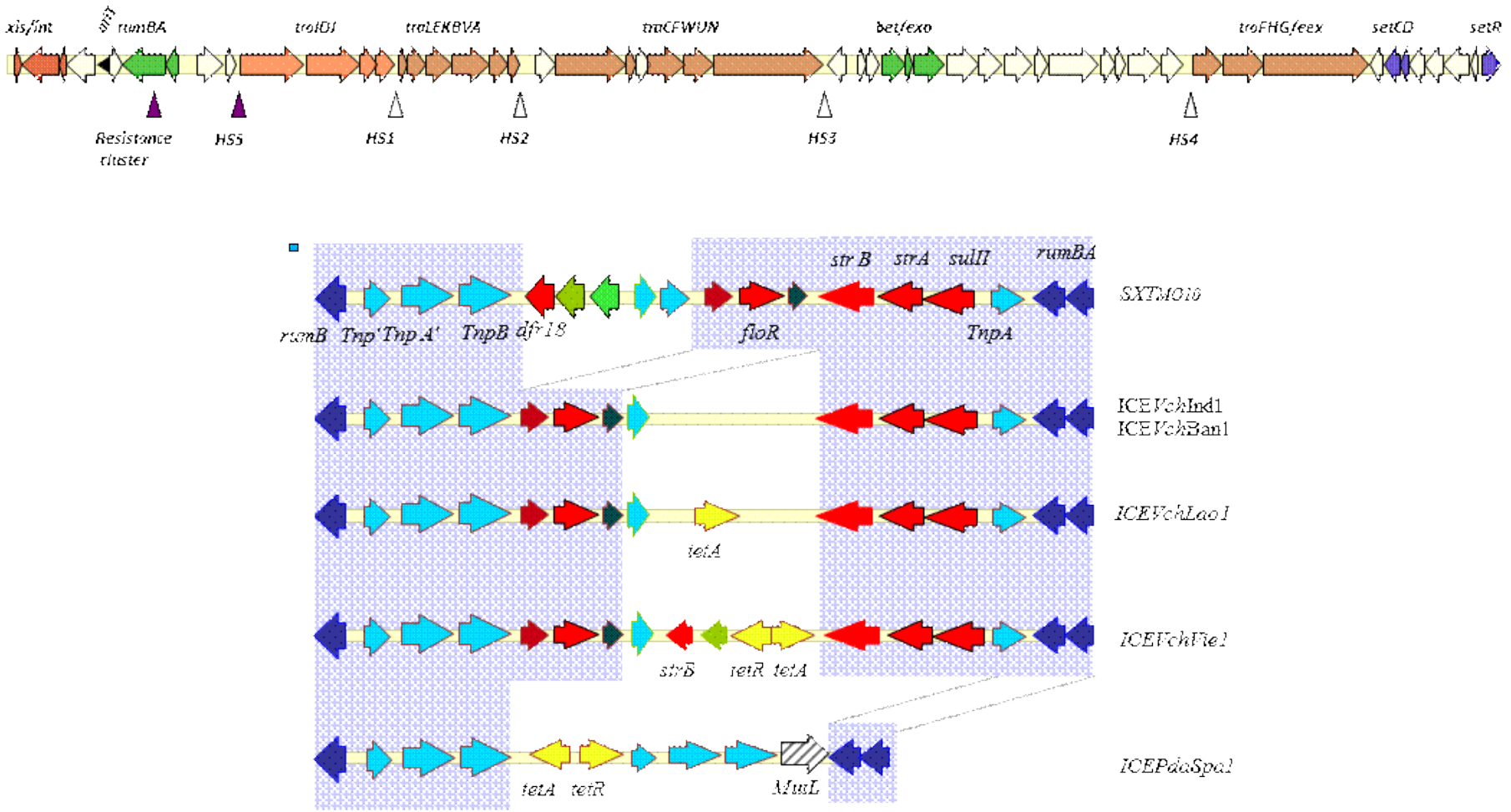


Figure 2.2. A. Plasticity of the genome of the ICE of the SXT/R391 family. B. Genetic organization of the resistance cluster of different ICEs. Conserved regions are highlighted.

Characteristic of some of the SXT-like elements is the presence of several resistance genes (responsible for the multiple resistance phenotype found in *V. cholerae* O139), grouped in a cluster with a transposon-like structure, probably acquired through horizontal gene transfer by the element and inserted at the *rumBA* locus (Beaber *et al.*, 2002b). The *rumAB* operon encodes proteins homologous to the UmuCD proteins of *E. coli* implicated in the repair of UV-damaged DNA, and the transfer induction of SXT/R391 by UV exposure (McGrath *et al.*, 2006). In SXT^{MO10} and ICE*Vch*Ind4, the resistance cluster is ca. 17 kb long and includes resistance genes for trimethoprim (*dfr18*), chloramphenicol (*floR*), streptomycin (*strAB*) and sulfamethoxazole (*sulII*) (Fig. 2.2b). The same genes are found in the resistance cluster of ICE*Vfl*Ind1, an ICE found in a *V. fluvialis* clinical isolate (Ahmed *et al.*, 2005). The presence of several transposases suggests a composite structure for this cluster, where genes may be acquired or depleted via homologous recombination. This is further supported by different arrangements observed in the resistance cluster of other ICE, for example, the ICE*Vch*Ind1 and ICE*Vch*Ban1, both elements found in *V. cholerae* O1 El Tor that lack the *dfr18* gene, inserted between two transposases. These elements carry trimethoprim resistance on an integron-like structure inserted at the *073/traF* hotspot, located at the 3'-end. ICE*Vch*Lao1, ICE*Vch*Vie1 and ICE*Pda*Spa1 also carry tetracycline resistence determinants in this cluster (Fig 2.2b) (see below). Other SXT-like elements characterized by various molecular rearrangements in the antibiotic resistance cluster have been described in *Vibrio spp.* and other bacterial taxa. R391, ICE*Vch*Hko1 and ICE*VchMex*1, ICE*Spu*PO1 are characterized by loss of the antibiotic cluster inserted in the *rumAB* operon (Wozniak *et al.*, 2009).

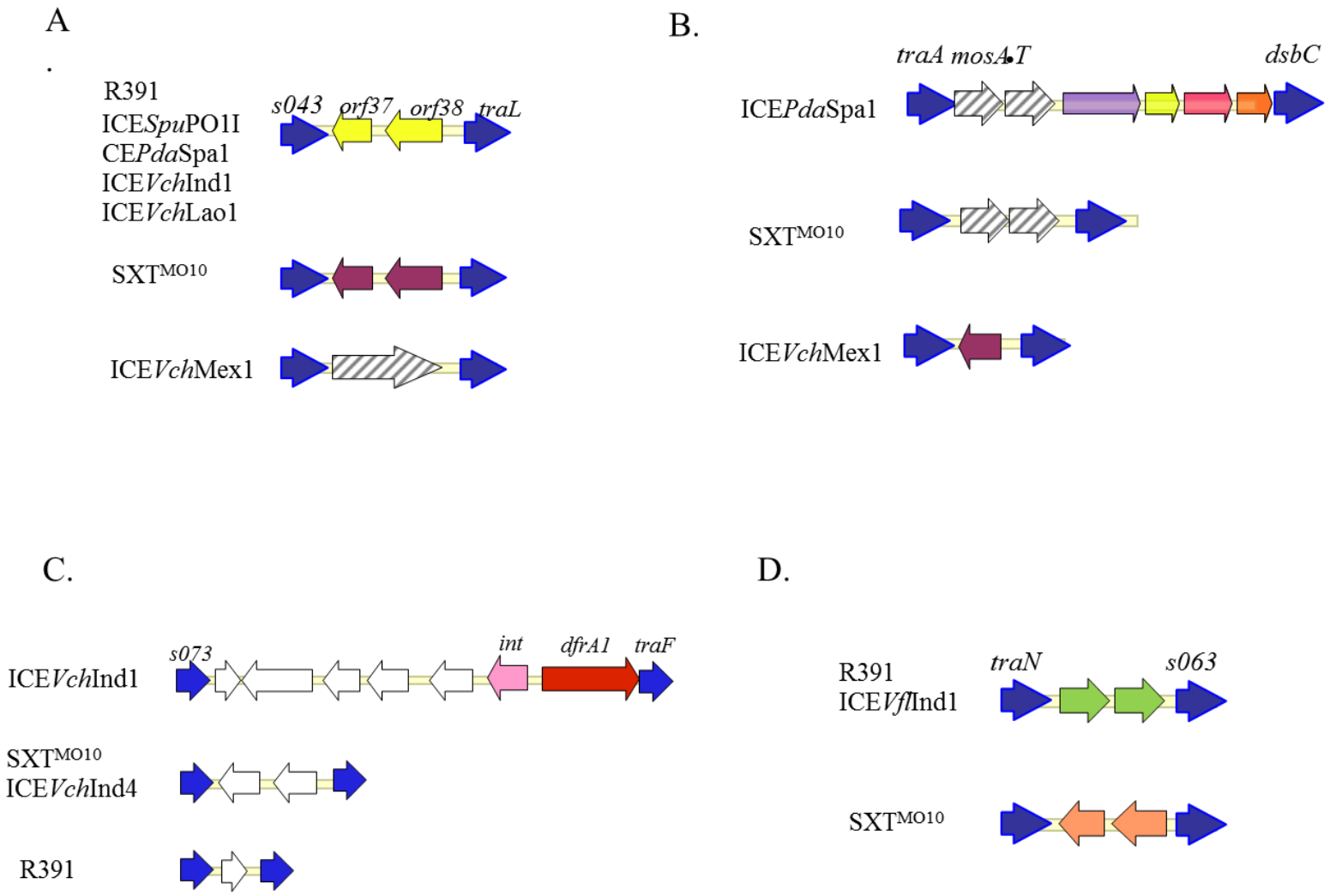


Figure 2.3.Genetic organization of ICE hotspots.A. Hotspot 1; B. Hotspot 2; C. Hotspot 3; D. Hotspot 4.

Five hotspots are located between genes *s043-traL* (HS1), *traA-s054* (HS2), and *s073-traF* (HS3), *traN-s063* (HS4) and *s026-traI* (HS5) (Fig. 2.3). Despite high genetic variability, the inserted sequences are confined to intergenic regions separating the four *tra* gene clusters, thus preserving the functional core. Among the various genes found in hotspots, most are of unknown function. Others encode adaptative systems, such as kanamycin and mercury resistance, class IV integron, enzymes involved in biofilm formation, genes involved in recombination and DNA processing, such as helicases, endonucleases, restriction-modification systems, and the toxin-antitoxin system (MotAB, HipAB) believed to protect against loss of the element (Wozniak *et al.*, 2009).

ICE in V. cholerae and geographical distribution

To date, besides SXT^{MO10}, several other ICEs have been described in *V. cholerae* (Fig 2.4).

Asia.

Almost all *V. cholerae* clinical isolates and various environmental *V. cholerae* non-O1/non-O139 in Asia contain SXT-related ICEs (Ehara *et al.*, 2004; Iwanaga *et al.*, 2004; Thungapathra *et al.*, 2002; Burrus *et al.*, 2006). The element has also been detected in species other than *V. cholerae* (Ahmed *et al.*, 2005).

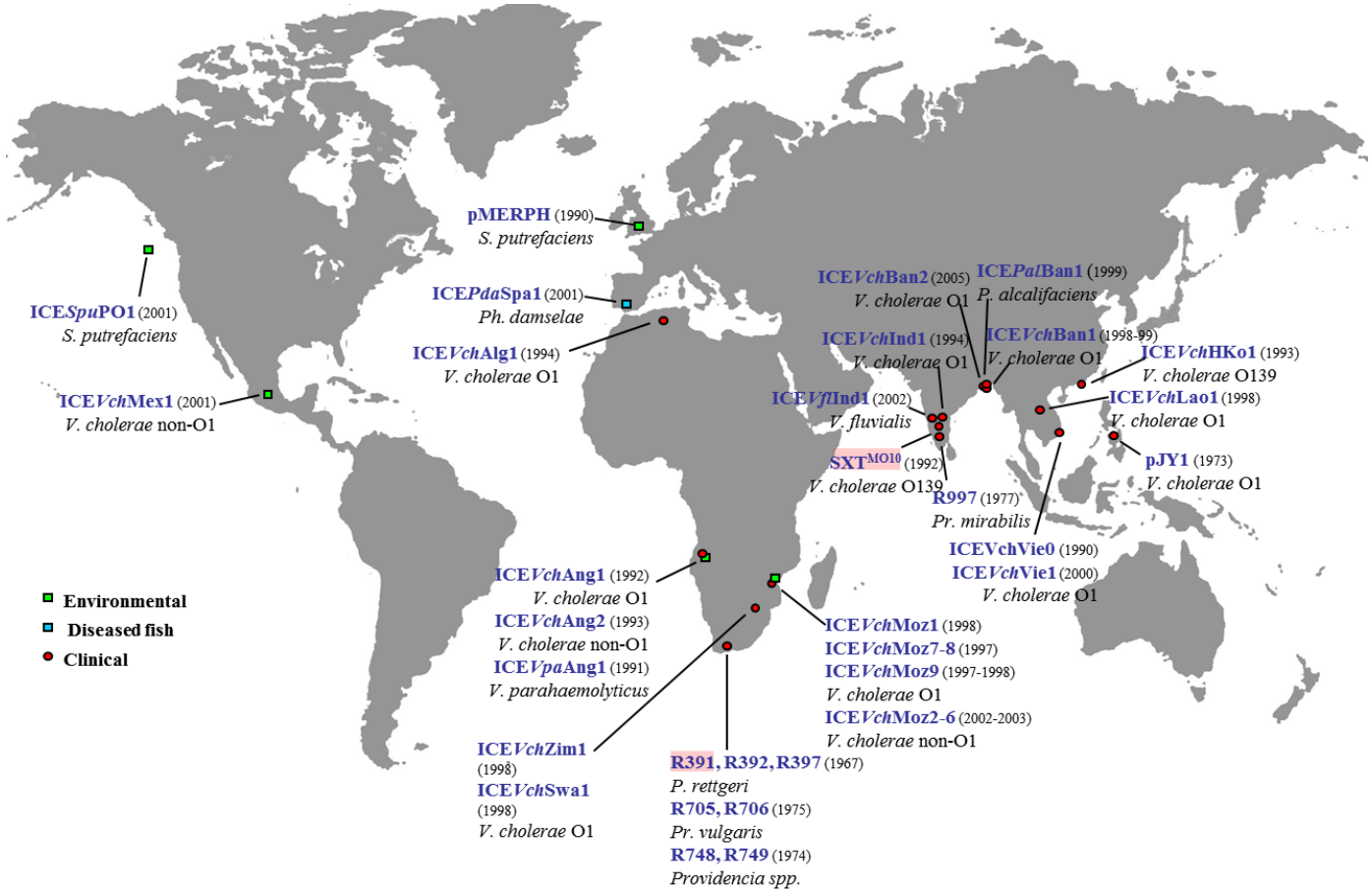


Figure2.4.Geographical distribution of ICE of the SXT/R391 family described to date.

ICE*Vch*Ind1 (SXT^{ET}) was isolated from strains of *V. cholerae* O1 El Tor in India, in 1994, after its re-appearance as an epidemic clone following the emergence of *V. cholerae* O139 in 1992 (Hochhut *et al.*, 2001b). It carries the same drug resistance as SXT^{MO10}, but with a different genetic organization, (Beaber *et al.*, 2002b).

ICE*Vch*Lao1 (SXT^{LAOS}) was identified in strains of *V. cholerae* O1 isolated in Laos after 1997 (Toma *et al.*, 2005). Its resistance cluster includes tetracycline, but not the trimethoprim resistance gene. Two new ORFs were identified in hotspot 3 (*s073-traF*), showing homology with the *s024* (exonuclease) and *s023* (helicase) genes of a genomic island associated with the multi-resistance cluster of *Salmonella enterica*. Thus, they appear to have been acquired by lateral transfer from this organism (Iwanaga *et al.*, 2004).

ICE*Vch*Hko1 (SXT^{S}) was described in a strain of *V. cholerae* O139, sensitive to cotrimoxazole (trimethoprim and sulfametoxazol), isolated in Hong Kong in 1994 (Yam *et al.*, 1998). Interestingly, its molecular characterization provides evidence of the absence of the cluster containing the typical resistance genes of SXT^{MO10} (Hochhut *et al.*, 2001b). This genetic organization suggests a deletion occurred, mediated by homologous recombination, between two extremities flanking the cluster, which made the strain sensitive. Alternatively, it can be suggested that this sensitive *V. cholerae* O139 strain never carried resistance genes and may be considered an ancestral form of the SXT-related ICEs (Hochhut *et al.*, 2001b). Since it is highly related to SXT^{MO10}, but depleted of the resistance cluster, it can be

hypothized that drug resistance is not the only selective pressure for acquisition and persistence of ICE genetic material (Burrus *et al.*, 2006).

Certain strains of *V. cholerae* O1 isolated in Vietnam in 2000 are characterized by **ICE*Vch*Vie,** an ICE encoding tetracycline resistance effector TetA and its repressor TetR. Like other SXT-related ICEs, ICE*Vch*Vie1 does not contain genes for trimethoprim resistance (Ehara *et al.*, 2004).

Interestingly, the presence of SXT elements was detected in *V. cholerae* non-O1/non-O139 isolated from water and sewage in Varanasi, India, before 1992 (Mohapatra *et al*., 2008). In six strains among 27 tested, the SXT integrase was detected by PCR. The strains tested were isolated between 1979 and 1989, but SXT-carrying strains were all isolated in 1986 and 1987. Presence of the resistance cluster typical of SXT^{MO10} and conjugative transfer were confirmed in four strains, two of which appeared to carry a sensitive element (Mohapatra *et al*., 2008).

Africa.

An outbreak of cholera in 1998 in South Africa was characterized by the SXT element (Daalsgard *et al.*, 2001). Nevertheless, the information about circulation of ICE and results of analysis of their role in emergence of drug resistance are limited, at least for Africa. In Angola, SXT integrase characterized most of the clinical *V. cholerae* O1 strains isolated in 1992, the same year of isolation, in India, of *V. cholerae* O139 MO10 containing SXT. This discovery is congruent with the observation of the ubiquitous presence of an ICE similar to ICE*Vch*Ind1 in *V. cholerae* strains isolated from Zimbabwe, Swaziland, and Mozambique, between

1997 and 1998 (Ceccarelli *et al.*, 2006) (Fig 2.4). Furthermore, ICEs were also characterized in environmental *V. cholerae* non-O1/non-O139 isolated in Mozambique between 2002 and 2003 in a cholera endemic area around Maputo (Taviani *et al.*, 2008) (Fig 2.4).

The widely dispersed presence of ICE in austral Africa is also supported by the discovery of the first member of the ICE, R391 in environmental strains of *Providencia spp.* in 1967, suggesting the hypothesis of independent origin and diffusion of SXT/R391 ICE in Africa (Burrus *et al.*, 2006) or that these elements originated in the African continent, spreading to the east.

Americas

Discovery of ICE*Vch*Mex1 provides the first and only report of an SXT-related ICE in the western hemisphere to date. It was described in an environmental strain of *V. cholerae* isolated in Mexico in 2001 (Burrus *et al.,* 2006b) (Fig 2.4). ICE*Vch*Mex1 integrates in *prfC* and is self-transmissible. Given its unique molecular organization, it is believed that ICE*Vch*Mex1 evolved independently, with respect to other ICEs related to SXT^{MO10} and ICE*Vch*Ind1. The *V. cholerae* O1 El Tor causing cholera epidemics in Latin America is related to pre-O139 Asian *V. cholerae* El Tor, which did not contain SXT, thereby suggesting acquisition of ICE occurs independently (Burrus *et al.*, 2006b).

Other ICEs

Since 1994, various other ICEs have been characterized in species other than *V. cholerae*.

ICE*Vfl*Ind1 was identified in a multiresistant strain of *Vibrio fluvialis* isolated in India in 2004 (Ahmed *et al.*, 2005). Although the resistance cluster appears to be conserved, as in SXT^{MO10} (*floR*, *sul2*, *strA*, *strB* and *dfrA18*), the integrase and the *attP* site involved in recombination showed some differences. The ability of ICE*Vfl*Ind1 to transfer between *V. cholerae* and *V. fluvialis* was demonstrated *in vitro* (Ahmed *et al.*, 2005).

ICE*Pda*Spa1 was identified in 2005 in a strain of *Photobacterium damselae* subsp. Piscicida from Spain (Juiz-Rio *et al*., 2005). This ICE presents all the characteristics necessary for inclusion in the SXT/R391 family of ICE, namely an integrase similar to int_{SXT}, integration at the *prfC* site, and presence of three conserved regions involved in the transfer of the element, *traI*, *traC* and *traN*. This is the first ICE of the SXT/R391 family isolated in Europe.

Identified as a result of the complete sequencing of the genome of *Shewanella putrefaciens*, **ICE*Spu*PO1** (110 kb), demonstrates high sequence similarity (> 96%) with R391. It contains an elevated number of insertions that confer resistance to copper and other heavy metals and various multi-function effusion pumps, similar to the effusion systems of antibiotic resistance. Integrated in the *prfC* gene of *S. putrefaciens*, ICE*Spu*PO1 carries an integrase identical to that encoded by R391 and has, therefore, been included in the SXT/R391 family of ICEs (Pembroke and Piterina, 2006).

Chapter 3. Complete genomic and functional analysis of ICEs identified in *V. cholerae* O1 El Tor.

Introduction

Integrative conjugative elements (ICEs) are a class of self-transmissible mobile elements that have been found in the genomes of prokaryotes (Burrus and Waldor, 2004; Burrus *et al*., 2006). As mediators of horizontal gene transfer in bacteria, their importance in bacterial evolution is increasingly appreciated. ICEs are, indeed, able to self transfer from a donor to a recipient cell and integrate into the host chromosome (Hochhut and Waldor 1999). Initially described as conjugative plasmids, namely R factors, they were assigned to the IncJ exclusion group of conjugative plasmids (Burrus *et al.*, 2002). Subsequently they were found unable to replicate and were classified as conjugative transposons that transfer to recipient cells via a non-replicative circular intermediate (Burrus *et al.*, 2002). ICEs are a genetic mosaic of plasmids, phages, and transposon-like elements, providing insight into mechanisms whereby genetic regions assemble and become incorporated in the evolutionary process of the bacterial genome.

Two ICEs, the SXT_{MO10} element described in a clinical strain of *Vibrio cholerae* serotype O139 isolated in India in 1992 and R391, a IncJ element from a strain of *Providencia rettgeri* isolated in South Africa, are functionally and

genetically related (Hochhut and Waldor 1999; Hochhut *et al.,* 2001; Beaber *et al.*, 2002; Boltner *et al.,* 2002; Burrus *et al.,* 2006).

Before the appearance of the *V. cholerae* O139 serogroup, the SXT-like elements had not been detected in *V. cholerae*. This serogroup is considered responsible for introduction of ICEs into *V. cholerae* and ICEs have since been found in the majority of *V. cholerae* O1 and O139 isolates from Asia. Furthermore, ICEs of the SXT/R391 family have been described in *V. cholerae* non-O1 non-O139, several related *Vibrio* spp., and other bacterial species (Iwanaga *et al.,* 2004; Ahmed *et al.*, 2005; Burrus *et al.,* 2006; Pembroke *et al.,* 2006; Osorio *et al.,* 2008; Taviani *et al.*, 2008). Members of this family carry a related integrase, int_{SXT}, and insert into the host chromosome at the same locus, i.e. the *prfC* gene (Burrus *et al*., 2006). From results of a molecular analysis of several ICEs, it is concluded that these elements share a conserved genetic scaffold that incorporates a plasmid-like transfer system and a phage-like integration and control cluster in which specific sequences are inserted. SXT^{MO10} and R391 carry different insertions, while newly described ICEs share insertions with both prototype elements, or carry unique sequences (Hochhut *et al.,* 2000; Beaber *et al.,* 2002; Boltner *et al.,* 2002).

In *V. cholerae*, several mobile elements have been reported and five have been fully sequenced to date. Evidence shows that ICEs found in *V. cholerae* are a polymorphic class of elements in genetic content, making comparative genomic analysis necessary to understand fully the extent to which these elements play a role in *V. cholerae* ecology and evolution.

In this study, a genomic and functional analysis was carried out on three SXT-related ICEs present in three clinical isolates of *V. cholerae,V. cholerae* O1 El

Tor MJ1236 isolated from Bangladesh in 1994, *V. cholerae* O1 El Tor B33 isolated from Mozambique in 2004, and *V. cholerae* CIRS101 isolated in 2002 in Bangladesh.

Objective of this study

The study was carried out to complete the genomic and functional analysis of three SXT-related ICEs found in three clinical isolates of pandemic *V. cholerae* and the comparative analysis was aimed at determining correlations and differences among ICEs of the R391/SXT family in *V. cholerae* and related species.

Material and Methods

Bacterial strains and DNA preparation

The three *V. cholerae* El Tor O1 strains used in this study are *V. cholerae* MJ1236, isolated in *Bangladesh* in 1994, *V. cholerae* CIRS101, isolated in 2002 in Bangladesh, and *V. cholerae* B33, isolated in Beira, Mozambique, in 2004.

The cultures were grown in Luria-Bertani (LB) medium and stored at -80C in LB enriched with added 25% glycerol. Total genomic DNA was extracted as described in Current Protocol in Molecular Biology (Ausubel *et al.,* 1990).

Genome sequencing

Draft sequences were obtained from a blend of Sanger and 454 sequences (Chun *et al.,* 2009). In total the libraries provided 6.5X coverage. Library construction and sequencing were performed at the Joint Genome Institute (JGI),

details can be found at http://www.jgi.doe.gov/. To finish the genomes, a collection of custom software and targeted reaction types were used. The Phred/Phrap/Consed software package (www.phrap.com) was used for sequence assembly and quality assessment. Reads were assembled with Phrap (Chun *et al.*, 2009). Possible mis-assemblies were corrected with Dupfinisher or transposon bombing of bridging clones (Epicentre Biotechnologies, Madison, WI). Gaps between contigs were closed by editing in Consed, custom primer walk or PCR amplification (Roche Applied Science, Indianapolis, IN). Gene-finding and annotation were achieved using the RAST server.

Comparative genomics

Genome to genome comparison was performed using three approaches, since completeness and quality of nucleotide sequences varied from strain to strain in the set examined in this study. Firstly, nucleotide sequences as whole contigs were directly aligned using the Mummer program. Secondly, ORFs of a given pair of genomes were reciprocally compared each other, using the BLASTN, BLASTP and TBLASTX programs (ORF-dependent comparison). Thirdly, a bioinformatic pipeline was developed to identify homologous regions of a given query ORF. Initially, a segment on a target contig homologous to a query ORF was identified using the BLASTN program. This potentially homologous region was then expanded in both directions by 2,000 bp. Then, nucleotide sequences of the query ORF and selected target homologous regions were aligned using a pairwise global alignment algorithm and the resultant matched region in the subject contig was extracted and saved as an

homolog (ORF-independent comparison). Orthologs and paralogs were differentiated by reciprocal comparison. In most cases, both ORF-dependent and independent comparisons yielded the same orthologs, though the ORF-independent method performed better for draft sequences of low quality, in which sequencing errors, albeit rare, hampered identification of correct ORFs (Chun *et al*., 2009).

ICEs comparative genomics, genetic organization and annotation

Genetic comparisons of the ICEs were accomplished by following four separate steps. First, the nucleotide sequence of ICE*Vch*B33 was aligned with SXT^{MO10} and R391, using NucMer (Kurtz *et al.,* 2004). Second, ORFs were identified and annotated using the RAST pipeline (http://rast.nmpdr.org). The ICE sequence and genetic organization were then compared with published ICEs using the Artemise Comparative Tool (ACT) (www.sanger.ac.uk/Software/ACT). Similarity in nucleotide and protein sequences for ICE*Vch*B33 was determined as % nucleotide or amino acid identity with other ICEs in GenBank, employing BLASTN and BLASTP.

ICE transfer

ICE transfer was confirmed by conjugation assays, as previously described (Beaber *et al*., 2002, Burrus and Waldor 2003). *V. cholerae* B33, *V. cholerae* CIRS101, and *V. cholerae* MJ1236 were used as donor and *Escherichia coli* CP126 resistant to Nalidixic Acid as recipient. Overnight cultures of marked donor and recipient cells were mixed on an LB plate in 1:1 ratio and incubated at 37°C for 5 h. The cells were resuspended in LB broth and dilutions were plated on selective media,

allowing enumeration of donors, recipients, and exconjugants. Trimethoprim resistance was used as a marker for ICEs. ICE transfer frequency was calculated as the number of Tp^R/Nal^R exconjugants observed per Tp^R donor cell.

PCR assays

The excision, transfer and integration of ICE*Vch*B33 and ICE*Vch*CIRS101 were confirmed by PCR. P4 (TGCTGTCATCTGCATTCTCCTG) and P5 (GCCAATTACGATTAACACGACGG) primers were used to amplify the circular form of the ICE. PCR with primers pair INT1 (GCTGGATAGGTTAAGGGCGG) and INT2 (CTCTATGGGCACTGTCCACATTG) confirmed the transfer of the ICE into the *E. coli* recipient and primers P3 (CAGCTACAACTGAGCATTGGC) and PVC-B (ACCACGCTCTTTTTCCATTTCCAT) confirmed its integration at the *prfC* locus in exconjugants (Hochhut *et al.*, 1999; Taviani *et al.*, 2008).

Results and discussion

Location of ICE in the genome of *V. cholerae* O1 El Tor MJ1236 and *V. cholerae* O1 El Tor B33 and assembly of ICE*Vch*B33

A BLASTN search was done using the SXT^{M010} integrase nucleotide sequence in the genomes of *V. cholerae* MJ1236 and B33, detecteding an homologous gene at nt 3061809 of chromosome 1 of MJ1236 and at position 12187 of contig 3 of B33. As was expected, in ICEs of the SXT/R391 family, in both strains, the integrase was found downstream of the truncated *prfC* gene encoding the peptide

chain release factor 3 (Burrus and Waldor 2003). Further analysis confirmed a complete 106,014 bp SXT-like ICE in *V. cholerae* MJ1236 and detected SXT-like nucleotide sequences located on three contigs of *V. cholerae* B33: nt 1–16349 of contig 3; the entire contig 8; and nt 1–12120 of the contig 4 (**accession no.** ACHZ0000000). Assembly of the SXT-related sequences in *V. cholerae* B33 was done using NucMer, with the nucleotide sequence of R391 and SXT^{M010} as reference. Alignment by NucMer revealed two gaps at junctions between the three contigs that were resolved by PCR and confirmed by sequencing. The two ICE-like sequences showed 99% sequence similarity and, therefore, were considered to be the same ICE. We refer to this element as ICE*Vch*B33, since *V. cholerae* B33 ICE had been named previously (Ceccarelli *et al.*, 2008).

Genomic organization of ICE*Vch*B33

The two ICE-like sequences identified in the genome of *V. cholerae* B33 and *V. cholerae* MJ1236 were submitted to the RAST pipeline for annotation (http://rast.nmpdr.org). RAST identified 99 ORFs in both elements (Table 3.1). Based on the significant similarity between integrase genes of the elements and SXT/R391 integrase (99% similar to the int gene of both SXT^{M010} and R391) and insertion at the *prfC* locus, ICE*Vch*B33 was confirmed to be an ICE belonging to the SXT/R391 family. Comparative analysis of ICE*Vch*B33 and other ICEs in GenBank, using the Artemis Comparative Tool (ACT) (http://www.sanger.ac.uk/Software/ACT), was done. Overall, the general organization of ICE*Vch*B33 was found to be highly similar to SXT/R391 ICEs, namely SXT^{M010}, R391, and ICE*Pda*Spa1 (Osorio *et al*., 2008)

ICE*Vch*B33, as all other related ICEs, possesses an highly conserved core of genes responsible for transfer, integration, excision, and control, that encompasses ca. 60 kb (Fig. 3.1). As noted for other ICEs, specific inserted genes were identified in four hot spots and two variable regions within the core backbone of ICE*Vch*B33. Of the 99 ORFs identified by RAST, 75 were present in SXT^{M010}, 65 in R391 and 73 in ICE*Pda*Spa1 (Table 3.1).

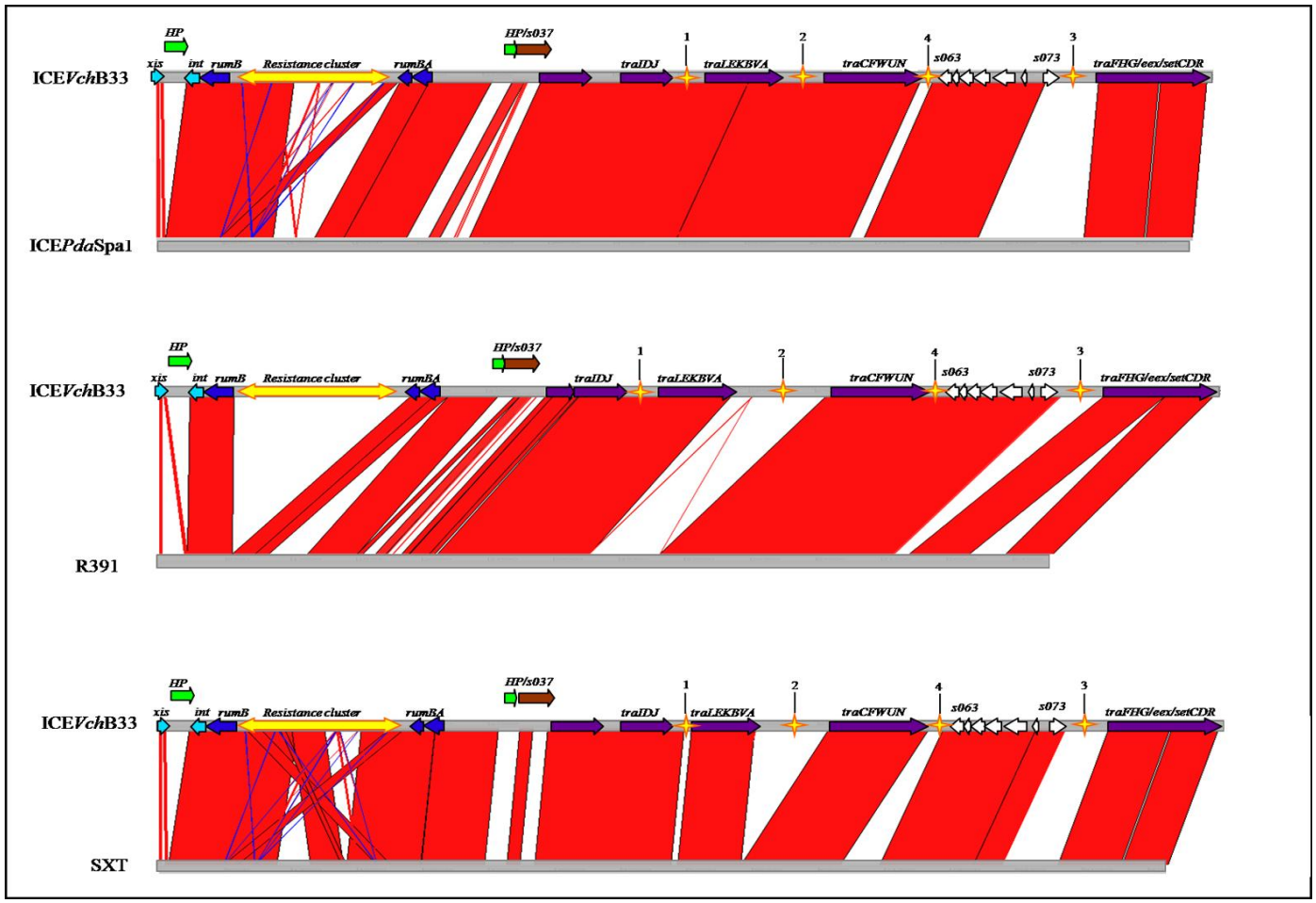


Fig. 3.1. *Alignment generated by Artemis Comparative Tool (ACT) of ICEVchB33 with: (A) ICEPdaSpa1; (B) R391; and (C) SXT^{MO10}. Red areas indicate homologous regions and blue lines indicate inversions. Conserved regions are schematically represented by arrows, int/xis, the resistance cluster, umuCD and tra genes are shown in blue, yellow and purple, respectively. Other genes are indicated by the white arrows. Two hypothetical proteins unique to ICEVchB33 are shown in green. Numbered yellow stars indicate hot spots 1–4.*

At the 5′ end of the element, two ORFs located between nt 384–491 and 872–2890, respectively, appeared to be specific to this element, since they were not found in eitherSXT^{M010} , R391, or ICE*Pda*Spa1 (Table 3.1, Fig. 3.1). The first ORF was 108 nt long and did not show any similarity when searched by BLASTN or BLASTP. A second ORF, located between nt 872 and 2890, is also unique to ICE*Vch*B33 and showed 73% similarity to a putative DNA mismatch repair protein in *Shewanella spp*., and 40% identity with the same protein family in *Vibrio splendidus* 12B01 and *Burkholderia thailandensis* E264. BLASTP analysis of the protein detected an histidine kinase-like ATPase motif, commonly associated with ATP-binding proteins, e.g. histidine kinase, DNA gyrase B, topoisomerases, heat shock protein, and DNA mismatch repair proteins involved in DNA replication, recombination, and repair (Spampinato and Modrich, 2000). The MutL ATPase is required for mismatch repair. A protein belonging to this family, the MutL ATPase of *Escherichia coli*, is involved in DNA repair, therefore a part of the mismatch repair system (Spampinato and Modrich, 2000). Interestingly, this ORF is located between the *xis* and *int* genes (Table 3.1) usually found adjacent in ICEs in an operon (Burrus and Waldor, 2003), but its presence did not interfere in the integration/excision of the ICE (see below), showing both *xis* and *int* to be functional. Since *xis* belongs to the recombination directionality factor (RDF) protein family, a highly diverse group of small proteins, the RAST pipeline failed to annotate the ICE*Vch*B33 *xis* gene. Through a BLASTN search, using the SXT^{M010}*xis* as query, the orthologous *xis* was located in the ICE*Vch*B33 genome, between nt 539 and 722, with 97% nucleotide similarity (Table 3.1).

Table 3.1 *ORFs annotated in ICEVchB33 by RAST and nucleotide sequence comparison with ICEs R391, SXT^{MO10}, and ICEPdaSpa1 unique ORFs in ICEVchB33 are in bold. Content of the four hotspots is shown in red.*

Function	Length (bp)	% Identity		
ICE*Vch*B33		R391	SXT	ICE*Pda*Spa1
Hypothetical protein	**108**	**0**	**0**	**0**
xis recombination directionality factor	183	97	97	97
Putative DNA mismatch repair protein	**2019**	**0**	**0**	**0**
Integrase-like protein	1242	99.27	99.03	99.27
Hypothetical protein	276	92.13	94.38	93.26
Rod shape determination protein	975	100	99.38	99.38
Hypothetical protein	444	97.28	99.32	99.32
Error-prone repair protein UmuC (RumB)	1173	92.82	97.44	97.44
Transposase	594	0	100	100
Transposase	2739	0	99.85	99.86
IS, transposon-related functions	1494	0	100	100
Type IV secretory pathway, VirD2 (relaxase)	885	0	100	0
Florfenicol/chloramphenicol resistance protein	1215	0	99.5	30.41
Glycine cleavage system transcriptional activator	306	0	100	0
Tetracycline efflux protein TetA	1275	0	0	47.11
Tetracycline repressor protein	678	0	0	45.77
Aminoglycoside-phosphotransferase	195	0	90.32	0
Hypothetical protein	**237**	**0**	**0**	**0**
Probable NreB protein	**219**	**0**	**0**	**0**
Cobalt–zinc–cadmium resistance protein czcD	**1023**	**0**	**0**	**0**
Transcriptional regulator, ArsR family	**378**	**0**	**0**	**0**
Streptomycin phosphotransferase	837	0	100	0
Aminoglycoside-phosphotransferase	804	0	100	0
Dihydropteroate synthase	816	0	100	0
Transposase	1134	0	99.72	99.35
Error-prone repair protein UmuC (RumB)	270	98.72	100	98.88
Error-prone repair protein UmuD (RumA)	450	98.66	100	97.32
Polymerase epsilon subunit	906	98.34	99.67	98.34
Hypothetical protein	315	99.04	99.04	98.53
COGs COG2378	927	97.4	99.68	100
Putative inner membrane protein	600	97.84	98.76	100
Putative cytoplasmic protein	585	99.48	100	100
Hypothetical protein	3678	97.88	99.52	99.43
Hypothetical protein	**1764**	**0**	**0**	**0**
Putative type II restriction enzyme methylase	3627	98.82		96.26
Putative cytoplasmic protein	2658	97.65	97.81	99.44
Putative ATP-dependent protease	2076	98.09	99.28	99.57
ATPase involved in DNA repair	2652	98.98	98.75	99.66
TraI	2151	94.27	93.02	93.3
TraD	1821	99.34	99.17	99.5
Hypothetical protein	561	96.63	98.72	97.31
Conjugative transfer protein s043	645	98.6	99.53	98.1
Hypothetical protein	621	94.17	0	92.23
TraL	282	97.85	100	98.92
TraE	627	99.04	99.04	99.52
TraK	903	97	97.32	95.32
TraB	1290	98.83	98.14	98.6
TraV	651	99.07	99.54	98.61
TraA	387	99.22	100	100
Ynd	834	0	98.92	98.56
Ync	891	0	97.97	96.96
DNA helicase, UvrD/REP family	2730	0	0	99.56
Hypothetical protein	1110	0	0	96.75
Cardiolipin synthetase domain protein	1536	0	0	96.48
Type 4 fimbriae expression regulatory protein pilR	1446	0	0	97.92
Thiol:disulfide interchange protein DsbC	693	100	100	99.57
TraC	2400	99.37	99.37	99.25
Hypothetical protein	348	95.65	99.13	98.26
TrhF	513	97.06	97.65	92.94
TraW	1185	96.53	96.38	94.67
TraU	1029	97.95	98.54	98.54
TraN	3642	98.27	98.36	98.6
Hypothetical protein	474	100	0	0
Hypothetical protein	663	100	0	0
Unknown	603	99	97.5	97.5
Hypothetical protein	327	99.07	98.02	96.3
Single-stranded DNA-binding protein	420	97.12	95.68	99.28
Putative factor phage or prophage related	819	100	99.63	100
Hypothetical protein	144	95.74	95.74	95.74
DNA recombination protein	1017	99.11	100	100
cobS protein, putative	960	98.43	99.06	99.37
Hypothetical protein	768	98.82	98.85	98.04
Cobalamine biosynthesis protein	954	99.68	99.68	100

ICE*Vch*B33 resistance cluster

Like SXT^{M010} and ICE*Pda*Spa1, ICE*Vch*B33 contains a cluster of resistance genes inserted at the *rumB* locus, while in R391 the *rumB* gene is intact (Boltner *et al* 2002). Insertion of the resistance cluster occurs at the same base pair as SXT^{M010} (Fig. 3.1). In ICE*Vch*B33 the cluster spans a region of 19,284 bp and its genetic organization is similar to two other ICEs described in *V. cholerae*: ICE*Vch*Vie1 and ICE*Vch*Lao1 (Fig. 3.2) (Iwanaga *et al*., 2004). Although sequences of the ICE*Vch*B33 resistance cluster and SXT^{M010} are highly related, significant genetic rearrangement in this region was observed (Fig. 3.1).

The first three ORFs at the 5′ end of the resistance cluster are shared by SXT^{M010}, ICE*Pda*Spa1, and ICE*Vch*B33, in the same order: the three transposases of 594, 2139, and 1494 bp, respectively. Downstream of the truncated IS *tnpB*, ICE*Vch*B33 shows a genetic organization that is similar to ICE*Vch*Ind1, an ICE in *V. cholerae* O1 El Tor (Hochhut *et al.*, 2001), ICE*Vch*Vie1, and ICE*Vch*Laos (Iwanaga *et al.*, 2004). All of these elements lack the *dfr18* dihydrofolate reductase gene that encodes resistance to trimethoprim, the deoxycytidine triphosphate deaminase gene, and two other ORFs in SXT^{M010} (Table 3.1, Fig. 3.2). Like ICE*Vch*Ind1, ICE*Vch*Vie1, and ICE*Vch*Lao1, ICE*Vch*B33 carries *virD2* (homologous to *s013* of SXT^{M010}), followed by *floR*, conferring resistance to chloramphenicol and an ORF encoding the LysR family transcriptional regulator (Fig. 3.2). These three genes are present in SXT^{M010}, having the same arrangement downstream of the *dfr18* minicluster (Fig. 3.2).

It is noteworthy that ICE*Vch*B33 carries the efflux protein TetA, an effector of resistance to tetracycline and its repressor TetR, which are not found in SXT^{M010}.

These two genes are present in ICE*Pda*Spa1 and are the only resistance determinants carried in the resistance cluster of this ICE (Osorio *et al*., 2008). *tetR* and *tetA* were also found in the resistance cluster of ICE*Vch*Vie1 in a co-oriented fashion, while ICE*Vch*Lao1 carries only the *tetA* gene (Fig. 3.2).

Two new genes were described in ICE*Vch*B33: *nreB* and cobalt–zinc–cadmium resistance protein *czcD*. By BLASTN search, *nreB* gene was 95% similar to *Pseudomonas aeruginosa* LESB58 putative *nrbE*-like protein and cation transporter, also annotated as cobalt–zinc–cadmium resistance protein and present in *Shewanellasp*. W3-18-1 (Pembroke *et al*., 2006). *czcD* encodes a regulator usually associated with a three component chemostatic antiporter that pumps cations outside the bacterial cell. A cluster, comprising proteins CzcCBA, is found at the 3′ end of ICE*Spu*PO1 (Pembroke *et al*., 2006).

The 3′ end of the resistance cluster showed the same genetic arrangement as SXT^{M010}, encompassing *strB*, *strA*, *sulII*, conferring resistance to streptomycin and sulfamethoxazole, respectively, and a transposase (Fig. 3.2). Like SXT^{M010} and ICE*Pda*Spa1, ICE*Vch*B33 lacks the kanamycin resistance cluster present in R391.

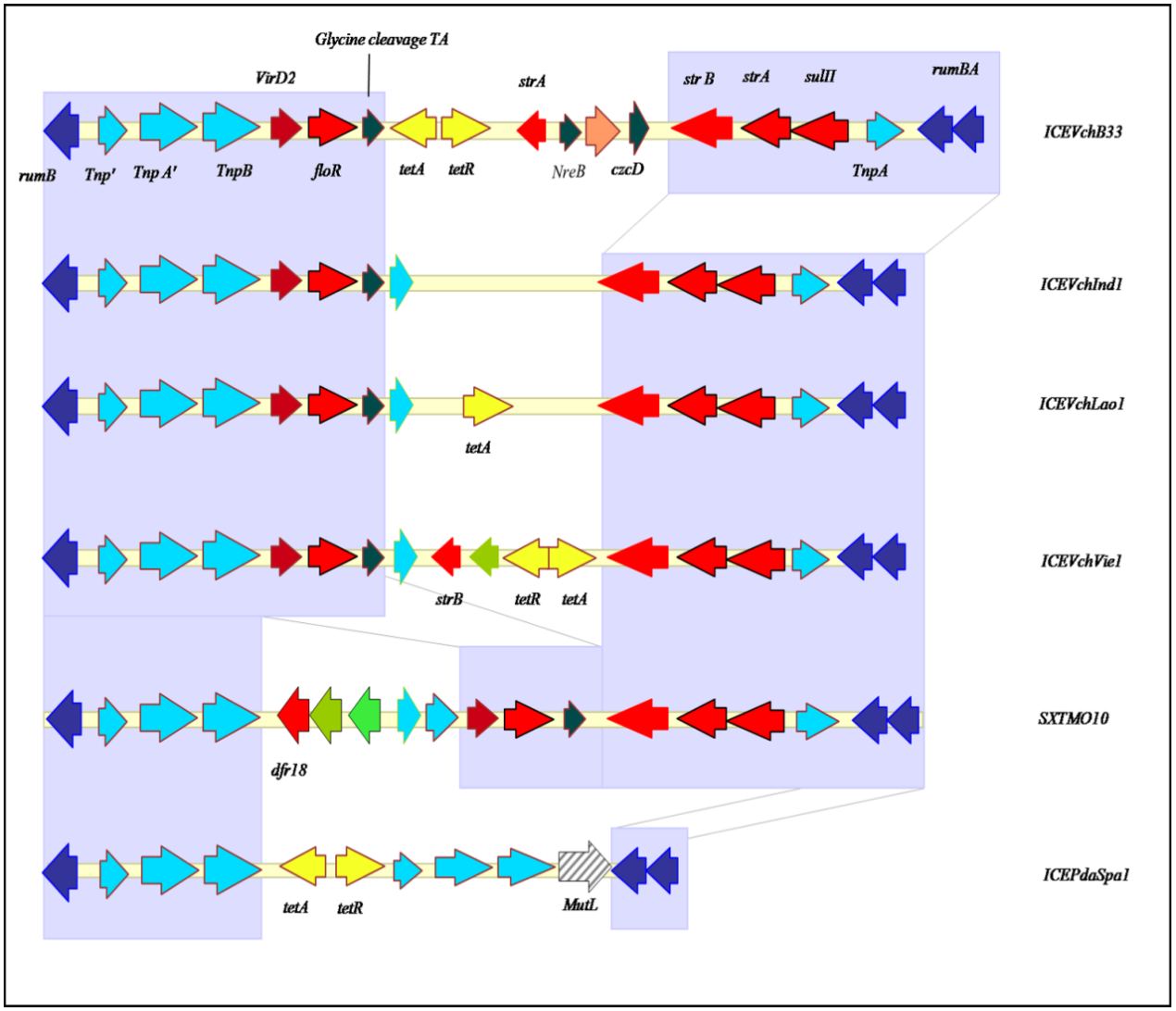


Fig. 3.2.*Genetic organization of the resistance cluster of different ICEs. Conserved regions are highlighted in blue.*

Genetic content of ICE*Vch*B33 hot spots

As in other ICEs, the ICE*Vch*B33 specific genes are inserted into five hot spots that confer specificity to the ICE.

Hot spot 1

Between genes *s043* and *traL*, the inserted sequence in ICE*Vch*B33 is identical to that of R391, ICE*Pda*Spa1, and several other ICEs, such as ICE*Spu*P01, ICE*Vch*Ind1, and ICE*Vch*Lao1 (Hochhut *et al*., 2001; Iwanaga *et al*., 2004), which contain *orf37* and *orf38*. SXT^{M010} and ICE*Vch*Mex1 are the only ICEs with a different insertion (Fig. 3.3A).

Hot spot 2

Inserted at the *traA-dsbC* locus, the element under study has the same inserted cluster as ICE*Pda*Spa1 (Fig. 3.3B). In ICE*Vch*B33, the cluster is 9480 bp long, between nt 56780 and nt 66260, and includes six ORFs with 97% similarity to ICE*Pda*Spa1. The first two genes downstream of *traA*, at the 5′ end of the inserted sequence, have 97% amino acid sequence similarity with the *mosA* and *mosT* genes in SXT^{M010}. These two genes, previously annotated as *ync* and *ynd* encoding two proteins of hypothetical function, have been recognized recently as a toxin–antitoxin module, essential for maintenance of SXT in its host (Wozniak *et al.*, 2009). At the 3′ end of the inserted sequence, ICE*Vch*B33 carries ORFs *spa01–spa04*, previously described in ICE*Pda*Spa1, encoding an helicase, an hypotetical protein and two proteins annotated as cardiolipin synthetase domain protein, and the type 4 fimbriae expression regulatory protein PilR

(Table 3.1). PilR amino acid sequence showed only 31% identity with an homologous PilR protein in *CellVibriojaponicus* Ueda10. PilR proteins are Sigma-54 dependent DNA-binding response regulators of the Fis family and involved in regulation of fimbriae expression.

Hot spot 3

As noted for SXT^{M010} and R391, the ICE*Vch*B33 hot spot 3 is located at the 3′ end of *s073*, between nt 89450 and 99747. The Artemis Comparative Tool was used to show sequence divergence between ICE*Vch*B33 and ICE*Pda*Spa1 that starts at nt 88629, since the latter ICE lacks the *s073* gene (Table 3.1, Fig. 3.1).

Although the 5′ and 3′ boundaries of ICE*Vch*B33 Hotspot 3 are the same as those of SXT^{M010} and R391, the insertion content differs. Inserted at this locus, ICE*Vch*B33 carries a class 4 integron encoding dihydrofolate reductase (*dfrA1*) responsible for trimethoprim resistance, occurring exclusively as an integron cassette (Hochhut *et al.*, 2001). The same insertion is present in ICE*Vch*Ind1 (Hochhut *et al*., 2001), the two sequences showing 99% nucleotide similarity. As in ICE*Vch*Ind1, the *dfrA1* and the *intI9* integrase genes are oriented in opposite fashion, a common arrangement of integrons. Four integron cassettes, each characterized by presence of an *attC* site, precede *dfrA1* and all are of unknown function (Fig 3.3C).

Hot spot 4

At the 3′ end of gene *traN*, between nt 76025 and 77377 of ICE*Vch*B33, two ORFs encoding hypothetical proteins are inserted and are also present in R391 (96% sequence similarity), but absent in SXT^{M010} and ICE*Pda*Spa1 (Fig. 3.3D).

Hot spot 5

In the region between the *rumAB* cluster and hotspot 1 (nt 33540–38360), there is a sequence discrepancy in the backbone of ICE*Vch*B33, with respect to SXT^{M010}, R391, and ICE*Pda*Spa1 (Fig. 3.1). This region encompasses SXT^{M010} ORFs *s031–s038*, also found in ICE*Pda*Spa1. At position 33540, a 1764 bp ORF was annotated by RAST as an hypothetical protein. BLASTN results showed that the entire sequence is 97% similar to a protein of unknown function found in *E. coli* plasmid p1658/97 and *Escherichia fergusonii* ATCC 35469T plasmid pEFER. A second sequence discrepancy was found in the type II restriction enzyme methylase subunit gene (*s037*). In ICE*Vch*B33, the gene is 3111 bp longer than in SXT^{M010} and R391. Based on results of the BLASTN analysis, this gene is most likely truncated in SXT^{M010}, R391, and ICE*Pda*Spa1 (Fig. 3.1).

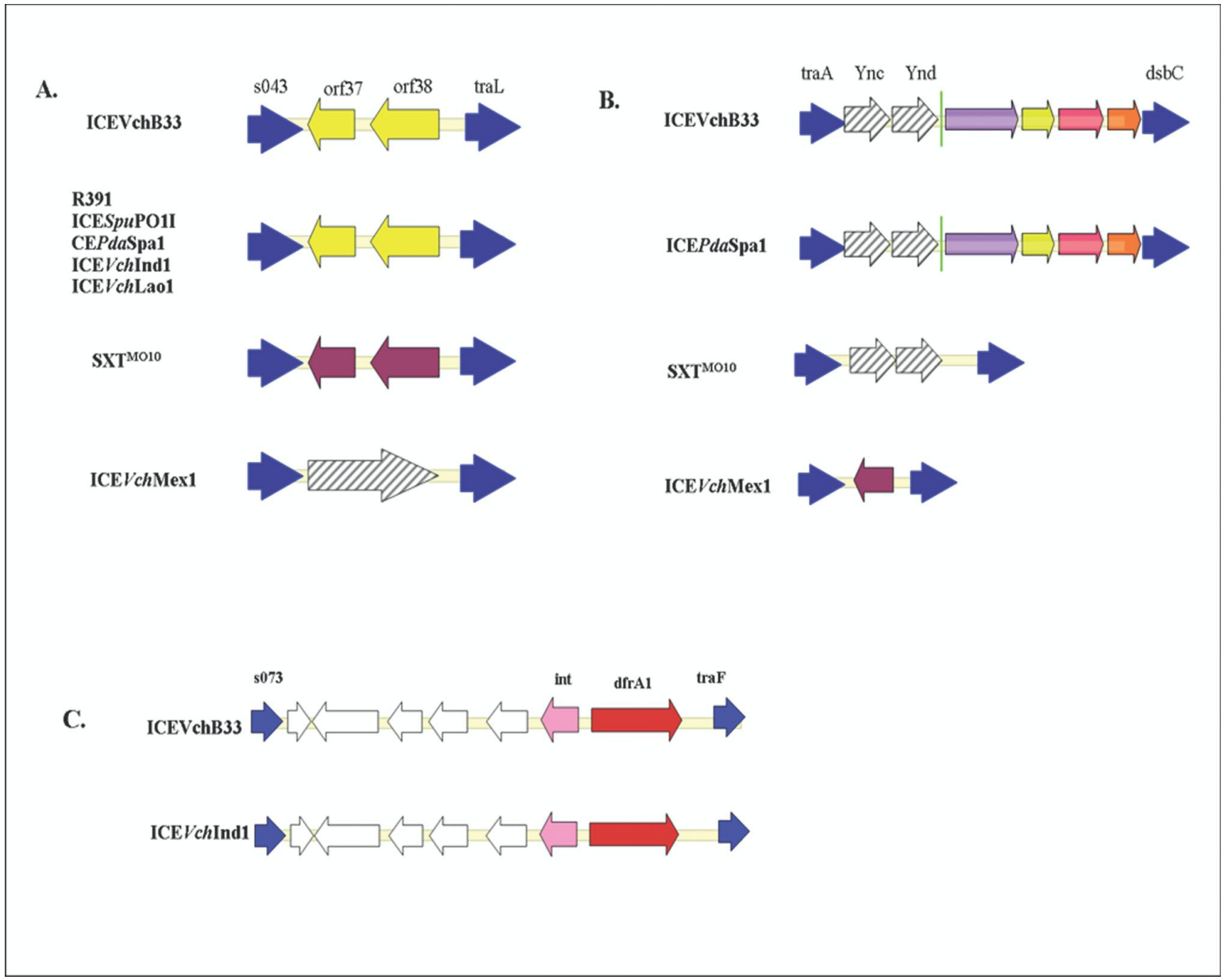


Figure 3.3. *(A) Genetic organization of hot spot 1; (B) genetic organization of hot spot 2; and (C) genetic organization of hot spot 3.*

Location of ICE in the genome of *V. cholerae* O1 El Tor CIRS101 and assembly of ICE*Vch*CIRS101

V. cholerae O1 El Tor CIRS101 also contains an ICE-like element (Fig. 3.2), which is 98 kb in size and inserted into the *prfC* gene. Based on significant similarity between the integrase gene (99% similar to the *int* gene of both SXT^{MO10} and R391) and insertion site, the element was confirmed as belonging to the SXT/R391 family and named ICE*Vch*CIRS101, according to the proposed nomenclature (Burrus *et al*., 2006). The element encompasses genes VCH_000014-67, 000179-188, and 000420-000433, which span three contigs of the draft genome, 77, 79, and 82, respectively.

Genomic organization of ICE*Vch*CIRS101

A comparative analysis of ICE*Vch*CIRS101 with ICE*Vch*B33 and other previously described ICEs, using the Artemis Comparative Tool (ACT) (www.sanger.ac.uk/Software/ACT), was performed. Overall, the general organization of ICE*Vch*CIRS101 was found to be highly similar to that of ICE*Vch*B33 and other members of the SXT/R391 ICEs. ICE*Vch*CIRS101, like all related ICEs, revealed a highly conserved core of genes responsible for transfer, integration and excision, and control that encompasses*ca.* 60 kb (Fig. 3.4A). As noted for other ICEs, specific inserted genes were identified in four hot spots and two variable regions within the core backbone. The 5' region of ICE*Vch*CIRS101, between nt 1-7543, shares 99% nucleotide similarity with ICE*Vch*B33 (Fig. 3.4A). The first two ORFs appear to be unique since they are found only in ICE*Vch*B33 and not in either SXT^{MO10}, R391, or ICE*Pda*Spa1 (Taviani *et al.*, 2009).

Gene function ICEVchClRS101	Length	%identity SXT	ICEVchB33	R391	ICEPdaSpa1	
xis	184	100	100	100	100	
putative DNA mismatch repair protein	***673***	***0***	***100***	***0***	***0***	
integrase	414	99.03	100	99.27	99.27	
hypothetical protein	90	94.38	100	92.13	93.26	
Rod shape determination protein	325	99.38	100	100	99.38	
hypothetical protein	***56***	***0***	***100***	***0***	***0***	
hypothetical protein	***45***	***0***	***100***	***0***	***0***	
hypothetical protein	148	99.32	100	97.28	99.32	
Error-prone repair protein UmuC	391	97.44	100	92.82	97.44	
putative transposase	*138*	*100*	100	*0*	*100*	
putative transposase	*120*	*95.69*	95.69	*0*	*0*	
hypothetical protein	*45*	*0*	0	*0*	*0*	
putative transposase	*713*	*100*	99.86	*0*	*100*	
RepA	*64*	*0*	0	*0*	*0*	
putative transposase	*498*	*9*	100	*0*	*100*	
hypothetical protein	***48***	***0***	***0***	***0***	***0***	
hypothetical protein	***152***	***0***	***0***	***0***	***0***	
sul1delta fusion protein	*47*	*15*	*97.83*	*0*	*0*	
Florfenicol/chloramphenicol resistance protein	*405*	*99.5*	100	*0*	*30.41*	
Transcriptional regulator	*93*	*98.91*	98.91	*0*	*0*	
streptomycin phosphotransferase	*277*	*100*	100	*0*	*0*	
Aminoglycoside -phosphotransferase	*268*	*99.63*	99.63	*29.62*	*0*	
Dihydropteroate synthase	*272*	*100*	100	*0*	*0*	
putative transposase	*697*	*99.5*	100	*0*	*99.5*	
Transposase	*404*	*0*	*0*	*0*	*0*	
Error-prone repair protein UmuC	90	100	100	98.72	98.88	
Error-prone repair protein UmuD	150	100	100	98.66	97.32	
COG0847: DNA polymerase III	302	97.01	97.34	98.34	96.68	
conserved hypothetical protein (s025)	100	94.67	96	96	89.74	
COGs COG2378 (s026)	307	63.64	63.29	62.94	63.29	
hypothetical protein	***267***	***0***	***0***	***0***	***0***	***HS5***
hypothetical protein	***183***	***0***	***0***	***0***	***0***	
hypothetical protein	*1230*	*22.28*	*19.39*	*22.78*	*19.39*	
hypothetical protein	*1247*	*27.78*	*23.16*	*23.02*	*27.27*	
hypothetical protein	*768*	*0*	*0*	*0*	*0*	
ATP-dependent protease La	*684*	*44.74*	*44.89*	*45.43*	*45.05*	
hypothetical protein	***257***	***0***	***0***	***0***	***0***	
hypothetical protein	***272***	***0***	***0***	***0***	***0***	
TraI	717	94.69	94.55	94.27	91.62	
Conjugative transfer protein TraD	607	99.67	99.5	99.17	99.34	
hypothetical protein	187	97.44	97.85	95.51	95.16	
Conjugative transfer protein s043	198	98.48	97.97	97.46	97.97	
hypothetical protein	*196*	*0*	*94.36*	*95.38*	*94.36*	***HS1***
IncF conjugative protein TraL	94	98.92	98.92	98.92	100	
IncF conjugative protein TraE	209	99.52	99.52	99.52	100	
IncF conjugative protein TraK	299	98.66	96.64	98.66	96.64	
IncF conjugative protein TraB	430	98.37	98.83	99.07	99.3	
IncF conjugative protein TraV	217	98.15	98.61	99.54	98.15	
TraA	129	97.66	97.66	96.88	97.66	
Ynd	*264*	*99.62*	*98.48*	*0*	*97.72*	***HS2***
Ync	*313*	*100*	*97.97*	*0*	*98.4*	
DsbC	231	99.57	99.57	99.57	99.13	
IncF conjugative protein TraC	800	99.5	99.37	99.5	99.37	

Gene function ICEVchCIRS101	Length	%identity SXT	ICEVchB33	R391	ICEPdaSpa1	
hypothetical protein	94	100	98.92	96.77	100	
Conjugative signal peptidase TrhF	171	97.06	94.71	97.65	97.06	
IncF conjugative protein TraW	375	95.72	96.52	97.33	94.92	
IncF conjugative protein TraU	343	99.42	98.54	98.25	98.83	
IncF conjugative protein TraN	1231	97.84	98.19	97.77	98.13	
SMC protein-like	***539***	***0***	***0***	***0***	***0***	***HS4***
transposase	***252***	***0***	***0***	***0***	***0***	
ISPsy20	***503***	***0***	***0***	***0***	***0***	
SMC protein-like	*339*	*0*	*0*	*0*	*0*	
endonuclease precursor	*228*	*93.39*	*0*	*0*	*0*	
hypothetical protein	201	98	96.5	97.5	98	
hypothetical protein	109	91.09	93.52	94.44	93.52	
Single-stranded DNA-binding protein	140	97.84	96.4	97.12	97.12	
phage recombination protein Bet	273	99.63	100	100	100	
hypothetical protein	48	93.62	89.36	93.62	93.62	
putative phage-type endonuclease	339	99.41	99.41	98.52	99.41	
Aerobic cobaltochelatase cobS subunit	320	99	99.69	99.69	98.75	
hypothetical protein	255	98.56	99.21	98.03	98.03	
Cobalamine biosynthesis protein	318	99.68	100	99.68	100	
hypothetical protein	147	96.58	100	97.95	95.21	
hypothetical protein	552	96.55	100	97.46	97.46	
DNA repair protein RadC	166	98.18	100	97.58	98.79	
hypothetical protein	114	95.58	100	92.92	96.46	
Phage P4 alpha	358	98.47	100	98.32	96.64	
hypothetical protein	236	93.19	100	90.21	0	
hypothetical protein	*82*	*0*	*0*	*0*	*0*	***HS3***
hypothetical protein	*234*	*0*	*100*	*0*	*0*	
hypothetical protein	*99*	*0*	*100*	*0*	*0*	
Membrane protein (xre/DNA-binding/HTH)	*137*	*0*	*100*	*0*	*0*	
integrase	*320*	*0*	*100*	*0*	*0*	
hypothetical protein	*47*	*0*	*0*	*0*	*0*	
IncF conjugative protein TraF	315	98.73	100	97.45	97.13	
IncF conjugative protein TraH	463	99.57	100	98.7	98.7	
IncF conjugative protein TraG	1190	98.74	100	98.57	98.65	
hypothetical protein	178	98.31	100	100	99.44	
hypothetical protein	100	98.99	100	98.99	98.99	
regulatory protein (LysM/invasin)	183	100	100	97.8	97.8	
hypothetical protein	221	98.88	100	98.18	98.18	
hypothetical protein	290	97.6	100	98.96	98.27	
hypothetical protein	84	101	100	100	100	
HTH-type transcriptional regulator (CI)	216	100	100	100	100	

Table 3.2 *ORFs annotated in ICEVchCIRS101 by RAST and nucleotide sequence comparison with ICEs SXT^{MO10}, R391, ICEVchB33 and ICEPdaSpa1 unique ORFs in ICEVchCIRS101 are in bold. Content of the five hotspots and resistance cluster is shown in red.*

ICE*Vch*CIRS101 resistance cluster

Like ICE*Vch*B33, SXT^{MO10} and ICE*Pda*Spa1, ICE*Vch*CIRS101 contains a cluster of resistance genes inserted at the *umuD* locus, while in R391 the *umuD* gene is intact. Insertion of the resistance cluster occurs at the same base pair as SXT^{MO10} and ICE*Vch*B33 (Fig. 3.4A). In ICE*Vch*CIRS101, the cluster is 16,445 bp in length compared to 19,284 bp for ICE*Vch*B33, with significant rearrangement (Fig. 3.4B).

The 5' end of the resistance cluster is 99% identical to ICE*Vch*B33, SXT^{MO10}, and ICE*Pda*Spa1. Like ICE*Vch*B33, ICE*Vch*Ind1, an ICE found in *V. cholerae* O1 El Tor (Hochhut *et al.*, 2001), ICE*Vch*Vie1, and ICE*Vch*Laos (Iwanaga *et al.*, 2004), ICE*Vch*CIRS101 lacks the *dfr18* dihydrofolate reductase gene that codes for resistance to trimethoprim, deoxycytidine triphosphate deaminase gene, and two other ORFs present in SXT^{MO10}. ICE*Vch*CIRS101 also lacks the efflux protein TetA, effector of resistance to tetracycline, and its repressor TetR, found in ICE*Vch*B33 and ICE*Vch*Vie1, NreB, and the cobalt-zinc-cadmium resistance protein, CzcD. The 3' end of the resistance cluster showed the same genetic arrangement as ICE*Vch*B33 and SXT^{MO10}, namely, *strB*, *strA*, *sulII*, conferring resistance to streptomycin and sulfamethoxazole, respectively, and a transposase (Table 3.2, Fig. 3.4B). In addition, ICE*Vch*CIRS101 carries a second transposase (VCH_000426) not found in any other ICE belonging to the mutator family of transposases that has 100% nucleotide sequence similarity with the same gene in *Shewanella oneidensi*s MR-1 (gb AE014299.1).

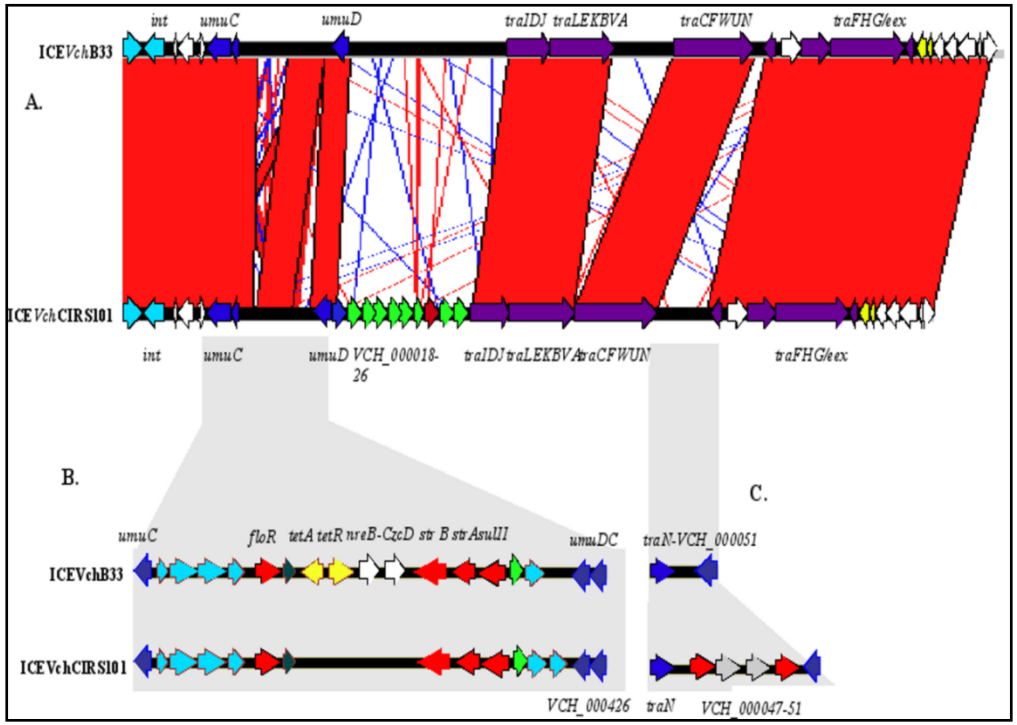


Figure 3.4*.Schematic representation of ICEVchCIRS101. (A) Alignment generated by Artemis Comparative Tool (ACT) of ICEVchCIRS10 with ICEVchB33. Red lines/areas indicate homologous regions and blue lines indicate inversions. Arrows indicate genes or clusters of genes (not scaled) and their orientation. (B) Genetic organization of the Resistance Cluster of ICEVchCIRS10 and ICEVchB33. (C) Genetic organization of the traN-VCH_000051 insertion of ICEVchCIRS10 and ICEVchB33.*

Genetic content of ICE*Vch*CIRS101 hotspots

Inserted at the *traA*- *dsbC* locus, ICE*Vch*CIRS101 has the same inserted cluster of genes as found in SXT^{MO10}, while ICE*Vch*B33 carries a larger insertion (Taviani *et al* 2009). At the 3' end of the *traN* gene (nt 64,641) ICE*Vch*CIRS101 has a 6,140 bp unique insertion comprising four genes (VCH_000047-50) (Fig 4C). Two of the ORFs are annotated as SMC-like proteins and two as transposases. The SMC (structural maintenance of chromosomes) proteins are a family of ATPases involved in higher-order chromosome organization and dynamics in prokaryotes and eukaryotes. An ABC motif is present in this protein, usually found in transporters and permeases and involved in drug resistance mechanisms. The two transposases are found between the SMC-like proteins, suggesting that the two ORFs might have been one protein interrupted by insertion of the transposase.

The 3' end of ICE*Vch*CIRS101 showed 99% nucleotide sequence similarity with ICE*Vch*B33 (Fig. 3.4A), indicating that this element has its integron inserted at the *traF* locus, like ICE*Vch*B33 and ICE*Vch*Ind1, two ICEs found in *V. cholerae* O1 El Tor.

New insertions in ICE*Vch*CIRS101 hotspot 5

Like ICE*Vch*B33, SXT^{MO10}, and ICE*Pda*Spa1, ICE*Vch*CIRS101 lacks the kanamycin resistance cluster present in hotspot 5 of R391. Interestingly, at this locus, a unique 15,000 bp sequence is inserted and is not found in any other ICE. The left flanking gene is a common protein among ICEs, which separates several members into

two groups. The one found in ICE*Vch*CIRS101 has 89% nucleotide similarity with that in R391 and ICE*Spu*PO1, while the same gene in ICE*Vch*B33 is 100% similar to SXT^{MO10} and ICE*Pda*Spa1. Interestingly, while ICE*Vch*B33, SXT, and ICE*Pda*Spa1 do not carry any insertions at this locus, R391 and ICE*Spu*PO1 have a kanamycin resistance cluster and a restriction modification cassette inserted, respectively. Thus, it appears that the two forms of this gene are associated with different insertions. The cluster found in the ICE*Vch*CIRS101 is composed of eight ORFs (VCH_000019-26), seven of which are annotated as hypothetical proteins and one as ATP-dependent protease La (VCH_000024), which is 97% similar, at the nucleotide level, to similar proteases found in *Pelodictyon phaeoclathratiforme* BU-1 (CP001110.1) and *Syntrophomonas wolfei* subsp. *wolfei* (CP000448.1). Proteins encoded by ORFs VCH_000021-26 appear to be shared by these two microorganisms in a similar arrangement. Whatever the function of this cluster, it appears to be new for ICEs and for *V. cholerae*.

Functional analysis of ICE*Vch*B33 and ICE*Vch*CIRS101

As genetic mobile elements and mediators of LGT, ICEs of the SXT/R391 family have been shown to be capable of excising from the host chromosome and to form an episomal, non-replicative intermediate required for conjugative transfer (Hochhut *et al.,* 2000; Beaber *et al.,* 2002; Osorio *et al.,* 2008; Burrus *et al.,* 2003). Using PCR, we tested the ability of ICE*Vch*B33 and ICE*Vch*CIRS101 to excise and form an extrachromosomal circular intermediate. Amplification of a circular ICE was

confirmed, suggesting ability of this element to excise from the host chromosome (Fig. 3.5). Furthermore, we determined whether ICE*Vch*B33 and ICE*Vch*CIRS101 transferred from donor strains (*V. cholerae* B33, *V. cholerae* MJ1236, and *V. cholerae* CIRS101) to an *E. coli* recipient strain. In plate mating assays, ICE*Vch*B33 transferred at frequencies of 3.8×10^5 and 2.4×10^5 exconjugant/donor, with *V. cholerae* B33 and *V. cholerae* MJ1236, respectively, as donor strains. The frequencies observed are comparable to those reported for other ICEs (Beaber *et al.*, 2002) and (Osorio *et al.*, 2008). Ten exconjugants for each transfer assay were confirmed by PCR, demonstrating the presence of int and insertion of ICE*Vch*B33 and *V. cholerae* at the *prfC* locus.

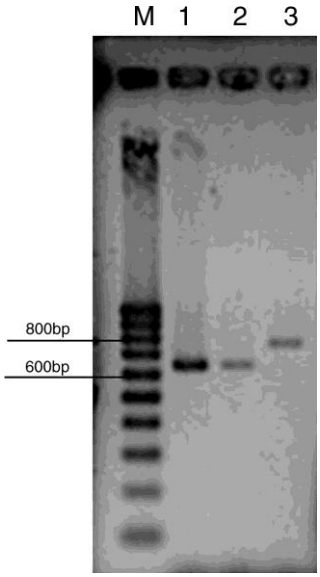


Fig. 3.5*.PCR with primer pair P4/P5 amplifying the extrachromosomal circular form of the ICE. ICEVchB33 gave an amplicon of 650 bp as expected from the nucleotide sequence in both strain harboring the element, while SXT^{M010} gave the expected amplicon of 785 bp. M: Hyperladder IV; 1: V. cholerae MJ1236; 2: V. cholerae B33; and 3: V. cholerae MO10.*

Conclusions

The three strains analyzed here possess the genomic backbone of the O1 El Tor/O139 group, phylocore genome - 1 (PG-1) and belong to a monophyletic group, the 7th pandemic clade, with other seventh pandemic *V. cholerae* strains, such as *V. cholerae* O1 El Tor N16961, RC9, and O139 MO10 (Chun *et al.,* 2009, Grim *et al.*, 2010). All three possess a hybrid or altered cholera toxin prophage associated with improved environmental fitness and possibly associated with a severe form of cholera. The 'hybrid' clones replaced the prototype seventh pandemic El Tor *V. cholerae* O1, typified by strain N16961 (Grim *et al.,* 2010). From an epidemiological viewpoint, among the three El Tor variants CIRS101 is the most successful since strains belonging to this type have replaced prototype El Tor and represent the most prevalent epidemic clone in Asia and part of Africa, notably east Africa (Grim *et al.*, 2010).

V. cholerae O1 MJ-1236 was isolated in Matlab, Bangladesh in 1994, while *V. cholerae* O1 El Tor CIRS101 was isolated in Dhaka, Bangladesh in 2002 and *V. cholerae* O1 El Tor B33 from Beira, Mozambique in 2004. Interestingly, the same ICE element, ICE*Vch*B33 was found in strains MJ1236 and B33, two *V. cholerae* El Tor O1 strains isolated from very different geographical locations and at different chronological times. The third strain, *V. cholerae* CIRS101, also bearing the hybrid CTX cluster, carries yet a different ICE (ICE*Vch*CIRS101). These results support the conclusion from the comparative genomic analysis of the three strains, namely despite geographic variation and time of isolation, *V. cholerae* O1 MJ-1236 is more similar in genomic content to *V. cholerae* O1 El Tor B33 than to *V. cholerae* O1 El Tor CIRS101 (Grim *et al*., 2010).

In addition to ICE*Vch*B33, both *V. cholerae* O1 MJ-1236 and B33 contain the same hybrid CTX cluster and have significant genomic similarity, i.e., 98% identical, suggesting significant relationship between pandemic *V. cholerae* El Tor O1 from *Bangladesh* and from Mozambique (Grim *et al.*, 2010). Furthermore, most of the differences between *V. cholerae* O1 MJ-1236 and *V. cholerae* O1 El Tor B33 versus *V. cholerae* O1 El Tor CIRS101 are presence/absence, content, and organization of several genomic islands. In addition to carrying a different ICE, *V. cholerae* O1 El Tor CIRS101 possesses a very different superintegron and lacks several of the genomic islands found in *V. cholerae* O1 MJ-1236 and O1 El Tor B33, including the kappa phage, the GI-14 genomic island, and a truncated variant of VSP-II (Grim *et al.*, 2010, Taviani *et al.*, 2010)

Overall, as described for other ICEs, ICE*Vch*B33 and ICE*Vch*CIRS101 possess a mosaic genetic structure, comprising a backbone common to other elements of the same family, most likely having evolved by acquisition of different gene clusters via serial recombination events. These clusters are shared with other elements, indicating an ability of ICE to modulate their adaptive potential.

Furthermore, ICE*Vch*B33 and ICE*Vch*CIRS101 were successfully excised from the host chromosome and transferred by conjugation, demonstrating the core conjugative apparatus to be functional.

Although the function of several of the inserted genes remains unknown, it can be concluded that they represent a mechanism of evolution that determines fitness of the genetic island within the host genome. The ICE, therefore, may well be an

important evolutionary agent involved in the emergence and epidemiological success of these seventh pandemic variants.

Based on these results, the evolutionary history of the new seven pandemic *V. cholerae* variants can be extrapolated. It appears that *V. cholerae* O1 Matlab hybrid emerged in Bangladesh and Mozambique, in the latter geographical location becoming established as the prevalent type. In contrast, the altered *V. cholerae* O1 El Tor in Bangladesh, typified by CIRS101, emerged to become the prevalent type.

Publications

Taviani E., Grim C., Chun J., Huq A., Colwell R.R. (2009). Genomic analysis of a novel integrative conjugative element in *Vibrio cholerae*. FEBS Letters, **583**(22): 3630-3636.

Grim, C. J., N. A. Hasan, E. Taviani, B. Haley, J. Chun, T. S. Brettin, D. C. Bruce, J. C. Detter, C. S. Han, O. Chertkov, J. Challacombe, A. Huq, G. B. Nair and R. R. Colwell (2010). "Genome Sequence of Hybrid *Vibrio cholerae* O1 MJ-1236, B-33, and CIRS101 and Comparative Genomics with *V. cholerae*." J. Bacteriol. **192**(13): 3524-3533.

Chapter 4. Genomic and functional analysis of an ICE identified in a clinical *V. cholerae* O37 isolate.

Introduction

Vibrio cholerae is an autochtonous inhabitant of riverine, estuarine, and marine environments and pathogenic forhumans. The current serotyping of *V. cholerae* includes 206 serogroups, among which to date only serougroups O1 and O139 have been cause of epidemic or pandemic cholera. Strains belonging to serogroups other than O1 and O139 are commonly referred to as non-O1, non-O139 *V. cholerae;* they are the more abundant in the aquatic environment (Sharma *et al.*, 2007). These organisms do not have epidemic or pandemic potential, as *V. cholerae* serogroups O1 and O139, but they have often been associated with sporadic cases of cholera and localized outbreaks of cholera-like diarrhea (Mukhopadhyay *et al.*, 1995; Sharma *et al.*, 2007). Several outbreaks caused by *V. cholerae* non-O1 non-O139 have been reported, including an outbreak caused by *V. cholerae* serogroups O10 and O12 in 1994 in Lima, Peru (Dalsgaard *et al*., 1995), O10 in Delhi, India (Rudra *et al*., 1996), O37 in Sudan and Czechoslovakia (Aldova *et al* 1968), non-O1 non-O139 in Kolkata (Sharma *et al*., 1998) and Thailand (Bagchi *et al*., 1993), and serogroup O141 in many other regions (Chatterjee *et al.,* 2009). Beside the potential to cause cholera-like disease, these organisms have also been associated with cases of extra-intestinal infections.

The pathogenic mechanisms and factors involved in the virulence of non-*V. cholerae* O1, non-O139 remain unclear. Among *V. cholerae* non-O1, non-O139 strains, some serogroups (e.g., O10, O11, O12, and O144) appear to be more frequently associated with cholera-like disease, despite the absence of virulence gene clusters (*ctxA, ctxB, zot, ace*, and other genes), suggesting that these serogroups have mechanisms of pathogenesis different from those of toxigenic *V. cholerae* (Lukinmaa *et al.*, 2006). Several virulence factors are associated with *V. cholerae* non-O1, non-O139, such as RTX, hemagglutinins, hemolysins, including the El Tor hemolysin (Saha *et al.*, 1996; Ichinose *et al.,* 1987), cytotoxins (Saha and Nair, 1996), mannose-sensitive hemagglutinin (Dalsgaard *et al.,* 1995), the *V. cholerae* non-O1 heat-stable enterotoxin (NAG-ST) (Iyer *et al*., 2000; Restrepo *et al.,* 2006) and a type three secretion system (TTSS) correlated with increased hemolytic titers and increased motility (Dziejman *et al*., 2005). All these factors may play some role in the pathogenicity of these organisms (Chatterjee *et al*., 2009). It has also been suggested that the pathogenic mechanism of *V. cholerae* non-O1, non-O139 in a way resembles that of enteropathogenic *E. coli* (Sharma *et al.*, 2007).

The strain of *V. cholerae* under study here belongs to serogroup O37 and was isolated from a cholera patient in Bangladesh. From the whole genome sequence it was found that *V. cholerae* O37 MZO-3 does not carry the CTX and TCP virulence clusters and both the *Vibrio* pathogenicity islands, VPI-1 and 2, are absent. Furthermore, it carries a variant of the *Vibrio* seventh pandemic 2 (VSP-2) and virulence genes such as RTX, HylA and the MSHA (Taviani *et al.,* 2009). Recently, a case appearing in the Journal of Clinical Microbiology described a fatal *V. cholerae* O37 enteritis in a patient

in Italy who suffered a cholera-like syndrome after consumption of contaminated food in Egypt (Farina *et al.,* 2010). The *V. cholarae* O37 was isolated from feces in pure culture and tested negative for the major viulence factors. However, it showed hemolytic and proteolytic activities and was able to colonize the intestine of suckling mice (Farina *et al.,* 2010). We found *V. cholerae* O37 MZO-3 of the study reported here carried an unusual genetic element showing high similarity to the SXT-like integrative conjugative element (ICE) of *V. cholerae* and is associated with putative virulence determinants and pathogenic cellular apparati.

ICEs are a class of self-transmissible mobile elements found in the genomes of prokaryotes, serving as mediators of horizontal gene transfer (Burrus and Waldor, 2004; Burrus *et al*., 2006). They are a genetic mosaic of plasmids, phages, and transposon-like features and are able to self transfer from a donor to a recipient cell and integrate into the host chromosome (Hochhut and. Waldor, 1999).

In *Vibrio cholerae,* these elements were described after appearance of *V. cholerae* O139 in 1992, with the SXT element. ICEs have since been found in a majority of *V. cholerae* O1 and O139 isolates from Asia and in *V. cholerae* non-O1 non-O139 and several related *Vibrio* spp. (Iwanaga *et al.,* 2004; Ahmed *et al.,* 2005; Burrus *et al.,* 2006; Taviani *et al.,* 2008). All ICEs described in this species belong to the same family, the SXT/R391 (Burrus *et al*., 2006) and share a conserved genetic scaffold that incorporates unique sequences (Hochhut *et al.,* 2000; Beaber *et al.,* 2002; Boltner *et al.*, 2002).

Previous studies showed how some characteristics of the pathogenicity islands of *Legionella pneumophila, Agrobacterium tumefaciens,* and *Bartonella tribocorum,*

with encoded type IV secretion systems and the presence of tyrosine-recombinases similar to Int_{SXT}, might be considered ICEs, or defective ICEs (Burrus *et al.*, 2004). This hypothesis is further supported by the report of a new element in a strain of *Escherichia coli*, ICE*Ec*1, which carries the high pathogenicity island HPI encoding the siderophore system "yersiniabactin", previously described in *Yersinia* (Shubert *et al.*, 2004). Furthermore, discovery of ICE*Kp*1 in a *Klebsiella pneumoniae* strain suggests it is a putative ICE containing the HPI (Lin *et al.*, 2008).

In this study, we undertook a genomic and functional analysis of the ICEs carried by *V. cholerae* O37 MZO-3 and showed features of the element to be different from any ICE described to date. Its correlation with pathogenicity islands may be a further step toward understanding virulence of the non-epidemic strains of *V. cholerae*.

Materials and Methods

The experimental method for culture and DNA extraction, genome sequencing, ICE comparative genomics, and genetic organization analysis and annotation were the same as described for ICE*Vch*CIRS101 and ICE*Vch*B33 (See Chapter 3).

PCR assays

To confirm accurate assembly of contigs containing ICE-like sequences and to resolve the gap between those contigs, a set of specific primers were designed and used in different combinations. Primers pMZO_1F (caagaacaagtgcgtccaca) and pMZO3_1R (ctttgcataccacaggttgac) were designed on contig 1 to determine correct assembly of the

contig. The expected amplicon size was 598 bp. Primers pMZO3_1F2 (tgctaaacaagcaggagtcagt) and pMZO3_54R (gaacactccgctgcgatgag) were designed to study the junction between contigs 1 and 54. A variable amplicon >500 bp confirmed the position of these two contigs. Two primers were designed to confirm the correct location of the 3'-end of the Hot Spot 4 insertion: EnterotoxinBF (ccaagaacagcgtttacatcca) and s063R (acaccgacccaattaccttca) which produced a 3431 bp amplicon. Moreover, primers AcriF (aatcctaatcctgccgcct) and AcriR (acataaacctctcgaatcagcca) with a 1469 bp amplicon confirmed presence of the acriflavine resistance gene. Similarly, to study and confirm insertion of the element at the tRNA-Ser, two primer pairs were designed, namely at the 5' end, primers pMZO3_5F (caaaatcgcacggtagtgaa) and pMZO3_5R (cagttggtcattgacggtga) (986 bp amplicon) and at the 3' end primers pMZO3_3F (ccacgagcacctccttagtc) and pMZO3_3R (ctccgcaaggagccttatta) (598 bp amplicon). Sequencing of selected amplicons was done at the DNA sequencing facility at the University of Maryland College Park, employing the Applied Biosystems DNA sequencer model 3730.

Acriflavine resistance assay

Since an ORF was discovered with high similarity to the acriflavine resistance gene found in other microorganisms, we tested *V. cholerae* O37 MZO-3 for ability to grow in a medium containing different concentrations of this acridine derivative.

Cultures of *V. cholerae* O37 MZO-3 and *V. cholerae* O1 N16961 grown in Luria Broth (LB) and on LB agar containing acriflavine at concentrations from 1.0

ng/ml to 10.0 ug/ml were tested in parallel in duplicate. Cultures were incubated at 30 C and 37 C for 3 days and growth was recorded daily. No difference was observed in growth of *V. cholerae* O37 MZO-3, compared to control, that is, the strain tested had the same ability to grow at concentrations up to 10.0 ng/ml of acriflavine as the control, both in liquid and agar media.

Results and discussion

Location of ICE in the genome of *V. cholerae* MZO-3

By comparative analysis of the *V. cholerae* O37 MZO-3 genome, sequences belonging to the genetic backbone of ICE were found on contigs 1 and 54. To confirm the presence of the genetic element, a Blastn search was done using the SXT^{M010} integrase nucleotide as query sequence. The result was no hits. Furthermore, analysis of the insertion site expected for ICEs of the SXT/R391 family at the 3'end of the *prfC* gene, showed, instead, an intact locus on contig 41 (Fig 4.1). This was surprising since all ICEs described to date in *V. cholerae* are part of the SXT/R391 family, whose distinguishing characteristics are the presence of a conserved integrase (int_{SXT}) and insertion at the *prfC* locus. A more detailed analysis of the insertion site and integrase gene of this genetic element is provided below.

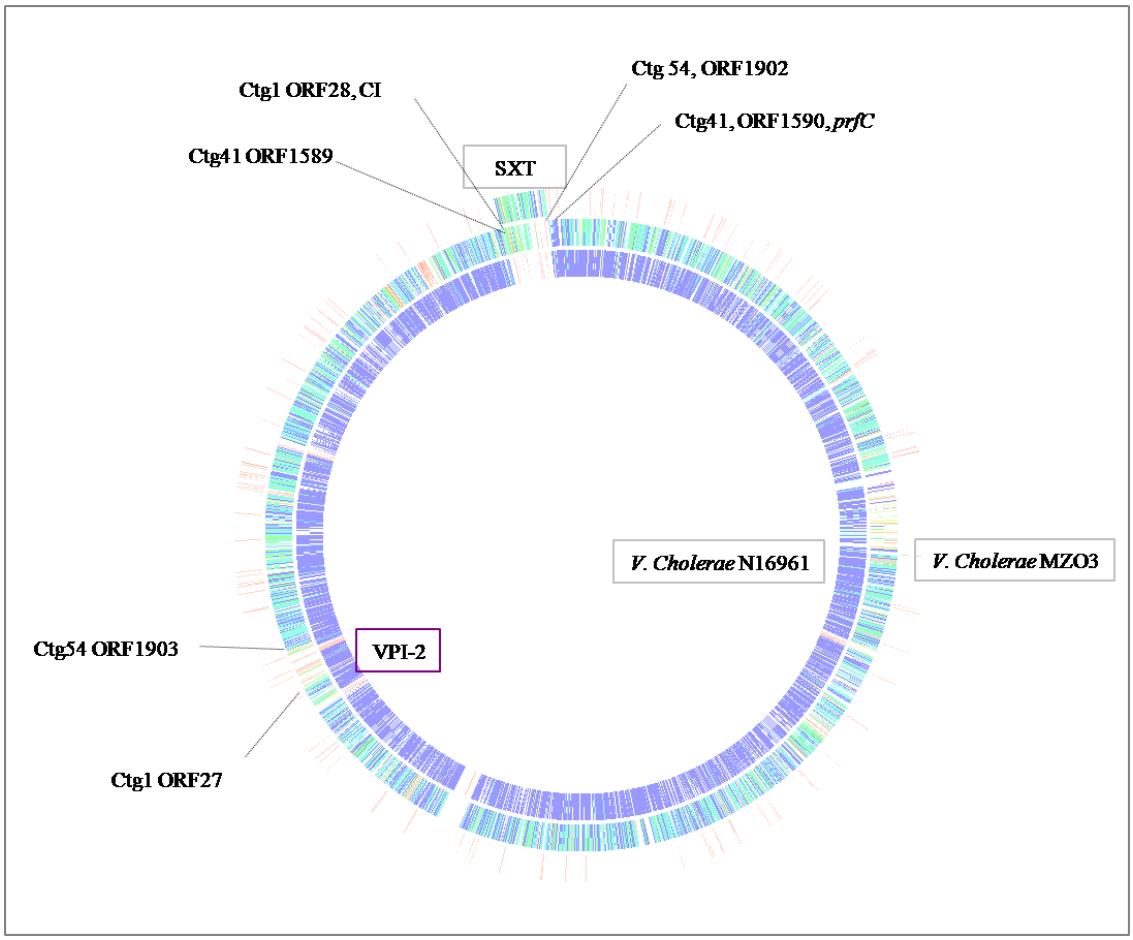


Fig 4.1. *RAST genome alignment. SXT element, V. cholerae O37 MZO-3, V. cholerae O1 El Tor N16961 were aligned to the genome of V. cholerae O1 MJ1236. The figure shows position of first gene (ORF28, CI, Ctg1) and last gene (ORF1902, Ctg 54) of ICE-like element in V. cholerae O37 MZO-3 and position of flanking genes (ORF27, Ctg1 and ORF 1903, Ctg 54) which are found at the borders of the VPI-2 insertion site.*

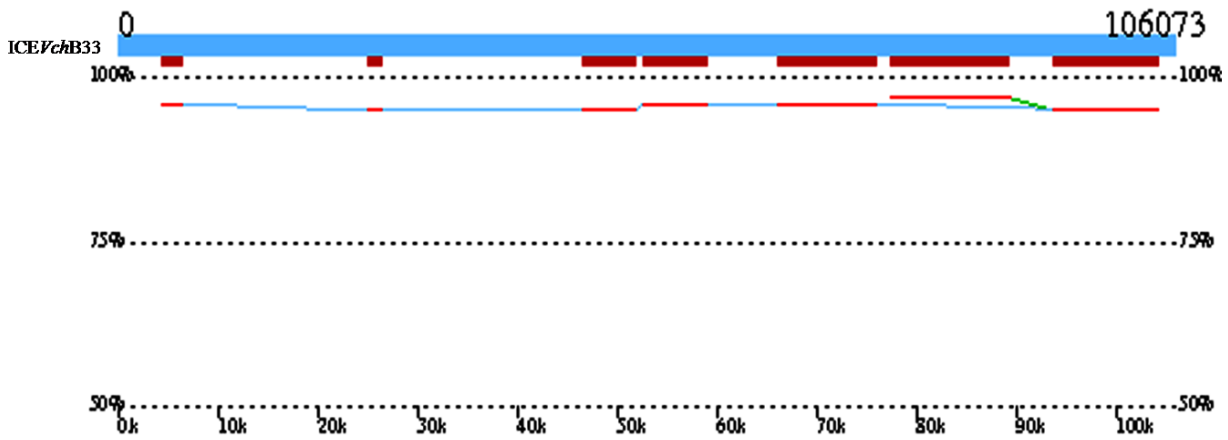
Assembly and annotation of the ICE

SXT-like nucleotide sequences were detected on two contigs of *V. cholerae* MZO-3: nt 27439–105212 of contig 1 (accession no. NZ_AAUU01000001.1); and nt 1–25409 of the contig 54 (accession no. NZ_AAUU01000001.54). Assembly of the SXT-related sequences in *V. cholerae* MZO-3 was done using NucMer with nucleotide sequence of ICEV*ch*B33 (Fig 4.2) and SXT^{M010} (Fig 4.3) as reference. Alignment by NucMer revealed one gap at the junction between the two contigs that was resolved by PCR and confirmed by sequencing. Surprisingly, only 196bp were missing to link the two contigs.

The assembled 103381bp ICE sequence was submitted to RAST pipeline for annotation and compared to SXT and other ICEs found in *V. cholerae* using the RAST sequence comparative tool (Table 4.1) (http://rast.nmpdr.org)**.**

Nomenclature

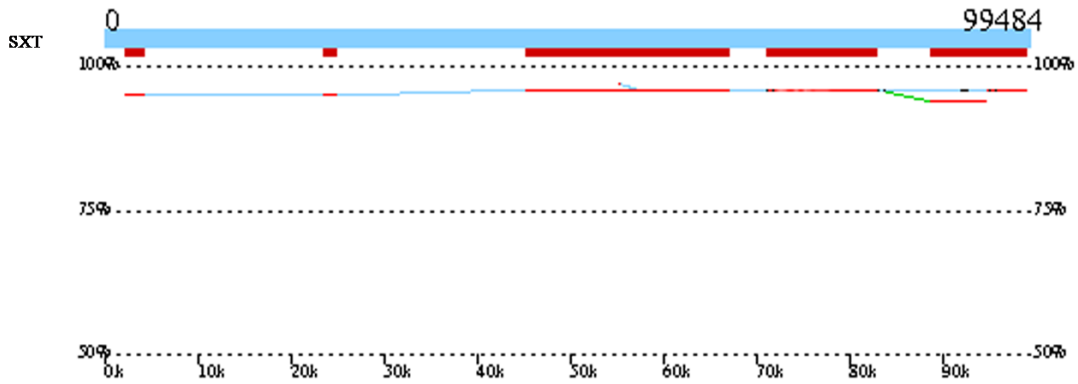
Since the ICE-like element found in *V. cholerae* O37 MZO-3 was unique and not part of the SXT/R391 family of ICEs, as previously discussed, it raised an issue of nomenclature. However, this genetic element was included in a *V. cholerae* ICE comparative genomics analysis in a previous work, and named ICE*Vch*Ban8, following the proposed nomenclature for SXT/R391 ICEs (Wozniak *et al*., 2009).



/home/elisa/Desktop/MUMmer3.21/ICEVchB33.fasta /home/elisa/Desktop/MUMmer3.21/ICELike_MZO3.fasta NUCMER

[S1]	[E1]	[S2]	[E2]	[LEN 1]	[LEN 2]	[% IDY]	[LEN R]	[LEN Q]	[COV R]	[COV Q]	[FRM]	[TAGS]	
4282	6505	32222	34445	2224	2224	96.22	106074	105212	2.10	2.11	1 1	ICE-B33	gn|VCMZO3|ctg|1|ref|
25023	26364	34642	35983	1342	1342	95.98	106074	105212	1.27	1.28	1 1	ICE-B33	gn|VCMZO3|ctg|1|ref|
46525	51788	37155	42418	5264	5264	95.16	106074	105212	4.96	5.00	1 1	ICE-B33	gn|VCMZO3|ctg|1|ref|
52563	59048	43095	49595	6486	6501	96.80	106074	105212	6.11	6.18	1 1	ICE-B33	gn|VCMZO3|ctg|1|ref|
66164	75987	50769	60592	9824	9824	96.46	106074	105212	9.26	9.34	1 1	ICE-B33	gn|VCMZO3|ctg|1|ref|
77439	89227	4851	16625	11789	11775	97.00	106074	57556	11.11	20.46	1 1	ICE-B33	gn|VCMZO3|ctg|54|ref|
93751	99791	16800	22839	6041	6040	95.45	106074	57556	5.70	10.49	1 1	ICE-B33	gn|VCMZO3|ctg|54|ref|
99816	104223	32021	27617	4408	4405	95.92	106074	105212	4.16	4.19	1 -1	ICE-B33	gn|VCMZO3|ctg|1|ref|

Figure 4.2. *NucMer alignment of V.cholerae O37 MZO3 contig 1 and 54 to ICEVchB33. Positive matches are shown in red. Output file of the NucMer alignment is shown. Text in black and red indicate position of positive matches for contigs 1 and 54, respectively.*



/home/elisa/Desktop/MUMmer3.21/SXT.txt /home/elisa/Desktop/MUMmer3.21/ICELike_MZO3.fasta NUCMER

[S1]	[E1]	[S2]	[E2]	[LEN 1]	[LEN 2]	[% IDY]	[LEN R]	[LEN Q]	[COV R]	[COV Q]	[FRM]	[TAGS]	
2098	4318	32225	34445	2221	2221	95.90	99485	105212	2.23	2.11	1 1	SXT.99485	gn|VCMZO3|ctg|1|ref|
23548	24889	34642	35983	1342	1342	95.98	99485	105212	1.35	1.28	1 1	SXT.99485	gn|VCMZO3|ctg|1|ref|
45141	57326	37155	49338	12186	12184	96.37	99485	105212	12.25	11.58	1 1	SXT.99485	gn|VCMZO3|ctg|1|ref|
55202	55299	49408	49505	98	98	97.96	99485	105212	0.10	0.09	1 1	SXT.99485	gn|VCMZO3|ctg|1|ref|
57295	67099	50786	60592	9805	9807	96.22	99485	105212	9.86	9.32	1 1	SXT.99485	gn|VCMZO3|ctg|1|ref|
71070	82867	4851	16630	11798	11780	96.70	99485	57556	11.86	20.47	1 1	SXT.99485	gn|VCMZO3|ctg|54|ref|
88684	94724	16800	22839	6041	6040	94.64	99485	57556	6.07	10.49	1 1	SXT.99485	gn|VCMZO3|ctg|54|ref|
94892	99155	31878	27617	4264	4262	96.20	99485	105212	4.29	4.05	1 -1	SXT.99485	gn|VCMZO3|ctg|1|ref|

Figure 4.3. *NucMer alignment of V.cholerae O37 MZO3 contig 1 and 54 to* SXT_{MO10}*. Positive matches are shown in red. Output file of NucMer alignment is also shown. Text in black and red indicate position of positive matches for contigs 1 and 54, respectively.*

Genomic organization of ICE*Vch*Ban8

Annotation resulted in identification of 99 ORFs in the element (Table 4.1). Comparative analysis of ICE*Vch*Ban8 and SXT, using the Artemis Comparative Tool (ACT) (http://www.sanger.ac.uk/Software/ACT), was done. Although it could not be assigned to the SXT/R391 family, the overall genetic organization of ICE*Vch*Ban8 was highly similar to the SXT/R391 ICEs, namely SXT^{M010} and ICE*Vch*B33 (Fig. 4.4).

ICE*Vch*Ban8, as all other ICEs, possesses an highly conserved core of genes responsible for transfer, integration and excision, and control (Table 4.1, Fig. 4.4). As noted for other ICEs, specific inserted genes were identified in five hot spots and two variable regions within the core backbone. Of the 99 ORFs identified by RAST, 47 were present in SXT^{M010} and ICE*Vch*B33 (Table 4.1).

Table 4.1.*ORFs annotated in ICEVchBan8 by RAST and nucleotide sequence comparison with SXT, and ICEVchB33. Content of the four hotspots is shown in red. First 4 ORFs are highlighted to indicate their unusual location at the 5'end of the ICE.*

function	***% identity***				
ICEVchBan8	***SXT***	***ICEVchB33***			
cI prophage repressor	95.35	95.35			
hypothetical protein	96	98.27			
hypothetical protein	100	98.51			
regulatory proteins (LysM/invasin)	98.35	98.35			
hypothetical protein	100	98.99			
hypothetical protein	98.06	98.06			
hypothetical protein	85.96	84.21			
Rod shape determination protein	99.07	99.07			
hypothetical protein	97.28	96.6			
UmuC	97.96	97.96			
UmuD	95.97	95.97			

hypothetical protein	96.6	96.6			
Transposase	0	0			
TraI	92.04	92.04			
TraD	99.01	99.5			
hypothetical protein	94.87	97.85			
s043	99.05	98.58			
hypothetical protein	100	0	HS1		
hypothetical protein	99.07	0			
TraL	98.92	98.92			
TraE	99.04	99.04			
TraK	97.32	95.97			
TraB	98.6	97.9			
TraV	98.61	99.07			
TraA	99.22	99.22			
hypothetical protein	99.64	98.92			
Ync	99.36	97.3			
hypothetical protein	0	0	HS2		
hypothetical protein	0	0			
Thiol:disulfide interchange protein	99.57	99.57			
TraC	99.5	99.37			
hypothetical protein	97.17	96.23			
TrhF	96.47	96.47			
TraW	96.52	97.06			
TraU	99.42	98.54			
TraN	96.89	97.6			
TPR repeat	0	0	HS4	1152	
Outer membrane lipoprotein A precursor	0	0		846	
hypothetical protein	0	0		90	OmpB/ABC transporter
Putative substrate-binding transport protein	0	0		168	OmpA-like
Transposase	0	0		186	
transposase and inactivated derivative	0	0		654	
Accessory colonization factor AcfD	0	0		4389	Acf
hypothetical protein	0	0		555	
Transposase	0	0		921	
hypothetical protein	0	0		108	
Accessory colonization factor AcfA	0	0		630	Acf
hypothetical protein	0	0		171	
transthyretin family protein	0	0		384	transthyretin
putative extracellular protein	0	0		513	Hemoglobin-domain
Bacterial surface protein	0	0		2094	Ig-like fold
hypothetical protein	0	0		786	hydrolase fold
hypothetical protein	0	0		204	
TcpA	0	0		732	Type IV pilus
LngB	0	0		1539	Type IV pilus
CofC	0	0		372	Type IV pilus
TcpC	0	0		1461	Type IV pilus

hypothetical protein	0	0		540	Type IV pilus
CofF	0	0		834	Type IV pilus
CofG	0	0		534	Type IV pilus
TcpT	0	0		1533	Type IV pilus
T2 secretory pathway, PulF	0	0		972	Type IV pilus
prepilin peptidase, type IV	0	0		765	Type IV pilus
TcpP	0	0		678	Type IV pilus
ToxT	0	0		798	Type IV pilus
hypothetical protein	0	0		720	s002-rumB HP
Accessory colonization factor AcfC	0	0		750	Acf
hypothetical protein	0	0		1020	
StcE	0	0		2922	Protease
methyl-accepting chemotaxis sensory transducer	0	0		405	
Protein ygiW precursor	0	0		228	
T3 secretion and flagellar regulator RtsA	0	0		804	
putative regulatory protein	0	0		762	ToxR
AcrB/AcrD/AcrF family protein	0	0		3066	Acriflavine resitance
hypothetical protein	0	0		1086	
transcriptional regulator	0	0		117	
CtxA	0	0		1806	CtxA-LTA
Sialidase (EC 3.2.1.18)	0	0		111	
Transposase	0	0		375	
Transposase	0	0		141	
hypothetical protein	0	0		627	
hypothetical protein	91.88	89.8			
hypothetical protein	96.23	96.23			
Single-stranded DNA-binding protein	94.96	97.12			
Recombination protein BET	98.9	99.26			
hypothetical protein	100	95.74			
DNA recombination protein	99.11	99.11			
ATPase	100	99.51			
hypothetical protein	98.56	97.65			
Cobalamine biosynthesis protein	99.68	100			
hypothetical protein	95.21	93.15			
hypothetical protein	97.46	95.46			
RadC	99.39	98.79			
Phage P4 alpha, zinc-binding domain protein	96.32	96.36			
hypothetical protein	82.76	85.34			
TraF	94.27	94.59			
TraH	98.7	98.27			
TraG	97.81	98.23			
Phage integrase	25.25	25.25			

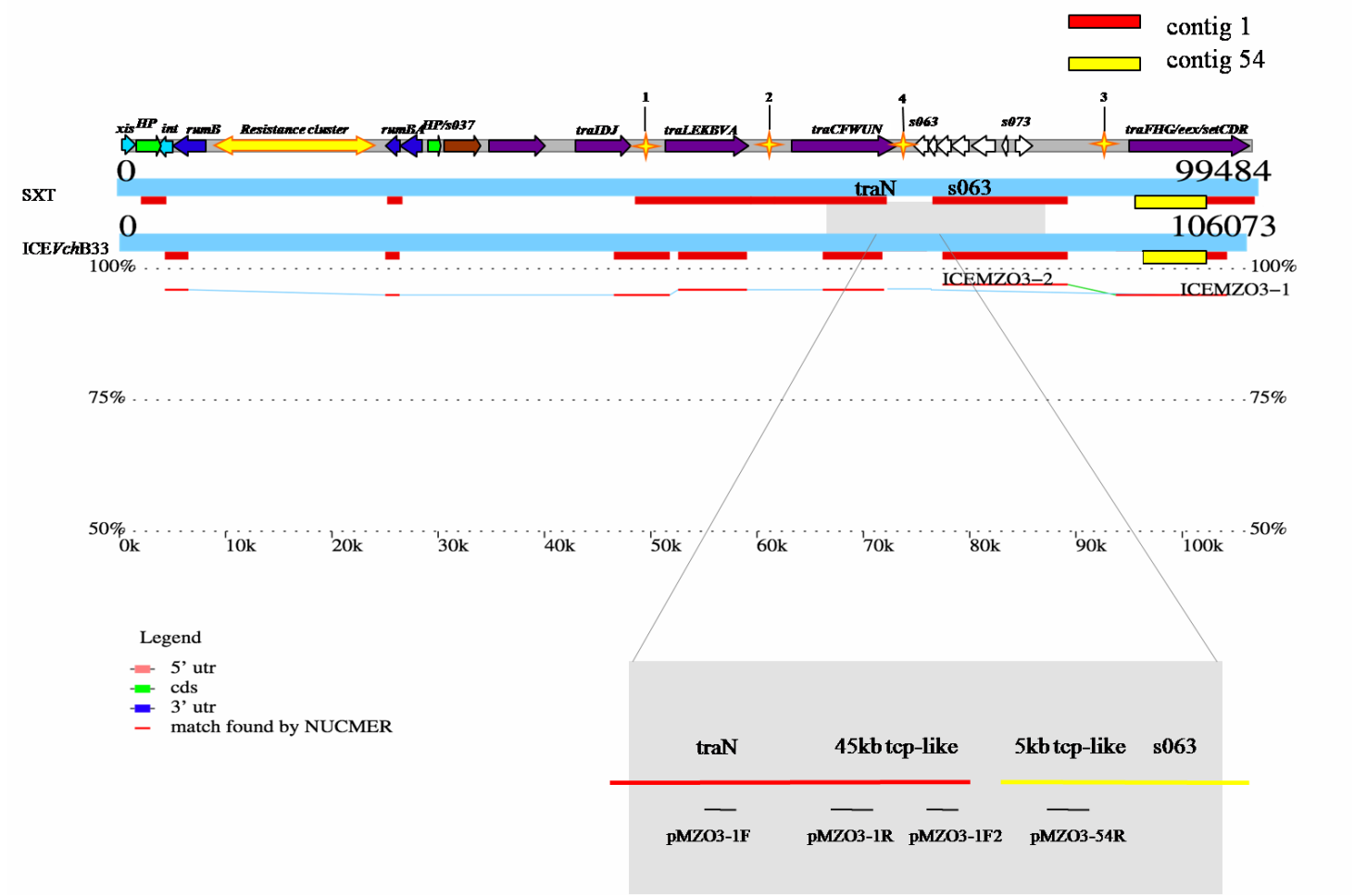


Figure 4.4. *NucMer alignment of V. cholerae O37 MZO3 contig 1 and 54 to SXT and ICEVchB33. Positive matches for contigs 1 and 54 are shown in red and yellow, respectively. Hotspot 4 is shown in detail and primers pairs used to resolve the gap between contigs are indicated.*

Integration machinery

Insertion site

ICEs found in *V. cholerae* to date belong to the same family of mobile elements (the SXT/R391 family) and all integrate at the same locus in the host genome: the 3' end of the *prfC* (Burrus and Waldor, 2006). In contrast, ICE*Vch*Ban8 is inserted at the t-RNA-Ser locus in the genome of *V. cholerae* MZO-3, leaving the *prfC* locus intact (Fig 4.1). The t-RNA-Ser is, instead, the insertion site of another important pathogenicity island in pandemic and epidemic strains of *V. cholerae*, the *Vibrio* Pathogenicity Island 2 (VPI-2) (Jermin and Boyd 2002; Murphy and Boyd 2008). VPI-2 is a 57.3 kb PAI encodings 52 ORFs (VC1758 to VC1809 in *V. cholerae* N16961) and includes several gene clusters, such as a type 1 restriction modification (RM) system (*hsdR* and *hsdM*), a sialic acid metabolism gene cluster (VC1773–VC1783), and genes required for the utilization of aminosugars, including the *nanH* neuramidase (VC1784), a gene cluster that shows homology to Mu phage. The neuramidase protein is involved in *V. cholerae* pathogenicity in two ways: on the one hand 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) and on the other hand, it appears to be part of the mucinase complex responsible for digestion of intestinal mucus, acting to enhance *V. cholerae* colonization of the gut (Murphy and Boyd, 2008). When it was discovered, VPI-2 showed all of the characteristics of an horizontally transferred element**.** The genomic island encodes a phage-like integrase (VC1758), responsible for insertion/excision of the island to and from the host genome

(Murphy and Boyd, 2008). It has been shown that VPI-2 forms a circular intermediate (CI) upon excision, which very likely represents the mobile form of this genetic element. Mobilization of VPI-2 has not yet been demonstrated nor the presence of a gene cluster encoding a conjugative apparatus (*tra* gene operon), or type IV secretion system, suggesting the island is mobilized by other genetic elements or is an ancient mobile element which has lost part of its genetic structure. To date, all toxigenic *V. cholerae* O1 serogroup isolates contain VPI-2, whereas non-toxigenic isolates lack the island (Murphy and Boyd, 2008).

By genomic analysis of 23 *V. cholerae* and *V. mimicus* genomes, VPI-2 was found to have significant polymorphism in its genetic structures. Boyd and colleagues found five variants of the VPI-2 present in *V. cholerae* non-O1. Despite its variable genetic content, which include a type three secretion system (T3SS), the integrative module, comprising the integrase gene, the *xis* gene for recombination directionality factor, and the *att* site, is identical in all of the variants (Jermin and Boyd, 2005; B. Haley, personal communication). According to these findings, the annotated genome sequence of *V. cholerae* O37 MZO-3 showed it did not possess the VPI-2 island (Chun *et al.*, 2009). Interestingly, it did carry a different insertion at the *t-RNA-Ser* locus, that is, the ICE*Vch*Ban8 element (Fig. 4.1).

The *attB* attachment site of ICE*Vch*Ban8 was identical to that found in VPI-2 when inserted at this locus and, upon integration, two 23 bp inverted sequences were restored flanking the element. These are nearly identical to *attL* and *attR* flanking VPI-2, with the only difference being in *attL* in *V. cholerae* MZO-3, differing at the 23rd base pair, with G instead of A. 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.

The integrase gene

Together with the insertion site, the integrase gene (int_{SXT}) is the main distinguishing feature for ICE elements of the SXT family (Burrus and Waldor, 2006). Above the cutoff of 95% similarity at the nucleotide level, the mobile island is considered related to its cognate ICEs.

The integrase gene found in ICE*Vch*Ban8 is not part of the int_{SXT} family, having high variance at the nucleotide sequence level. In ICE*Vch*Ban8, the *int* gene is 1233bp long. Its nucleotide sequence showed, by low stringency Blastn search, low similarity only at the 3' end of the gene. About 37% of its sequence, 326 nucleotides out of 1233, matched with 69% similarity to the integrase of *Vibrio parahaemolyticus* RIMD 2210633, 64% with the *V. cholerae* VPI-2 integrase and 63% with different integrases found in *Shewanella spp*. In *Vibrio parahaemolyticus* RIMD 2210633 the integrase is found on the VPaI-3 island, a region unique to the highly virulent *V. parahaemolyticus* O3:K6 pandemic clone.

At the amino acid level, Blast-PSI showed 75% (303/401) amino acid similarity of the protein to the phage integrase family protein of *Shewanella sp*. MR-7 and 74% (298/399) to the integrase phage family of *V. cholerae* VPI-2.

The 3'-end of the integrase contained a tyrosine recombinase domain common to all site-specific recombinases, enzymes involved in DNA breaking-rejoining, with a C-terminal catalytic domain. Members of this diverse protein superfamily include enzymes like the human topoisomerase I, the bacteriophage lambda integrase, the bacteriophage P1 Cre recombinase, the yeast Flp recombinase and the bacterial XerD/C recombinases which include the integrase found in ICEs.

Notably, there was no significant similarity at either the nucleotide or amino acid level with any of the ICE integrases, showing that this is a new genetic element not previously described in *V. cholerae*, and having high variation in genetic composition, suggesting a lenghty independent evolution.

In a recent study, Boyd and colleagues reported results of a phylogenetic analysis of integrase enzymes of several bacteria groups (Boyd *et al.,* 2009). Following this particular genetic marker, they wished to infer diverse evolutionary paths taken by genetic mobile elements and suggest genomic islands (GI) are genetic entities with a different history relative to other integrative and mobile elements (Boyd *et al.,* 2009). Their findings showed that, except for a few exceptions, GI integrases clustered together and separated from phage, plasmid, integron, and ICE integrases. As confirmed by our results, the ICE*Vch*Ban8 *int* included in their study, although not annotated in an ICE, did not clustered within the SXT-like *int*, but in the GI group, though on a branch separate from all *V. cholerae* VPI-2 integrases (Boyd *et al.,* 2009). According to our analysis, the element inserted in t-RNA-Ser is an ICE with a GI integrase, suggesting that recombination generates 'hybrid' islands, thus compromising

boundaries and diminishing evolutionary distances between different classes of mobile elements.

An interesting observation is that the int_{VPI2} is associated with the *t-RNA-Ser* in different genetic elements and in different species (Fig. 4.5). In our case this structure suggests two possible scenarios: int_{VPI2} is found on the element itself before integration at the *t-RNA-Ser* locus or insertion of the ICE occurs in tandem with a genetic element carrying the int_{VPI2} successively excised. From preliminary analysis of the insertion site and integrase gene, we conclude that ICE*Vch*Ban8 of *V. cholerae* O37 MZO3 is a new genetic element, with a mixture of different features of genomic islands and ICEs. According to these results, it can be inferred that mobile elements previously thought to be separated by evolution, may have followed the same path in genome selection and recombination occurred also between elements currently ascribed to different families.

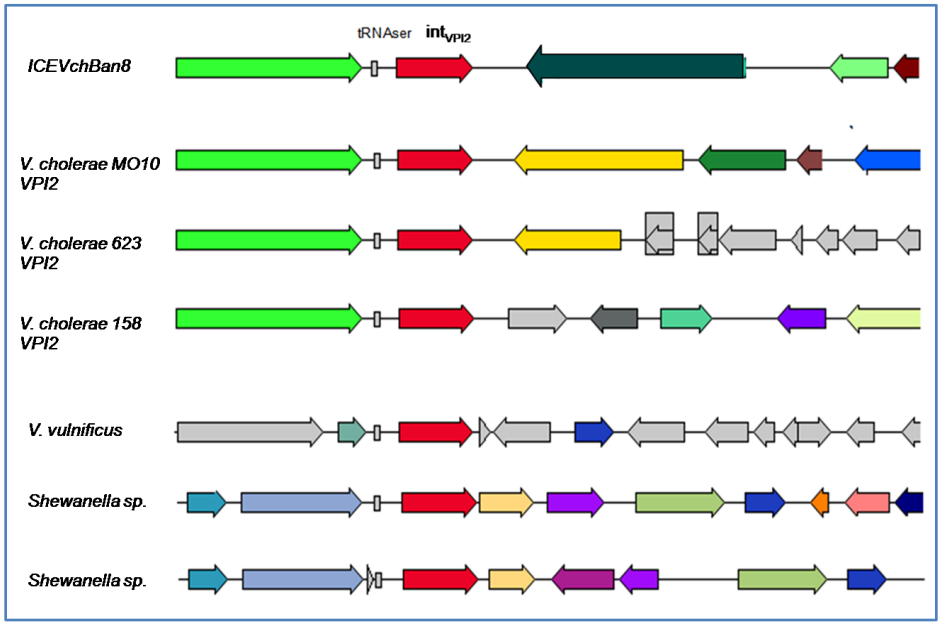


***Fig 4.5**. RAST alignement of the t-RNASer locus in V. cholerae and other related species. The genetic association of t-RNASer with the int$_{VPI2}$ is shown for different strains carrying different inserted elements.*

Xis gene

No recombination directionality factor was identified in the genome of ICE*Vch*Ban8.

Conjugative machinary: *tra* gene clusters

Genes encoding enzymatic functions involved in ICE transfer (*tra* genes) are grouped in four clusters or operons transcriptionally linked. The *tra* genes encode

proteins required for DNA processing, mating pair formation, and conjugation. The conjugative apparatus encoded by the SXT family of ICE is structurally and genetically related to the F- factor and is considered a variant of the type IV secretion system family (T4SS) (Lawley *et al.*, 2003).

ICE*Vch*Ban8 possesses all four clusters of genes encoding the conjugative apparatus that mobilizes the element. These are represented by the *traIDJ*, *traLEKBA*, *traCFWUN* and *traFHG* operons. All of the genes have at least 97% similarity at the nucleotide level with the *tra* clusters found in the conserved backbone of ICE elements, including the SXT/R391 family of ICE found in *V. cholerae* (Wozniak *et al.*, 2009). From the molecular aspect, ICE*Vch*Ban8 appears to be capable of transfer by conjugation.

Regulatory cluster

Regulation of integration/excision and transfer of the circular intermediate of the element resembles the pathway that regulates the lytic cycle of lambdoid phages (Beaber *et al.*, 2004). ICE regulation is governed by three ICE encoded genes, *setC*, *setD* and the repressor *setR*. It is thought that the regulatory cascade is triggered by events that activate the host SOS response, including damage to the DNA, by cleavage and inactivation of the repressor SetR. This protein is a phage lambda cI-like repressor and its inactivation permits expression of SetC and SetD proteins, which are transcriptional activators of the *int* and *tra* genes (Beaber *et al.*, 2004).

In ICE*Vch*Ban8, the regulatory cluster is present along with the *setC*, *setD*, and *setR* genes, but it is unusual in its location at the 5'-end of the element, which is different from all other SXT/R391 elements described to date where the regulatory cluster is found in the last genetic region at the 3'-end of the element.

In particular, in ICE*Vch*Ban8, the *setR* gene is the first ORF of the element, located downstream to the *attL* insertion site, at position 174-821 (Fig 4.4). *SetC* and *setD* genes are located in the same cluster.

The genetic structure observed in ICE*Vch*Ban8, featuring a large region usually present at the 3'-end of ICEs of the SXT/R391 family at the opposite end of the element and absence of int_{SXT}, together suggests this element has undergone several genetic rearrangements. Alternatively, ICE*Vch*Ban8 may be considered an ancient form of mobile element that originated the SXT-like ICE commonly described in *V. cholerae*. Again, as previously mentioned regarding *int*, since we have a partial picture of the variability of mobile genetic elements, it is not possible to prepare an accurate evolutive history for this ICE.

Genetic content of ICE*Vch*Ban8 insertions and Hot Spots

As in other ICEs, the ICE*Vch*Ban8 specific genes are inserted into five hot spots (HS) and two variable regions that confer specificity to the ICE. The resistance cluster inserted at the *umuCD* locus was not present in this ICE (Table 4.1, Fig 4.6). Between genes *s043* and *traL* (HS1), ICE*Vch*Ban8 has two hypothetical proteins inserted, the same as found in SXT^{M010}. ICE*Vch*B33, R391, ICE*Pda*Spa1, and several

other ICEs, such as ICE*Spu*P01, ICE*Vch*Ind1, and ICE*Vch*Lao1 have a different insertion which contains *orf37* and *orf38* (Hochhut *et al*., 2001; Iwanaga *et al*., 2004, Wozniak *et al.*, 2009) (Table 4.1). Hotspot 1 and Hotspot 2 have the same molecular rearrangement as in SXT, with an additional gene coding for an hypothetical protein in the latter (Table 4.1, Fig 4.6). Hotspot 3, and Variable Region I, II and III do not show any insertion in ICE*Vch*Ban8. Variable region IV is disrupted by the genetic rearrangement of the *setCDR* cluster previously discussed (Table 4.1, Fig 4.6).

HotSpot 4

At the 3' end of the *traN gene*, *V. cholerae* MZO-3 contains an unique insertion 49554 bp long. This inserted sequence includes 45 ORFs, as annotated by RAST (Table 4.1, Fig 4.6). Forty ORFs were sufficiently similar to known proteins to be annotated with suggested functions and five had insufficient similarity and were annotated as hypothetical proteins. Interestingly, the insertion included 12 genes resembling a type IV pilus assembly operon, three with similarity to accessory colonization factors present in *V. cholerae* VPI-1, three genes putatively related to known toxins, three ORFs encoding trans-membrane transporters, and several other genes, of which five were proteins of unknown function (Table 4.1). The overall genetic organization resembled that found in the Toxin Coregulated Pilus (TCP) of *V. cholerae* and Pilus Longus of *E. coli* ETEC, both being type IV pili (Giron *et al.,* 1994).

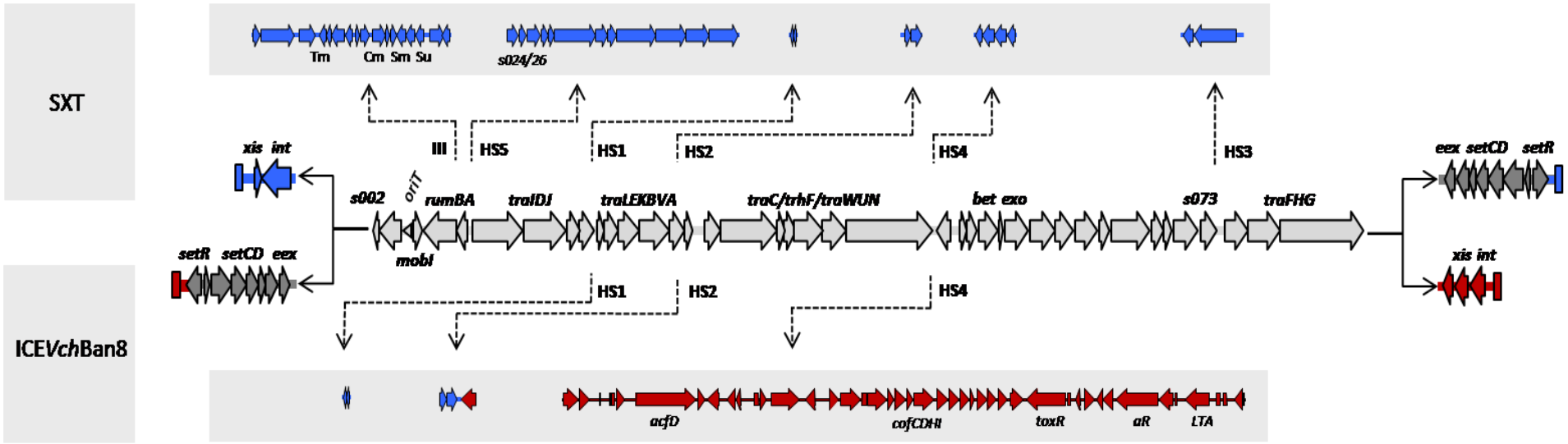


***Figure 4.6**. Structural comparison between SXT and ICEVchBan8. Genetic map of core genes shared between the two ICEs (light gray). Specific regions for SXT and ICEVchBan8 are shown in blue and in red, respectively. In dark gray: inverted genes setR/eex. Hotspot insertions are indicated by dotted arrows and are shown above (SXT) and below (ICEVchBan8) the shared backbone.*

Genetic content of HS4

TPR repeat

This is a 1152 bp gene with no significant nucleotide similarity with any sequence in the database. A BLASTP search showed that the predicted protein contained a DUF560 domain, characteristic of many chromosomal replication initiation proteins in bacteria. The amino acid sequence for the TPR repeat protein has a maximum of 28% identity with a replication initiation protein in *Aggregatibacter actinomycetemcomitans* and 26% with the same protein in *Haemophilus influenzae*.

Outer membrane lipoprotein A precursor

A 846 bp gene has no significant nucleotide similarity with any sequence in the database, 31% amino acid similarity with an hypothetical protein of *Salmonella enterica,* and 27% with the outer membrane lipoprotein of *Actinobacillus pleuropneumoniae* (formerly *Haemophilus pleuropneumonia*), which causes disease in swine.

Putative substrate-binding transport protein

Of the 168 bp of this ORF, 144 bp showed 88% nucleotide identity with the oligopeptide transport system permease protein *ompB* gene of *V. cholerae* MJ-1236 and other ABC transporters-encoding genes in *V. cholerae*. The protein showed 88% amino acid identity with the putative substrate-binding transport protein in several *V. cholerae* strains.

Hypothetical protein

This 555 bp ORF encoded an hypothetical protein containing an OmpA/MotB domain. The protein showed 64% positive amino acid hits with the OmpA-like trans-membrane domain protein of *Vibrio spp*. Ex25. The same protein is also found in *V. parahaemolyticus* AQ3810 and *V. harveyi* 1DA3. In these organisms the protein is inserted in the same order in what seems to be an operon or genetic cluster shared by these organisms in MZO3 not in the same order.

Accessory colonization factor AcfA

This 630 bp gene encoding a putative accessory colonization factor, has no significant similarity at the nucleotide level with any sequence present in the database. The amino acid sequence has 53% positive hits with the accessory colonization factor AcfA of *Grimontia hollisae* CIP 101886, 50% with *Vibrio harveyi* HY01, and 53% with accessory colonization factor AcfA-like protein of *Vibrio fischeri* ES114, *Vibrio alginolyticus*, *V. parahaemolitycus*, and *Vibrio spp.* EX25.

TLP protein

The nucleotide sequence had 70% (229/325) similarity with *Vibrio vulnificus* pR99 plasmid and 68% (152/222) with *V. vulnificus* pC4602-2 plasmid. The amino acid sequence is 98/126 (77%) similar to the transthyretin of *V. vulnificus.* Both pR99 and pC4602-2 are virulence plasmids in *V. vulnificus* biotype-2 and have been showed to be conjugative plasmids virulent for eels (Lee *et al.,* 2008).

Putative extracellular protein

This ORF encodes a protein containing a predicted 35 amino acid signal peptide. Its 513 bp nucleotide sequence has 196/288 (68%) similarity with *Campylobacter coli* Cgb single-domain hemoglobin. The amino acid sequence had 67% positive hits with the homologous protein in *C. coli* RM2228.

This protein family contains many conserved motifs that are suggestive of cofactor binding and enzymatic activity, including several GlcG proteins. The Cgb protein in *C. coli* is associated with a protective role against NO-related stress. According to Elvers and colleagues (2004), *cgb* expression was induced after exposure to nitrosative stress, indicating that Cgb represents a specific and inducible defense against NO and nitrosating agents (Elvers *et al.*, 2004).

Bacterial surface protein

Interestingly, this rather large ORF of 2049 bp encoding a surface protein had no nucleotide significant hits in a BLASTN search. The amino acid sequence revealed weak similarity to surface proteins in other *V. cholerae* strains, *V. parahaemolyticus* and *Shewanella baltica*. Members of this family consist of bacterial proteins with an Ig-like fold and are found in bacterial and phage surface proteins such as intimins.

Hypothetical protein

This 786 bp ORF had no significant similarity at the nucleotide level with known sequences in the database. At the amino acid level, the protein showed similarity

to alpha/beta hydrolase in *V. cholerae* V51 and *V. cholerae* AM-19226, with 59% and 61% similarity respectively. An homologous protein is also found in avian pathogenic *E. coli* (APEC).

Type IV pili biogenesis cluster

The next 12 ORFs in HS4 of ICE*Vch*Ban8 form a genetic cluster that resembles the type IV pili biogenesis cluster found in *V. cholerae, E. coli,* and other pathogenic bacteria. In ICE*Vch*Ban8, the twelve gene operon encodes a new type IV pilus with high structural and functional similarity, not always reflected at the nucleotide level, with type IV pili and proteins of the general secretion pathway (GSP), also known as the"Type II" secretion pathway, that are molecular structures involved in virulence and host colonization of *V. cholerae* and *E. coli* (Table 4.2). The amino acid sequences of the proteins encoded in this region share similarities with class B type IV pili such as the Longus, a T4P composed of a repeating structural subunit called LngA of 22 kDa, the CFA III pilin subunit CofA of ETEC, the toxin-coregulated pilin (TCP) of *V. cholerae,* and the bundle-forming pilin (BFP) found in enteropathogenic *E. coli* (EPEC) and, in a small percentage, in other Gram-negative pathogens. Many functions associated with pathogenicity in Gram-negative bacteria are attributed to T4P, including twitching motility, adherence to host cells, biofilm formation, DNA uptake, phage attachment, cell signaling, invasion, and evasion of the immune system (Giron and *et al.,* 1994**;** Mazariego-Espinosa *et al.,* 2010). Several of these species are also known to form biofilms and mutations in the *eps* genes (the T2SS found in *V. cholerae*)

interfere with polysaccharide production and biofilm formation in *V. cholerae* (Sandkvist**,** 2001).

TcpA

The 732 bp ORF was annotated by RAST as the major subunit of the toxin co-regulated pilus of *V. cholerae*. The nucleotide sequence had 75% similarity to the *tcpA* gene of *V. cholerae* and 67% identity to the *E. coli* Longus pilus major subunit. Interestingly, at the amino acid level, the protein showed 76% similarity to the Longus pilus structural subunit of *E. coli* and only 53% with the structural subunit of the toxin co-regulated pilus of *V. cholerae*.

LngB, CofC, CofG

Interestingly, none of the following three ORFs of the pilus assembly operon found in HS4 showed any significant similarity at the nucleotide level when searched by BLASTN. However, the amino acid sequences revealed proteins with some similarity to known components of pili or T2SS structures. LngB showed low (46%) similarity with the LngB pilus biosynthesis protein of *E. coli* ETEC. CofC had 60% positive hits with the exported pilus biosynthesis protein LngC of *E. coli and* CofG was only 41% similar with the putative Type IV pilus biogenesis protein of *Citrobacter rodentium* ICC168.

TcpT

This 1533 ORF was 71% identical to a type IV pilus gene cluster in *C. rodentium* strain DBS100 and 70% identical to the *cofR* gene of the pilus assembly colonization factor antigen III (CFA/III) of *E. coli.* The amino acid sequence showed 82% positive hits (68% identity) with the LngH protein of E. coli and 78% (64% identity) with the Type IV pilus biogenesis protein CfcH of *C. rodentium* ICC168. This protein is part of the PilT subfamily of proteins. PilT is a putative cytosolic ATPase associated with type IV pilus systems, probably involved in twitching motility. Members of this family are related to the GspE protein involved in type II secretion (also called the General Secretion Pathway) and may be found in type IV pili and related structures for DNA uptake and natural transformation.

PulF/CofI

The 972 bp ORF encodes for the Type II secretory pathway component PulF. Its nuleotide sequence was 70% similar to the *tcpE* gene of the synthetic construct *V. cholerae* clone FLH234732.01F and 64% similar to the *E. coli* plasmid pE9034A Longus operon. The PSI-BLAST search showed 80% positive hits (60% identity) with the CofI and 79% (60% identity) with LngI of *E. coli.* These are membrane proteins and components of the terminal branch of the general secretion pathway (GSP) or Type II secretion pathway.

Pilin peptidase leader peptidase PilD

No similarity was found at the nucleotide level for this ORF. Low identity was observed also at the amino acid level, with only 54% positive hits (36% identity) to the CofP protein and 53% (33% identities) with LngP putative transmembrane pilus biosynthesis protein of *E. coli* ETEC 1392/75. Interestingly, the search also hit the type II secretory pathway prepilin signal peptidase PulO of *Yersinia ruckeri* ATCC 29473 with 52% similarity (38% identity).

Hypothetical protein

This ORFs had 100% nucleoide and 99% amino acid sequence identity with all ICE of SXT/R391 family in the database. In SXT, this ORF is located at the 5'-end of the element, between ORFs *s002* and *rumB*. Interestingly, the exact same ORF is found at this location (upstream of the *umuC* gene) in ICE*Vch*Ban8, suggesting genetic duplication of this region.

AcfC

At the nucleotide level, 68% similarity was observed with all *acfC* genes found in *V. cholerae* strains and with the *Campylobacter lari* RM2100 major antigenic peptide PEB3. The identity is reflected at the amino acid level with 75% positive (58% identity) with all AcfC proteins in *V. cholerae* TCP clusters and the major antigenic peptide PEB3 of *C. lari* RM2100.

Hypothetical Protein

This 1020 bp ORF did not show any significant similarity at the nucleotide or amino acid level in the database.

StcE

This ORF was annotated as *tagA* gene, encoding ToxR-activated A protein. Only a short stretch of 121 bp of the 2922 bp showed 69% similarity to *E. coli* O157:H7 ATCC 43894 plasmid pO157 and *Shigella boydii* plasmid p3053-94 *stcE* genes. The amino acid sequence revealed higher similarity with an hypothetical ToxR-regulated lipoprotein of *P. profundum* 3TCK and the ToxR-activated gene A (TagA protein) found on the VPI-1 in pathogenic *V. cholerae*, with 56% and 52% similarity, respectively.

StcE is a secreted zinc metalloprotease and a virulence factor of *E. coli* EHEC, and is part of the metallopeptidase family. StcE peptidase cleaves the glycoproteins protecting the intestinal epithelium, associates and cleaves the C1 esterase inhibitor, and contributes to intimate adherence of this bacterium to host cells, a process essential for pathogens to form attaching and effacing (A/E) lesions of the microvilli in mammalian gut colonization. Later during the infection of EHEC, virulence factors including Shiga toxins are secreted and affect the intestinal epithelium at the point that blood is released into the intestinal lumen. Part of this family is also the ToxR-regulated lipoprotein (TagA) in *V. cholerae* and *P. profundum*. In *V. cholerae* the role of TagA and other Tag proteins in the pathogenesis of the organism is not yet clarified. It is known that they are proteins residing in the VPI-1 and are under control of the ToxR regulon.

Interestingly, the StcE protease produced by EHEC is secreted by the type II secretion system encoded at the 3′ the pO157 virulence plasmid (Burland *et al*., 1998; Grys *et al.,* 2005). Moreover, pO157 was reported to affect biofilm formation of *E. coli* O157:H7 Sakai strain through the EPS, generating hyper-adherent variants (Lim *et al.*, 2010). Type II secretion apparatus was associatedwith the secretion of proteins required for efficient adherence under control of the *toxB* gene, also encoded on pO157 (Lim *et al.,* 2010).

Methyl-accepting chemotaxis sensory transducer

No significant nucleotide similarity was observed for this ORF. PSI-BLAST search showed weak similarity with the methyl-accepting chemotaxis sensory transducer of *Marinobacter aquaeolei* VT8 and *Geobacter lovleyi* with 54% positive (37% identity).

Protein ygiW

Sixty-six % of the nucleotide sequence was similar to the *Vibrio fischeri* ES114 chromosome I. The amino acid sequence showed 56% identities and 77% positive with the YgiW precursor of other *Vibrio spp* and *P. profundum*.

Type III secretion and flagellar regulator RtsA

This ORF encodes a transcriptional regulator containing an amidase domain and an AraC-type DNA-binding HTH domain. Its amino acid sequence showed 64% of similarity with the hypothetical protein VP1765 of *V. parahaemolyticus* RIMD

2210633 and other *V. parahaemolyticus* strains and also 46% with the transcriptional regulator HilD of *Salmonella enterica*.

Putative regulatory protein

This 762 bp ORF had no significant similarity at the nucleotide level. The encoded protein contained a domain found in trans-C regulator superfamily common to the response regulator of the two-component signal transduction system used by microorganisms to detect and respond to changes in the environment. Part of this family are also the DNA-binding transcriptional regulator RstA and the two-component response regulator VirG found in the conjugative apparatus and in T4SS. The amino acid sequence showed 60% similarity with the transcriptional regulator ToxR of *V. cholerae* and 58% with the transcriptional regulator ToxR of *V. parahaemolyticus.*

AcrB/AcrD/AcrF family protein

Members of this family are integral membrane proteins. Some are involved in drug resistance, including various prokaryotic SecD and SecF export membrane proteins. The protein encoded by this ORF was found to be 71% similar to the acriflavin resistance protein in *Shewanella baltica* OS223 and *Shewanella putrefaciens* CN-32. Furthermore, it was ca. 71-78% similar to the AcrB/AcrD/AcrF family protein found in *Grimontia hollisae* CIP 101886, *Vibrio angustum* S14, *P. profundum* 3TCK. *Vibrio splendidus* LGP32, *Vibrio shilonii* AK1, and *Vibrio coralliilyticus* ATCC BAA-450.

Acriflavine is a topical antiseptic and a derivative of acridine. Originally it was used against the sleeping sickness agent in WWI. This compound is also used as treatment for external fungal infections of aquarium fish. It was shown that acridine compounds selectively inhibit plasmid replication, but not chromosomal replication, in Prokaryotes and were found to have anti-cancer activity in higher organisms (Nakamura, 1982). We tested *V.cholerae* MZO-3 for resistance to acriflavine, with no significant difference with respect to the control strain (see Matherials and Methods).
The only nucleotide identity was 63% with the hypothetical AcrB/AcrD/AcrF family protein of *P. profundum* SS9 and 263 nt over 3066 with genes of other *Vibrio spp.* Interestingly, in all *Photobacterium*, *Vibrio* and *Shewanella spp.*, this protein is part of a conserved cluster composed of the AcrB/AcrD/AcrF family protein, a conserved hypothetical protein and a transcriptional regulator of the TetR family. In *V. cholerae* MZO3, the same cluster was found but the transcriptional regulator was annotated as a short 117 bp ORF by RAST, showing that this gene is probably truncated.

The following two ORFs had the same degree of identity at the amino acid level as the acriflavine resistance protein. The first, an hypothetical protein,was found to contain a domain common to the MFP (membrane fusion protein) component of the RND family of transporters (resistance, nodulation, and cell division),\ and related HlyD secretion protein. The primary functions of this protein family include outward transport, notably nodulation, acriflavin resistance, heavy metal efflux, and multidrug resistance proteins found in Gram-negative bacteria. The proposed function of MFP proteins is to bring the inner and outer membranes together and enable transport to the outside of the outer membrane.

As previously reported, the transcriptional regulator found in this cluster was truncated or not properly annotated. The same degree of acriflavine resistance proteins was found for this ORF, with the transcriptional regulator of the TetR family.

CtxA

This 1806 bp ORF encodes a domain found in enterotoxins such as the heat labile (LT) toxin of *E. coli* and the CTX subunit A in *V. cholerae*. However, only 300 bp had 66% similarity with the *ctxA* gene of toxigenic *V. cholerae* strains, and 215 bp were 73% similar to the *E. coli* (strain 41) heat-labile enterotoxin type IIb (LT-IIb) A and B chain genes.

At the amino acid level, from aa 1 to 240, the protein was found to be 70% similar to the heat-labile enterotoxin IIB A chain of *E. coli* and 67% with CtxA of *V. cholerae*. Interestingly, the second part of the protein, from aa 297, had only weak similarity to the surface proteins PspC and CbpA of *Streptococcus pneumoniae* Taiwan19F-14 (45%), and the choline binding protein A of *S. pneumoniae* CGSP14 (43%).

Sialidase

This 111 bp ORF was annotated by RAST as a sialidase. Its nucleotide sequence is 94% (133/141) similar to the neuramidase gene (*nanH*) found in the VPI-2 cluster in *V. cholerae*. The latter is 2424 bp long, thus the ORF present in ICE*Vch*Ban8 is likely a truncated gene. This idea is supported by the presence of two transposases downstream from the ORF, indicating an insertion event that disrupted the gene.

Interestingly, the first 375 bp transposase was 98% similar to a conserved insertion sequence (IS) found in the *V. cholerae Vibrio* Seventh Pandemic Island 2 (VSP-2), a genomic island originally considered to be unique to seventh pandemic *V. cholerae* and recently shown to be present in a variety of strains of different epidemic importance and also in this study (See Chapter 7, Taviani *et al.*, 2009).

At the amino acid level the short 36 amino acid sequence encoded was 97% identical (35/36) and 100% positive (36/36) with the neuramidase or sialidase genes of *V. cholerae*. The NanH domain is peculiar to the enzyme involved in carbohydrate transport and metabolism and is considered to be of pivotal importance in *V. cholerae* pathogenesis (Jeremy and Boyd, 2002). As previously discussed, the neuramidase is involved in *V. cholerae* pathogenicity in two ways. It cleaves sialogangliosides, generating the GM1 ganglioside receptor for cholera toxin with release of sialic acid. On the other hand, it takes part in the mucinase complex responsible for digestion of intestinal mucus and, thereby, enhancing *V. cholerae* colonization of the gut (Murphy and Boyd, 2008).

Hypothetical Protein.

The last ORF of the 49554 bp insertion found in the HS4 of ICE*Vch*Ban8 was annotated as an hypothetical protein with an interesting 72% similarity of 576 bp to the *V. mimicus* thermostable direct hemolysin gene and 350 bp 71% similar to an hypotetical protein found on chromosome 1 of *Brucella microti* CCM 4915 and also found in *Brucella melitensis*, *Brucella abortus* S19, *Brucella suis* ATCC 23445, and generally conserved in *Brucella spp.* In *Brucella*, this gene encodes a probable signal

peptide protein and contains a nitrobinding heme-binding domain. This protein may reversibly bind nitric oxide and members of this family are fatty acid-binding proteins. In *Brucella spp*., this gene is found in a cluster including three ABC transporters and a lead, cadmium, zinc and mercury transporting ATPase likely involved in heavy metal resistance in these organisms.

Conclusion

V. cholerae O37 MZO-3 was found to carry a mobile genetic element with high similarity to the SXT-like integrative conjugative element (ICE) of *V. cholerae*. The *V. cholerae* O37strain under study was isolated from a cholera patient in Bangladesh and was devoid of important genetic clusters involved in *V. cholerae* pathogenicity (CTX, TCP, VPI-1 and 2, and VSP-1) but carried virulence genes, such as RTX, HylA, and MSHA. The pathogenic mechanisms and factors involved in the virulence of *V. cholerae* non-O1, non-O139 remain unclear and it is hypothesized that these serogroups have different mechanisms of pathogenesis, with respect to toxigenic *V. cholerae.*

According to our results, the ICE present in *V. cholerae* O37 MZO-3 displays an unusual genetic organization. The element inserted at a different locus of the *V. cholerae* genome, other than the *prfC* locus, and appears not to carry an integrase gene related to int_{SXT}. Therefore, this ICE is new and cannot be classified as a member of the SXT/R391 family that includes all ICEs described in *Vibrio cholerae* to date. Furthermore, the ICE showed significant genetic rearrangement in genetic regions that

are usually conserved in the SXT/R391 ICEs, further suggesting it belongs to a novel class of ICE or is an ancient element that has lost the typical genetic structure of SXT/R391 ICEs.

The ICE found in *V. cholerae* O37 MZO-3 is associated with *Vibrio* pathogenicity island 2 (VPI-2) in its integrative capability. ICE*Vch*Ban8, in fact, carries an integrase related to the genomic island integrase and inserted at the VPI-2 locus. Neither VPI-1 or VPI-2 were present in this strain. The results of our analysis add further evidence supporting the hypothesis that ICEs are progenitors or potential dissemination tools for PAIs in *V. cholerae*, like ICE*Ec*1 and ICE*Kp*1 which carry the high pathogenicity island HPI in *E. coli* and *K. pneumoniae*.

ICE*Vch*Ban8 includes a 50 kb large insertion at HS4. The cluster containes genes related to type II secretory systems and type IV pili, both cellular apparati largely involved in Gram negative pathogenesis. Further, outer membrane transporters, accessory colonization factors and putative virulence genes were annotated, including a metalloprotease similar to the StcE protein secreted as virulence factor by highly pathogenic *E. coli.* All together, these virulence factors may take part in the molecular mechanism underlying pathogenesis and provide the ability to cause cholera-like diarrhea by *V. cholerae* O37.

Results of this study improve our knowledge of the variability of ICEs found in *V. cholerae* and their relationship with other genetic elements, such as pathogenicity islands and adds insights leading to a better understanding of the role of these elements in *V. cholerae* non-O1 non-O139 pathogenesis and evolution.

Chapter 5. Geography of ICEs

Introduction

To date, besides SXT^{MO10} several other ICEs have been described in *V. cholerae*. Almost all *V. cholerae* clinical isolates and various environmental *V. cholerae* non-O1 non-O139 in Asia contain SXT-related ICEs (Ehara *et al.*, 2004; Iwanaga *et al.*, 2004; Thungapatra *et al.*, 2002; Burrus *et al.*, 2006) (Fig 5.1).

In Africa, although the element appears to be widespread among clinical strains, information concerning ICE circulation and their role in drug resistance emergence is limited. An outbreak of cholerain 1998 in South Africa was associated with *V. cholerae* O1 carrying the SXT element (Daalsgard *et al.*, 2001). In Angola, SXT integrase was detected in most of the clinical *V. cholerae* O1 strains isolated in 1992, the same year of isolation in India of *V. cholerae* O139 MO10 containing SXT. This discovery was further confirmed by the observation of the ubiquitous presence of an ICE similar to ICE*Vch*Ind1 in *V. cholerae* strains isolated from Zimbabwe, Swaziland and Mozambique, between 1997 and 1998 (Ceccarelli *et al.*, 2006) (Fig 5.1). Furthermore, ICEs were also characterized in environmental *V. cholerae* non-O1 non-O139 isolated in Mozambique between 2002 and 2003 in a cholera endemic area around Maputo (Taviani *et al.*, 2008). The extensive distribution of ICE in austral Africa is also supported by discovery of the first ICE, R391, in environmental strains of *Providencia spp.* in 1967, supporting the hypothesis of independent origin and

diffusion of SXT/R391 ICE in Africa (Burrus *et al.,* 2006) or the supposition that these elements originated in this continent and then spread to the east.

With respect to the Americas, discovery of ICE*Vch*Mex1 provides the first and only report of an SXT-related ICE in the western hemisphere to date. It was found in an environmental strain of *V. cholerae* isolated in Mexico in 2001 (Burrus *et al.,* 2006b) (Fig 2.4). ICE*Vch*Mex1 integrates in *prfC* and is self-transmissible. Given its unique molecular organization, it is believed that ICE*Vch*Mex1 evolved independently with respect to other ICEs related to SXT^{MO10} and ICE*Vch*Ind1. The *V. cholerae* O1 El Tor associated with cholera epidemics in Latin America is related to pre-O139 Asian *V. cholerae* El Tor, which did not contain SXT, thus suggesting that acquisition of an ICE occurs independently from an environmental reservoir.

ICEs in other species have been found in other geographical areas, especially in organisms of environmental origin. For example the ICE*Pal*Ban1 element was detected in *Providencia alcalifaciens* in Bangladesh in 1999 and the ICE*Pda*Spa1 in *Photobacterium damsela* subsp. *piscicida*, a *Vibrio*-related fish pathogen, in Spain (Hochhut *et al.,* 2001; Ahmed *et al.,* 2005; Juiz-Rio *et al.,* 2005). The presence of an ICE in the SXT/R391 family was also detected in *V. parahaemolyticus* environmental strains in Angola and Mozambique, Africa (Ceccarelli *et al.,* 2006; Taviani *et al.,* 2008). The observation of related elements in species from the same natural environment as *V. cholera* suggests a broader host range and transfer capability of this class of mobile genetic elements.

Here we report results of screening a large, well-characterized collection of *V. cholerae* and other *Vibrio* strains isolated from different geographical areas and at

different chronological times. The study was designed to assess the distribution of ICEs among organisms isolated from geographical locations where the presence of ICEs had not yet been reported, namely North and South America, the Republic of Georgia, and Iceland. The analysis was done to determine the geographical distribution of ICEs in *V. cholerae* and not to provide a detailed analysis of the nucleotide sequence of each element. The presence of conserved clusters was compared to elements that had been shown to occur in other geographical regions. Through presence/absence denotation of the conserved regions and known variable regions, the ICEs were profiled and geographical distribution, if any, assessed for each profile.

Materials and Methods

Bacterial strains and DNA preparation

V. cholerae strains used in this study are presented in accordance with the geographical location of origin in Fig 5.1 as follow:

North America: 284 *V. cholerae* strains isolated from the Chesapeake Bay during 1998-2000 and 23 *V. cholerae* strains isolated across North America at various chronological times;

Central and South America: 227 strains isolated in central and south america at different times. This collection included 101 *V. cholerae* strains from Mexico, 77 *V. cholerae,* five *Aeromonas spp.,* four *V. metschnikovii,* one *V. natriegens* and one *Enterobacter*from Peru, 43 *V. cholerae* from Brazil, and five *V. cholerae* and two *V.*

parahaemolyticus from Chile, isolated at different chronological times;

Republic of Georgia: 500 *V. cholerae* strains isolated during 2004-2007.

Iceland: 200 *V. cholerae* strains isolated during 2006-2007.

All cultures were grown in Luria-Bertani medium and stored at -80C in LB, to which was added 25% glycerol. Total genomic DNA are prepared, according to methods described in *Current Protocol in Molecular Biology* (Ausubel *et al.*, 1990).

PCR assay

Screening was based on PCR assays with primers previously designed to test for the presence, integration, and general genetic organization of ICEs. Detection of ICEs belonging or related to the SXT/R391 family was preliminarily accomplished using specific primer pairs for int_{SXT} (integrase gene). Strains positive for int_{SXT} were further investigated for integration into *prfC* in the *V. cholerae* chromosome by amplification of the *prfC*/SXT right junction of the chromosomal integration region (P3/P1 primer pair) (Table 5.1). A specific backward primer was used to assess integration of ICE on the *V. parahaemolyticus prfC* gene (Table 5.1) coupled with the P3 forward primer on the SXT right junction. To confirm the integration locus, specific primers for PCR amplification of the *prfC* intact gene in both *V. cholerae* and *V. parahaemolyticus* (Table 5.1).

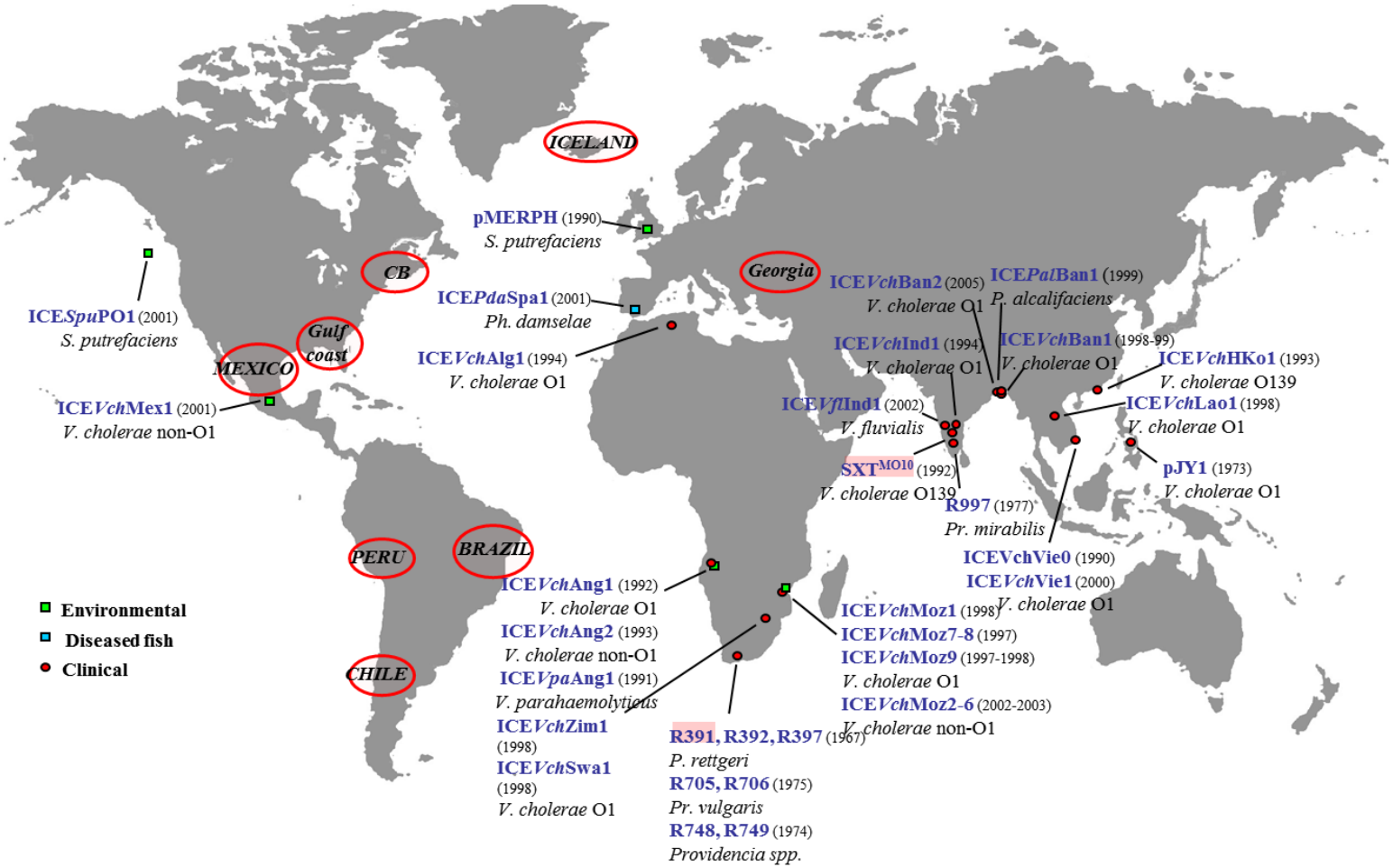


Fig. 5.1. *Worldwide distribution of SXT/R391 ICEs. Geographical regions investigated in this study are shown in red circles.*

All *int*$_{SXT}$-positive strains were further tested for genetic arrangement of the *rumAB* operon and for resistance genes typically associated with SXT^{MO10} or R391: *floR, strA* and *strB, sul2, dfrA18* and *dfrA1* and *kan* (Hochhut *et al.*, 2001; Iwanaga *et al.*, 2004; Bani *et al.*, 2007). Moreover, the conserved core of the SXT/R391 family was studied by amplification of three conserved genes present in the four *tra* clusters shared by SXT^{MO10} and R391: traI-R/orf171 (*traI* gene), traC-B/traC-F (*traC* gene) and SetRp-R/SetRp-F (*setR* gene) (Table 5.1). We further characterized the ICE molecular profile by investigation of five additional regions allowing distinction between the SXT^{MO10} and R391 elements: *s026/s027* bearing the *kan* gene in R391, *traG/s079* bearing the *mer* conferring mercury resistance in R391, *s043/traL* (Hotspot 1), *traA/s054* (Hotspot 2) and *s073/traF* (Hotspot 3) (Table 5.1, Fig 5.2). All PCR reactions were carried out in a 25 μL reaction containing 1U of Taq polymerase, as directed by the manufacturer (Promega).

ICEs identified were described and profiled by the results of amplifications and named according to place of isolation. It should be kept in mind that this comprised a preliminary survey limited in accurate characterization of a diverse genetic structure. Thus, elements that appeared to have a similar genetic profile in this analysis might include important structural rearrangement in regions not investigated, resulting in diverse functionality. Similarly, a discrepancy between two different genetic profiles can be caused by a single mismatch affecting amplification.

***Table 5.1**: primer pairs used in this study.*

Primer	***Sequence (5' to 3')***	***Tm***	***Target gene***	***Reference or source***
P3	CAGCTACAACTGAGCATTGGC	68.9	*insertion site*	AE004152, bp 269-250
VpPrfC-R	CCTGCTTGATACTCCAGGACA	68.8	*V. parahaemolyticus*	Taviani et al., 2008
P4	TGCTGTCATCTGCATTCTCCTG	67.6	*circular form*	Hochhut et al, 2000
P5	GCCAATTACGATTAACACGACGG	67.8		Hochhut et al, 2000
PVC-F	TCCTGCACCTTGCTCTGCTCT	60.3	*prfC intact*	Taviani et al., 2008
PVC-B	ACCACGCTCTTTTTCCATTTCCAT	57.1		Taviani et al., 2008
VisLF	GAGTACAAATTCCGTTTTAG	53	*attL-xis*	Burrus et al., 2006
VisLR3	GCATTCTCCTGAAAATCAATG	61		Burrus et al., 2006
INT3	TGAGCCAACCCAATCTATCC	66	*s002-xis*	This study
INT4	TATACGACGCTCTGGCGAAG	69		This study
INT1	GCTGGATAGGTTAAGGGCGG	63	*ICE integrase*	Hochhut et al, 2000
INT2	CTCTATGGGCACTGTCCACATTG	63.9		Hochhut et al, 2000
nint-F	TACCTACAGCAGGAACGGGC	64	*ICE integrase*	Dalsgaard et al., 2001
nint-B	GCAGCACAGACACCAGACGT	64		Dalsgaard et al., 2001
LEFTF3	GGTGCCATCTCCTCCAAAGTGC	68.1	*rumAB*	AB114188, bp 23369-23350
RUMA	CGAGCAATCCCCACATCAAG	64.7		Hochhut et al, 2000
RIGHTA	GCACTTGCCTTCATCAAAGC	64.7		
LEFTF3	GGTGCCATCTCCTCCAAAGTGC	68.1	*rumB-traD*	
TraIR	CCAGGGCATCTCATATGCGT	67.1		Burrus et al., 2006
aphF	GGCAATCAGGTGCGACAAT	66.1	*kan*	Hochhut et al, 2000
aphR	GTGACGACTGAATCCGGTGA	66.3		Hochhut et al, 2000
SO26F	GAGCAATGGGCGAGAGTTCC		*s026-s027*	Burrus et al., 2006
SO27R	TCAGCGACAACCGGAGAATG			Burrus et al., 2006
s026-F	AAGCAATGGAACCGAATCGTT	66.5	*s026-s027/kan*	AY055428, bp 26896- 26914
s027-R	ACCATGCATCAGCGGTTAAAG	65.9		AY055428, bp 27549-27528
HS1F	GGCTATTCCACCGGTGGTG	67.5	*s043-traL*	Burrus et al., 2006
HS1R	TGCCGATCACTAGCCCCAAC	69.4		Burrus et al., 2006
s045-F	TCAAATAACCCATCGCCCAT	66	*s045-traL (SXT)*	Hochhut et al, 2000
TraL-R	TGGGTTAGGGATCGTCTTGG	66		Hochhut et al, 2000
cds38-F	GCTAAAGTGGGCAGCTGGTC	66	*cds38-traL (R391)*	Hochhut et al, 2000
HotS2F	TTCCAGCAATCAGCGCCG	70.9	*traA-s054*	Burrus et al., 2006
HotS2R	CAGTTGTCCTATGTGGACTCGG	65.6		Burrus et al., 2006
s052-R	TTTATATCGCCCTGCGCATT	66.2	*traA-s052 (SXT)*	AY055428, bp 55699-55719
TraA-F	GGTGGCGGTATGGGTCTCTA	66		AY055428, bp 55163-55183
cds45-R	CAGTTTCCAGAGAGCAAGCGT	65.8	*traA-cds45 (R391)*	AY090559, bp 42916-42937
HotS3F2	CCTAAGCATCCTTGGAAGGCT	65.4	*s073-traF*	Burrus et al., 2006
TraF2R	TGGGATGGTCACCCATAGGA	67.4		Burrus et al., 2006
s075-F	TCGCTTTGCTTGGGTTTGTT	66.6	*s075-traF (SXT)*	AY055428, bp 87904-87924
TraF-R	CCACGCTGATAGAATTGCCC	66.6		AY055428, bp 88892-88872
cds78-F	TGAAGGCAATCCGGATAACAG	65.9	*cds78-traF (R391)*	AY090559, bp 73534-73555

TraG-F	GGGACGCGTATTCAACTCTGA	66.2	*traG-s079 (SXT)*	AY055428, bp 94436-94457
s079-R	CAAGTTAGCGGCTCTGCCAT	66.4		AY055428, bp 95036-95016
MerA-F	AGCTTTTCCCTTACCTGACTATG	64.9	*mer operon (R391)*	AY090559, bp 83571-83596
13TA4	TTGCGGGAGATTATGCTC	61.3	*traG-setC*	Hochhut et al, 2000
MerR	TGACCATCAATGAAGGTTG	59.4		Hochhut et al, 2000
pE1	TGACCATCAATGAAGGTTG	59.4		
pE2	TTGCGGGAGATTATGCTC	61.3		
pT1	CATCTAGCGCCGTTGTTAATC	69.4		
pT2	ATCGCGATACTCAGCACGTCG	74.1		
In-F	GGCATCCAAGCAGCAAG		*Class I integron*	Iwanaga et al., 2004
In-B	AAGCAGACTTGACCTGA			Iwanaga et al., 2004
dfr1-F	CGAAGAATGGAGTTATCGGG		*dfrA1*	Iwanaga et al., 2004
dfr1-B	TGCTGGGGATTTCAGGAAAG			Iwanaga et al., 2004
FLOR-F	TTATCTCCCTGTCGTTCCAGCG	62.1	*floR*	Iwanaga et al., 2004
FLOR-2	CCTATGAGCACACGGGGAGC	63.4		Iwanaga et al., 2004
TMP-F	TGGGTAAGACACTCGTCATGGG	62.1	*dfrA18*	Hochhut et al., 2001b
TMP-B	ACTGCCGTTTTCGATAATGTGG	58.4		Hochhut et al., 2001b
SUL2-F	AGGGGGCAGATGTGATCGC	64	*sul2*	Hochhut et al., 2001b
SUL2-B	TGTGCGGATGAAGTCAGCTCC	66		Hochhut et al., 2001b
STRA-F	TTGATGTGGTGTCCCGCAATGC	62	*strA*	Dalsgaard et al., 1999
STRA-B	CCAATCGCAGATAGAAGGCAA	62		Dalsgaard et al., 1999
strB-F	GGCACCCATAAGCGTACGCC	66	*strB*	Dalsgaard et al., 1999
strB-R	TGCCGAGCACGGCGACTACC	70		Dalsgaard et al., 1999
TET(A) 1	GCTACATCCTGCTTGCCTTC	64	*tetA*	Hochhut et al., 2001b
TET(A) 2	CATAGATCGCCGTGAAGAGG	63		Hochhut et al., 2001b

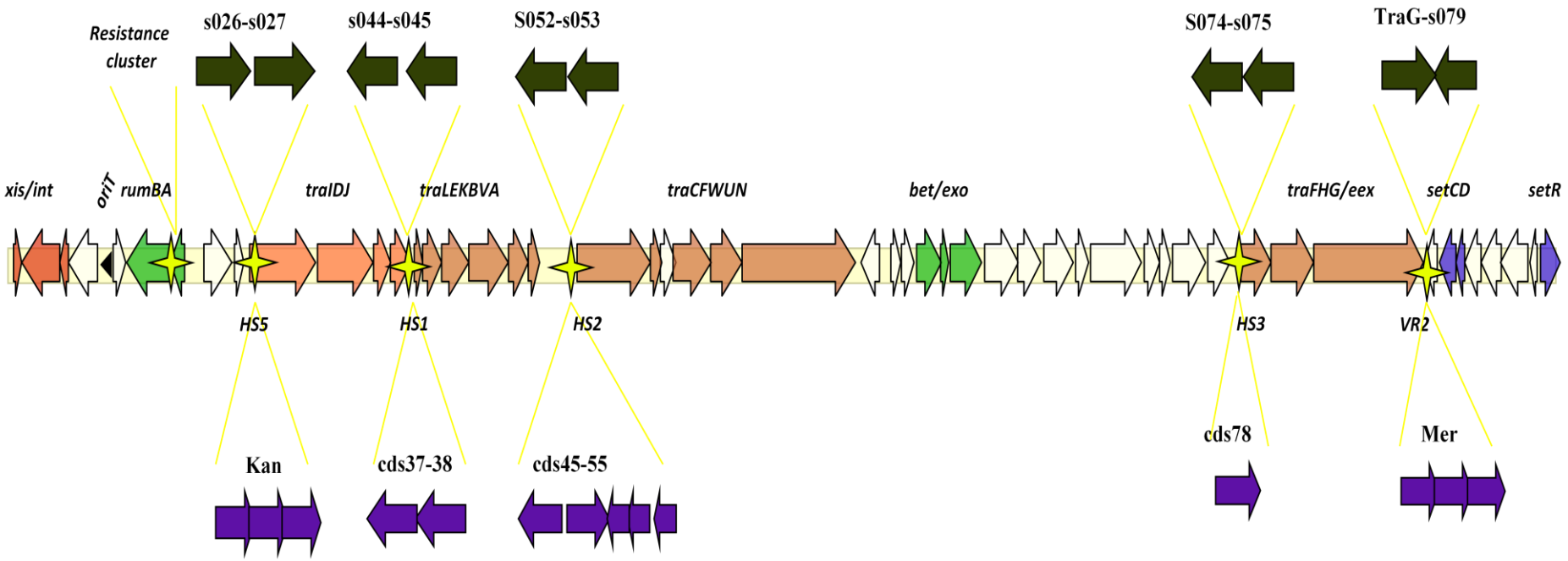


Fig 5.2: *Linear map of the common backbone of SXT/R391. Genes indicated are specific for insertions in SXT (green) or R391 (purple) in hotspots (HS), and conserved regions (CR), investigated using primer pairs listed in Table 5.1.*

Results and Discussion

North America

Of the 507 *V. cholerae* strains isolated from North America, only strains from the Chesapeake Bay were positive for presence of ICE (see below). None of the 12 strains isolated from the Gulf coast in Louisiana were positive for ICE.

Chesapeake Bay

V. cholerae non-O1, non O139 was isolated from water and plankton from five **s**hore sampling sites during January 1998 to February 2000 throughout the northern Chesapeake Bay: Susquehanna River Flats (F); Baltimore Harbour (B); Smithsonian Environmental Research Center (SERC) (S); Kent Island (K); and Horn Point Laboratory (H) (Fig. 5.3) (Louis *et al.*, 2003).

Of 284 *V. cholerae* non-O1 isolated from the Chesapeake Bay between 1998 and 2000, 26 (9%) were positive by amplification of the ICE integrase. Identity of the amplicons was confirmed by nested PCR with the nint-F/nint-B internal primer pair. Furthermore, integration into the *prfC* locus was confirmed in all 26 strains, showing that all belonged to the SXT/R391 family (Table 5.2; Fig. 5.4a).

***Table 5.2**: ICEs detected in V. cholerae isolated from the Chesapeake Bay.*

strain	*species*	*site*	*int*	*prfC*	*rumAB*	*VR1*	*HS1*	*HS2*	*HS3*	*VR2*	*traI*	*traC*	*setR*	*ICE*
523	*V. cholerae*	B	1	1	1	0	444	0	0	601	1	1	1	ICE*Vch*CB2
342	*V. cholerae*	H	1	1	1	0	444	0	0	601	1	1	1	ICE*Vch*CB2
473	*V. cholerae*	H	1	1	1	0	444	0	0	601	1	1	1	ICE*Vch*CB2
486	*V. cholerae*	H	1	1	1	0	444	491	0	601	1	1	1	ICE*Vch*CB1
488	*V. cholerae*	H	1	1	1	0	444	0	0	601	1	1	1	ICE*Vch*CB2
492	*V. cholerae*	H	1	1	1	0	0	491	0	601	1	1	1	ICE*Vch*CB3
493	*V. cholerae*	H	1	1	1	0	0	0	0	601	1	1	1	ICE*Vch*CB4
497	*V. cholerae*	H	1	1	1	0	444	491	0	601	1	1	1	ICE*Vch*CB1
533	*V. cholerae*	H	1	1	1	0	444	491	0	0	0	0	0	ICE*Vch*CB5
504	*V. cholerae*	K	1	1	1	0	0	0	0	601	1	1	1	ICE*Vch*CB4
509	*V. cholerae*	K	1	1	1	0	599	0	0	601	1	1	1	ICE*Vch*CB6
401	*V. cholerae*	S	1	1	1	0	444	491	0	601	1	1	1	ICE*Vch*CB1
402	*V. cholerae*	S	1	1	1	0	0	491	0	601	1	1	1	ICE*Vch*CB3
404	*V. cholerae*	S	1	1	1	0	444	491	0	601	1	1	1	ICE*Vch*CB1
410	*V. cholerae*	S	1	1	1	0	444	491	0	601	1	1	1	ICE*Vch*CB1
450	*V. cholerae*	S	1	1	1	0	444	0	0	601	1	1	1	ICE*Vch*CB2
462	*V. cholerae*	S	1	1	1	0	444	491	0	0	0	0	0	ICE*Vch*CB5
514	*V. cholerae*	S	1	1	1	0	444	491	0	601	1	1	1	ICE*Vch*CB1
515	*V. cholerae*	S	1	1	1	0	444	491	0	601	1	1	1	ICE*Vch*CB1
516	*V. cholerae*	S	1	1	1	0	0	491	0	601	1	1	1	ICE*Vch*CB3
519	*V. cholerae*	S	1	1	1	0	599	0	0	0	1	1	1	ICE*Vch*CB7
520	*V. cholerae*	S	1	1	1	0	444	491	0	601	1	1	1	ICE*Vch*CB1
526	*V. cholerae*	S	1	1	1	0	444	491	0	601	1	1	1	ICE*Vch*CB1
527	*V. cholerae*	S	1	1	1	653	444	491	0	601	1	1	1	ICE*Vch*CB8
528	*V. cholerae*	S	1	1	1	0	444	491	0	601	1	1	1	ICE*Vch*CB1
530	*V. cholerae*	S	1	1	1	0	444	491	0	601	1	1	1	ICE*Vch*CB1
847	*V. cholerae*	B	1	1	1	0	599	0	0	0	0	0	0	ICE*Vch*CB77
928	*V. cholerae*	K	1	1	1	0	599	0	0	0	0	0	0	ICE*Vch*CB76

The 26 integrase-positive strains were tested for the presence of the resistance gene cluster disrupting the *rumAB* operon and it was found that all isolates carried an intact *rumAB*, suggesting absence of any insertion. Thus, ICEs present in the Chesapeake Bay *V. cholerae* lacked resistance determinants typically associated with SXT^{MO10} or R391 (Fig. 5.4a). To classify these ICEs, amplification of the conserved backbone and variable insertions was performed and results are shown in Table 5.2. We identified eight different profiles in the 26 isolates. The most common profile was ICE*Vch*CB1, found in 11 strains, ICE*Vch*CB2 in five strains, ICE*Vch*CB3 in three strains, ICE*Vch*CB4 and ICE*Vch*CB5 in two strains and ICE*Vch*CB6, ICE*Vch*CB7, ICE*Vch*CB8 present in only one strain.

Detection of ICE in the strains under study was not successful with comparable frequency at every sampling site. For instance, none of the 15 strains isolated at the Susquehanna River Flats (site F) carried ICEs (Table 5.2, Fig 5.3). ICEs were isolated from one of six *V. cholerae* strains (17%) at the Baltimore Inner Harbor (site B), 15 of 76 (20%) at the Smithsonian Environmental Research Center (SERC) (site S), two of 74 (3%) at Kent Island (site K) and eight of 89 (8%) at Horn Point Laboratory (site H) (Table 5.2, Fig. 5.3). Interestingly, since Susquehanna River is the only fresh water sampling site with salinity equal to 0 ppt and the only site where no ICE were isolated, whereas all other were brackish water sites with salinity ranging from 2 to 15 ppt (Louis *et al.*, 2003), this environmental parameter may influence genetic transfer within the bacterial populations. However, temperature, pH, and chlorophyl A measured at the four other sampling sites were comparable. Thus the different

isolation frequencies observed for ICE positive *V. cholerae* strains is concluded to be independent of these variables.

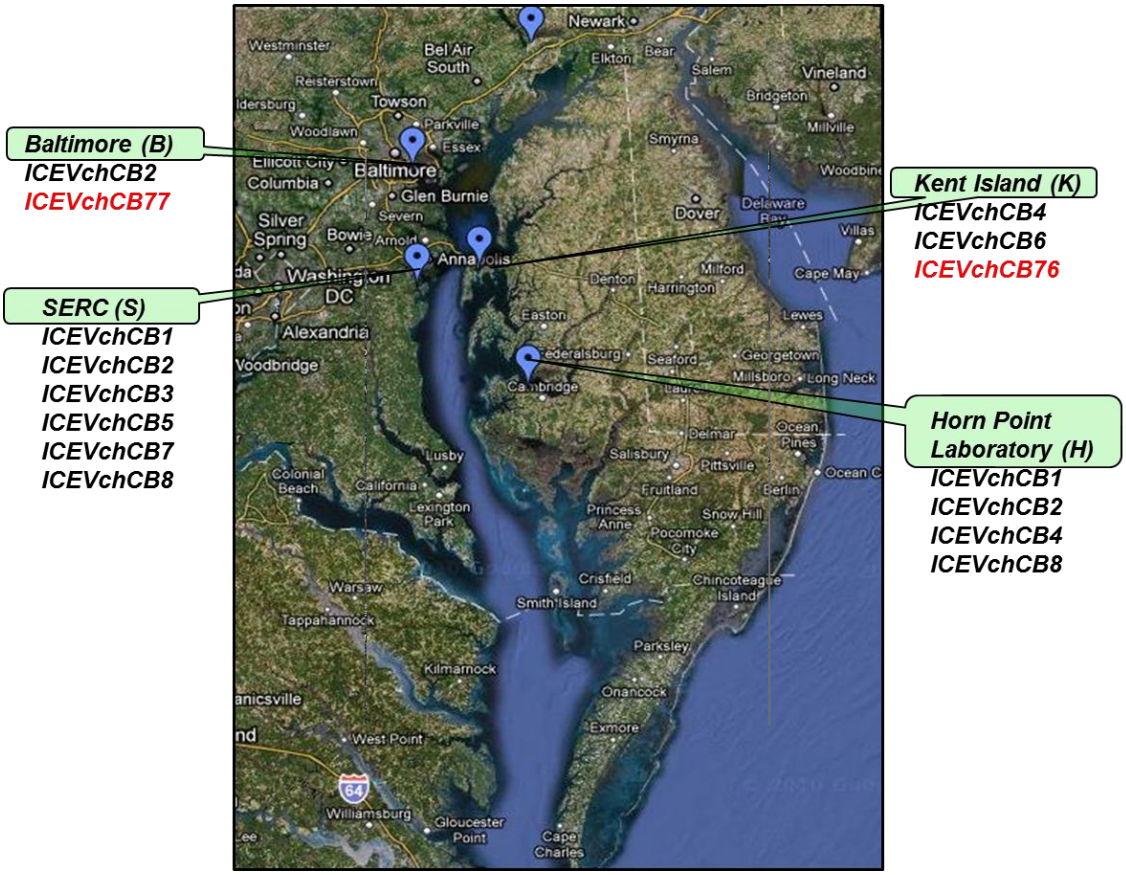


Fig. 5.3*. Chesapeake Bay, US, map shoving sampling sites and ICE profiles detected at each site*

The three genes included in the conserved genetic core of SXT/R391 were amplified in all ICE except ICE*Vch*CB8, present in two strains isolated from the SERC and Horn point Laboratory sites, respectively (Table 5.2). In ICE*Vch*CB8, *traI*, *traC*, and *setR* genes were not detected, suggesting the presence of genetic polymorphisms or rearrangements in these regions, affecting amplification with the primer pairs used. Amplification of variable regions and hotspots revealed that the elements carry a combination of insertions common to SXT and to R391 (shown in blue and grey, respectively in Table 5.2 and Fig 5.4a), and to R391. Variable region 2 (VR2), found between *traG* and ORF *s079* showed the typical genetic content as found in SXT, missing the mercury resistance operon present in R391. Interestingly, the variable region 1 (VR1) was not amplified in any of the strains with the primer pairs used, except for *V. cholerae* CP527 carrying a ICE with a unique profile, ICE*Vch*CB8, with the typical SXT insertion (Table 5.1). This suggested absence of the kanamycin resistance cluster found in R391, in the ICE circulating among environmental *V. cholerae* strains in Chesapeake Bay.

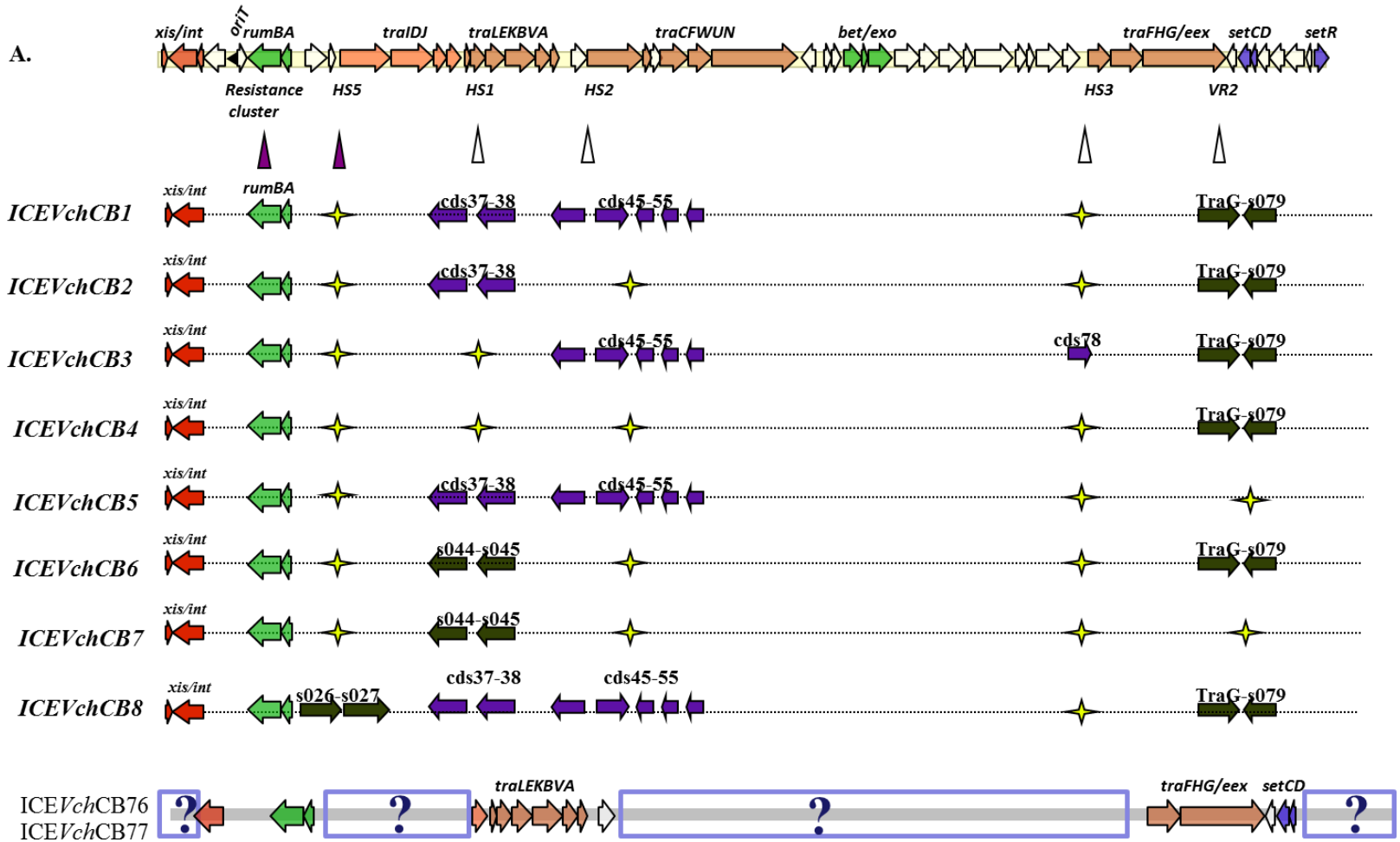


Fig. 5.4.*Genomic profiles of ICE discovered in V. cholerae isolates from the Chesapeake Bay, US.A. schematic representation of ICE conserved backbone. Content of each hotspots is shown. Green and purple arrows indicate SXT-like and R391-like insertions.*

.

ICE*Vch*CB76 and ICE*Vch*CB77

Interestingly, among the north American strains assayed, the other two *V. cholerae* isolated from Chesapeake Bay, MD, USA in 1976 and 1977 were positive for SXT integrase, namely *V. cholerae* non-O1/O139 CP847 isolated from the Chester River in May 1977 and *V. cholerae* non-O1/O139 CP928 from Jones Falls (Baltimore) in September, 1976 (Kaper *et al.,* 1979). These strains were part of a second collection of 65 *V. cholerae* non-O1/O139 isolated from water, sediment, and shellfish from the Chesapeake Bay in 1976-78 (Kaper *et al.,* 1979). Only 5 of the 65 strains were screened since they were the strains included in our current laboratory collection. To date, ICEs were found in *V. cholerae* strains isolated after 1992. There is only one report of SXT-like elements in *V. cholerae* non-O1, non-O139 strains isolated from Varanasi, India (Mohapatra *et al.,* 2008). Interestingly, the two ICE positive *V. cholerae* strains were isolated from the same sites where we found these elements more than 20 years later (Fig. 5.3). The Chester River sampling site of Kaper *et al.* (1979) coincides with the Kent Island sampling site of Louis *et al*, 2003 and Jones Fall sampling was done at the confluence of this tributary at the Baltimore Inner Harbor (Fig. 5.3).

PCR assays targeting the integrase gene and the *prfC* insertion locus confirmed its presence and the integration site of the ICE in both strains. Thus, the elements were confirmed as belonging to the SXT/R391 class of ICE and named ICE*Vch*CB76 and ICE*Vch*CB77 to emphasize the early isolation date of the host. Little characterization of the genetic backbone was possible because not all the primer pairs for the conserved

regions gave the expected amplification (Table 5.2, Fig 5.4b). Long PCR was performed with ExTaq Polymerase and generated up to 10 kb amplicons covering the regions at the 5' and 3' ends of the element, suggesting a different genetic arrangement at the borders of these ICEs, not observed in any other ICE described to date. The amplicons were submitted for sequencing but yelded mixed sequences in different attempts. Genetic characterization of these ancient ICEs is currently under study.

The results confirm the hypothesis that integrative conjugative elements were present in environmental *V. cholerae* long before the first report in 1992 and in geographical areas distant from the Indian subcontinent, suggesting that ICE are a powerful genetic tool for bacterial adaptation, rapidily spreading among environmental bacterioflora and that these elements evolved in autochthonous *V. cholerae* populations independently of their clinical importance.

Central and South America

Among strains from Central and South America, ICEs were found with differing frequency, depending on geographical origin.

Peru

In Peru the highest incidence of *V. cholerae* strains carrying ICEs was observed (Table 5.3): 40% (n=27) of *V. cholerae* non-O1/O139, 100% of *Aeromonas spp.* (n=5), and a single *V. natriegens* (n=1) and *Enterobacter*(n=1) were positive for the SXT

integrase and insertion at the *prfC* locus. The strains were collected from water and plankton at four sampling sites along the coast of Peru between October 1997 and June 2000 (Gil *et al.*, 2004) (Table 5.3, Fig. 5.5). From molecular characterization of the SXT/R391 ICEs found in the Peruvian strains emerged a high genetic variability that often resulted in negative amplification of the three conserved regions of the core (Table 5.3). Only ICEs with genetic profiles positive for these conserved regions were classified. With this criterion, eleven different profiles were identified in *V. cholerae* and named ICE*Vch*Peru1 to ICE*Vch*Peru11, one in *Enterobacter,* ICE*Ent*Peru1, one in *V. natriegens*, ICE*Vnat*Peru1, and two in *Aeromonas spp*. ICE*Aer*Peru1, ICE*Aer*Peru2 (Table 5.3, Fig. 5.6). Interestingly, ICE*Vch*Peru2 and both the ICE*Aer*Peru1 and the ICE*Aer*Peru2 were characterized by the presence of an insertion at the *rumAB* locus, the typical insertion site of the resistance cluster found in SXT-related ICEs. To characterize this insertion, the presence of *sul2*, *strA*, *strB*, *dfrA18*, *floR resistance genes* was tested and none were detected, suggesting a polymorphism in the *rumAB* sequence affecting amplification (Table 5.3, Fig. 5.6). Furthermore, the elements ICE*Vch*Peru1 and ICE*Vch*Peru7 appeared to carry the mercury resistance cluster found in R391, inserted downstream from *traG* (Table 5.3, Fig. 5.6). This locus appeared to be more variable, with two elements carrying either SXT or R391 insertions (515 bp amplicon), seven carrying the SXT-like profile (601 bp amplicon), and five with an amplification profile (300 and 450bp amplicons) never observed before (Table 5.3, Fig. 5.6).

***Table 5.3**: ICEs detected in V. cholerae isolated from Central and Latin America.Strains with negative amplification at the rumAB locus (shown in red, bold) were tested for antibiotic resistance cluster, none was positive. SXT-like and R391-like insertions are shown in blue and grey, respectively.*

						rum				*VR1*	*HS2*	*HS1*	*HS3*	*VR2*	
Strain	***Species***	***Serogroup***	***Location***	***int***	***prfC***	***AB***	***traI***	***traC***	***setR***	***s026***	***traA***	***traL***	***traF***	***traG***	
140	*V. cholerae*	O1	Chile	1	1	1	1	1	1	0	0	0	0	601	ICE*Vch*Chi1
156	*V. cholerae*	non-O1	Chile	1	1	1	1	1	1	0	0	0	0	601	ICE*Vch*Chi1
207	*V. parahaemolyticus*	O3:K58	Chile	1	1	**0**	1	1	1	0	0	0	0	350	ICE*Vpar*Chi1
257	*V. cholerae*	non-O1	S. Paulo	1	1	1	1	1	1	0	0	0	0	0	ICE*VchBra*1
314	*V. cholerae*	non-O1	Brazil	1	1	1	1	1	1	0	0	0	0	0	ICE*VchBra*1
319	*V. cholerae*	non-O1	Brazil	1	1	1	1	1	1	0	0	0	0	0	ICE*VchBra*1
638	*V. cholerae*	non-O1	Veracruz	1	1	1	1	1	1	0	0	0	0	601	ICE*Vch*Mex5
652	*V. cholerae*	non-O1	Yucatan	1	1	1	1	1	1	0	0	0	0	0	ICE*Vch*Mex6
665	*V. cholerae*	non-O1	Chiapas	1	1	1	0	0	0	0	0	0	0	515	ICE*Vch*Mex7
668	*V. cholerae*	non-O1	Colima	1	1	1	1	1	1	0	0	0	0	0	ICE*Vch*Mex5
767	*V. cholerae*	O1	Yucatan	1	1	**0**	1	1	1	0	0	0	0	0	ICE*Vch*Mex2
768	*V. cholerae*	O1	Yucatan	1	1	**0**	1	1	1	0	0	0	0	0	ICE*Vch*Mex2
769	*V. cholerae*	O1	Yucatan	1	1	**0**	1	1	1	0	0	0	0	0	ICE*Vch*Mex3
770	*V. cholerae*	O1	Yucatan	1	1	**0**	1	1	1	0	0	0	0	0	ICE*Vch*Mex3
771	*V. cholerae*	O1	Mexico	1	1	**0**	1	1	1	0	0	0	0	0	ICE*Vch*Mex3
774	*V. cholerae*	O1	Veracruz	1	1	**0**	1	1	1	0	0	0	0	0	ICE*Vch*Mex3
695	*V. cholerae*	O1 ET	Tamaulipas	1	1	1	0	0	0	653	0	0	0	0	ICE*Vch*Mex4
699	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	515	ICE*Vch*Peru1
700	*V. cholerae*	non-O1	Peru	1	1	**0**	1	1	1	0	0	0	0	0	ICE*Vch*Peru2
701	*V. cholerae*	non-O1	Peru	1	1	**0**	1	1	1	0	0	0	0	0	ICE*Vch*Peru2
702	*V. cholerae*	non-O1	Peru	1	1	**0**	1	1	1	0	0	0	0	0	ICE*Vch*Peru2
703	*V. cholerae*	non-O1	Peru	1	1	1	1	0	0	0	0	0	0	0	
705	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	601	ICE*Vch*Peru3
706	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	601	ICE*Vch*Peru3
707	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	601	ICE*Vch*Peru3
708	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	0	
709	*V. cholerae*	non-O1	Peru	1	1	**0**	0	0	0	0	0	0	0	0	ICE*Vch*Peru2
717	*V. cholerae*	non-O1	Peru	1	1	1	0	0	0	0	0	0	0	0	
718	*V. cholerae*	non-O1	Peru	1	1	**0**	1	1	1	0	0	0	0	0	ICE*Vch*Peru2

719	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	601	ICE*Vch*Peru3
720	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	444	0	601	ICE*Vch*Peru4
721	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	0	ICE*Vch*Peru5
722	*V. cholerae*	non-O1	Peru	1	1	1	0	1	1	0	0	0	0	0	
723	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	946	601	ICE*Vch*Peru6
725	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	515	ICE*Vch*Peru7
737	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	300	ICE*Vch*Peru8
738	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	653	0	0	0	300	ICE*Vch*Peru9
739	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	300	ICE*Vch*Peru8
740	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	0	ICE*Vch*Peru5
741	*V. cholerae*	non-O1	Peru	1	1	1	1	0	1	0	0	0	0	0	
742	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	0	ICE*Vch*Peru5
743	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	450	ICE*Vch*Peru10
744	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	601	ICE*Vch*Peru3
747	*V. cholerae*	non-O1	Peru	1	1	1	1	1	1	0	0	0	0	450	ICE*Vch*Peru10
757	*Enterobacter*		Peru	1	na	na	1	1	1	0	0	0	0	0	ICE*Ent*Peru1
760	*V. natriegens*		Peru	1	na	na	1	1	1	0	0	0	0	0	ICE*Vnat*Peru1
786	*Aeromonas sp.*		Peru	1	na	na	1	1	1	0	0	0	0	0	ICE*Aer*Peru1
787	*Aeromonas sp.*		Peru	1	na	na	1	0	1	0	0	0	0	0	ICE*Aer*Peru1
788	*Aeromanas sp.*		Peru	1	na	na	1	1	0	0	0	0	0	0	ICE*Aer*Peru1
789	*Aeromonas sp.*		Peru	1	na	na	1	1	1	0	0	0	0	0	ICE*Aer*Peru1
790	*Aeromonas sp.*		Peru	1	na	na	1	1	1	0	0	0	0	0	ICE*Aer*Peru1

Fig. 5.5*. South America Map.Regions where ICEs were detected are indicated and ICE profiles found at each site are shown.*

Brazil and Chile

ICEs were detected in three (2%) *V. cholerae* non-O1 non-O139, two strains of which had been isolated from sewage in the Sao Paulo region and one from marine sediment in Brazil. Furthermore, ICEs were found in two (40%) *V. cholerae* isolated from a river in Chile in 1992 and one (100%) *V. parahaemolyticus* from Chile (Table

5.3, Fig. 5.5). All ICEs belonged to the SXT/R391 family as confirmed by presence of the SXT integrase and insertion at the *prfC* locus. Three genetic profiles were identified and named ICE*VchBra*1, ICE*Vch*Chi1 and ICE*Vpar*Chi1. All were positive when investigated for the conserved regions and carried unknown insertions at HS1, HS2 and HS3. The VR2 insertion was characterized by presence of the typical SXT insertion in the two Chilean *V. cholerae* and by a new insertion (350 bp amplicon) in the ICE*Vpar*Chi1 found in *V. parahamolyticus.* The latter also appeared to carry the resistance cluster, inserted at the *rumAB* locus found in SXT-related ICEs described in *V. cholerae* and never observed in *V. parahamolyticus* (Table 5.3, Fig. 5.6 b,c).

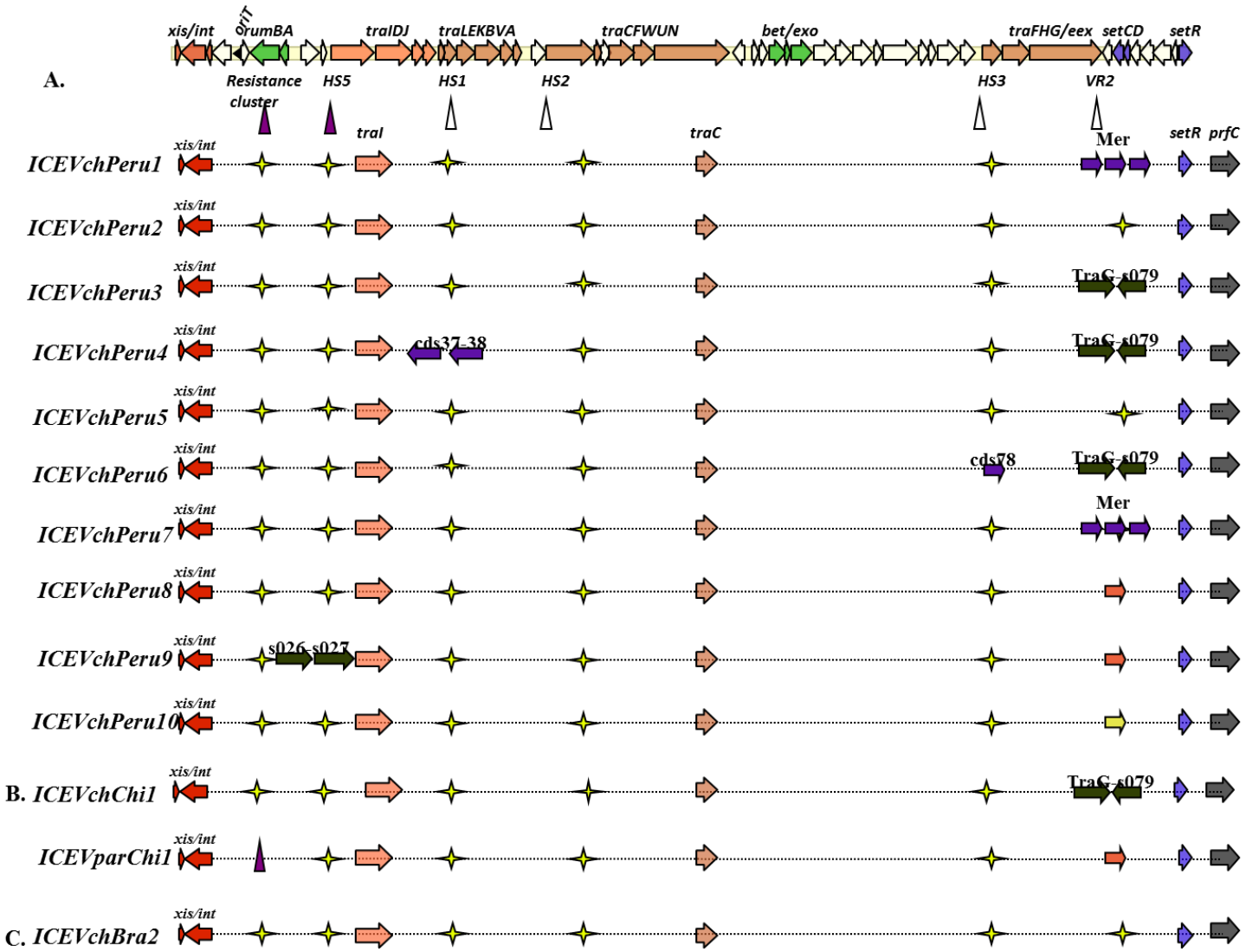


Fig. 5.6.*Genomic profiles of ICE discovered in V. cholerae isolated from Peru, Chile, and Brazil**. A. schematic representation of ICE conserved backbone. Content of each hotspots is shown. Green and purple arrows indicate SXT-like and R391-like insertions.*

Mexico

Mexico is the country where the first and only new world SXT-related ICE has been characterized. The element was detected in an environmental *V. cholerae* strain isolated from sewage in the Louis Potosi' State in 2002 and has been designated ICE*Vch*Mex1 (Burrus *et al.*, 2006).

Of the 101 *V. cholerae* strains from Mexico, 11% (n=11) were positive for the SXT integrase and *prfC* insertion site (Table 5.3, Fig. 5.7b). Seven *V. cholerae* O1 El Tor were isolated from cholera patients between 1991 and 1995 from the Tampico, Veracruz, Yucatan, and Mexico City provinces (Table 5.3). Interestingly, ICE was found in the *V. cholerae* O1 El Tor isolated from the first cholera case in Mexico in 1991. Among the *V. cholerae* O1 El Tor, three ICE profiles were detected, ICE*Vch*Mex2, ICE*Vch*Mex3 and ICE*Vch*Mex4, with the first two positive for the conserved regions investigated and not showing amplification for any of the inserted sequences tested (Table 5.3, Fig. 5.7b). Further, the ICE*Vch*Mex2 and ICE*Vch*Mex3 elements apparently are negative for intact *rumBA* locus, suggesting presence of the resistance cluster. When assayed for the typical antibiotic resistance genes found in this region, *strA*, *strB*, *dfrA18*, *floR* and *tetA* were negative, perhaps for two reasons, namely that the *rumAB* gene was not detected by PCR due to primer annealing problems or because of an insertion not containing the genes(Table 5.3, Fig. 5.7b). Interestingly, ICE*Vch*Mex3 showed amplification for the *sul2* gene, resistance to sulfamethoxazole, but generated a band of about 800bp instead of the expected 625 bp of the control. The three strains carrying ICE*Vch*Mex3 were isolated from different

sources, two clinical and one environmental, and at different times, 1991 and 1995 (Table 5.3). Further work is needed to determine if the gene is located in the ICE or elsewhere in the chromosome.

ICEs were also detected in four *V. cholerae* non-O1/O139 of environmental origin isolated from Chiapas, Veracruz, Yucatan, and Colima States (Table 5.3, Fig. 5.7a). Three different ICE genetic profiles were identified and named ICE*Vch*Mex4, ICE*Vch*Mex5 and ICE*Vch*Mex6. Interestingly, ICE*Vch*Mex6 found in a *V. cholerae* nonO1/O139 isolate from Colima, carries the mercury resistance cluster inserted at the *traG* locus (VR2). The same element was characterized by genetic rearrangement in the conserved backbone since none of the three conserved regions amplified with the primer pairs that were used (Table 5.3, Fig. 5.7b).

***Fig. 6.7. A**. Mexico Map. Sampling sites and ICE profiles found at each site are shown.*

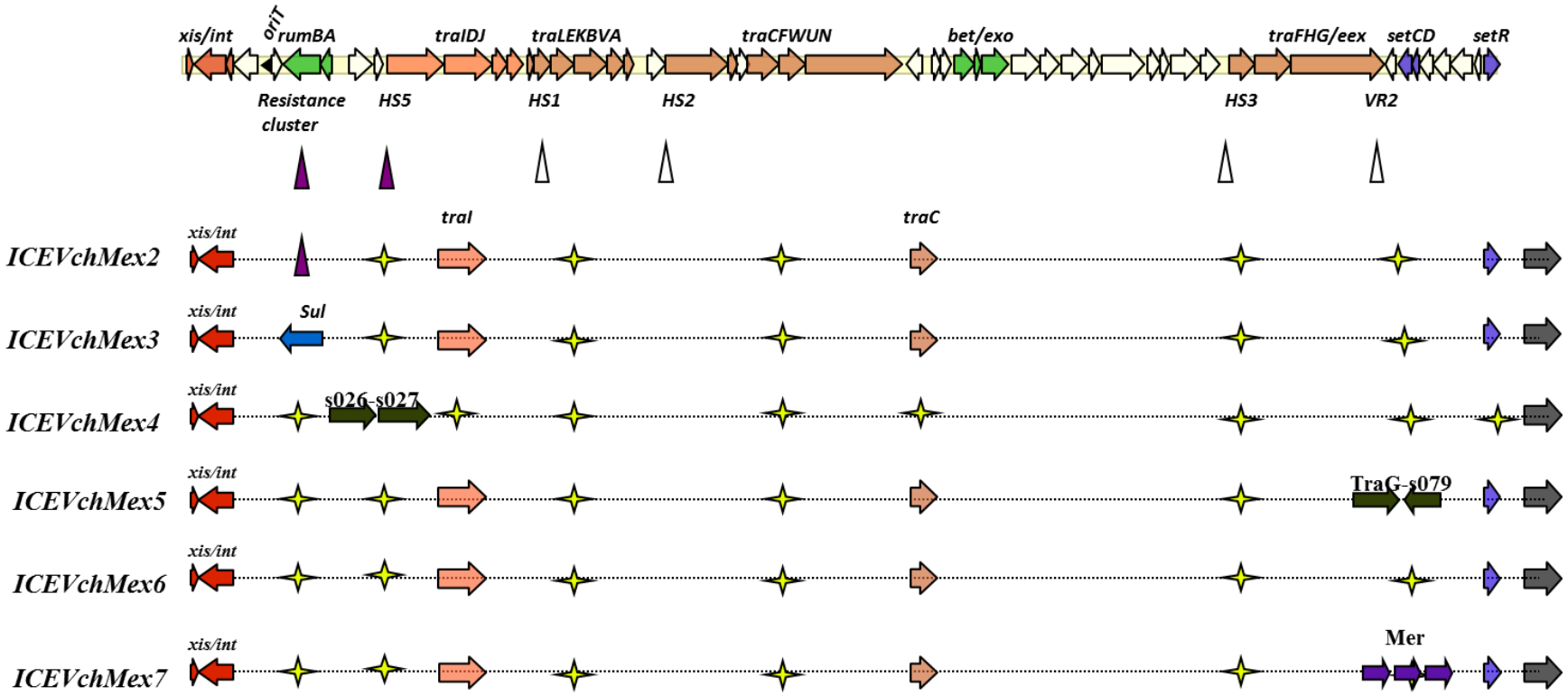


Fig. 5.7. B. *Genomic profiles of ICE discovered in Mexican V. cholerae isolates.*

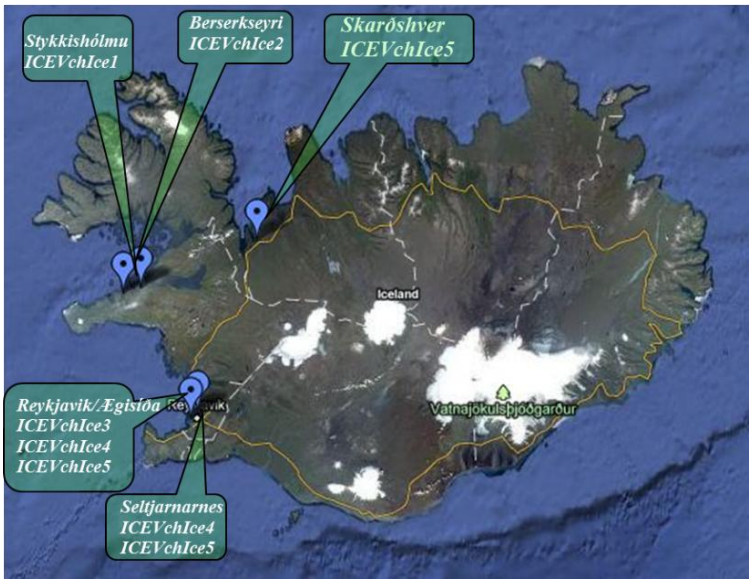


Fig. 5.8.A. *Iceland Map. Sampling sites and ICE profiles found at each site.*

Iceland

ICEs were detected in 88 of 200 (44%) *V. cholerae* non-O1 non-O139 strains isolated in Iceland in 2006-2007. The high frequency of ICE isolation observed was surprising, considering that *V. cholerae* had never before been isolated in Iceland and cholera cases had never been reported there (Table 5.4). All of these ICEs were confirmed as members of the SXT/R391 family by amplification of the SXT integrase and insertion at the *prfC* gene in the host chromosome. Results of molecular characterization revealed six different genetic profiles, designated ICE*Vch*Ice1 to ICE*Vch*Ice6 (Table 5.4). As observed for ICEs isolated in other geographical areas, most of the regions investigated gave negative amplification with the primer pairs

used. This is not surprising, since these primers were designed based on nucleotide sequences of the few available ICEs in the database, most of which had been isolated from highly related *V. cholerae* O1 or O139 clinical strains, thus not representative of the high genetic variability of ICEs circulating in the environment. All ICEs characterized in Icelandic *V. cholerae* lacked the resistance cluster (Table 5.4). Most, but not all, were positive for the conserved regions investigated (Table 5.4, Fig 5.8b). Interestingly, ICE*Vch*Ice1, ICE*Vch*Ice2, and ICE*Vch*Ice3 carried in the HS2 the typical insertion found in SXT. This cluster includes two genes, *mosT* and *mosT*, which encode for a toxin-antitoxin system that promotes maintenance of the genetic element by severely impairing growth of cells that have lost the element (Table 5.4, Fig 5.8b).

***Table 5.4**: ICEs detected in V. cholerae isolated from various samples collected in Iceland*

Site	*Source*			*rum*					*VR1*	*HS2*	*HS1*	*HS3*	*VR2*		*#*
			int	*AB*	*prfC*	*traI*	*traC*	*setR*	*s026*	*traA*	*traL*	*traF*	*traG*	*Profile*	
Ægisíða	Water	1/09/2007	1	1	1	1	1	1	1	536	0	0	601	ICE*Vch*Ice2	1
Ægisíða	Mussel	1/09/2007	1	1	1	0	0	0	0	536	0	0	0	ICE*Vch*Ice3	25
Ægisiða	Water	2/09/2007	1	1	1	0	0	0	0	0	0	0	0	ICE*Vch*Ice4	23
Ægisiða	Water	2/09/2007	1	1	1	1	1	1	0	0	599	0	0	ICE*Vch*Ice5	2
Berserkseyri	Water	12/08/2006	1	1	1	1	1	1	1	536	0	0	601	ICE*Vch*Ice2	11
Seltjarnarnes	Kelp	1/09/2007	1	1	1	0	0	0	0	536	0	0	0	ICE*Vch*Ice3	4
Seltjarnarnes	Water	1/09/2007	1	1	1	0	0	0	0	0	0	0	0	ICE*Vch*Ice4	11
Seltjarnarnes	Water	2/09/2007	1	1	1	1	1	1	0	0	599	0	0	ICE*Vch*Ice5	6
Skarðshver	Water	9/24/2008	1	1	1	1	1	1	0	0	599	0	0	ICE*Vch*Ice5	1

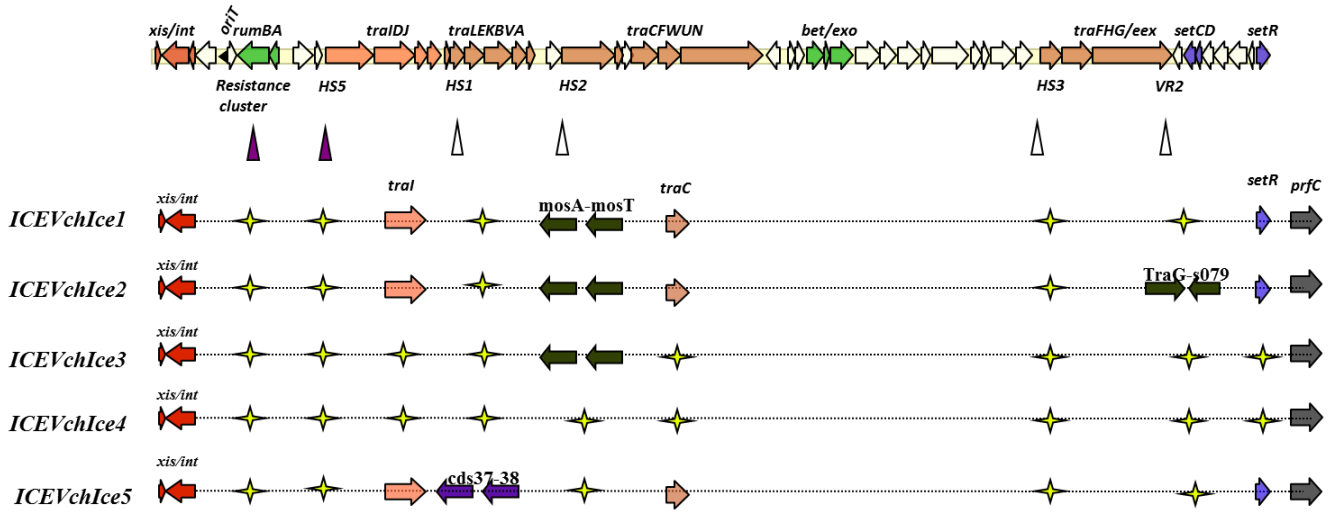


***Fig. 5.8.B.** Genomic profiles of ICE discovered in V. cholerae isolates from Iceland*

Republic of Georgia

Among the 500 *V. cholerae* and other *Vibrio spp.* strains isolated during an environmental surveillance carried out in Republic of Georgia between 2004-2007, eleven *V. cholerae* non-O1/non-O139 (2%) contained ICEs, confirmed by amplification of the SXT integrase and insertion in the *prfC* locus. All positive strains were isolated between September, 2006, and October, 2007, from lake water and seawater samples. Molecular characterization of these ICEs did not provide sufficient information since all of the conserved and variable loci assayed did not give any amplification. None of the Georgian ICEs were profiled and referred to them as a generic ICE*Vch*Geo1. Further genomics studies are needed to determine the genetic structure and content of these highly variable ICEs.

Conclusions

Here we report the detection of integrative conjugative elements of the SXT/R391 family in a large collection of *V. cholerae* strains and isolates from other related species. Our results revealed the presence of this class of genetic elements in geographical areas where they had never been discovered to date. The screening of a large collection of *V. cholerae* belonging to different serougroups and of clinical and environmental origin, as well as other related species isolated, revealed ICEs widely distributed among *V. cholerae* in the American continent. The element was detected for the first time in two *V. cholerae* non-O1/non-O139 isolated in 1976 and 1977,

from water samples in the Chesapeake Bay, US, long before its first appearance in this species in India. Related ICEs were, then, found in strains isolated from the same environment, twenty years later, suggesting a cryptic ecological role associated to these elements that confer positive selection to *V. cholerae* strains to maintain ICE as part of their genetic asset.

The American continent has been free from cholera for more than a century, until an outbreak hit Peru in 1991 (Tauxe *et al.*, 1994) reintroduced the disease in this continent. Since then, the ongoing cholera epidemic in Latin and Central America was associated with two *V. cholerae* O1 El Tor strains that coexisted in this region. It is thought that the epidemic was characterized by two waves, caused by two El Tor strains. The first wave in Latin America was caused by *V. cholerae* O1 El Tor Inaba related to pre-O139 Asian *V. cholerae* El Tor and grouped in a ribotype that clustered separately from other seventh pandemic *V. cholerae* El Tor clones. This *V. cholerae* O1 El Tor Inaba clone apparently was replaced by a *V. cholerae* O1 El Tor Ogawa related to the Asian seventh pandemic *V. cholerae* O1 El Tor and *V. cholerae* O139 (Beltran *et al.*, 1999). Based on results of our analysis, and consistent with those of Burrus *et al.* (2006), it is concluded that ICEs isolated in America had an independent origin and evolution with respect to the ICEs circulating in Asia and Africa. As demonstrated by the presence of ICEs in epidemic *V. cholerae* O1 El Tor in Mexico, including the first case, which appeared to be genetically diverse from any of the other SXT-related elements found in African and Asian epidemic *V. cholerae* strains, these elements most likely had been evolutionarily separated from their cognate in the Eastern hemisphere. Furthermore, the abundance of *V. cholerae* strains carrying ICEs

that have been detected in water around both Northern (Chesapeake Bay) and Southern (Mexico, Peru, Brazil, Chile) regions of this continent, indicates that these mobile elements are present in the environment where they constitute a genetic reservoir for the assembly of diverse genetic profiles.

A remarkable result from the survey reported here was the detection of ICEs in 44% of of *V. cholerae* non-O1/non-O139 isolated in Iceland, where cholera has never been recorded and the organism has never been isolated from a cholera victim (B. Haley in preparation). In Iceland, *V. cholerae* was isolated from water, mussels, and kelp year around, in coastal water characterized by elevated surface water temperatures associated by geothermical activity and hot-water overflow from urban and industrial geothermal operations. Such elevated temperatures occur also during the winter months. The presence of ICEs and virulence factors, such as *toxR, ompU, hlyA, luxO*, *rtxA,* and *hapA,* in a region geographically isolated, where cholera has not yet been documented, support the hypothesis that these elements have a primary role in the ecology of microorganisms, such as *V. cholerae,* in the environment. In fact, different functions encoded by these genetic elements have previously been associated with cellular mechanisms responsible for an improved environmental fitness. For example, restriction modification systems appear to be involved in the resistance to phages, which are one of the main factors influencing trends in bacterial populations, or cellular messangers of pathways regulating motility and biofilm formation, are likely to give selective advantages to microorganisms in the environment.

ICEs of the SXT/R391 family were sporadically detected in *V. cholerae* strains from Georgia. Only 2 % of the strains carried this class of mobile element and all had

high genetic variability in the loci that were studied. The low incidence of ICE in Georgia is surprising, considering that *V. cholerae* non O1/non-O139 are abundant and readily detected in surface water in this region, which is geographically closer to Asian countries where ICEs are commonly found in *V. cholerae*. This result further confirms the hypothesis that *V. cholerae* is a natural inhabitant of the aquatic environment and that local populations evolve separately, continuously adjusting to a changing environment via the plasticity of its genome. ICEs are important genetic tools contributing fitness in the population.

Interestingly, from analysis of the conserved and variable regions of the ICEs that were discovered, it is concluded that all profiles indicate genetic hybrids, in which sequences belonging to R391, SXT, and unknown genetic regions are present in the same element. These results support the hypothesis that the SXT/R391 family of ICEs is widely distributed in environmental bacterial populations. Even though they share a common backbone, they are of high genetic plasticity and variable genetic content that most likely contributes to a genomic evolution that is responsible for successful adaptation of the bacterial host to new ecological niches.

Presentation

Abstract of the poster presented at the "*Vibrio* in the Environment 2010" conference held in Biloxi, MS, November 7-12, 2010. (Judged one of the top ten poster presentations at the Conference and receiving an award).

Taviani E., Chen A., Haley B., Hasan N. A., Huq A. and. Colwell R. R. Geographical distribution and molecular characterization of Integrative Conjugative Elements in environmental *Vibrio cholerae*. *Vibrio* in the Environment 2010, Biloxi, MS, November 7-12, 2010.

Abstract of the poster selected for presentation at the ASM General Meeting (2009) in Philadelphia, PA, May, 2010.

Taviani E., Grim C., Huq A., Colwell RR. Discovery of ICEs circulating in *Vibrio cholerae* population in the Chesapeake Bay: new genetic elements in the New World. Amer. Soc. Microbiol. General Meeting. Philadelphia, PA.

Chapter 6. Discovery of novel *Vibrio cholerae* VSP I and VSP II genomic islands using comparative genomic analysis.

Introduction

Vibrio cholerae, an autochthonous aquatic bacterium, is the causative agent of cholera, a severe, watery, life-threatening diarrheal disease (Colwell, 1996). Cholera bacteria are serogrouped based on the variable somatic O antigen, with more than 200 serogroups identified (Chatterjee *et al.*, 2003). Although strains of most serogroups of *V. cholerae* are capable of causing a mild gastroenteritis and sporadic local outbreaks of cholera, only toxigenic strains of *V. cholerae* O1 and O139 have been linked to epidemics and pandemics. Genes encoding cholera toxin, *ctxAB*, and other pathogenic factors have been shown to reside in various mobile genetic elements.

The epidemic potential of cholera has been a threat throughout human history with seven recorded pandemics and the disease is persistent in many developing countries. Isolates from the sixth pandemic are almost exclusively of the Classical biotype. However, the seventh, current pandemic has been dominated by *V. cholerae* O1 El Tor (Kaper *et al.*, 1995). Isolates of all previous pandemics originated in the Indian subcontinent, whereas those associated with the seventh pandemic have their origin in the Indonesian island of Sulawesi, with subsequent isolation from Asia, Africa, and Latin America. In 1992, a new serogroup, *V. cholerae* O139, was identified as the

cause of massive cholera outbreaks, first in India and then in Bangladesh (Ramamurthy *et al.*, 1993).

Two gene clusters associated with seventh pandemic strains were identified by comparative genomics using microarray analysis and named *Vibrio* Seventh Pandemic (VSP) I and II (Dziejman *et al*., 2002). These clusters were absent in Classical and pre-pandemic *V. cholerae* El Tor strains and showed an unusual G+C content (40%), compared with the entire *V. cholerae* genome (47%) (Dziejman *et al.*, 2002).

VSP-I is a 16-kb region inserted in chromosome I between ORFs VC0174 to VC0186 and comprising 11 genes, VC0175–0185. Recently, VSP-I has been reported in *V. cholerae* non-O1/non-O139 (Chatterjee *et al*., 2009, O'Shea *et al*., 2004).

VSP-II was originally identified as a 7.5-kb island, spanning genes VC0490 to VC0497 in *V. cholerae* O1 El Tor N16961 (Dziejman *et al.*, 2002), and, subsequently, found to include a larger 26.9-kb region, spanning from VC0490 to VC0516 (O'Shea *et al.*, 2004). Its site of integration is a tRNA-methionine locus, VC0516.1. As described in *V. cholerae* O1 El Tor N16961, VSP-II encodes type IV pilin, two methyl-accepting chemotaxis proteins, an AraC-like transcriptional regulator, a DNA repair protein and a P4-like integrase (VC0516) at the 3' end of the island. Murphy and Boyd (Murphy and Boyd 2008) found that both VSP-I and VSP-II are able to excise from the chromosome, forming an extra-chromosomal circular intermediate through a site-specific recombination mediated by the integrase encoded in the island.

To date, two variants of VSP-II have been described in the literature, one in a *V. cholerae* non-O1 strain from Bangladesh and one in a *V. cholerae* O1 El Tor strain isolated in Peru during 1991-2003; moreover the cluster was detected in several *V.*

cholerae non-O1 non-O139 strains (Dziejman *et al.*, 2002; Dziejman *et al.*, 2005; Nusrin *et al.*, 2009).

In this study, comparative genomic analysis was employed to determine the presence and the genetic composition of VSP-I and VSP-II islands among 23 strains of *V. cholerae*. Polymerase chain reaction (PCR) primers designed to identify seventh pandemic VSP-I and the newly sequenced environmental variant were used to survey clinical and environmental isolates of *V.cholerae*. Here, we report the sequence of a VSP-I variant from a bioluminescent non-O1/non-O139 isolate of *V.cholerae*, VL426. Further, in our analysis, we re-annotated the VSP-II present in *V. cholerae* O1 El Tor N16961 and analyzed the VSP-II previously described in *V. cholerae* O37 MZO-3 (Dziejman *et al.*, 2005). Three new variants with significant genetic polymorphisms were discovered and their distribution among a large *V. cholerae* collection was assessed.

From this study it is concluded that *Vibrio* seventh Pandemic Islands I and II are not as conserved as has been reported and that VSP-II can be considered a molecular tag in epidemic *V. cholerae*.

Aim of the research

Characterize and describe new genomic islands in *V. cholerae* through comparative genomics and their distribution among different strains by genetic screening of the laboratory strains of *V. cholerae* originated from a diversified geographical locations.

Materials andMethods

Strains and media

Twenty-three *V. cholera* strains were included in a comparative genomics analysis and screened for VSP-I and VSP-II (Table 6.1), along with 188 well characterized laboratory collection strains and 190 *V. cholerae* isolates from two cholera-endemic regions of Bangladesh. All cultures were grown in Luria-Bertani medium (Difco/BD, Sparks, MD) and stored at -80°C in LB broth amended with 25% glycerol.

Comparative genomics

All the strains in this study were fully sequenced and genomic comparison was carried out and published by Chun *et al*. (2009). New VSP-I and II variants were discovered and annotated by RAST and their genetic organization analyzed and compared using MUMmer (Delcher *et al.*, 1999) and Artemis Comparative Tool (ACT) (Carver *et al.*, 2005). Individual gene polymorphisms were analyzed by ClustalX alignments and homology was attributed after BLASTN search in the non-redundant database (Larkin *et al*., 2007).

Strain	*Serogroup*	*Biotype*	*Geographical origin*	*Source of isolation*	*Year of isolation*	*Accession*
N16961	O1 Inaba	El Tor	Bangladesh	Clinical	1975	NC_002505/NC_002506
RC9	O1 Ogawa	El Tor	Kenya	Clinical	1985	ACHX00000000
MJ-1236	O1 Inaba	El Tor	Matlab, Bangladesh	Clinical	1994	CP001485/CP001486
B33	O1 Ogawa	El Tor	Beira, Mozambique	Clinical	2004	ACHZ00000000
MO10	O139		Madras,India	Clinical	1992	AAKF03000000
CIRS101	O1 Inaba	El Tor	Dhaka, Bangladesh	Clinical	2002	ACVW00000000
2740-80	O1 Inaba	El Tor	US Gulf Coast	Water	1980	NZ_AAUT01000000
BX330286	O1 Inaba	El Tor	Australia	Water	1986	ACIA00000000
MAK757	O1 Ogawa	El Tor	Celebes Islands	Clinical	1937	NZ_AAUS00000000
NCTC 8457	O1 Inaba	El Tor	Saudi Arabia	Clinical	1910	NZ_AAWD01000000
O395	O1 Ogawa	classical	India	Clinical	1965	NC_009456/NC_009457
V52	O37		Sudan	Clinical	1968	AAKJ02000000
12129(1)	O1 Inaba	El Tor	Australia	Water	1985	ACFQ01000000
TM 11079-80	O1 Ogawa	El Tor	Brazil	Sewage	1980	ACHW00000000
VL426	non-O1/O139	albensis	Maidstone, Kent, UK	Water	Unknown	ACHV00000000
TMA21	non-O1/O139		Brazil	Seawater	1982	ACHY00000000
1587	O12		Lima, Peru	Clinical	1994	NZ_AAUR01000000
RC385	O135		Chesapeake Bay	Plankton	1998	NZ_AAKH02000000
MZO-2	O14		Bangladesh	Clinical	2001	NZ_AAWF01000000
V51	O141		USA	Clinical	1987	NZ_AAKI02000000
MZO-3	O37		Bangladesh	Clinical	2001	NZ_AAUU01000000
AM-19226	O39		Bangladesh	Clinical	2001	NZ_AATY01000000
623–39	non-O1/O139		Bangladesh	Water	2002	NZ_AAWG00000000

Table 6.1. *Vibrio cholerae* strains used for the comparative genomics analysis conducted in this study.

Primer design and PCR analysis

PCR primers for group-specific target regions, identified by multiple allignement of VSP-I and II sequences using ClustalX2 software (Larkin *et al*., 2007), were designed using FastPCR Molecular Biology Software (Kalendar *et al*., 2009). PCR primers are listed in Table 6.2. PCR was performed using designed primers on 378 isolates of *V. cholerae* to determine the presence of the two VSP-I targets, a seventh pandemic group-specific marker, VC2346, annotated as ‘‘SMP-like protein,’’ and chromosomal insertion in both chromosomes. PCR was done using primers listed in Table 6.2, the same isolates of *V. cholera* for the five VSP-II variants.

VSP-I variants							
Primer	Sequence	Target			Amplicon size (bp)		References
dcd-820F	gcttattcagcagccctcaggtc	VC0175			584		This study
dcd-1403R	tagcgtcgagatgacacaccttcg						This study
VSPI-F	gccgagaactctaaagcgcttctc	VC0180-0181			331		This study
VSPI-R	ccaaggtacagatgagtaccagca						This study
smp-F	gcaactgtgctagcagttgccgtg	VC2346			405		This study
smp-R	gcctgtttcaaacgtgatgcgta						This study
VC0174-F	aaactggcgacctttgagcaagc	Chromosome I insertion site			1321		This study
VC0186-R	gatggtagcctgacgctgcatctg						This study
VCA0695-F	atagcgggagttggctctgca	Chromosome II insertion site			957		This study
VCA0697-R	ggtgacttggtgcccatcgta						This study
		VSP-II variant amplicon length (bp)					
		N16961	MZO3	TMA21	CIRS101	RC385	No island	References
pVSP2-IF	cctgtcatgttatgaggtgca	361	678	-	-	-		This study
pVSP2-IR	caggtctcttatcggctttgc							This study
pVSP2-IIF	acaacttgtaagatagccttgc	570	570	570	570	-		This study
pVSP2-IIR	gcaagacaaaactacagcttgc							This study
pVSP2-IIIF	agcaaacggtcattcgct	451	451	451	-	451		This study
pVSP2-IIIR	gttggaaggtgggttgtgt							This study
p489F	gatcaactacgatcaagcc	-	-	-	-	-	3532	O'Shea et al
p517R	cagtcacagcttaaac							O'Shea et al

Table 6.2*. Primer design and PCR conditions. Conserved and group-specific regions of VSP-II were identified by examining aligned and unaligned sequences, using ClustalX.*

Results and discussion

VSP-I: Genomic analysis

Consistent with previous studies, all strains in the seventh panedemic (7P) monophyletic group (Chun *et a.,l* 2009), showed the presence of VSP-I island inserted between genes VC0174 and VC0186 (Fig. 6.1) (Grim *et al.,* 2010). Interestingly, *V.cholerae* O1 MJ-1236 harbored a second copy of VSP-I that was detected between VCD_000620 and VCD_000630 genes on the smaller chromosome, and including 9 ORFs (Grim *et al.,* 2010). The insertion site of this second copy of the seventh pandemic VSP-I correspond to *V cholerae* N16961 genes VCA0095 and VCA0096; except for MJ1236, no other insertions were found at this site in any other 7P group strains. Despite RAST failed to annotate three ORFs in the smaller chromosome VSP-I in *V.cholerae* MJ-1236, based on our comparative analysis, the island is actually highly conserved. Only three nucleotide differences were found between the VSP-I islands of *V.cholerae* O1 El Tor N16961 and *V.cholerae* O1 MJ-1236 (Table 6.3) (Grim *et al.,* 2010).

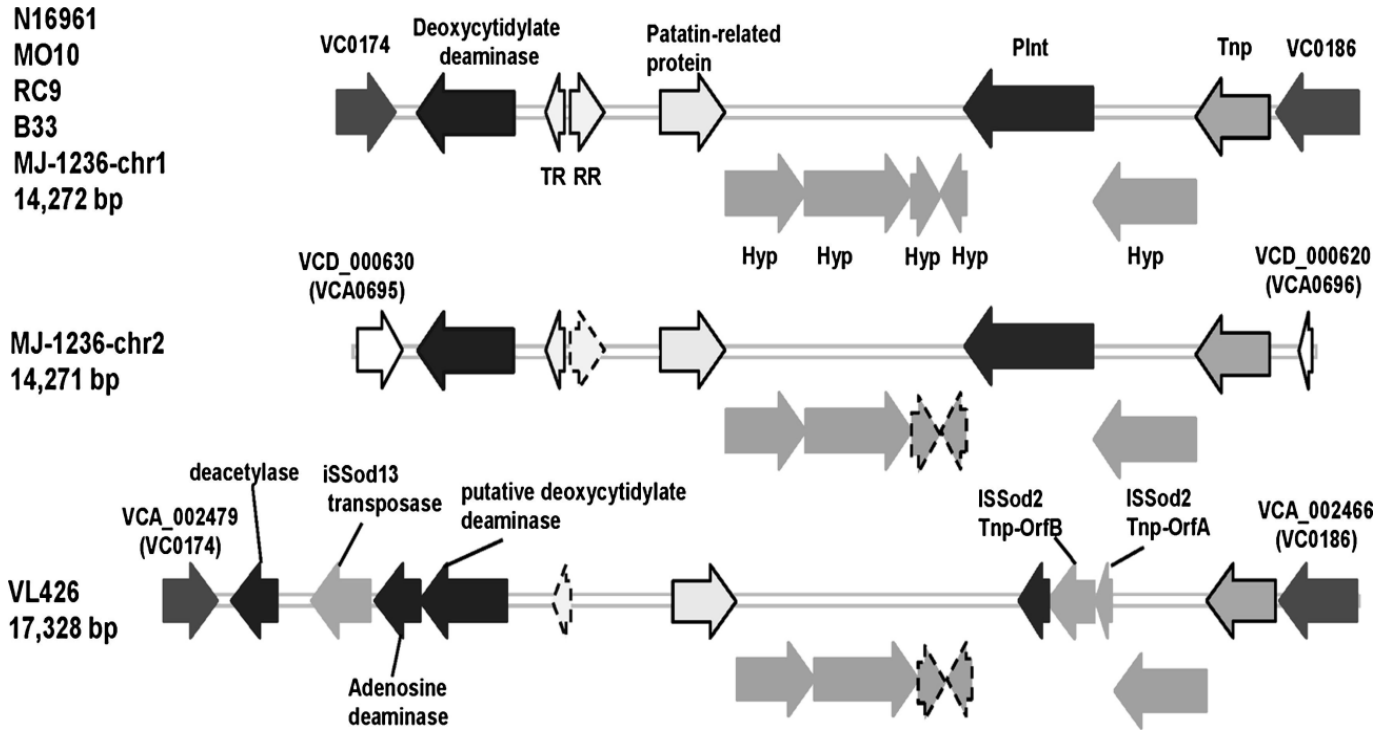


Fig. 6.1.*Schematic representation of the seventh pandemic VSP-I and VL426 variant. Flanking genes are indicated by their O1 El Tor N16961 VC locus. VSP-I island genes are indicated by their annotated function. Genes outlined with dashed lines indicate those not present in deposited annotation in GenBank, but which were found by sequence comparison with that of V.cholerae O1 El Tor N16961. TR, transcriptional regulator; RR, response regulator; Hyp, hyptothetical protein; PInt, phage integrase; Tnp, transposase .*

Identification of the VSP-I variant

Comparative genomics using 23 complete and draft genomes of *V.cholerae*, revealed a 17,328-bp genomic island in the genome of *V.cholerae* non-O1/O139 VL426 inserted at the same locus on chromosome 1 (Fig. 6.1). The island showed the same genetic organization and an high sequence similarity to the seventh pandemic VSP-I (Grim *et al* 2010). The island in VL426 encompasses genes VCA_002467 to VCA_002478, 9 of the 11 genes of the seventh pandemic VSP-I have homologues in VL426 VSP-I variant, while the other two genes have been disrupted by insertion of transposons in two regions (Fig. 6.1 and Table 6.3), making the new variant 3 kb larger than the seventh pandemic VSP-I. Nucleotide–nucleotide alignment revealed an overall highly conserved organization interrupted by three major regions of sequences discontinuity. These are located at the VC0175 gene, which in the VL426 VSP-I variant is interrupted by the insertions of a transposase; in the region between genes VC0177 and VC0178, where the response regulator seems to be deleted in the new variant; and, finally, at the VC0183 gene, also interrupted by the insertion of two transposases in the VL426 VSP-I variant (Fig. 6.1 and Table 6.3) (Grime *et al.*, 2010). Interestingly, the VC0183 gene encodes the phage integrase which supposedly should be highly conserved.

VSP-I chromosomal insertion sites

In addition to the seven strains, shown in Table 6.3, with a VSP-I variant at the insertion site, VC0174-0186, on the large chromosome, three others had alternate genomic islands. *V.cholerae* 623-39 contains an insert for which one gene was identified and annotated as ‘‘Chromosome segregation ATPase’’

(NZ_AAWG01000125, nt 8300 to 7857). Similarily, *V.cholerae* RC385 also contained a small genomic island at the chromosome I insertion site, approximately 3.7 kb in size, with a single 1,536-bp annotated gene, "deoxycytidylate deaminase-related protein". This gene had low sequence identity, 67%, with gene VC0175 of *V.cholerae* O1 El Tor N16961, although the annotation is similar. *V.cholerae* MZO-3 contained a putative prophage at the insertion site on the large chromosome. In addition to *V.cholerae* MJ-1236, several *V. cholerae* had VSP-I or another genomic island inserted into both chromosomes. Six clinical seventh pandemic *V.cholerae* O1 and O139 isolates had VSP-I in both chromosomes and several *V.cholerae* non-O1/O139 contained an insert in both chromosomes (Grim *et al*., 2010).

VSP-II Genomic analysis

From RAST annotation, the 26.9 Kb VSP-II found in the *V. cholerae* N16961 encompasses 30 ORFs, compared with 24 ORFs previously annotated (O'Shea *et al.*, 2004). Specifically, six putative transposases were newly annotated by RAST (Fig. 6.2).

Results of comparative genomics, using 23 complete and draft genomes of *V. cholerae* and the *V. cholerae* O1 El Tor N16961 VSP-II sequence as reference, revealed the presence of a VSP-II island with 99% nucleotide sequence similarity in four of the *V. cholerae* seventh pandemic strains: *V. cholerae* O1 El Tor B33; *V. cholerae* O1 El Tor MJ-1236; *V. cholerae* O139 MO10; and *V. cholerae* O1 El Tor RC9 (Fig. 6.2). Results of a phylogenetic analysis of the 23 *V. cholerae* studied showed that these five strains formed a monophyletic clade, termed the seventh phylopandemic clade (Chun *et al.*, 2009).

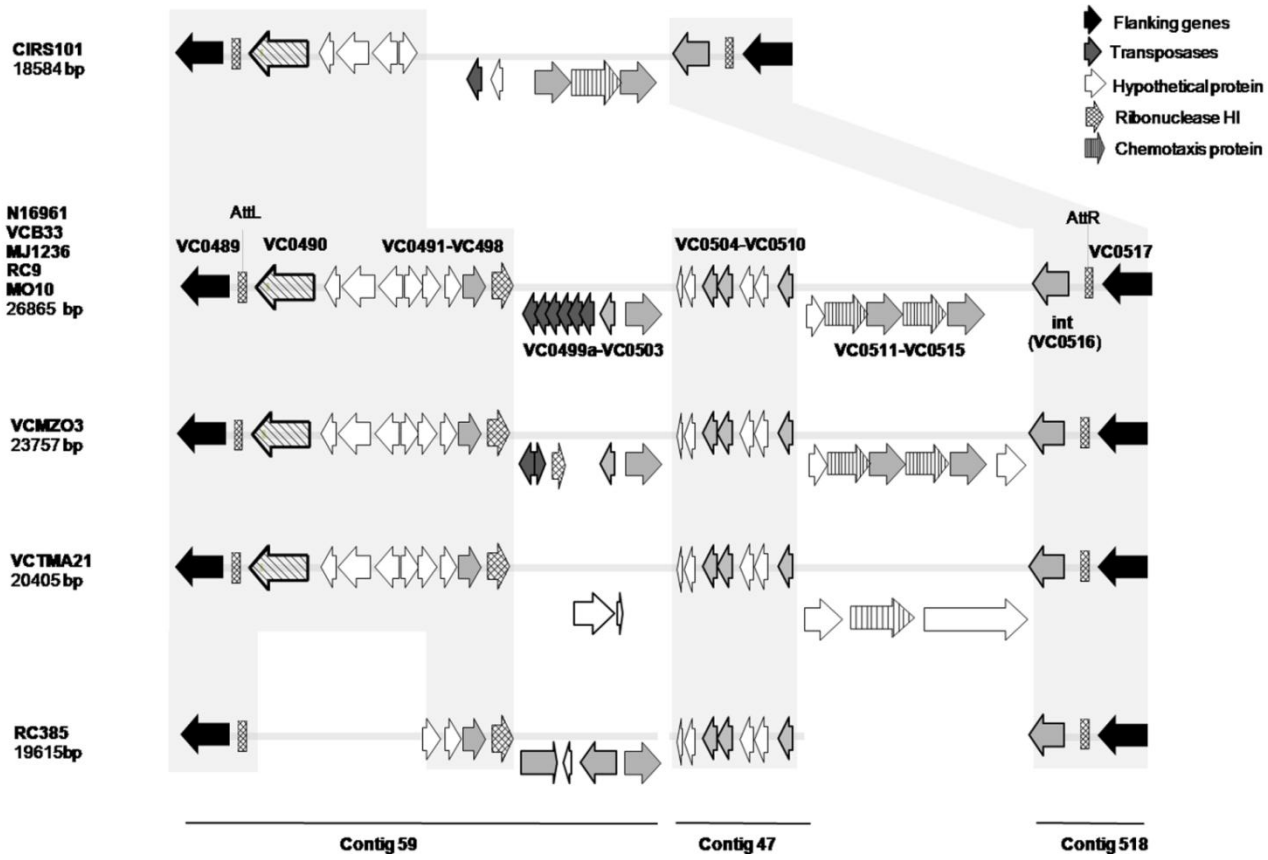


Fig. 6.2 *.**Genetic organization of the five variants of VSP-II in V. cholerae**. The direction of transcription of the ORFs is indicated by direction of the arrows. The numbers refer to the genetic organization of genes along the genome of V. cholerae N16961 (O'Shea et al., 2004). Genes are pattern-coded, according to function. Homologous regions are indicated by grey shadow. For V. cholerae RC385, the contigs number where the VSP-II island sequence resides are indicated.*

***Table 6.4**: Nucleotide-nucleotide comparison between the seventh pandemic VSP-II island and other variants described in this study, including flanking genes. Shaded regions indicate absent ORFs. 7P: Prototypical seventh pandemic VSP-II.*

N16961				7P	CIRS101	TMA21	MZO3	RC385
locus		**size**	**gene**	**% similarity**				
VC0489		1761,-	Putative hemolysin	100	100	99	82	99
VC0490		1962,-	Plasmid-related protein	100	100	99	100	n/a
VC0491		537,-	hypothetical protein	100	100	100	100	n/a
VC0492		1167,-	hypothetical protein	100	100	99	100	n/a
VC0493		876,-	hypothetical protein	100	100	99	99	n/a
VC0494		645,+	hypothetical protein	100	100	95	99	55
VC0495		675,+	hypothetical protein	100	55	92	92	92
VC0496		600,+	hypothetical protein	100	n/a	98	99	98
VC0497		201,+	Transcriptional regulator	100	n/a	98	99	99
VC0498		441,+	Ribonuclease HI	100	n/a	97	76	97
VC0499a		207,-	transposase	100	78	n/a	78	n/a
VC0499b		600,-	transposase	100	52	n/a	80	n/a
VC0500a		147,-	transposase	100	79	n/a	91	n/a
VC0500b		153,-	Transposase	100	n/a	n/a	85	n/a
VC0501a		489,-	Transposase	100	63	99	99	56
VC0501b		450,-	Transposase	100	72	98	99	n/a
VC0502		525,-	type IV pilin, putative	100	n/a	n/a	99	n/a
VC0503		1281,+	Cell wall endopeptidase,	100	n/a	n/a	99	n/a
VC0504		228,-	hypothetical protein	100	n/a	89	100	89
VC0505		369,-	hypothetical protein	100	n/a	92	100	94
VC0506		735,-	Transcriptional factor MdcH	100	46	95	99	96
VC0507		177,-	hypothetical protein	100	n/a	92	100	96
VC0508		444,-	hypothetical protein	100	64	94	99	94
VC0509		444,-	hypothetical protein	100	n/a	92	96	90
VC0510		474,-	DNA repair protein RadC	100	59	92	93	93
VC0511		120,+	hypothetical protein	100	n/a	n/a	100	n/a
VC0512		1590,+	chemotaxis protein	100	52	n/a	100	n/a
VC0513		816,+	AraC- protein	100	100	n/a	99	n/a
VC0514		1881,+	chemotaxis protein	100	100	n/a	100	n/a
VC0515		1233,+	EAL domain protein	100	100	n/a	99	n/a
VC0516		1242,-	Phage integrase	100	100	89	98	92
VC0517		1878,-	RNA pol-sigma factor RpoD	100	100	99	99	99

V. cholerae O1 El Tor CIRS101 variant

Interestingly, a sixth strain included in this clade, *V. cholerae* O1 El Tor CIRS101 (Nair *et al.*, 2006), isolated in 2002 in Bangladesh, carried yet another variant of VSP-II (Fig. 6.3). The VSP-II cluster found in *V. cholerae* CIRS101 is 18.5 kb long and 99% similar over the 13 kb homologous region (Figs. 6.2 and 6.3) to the *V. cholerae* N16961 VSP-II, with a 14.4 kb deletion at nt 118 of VC0495, spanning ORFs VC0495 to VC0512 (Fig. 6.3). Inserted downstream of VC0494 in VSP-II of *V. cholerae* CIRS101 is a 1260 nt transposase-encoding ORF (Fig. 6.3). The 3' region of the *V. cholerae* CIRS101 VSP-II island is identical to the prototypical seven pandemic VSP-II (Fig. 6.3).

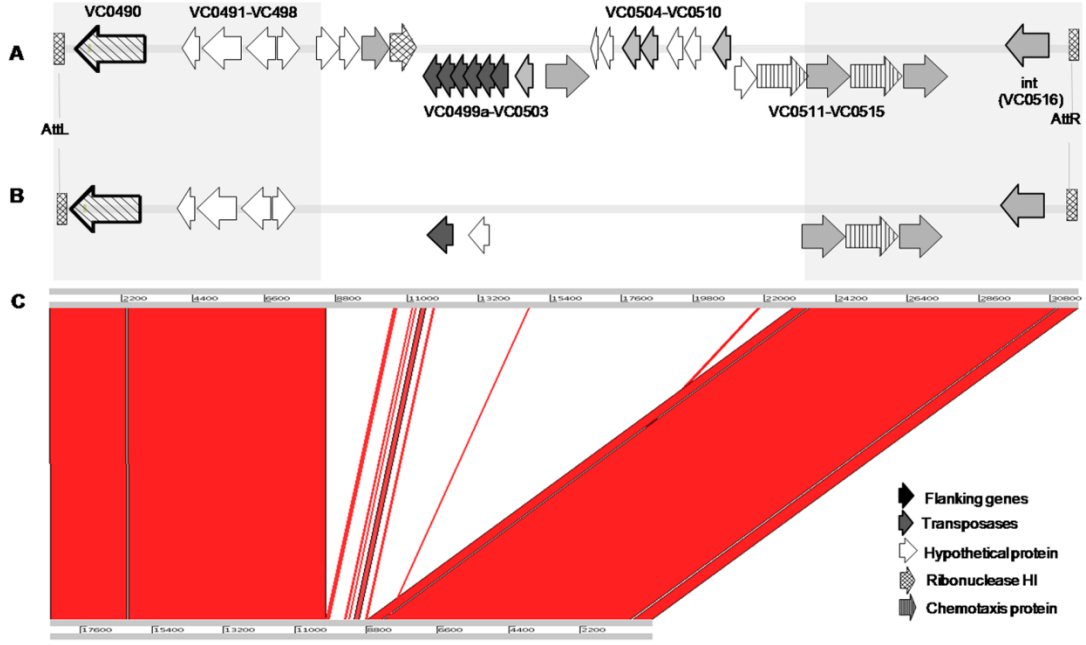


Fig. 6.3*.Comparative analysis of seventh pandemic VSP-II and CIRS101 VSP-II. (A) Schematic representation of seventh pandemic VSP-II (B) Schematic representation of CIRS101 VSP-II; and (C) Alignment generated by Artemis Comparative Tool (ACT) of seventh pandemicl VSP-II and V. cholerae CIRS101 VSP-II.*

V. cholerae O37 MZO-3 variant

VSP-II genes were present in *V. cholerae* strains other than the seven pandemic. As previously reported, *V. cholerae* MZO-3 O37 has a 26.5 kb VSP-II inserted at the same locus as in *V. cholerae* N16961 (Figs. 6.2 and 6.4) (Dziejman *et al.*, 2005). Our analysis and annotation showed this island contained 28 ORFs (Fig.6.4) and regions VC0490-VC0497 and VC0502-VC0516 of the island were 98% similar to the VSP-II in *V. cholerae* N16961 (Table 6.4, Fig.6.4). However, along the island there were two major regions of sequence discontinuity and/or rearrangement (Fig.6.2 and 6.4): two transposases were inserted within the VC0498 gene and a putative transposase is located between the VC0515 gene and the integrase at the 3' end of the island, which has 99% similarity with a putative transposase in *V. cholerae Vibrio* Pathogenicity Island I (VPI-I) (Karaolis *et al.*, 2001). Despite significant sequence similarity, from a phylogenetic point of view, the VSP-II variant found in *V. cholerae* O37 MZO-3 appears to have diverged with respect to the VSP-II evolutionary path (Fig.6.4). All three phylogenetic trees generated using the entire island, three conserved concatenated genes and two flanking genes of the island concluded that *V. cholerae* MZO-3 VSP-II lies outside the VSP-II of the seventh pandemic clade (Fig.6.6).

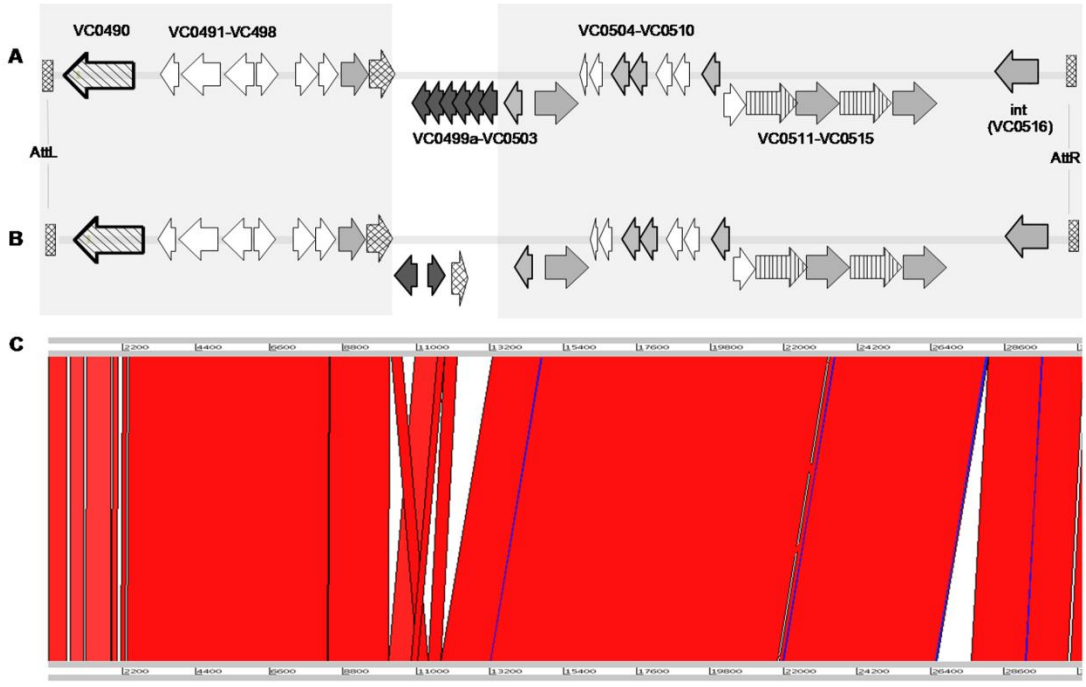


Fig. 6.4.Comparative analysis of seventh pandemic VSP-II and MZO-3 VSP-II*. (A) Schematic representation of seventh pandemic VSP-II (B) Schematic representation of MZO-3 VSP-II; and (C) Alignment generated by Artemis Comparative Tool (ACT) of seventh pandemic VSP-II and V. cholerae MZO-3VSP-II.*

V. cholerae non-O1/non-O139 TMA21 variant

A VSP-II variant was identified in *V. cholerae* non-O1/non-O139 TMA21, isolated from a sewage sample collected in Brazil in 1982 (Table 6.4, Fig. 6.2). The cluster found in this strain is 20.4 kb long, integrated at the same locus and shares 99% sequence similarity over homologous regions with the prototypical seventh pandemic VSP-II island (Fig.6.5). As in the case of the *V. cholerae* MZO-3 variant, significant genetic rearrangement was detected in the region downstream of VC0498 where ORFs

VC0499a-VC0500b and VC0502-VC0503 are deleted. In contrast, at this locus, we annotated two ORFs encoding hypothetical proteins not found in the prototypical seventh pandemic island. These ORFs have 92% and 85% nucleotide sequence similarity with two hypothetical proteins similar to that of *Vibrio vulnificus* YJ016, VV0516-VV0517, in the same arrangement (dbj|BA000036.2|). As reported by O'Shea and colleagues (2004), the 5' region of the prototypical *V. cholerae* VSP-II shows homology to the 5' end of the 43.4 kb *V. vulnificus* island-I (VVI-I), but ORFs VC0499-VC0503 of VSP-II are absent in VVI-I (O'Shea *et al.*, 2004). Therefore, in this region, *V. cholerae* TMA21 VSP-II appears to have an organization identical to VVI-I, i.e., ORFs VC0499-VC0503 are substituted by two hypothetical proteins (Fig. 6.5). Another major genetic rearrangement in *V. cholerae* TMA21 VSP-II was observed downstream of ORF VC0511, which is a deletion encompassing ORFs VC0512 to VC0516 substituted with three ORFs encoding two hypothetical proteins and a nucleotidyltransferase (Table 6.4, Fig. 6.2). Interestingly, the same deletion was noted in the VSP-II variant in *V. cholerae* O1 El Tor strains from Peru (Nusrin *et al.*, 2009). Two of the ORFs present in *V. cholerae* TMA21 VSP-II have 69% sequence similarity with two ORFs encoding hypothetical proteins in *Nitrosomonaseuropaea* ATCC 19718 (emb|AL954747.1|), arranged in the same order. The third ORF did not share significant similarity with any sequence in GenBank.

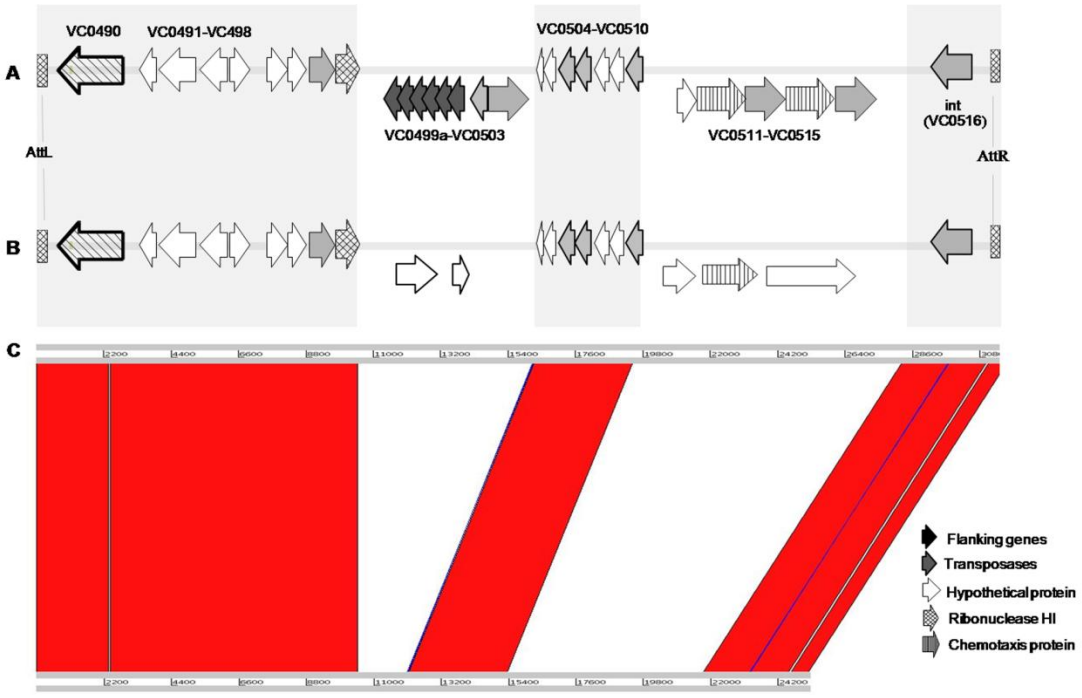


***Fig. 6.5*.*Comparative analysis of seventh pandemic VSP-II and TMA21 VSP-II*.** *(A) Schematic representation of seventh pandemic VSP-II (B) Schematic representation of TMA21 VSP-II; and (C) Alignment generated by Artemis Comparative Tool (ACT) of seventh pandemic VSP-II and V. cholerae TMA21 VSP-II.*

V. cholerae RC385 O135 variant

A fourth variant of the VSP-II island was found in the genome of *V. cholerae* RC385 O135, an isolate from Chesapeake Bay, MD, USA (Fig.6.2). Because of low sequence coverage resulting in a large number of contigs in this draft genome, we were only able to reconstruct the 5' region of the island. VSP-II sequences were present in three contigs: ctg 59; ctg 47; and ctg 518. The 5' region of the island resides on contig 59 and, according to the sequence in this contig, the island is inserted in the same location as

all other VSP-II islands described in this study. The rest of the contig comprises 19,615 bp (Fig.6.2). There is a significant deletion in this region, conserved in the prototypical seventh pandemic VSP-II, *V. cholerae* MZO3 and TMA21 variants. ORFs VC0490 to VC0494 are absent in VSP-II of *V. cholerae* RC385 (Fig.6.2). Furthermore, three new ORFs are inserted after the VC0498 gene, indicating that this locus represents a hot spot for recombinational events within the island. Genes VC0504 to VC0510 and the integrase are conserved, as had been found in the other VSP-II variants (Fig.6.2).

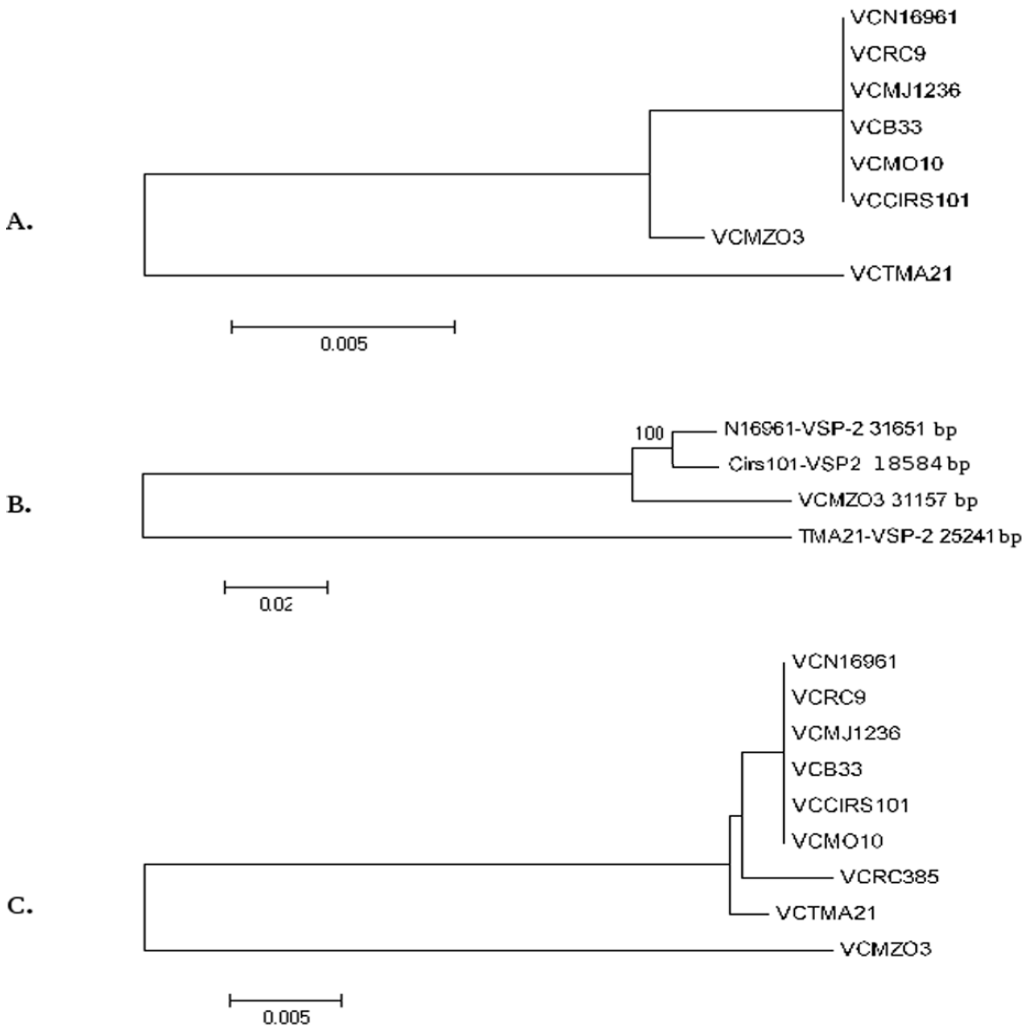


Fig. 6.6.Phylogenetic analysis of VSP-II variants*.(A) Neighbor-Joining tree based on three VSP-II concatenated genes: plasmid related protein, hypotetical protein, and phage integrase; (B) Neighbor-Joining tree based on the whole VSP-II and (C) Neighbor-Joining tree based on concatenated VSP-II flanking genes: putative hemolysin and Rpo sigma factor.*

PCR screening and ecology of VSP-I and VSP-II genomic islands

To assess the distribution of the VSP-II variants identified by comparative genomics, a well-characterized collection of 188 clinical and environmental isolates of *V. cholerae* representing different serogroups and biotypes and featuring diverse virulence pattern and 190 recent isolates from two cholera endemic sites in Bangladesh, were screened by PCR. To determine the presence and distribution of VSP-I variants within *V.cholerae*, two PCR primer pairs were designed (Table 6.2).

To distinguish the five VSP-II variants, three primers pairs were designed and incorporated into a multiplex PCR. Amplification patterns associated with specific VSP-II variants are shown in Table 6.2. Furthermore, the insertion site of the island was confirmed by amplification of a primer pair designed using flanking genes (Table 6.2). Positive amplification with the primer pair was considered evidence of an intact insertion site or absence of the island. In total, 378 isolates of *V. cholerae* were screened for VSP-I and VSP-II; results are reported in Table 6.5 and Table 6.6.

Classical, and pre-seventh pandemic *V.cholerae* O1 El Tor

As expected, all the *V. cholerae* O1 Classical and El Tor pre-seventh pandemic isolates from the laboratory collection did not contain the VSP-II island (Table 6.6). All classical and pre-seventh pandemic *V.cholerae* O1 El Tor were negative for VSP-I and VSP-II

Table 6.5. Results of PCR screening for VSP-I. *Target VC0180-0181 is a conserved region found in both variants of VSP-I, while VC0175 is unique to the seventh pandemic group VSP-I.Gene VC2346 is unique to the seventh pandemic V. cholerae. Chromosomal insertion sites are designated as Chr 1 and Chr 2.*

Strain group	Total	N16961	CIRS101	MZO3	TMA21	RC385	Negative
Lab culture collection	188						
*V. cholerae*O1 classical	8	0	0	0	0	0	8
V. cholerae O1 El Tor seventh pandemic	31	29	2	0	0	0	0
V. cholerae O1 El Tor pre-seventh pandemic	3	0	0	0	0	0	3
V. cholerae O1, environmental	23	0	0	0	0	2	21
V. cholerae O139	17	16	0	0	0	1	0
V. cholerae non-O1/non-O139	91	0	0	1	1	8	81
V. mimicus	15	0	0	0	1	4	11
Bangladesh isolates	190						
V. cholerae O1 clinical	101	1	96	0	0	1	3
V. cholerae O1 environmental	17	0	16	0	0	0	1
V. cholerae O139 clincal	10	10	0	0	0	0	0
V. cholerae O139 environmental	15	15	0	0	0	0	0
V. cholerae non-O1/non-O139 environmental	47	0	0	0	2	12	33

Seventh pandemic *V. cholerae*

Seventh pandemic *V. cholerae* O1 El Tor isolates belonging to our culture collection assayed for VSP-I and VSP-II gave expected results for VSP-I: all were positive when tested with the seventh pandemic groupspecific PCR of *V.cholerae* O1 and all carried the seventh pandemic VSP-I cluster. In regard to the VSP-II island, twenty-nine of 31 seven pandemic *V. cholerae* O1 El Tor strains (93.5%) harbored the prototypical seventh pandemic VSP-II island. In addition to *V. cholerae* CIRS101, only one other strain, a clinical isolate from Bangladesh, yielded an amplification pattern

corresponding to the *V. cholerae* CIRS101 VSP-II variant, (Table 6.6); both harbored the typical seventh pandemic VSP-I (Grim *et al.*, 2010; Taviani *et al*., 2010).

Interesting results emerged from screening the collection of *V. cholerae* isolates from two cholera endemic sites about 150 miles apart, in Bangladesh, collected from 2004 to 2007: the scenario becomes more complex. As expected, all clinical isolates were confirmed to be seventh pandemic strains by the 7P group specific PCR and all carried a seventh pandemic VSP-I. In contrast, the VSP-II cluster found in these strains was the CIRS101 VSP-II variant. Among the clinical *V. cholerae* O1 El Tor, a total of 96 carried the *V. cholerae* CIRS101 VSP-II variant and only one harbored the typical seventh pandemic VSP-II (Table 6.6). Moreover three isolates did not contain VSP-II and one was positive for *V. cholerae* RC385 VSP-II (Table 6.6), which was negative for VSP-I and *ctxA* (Grim *et al.*, 2010; Taviani *et al*., 2010).

However, among the seventh pandemic *V. cholerae* O139 the same shift between different VSP-II variants was not observed. The *V. cholerae* O139 strains in our collection and in the Bangladesh collection all had typical seventh pandemic VSP-I and VSP-II genomic islands (Grim *et al.*, 2010; Taviani *et al*., 2010).

The environmental *V.cholerae* O1 and O139 isolates from Bangladesh were isolated at typical epidemic periods during the year. Most were confirmed by the seventh pandemic group PCR and were *ctx* and *tcpA* positive strains indicating a degree of relatedness to clinical *V.cholerae* O1 and O139 isolates. Sixteen of 17 environmental (94%) *V. cholerae* O1 from Bangladesh were positive for seventh pandemic VSP-I and carried *V. cholerae* CIRS101 VSP-II, except one strain that did not have *V. cholerae* VSP-II or VSP-I (Table 6.5 and 6.6) (Grim *et al.*, 2010, Taviani *et al*., 2010). The latter

was also negative for *ctxA* by PCR (Grim *et al* 2010). In contrast, all *V. cholerae* O139 of environmental origin, confirmed as seventh pandemic by PCR, contained the canonical seventh pandemic VSP-II and VSP-II (Table 6.6), suggesting this serogroup is genetically isolated from the dominant *V. cholerae* O1 pandemic clones.

Table 6.5. Results of PCR screening for VSP-II. *Distribution of VSP-II variants in our laboratory V. cholerae culture collections and Bangladesh 2004-2007 V. cholerae collection.*

Strain group	Total	N16961	CIRS101	MZO3	TMA21	RC385	Negative
Lab culture collection	188						
*V. cholerae*O1 classical	8	0	0	0	0	0	8
V. cholerae O1 El Tor seventh pandemic	31	29	2	0	0	0	0
V. cholerae O1 El Tor pre-seventh pandemic	3	0	0	0	0	0	3
V. cholerae O1, environmental	23	0	0	0	0	2	21
V. cholerae O139	17	16	0	0	0	1	0
V. cholerae non-O1/non-O139	91	0	0	1	1	8	81
V. mimicus	15	0	0	0	1	4	11
Bangladesh isolates	190						
V. cholerae O1 clinical	101	1	96	0	0	1	3
V. cholerae O1 environmental	17	0	16	0	0	0	1
V. cholerae O139 clincal	10	10	0	0	0	0	0
V. cholerae O139 environmental	15	15	0	0	0	0	0
V. cholerae non-O1/non-O139 environmental	47	0	0	0	2	12	33

Environmental *V.cholerae* O1 and O139

In the environment we found that 91% (21/23) of the *V.cholerae* O1 isolates, including three U.S. Gulf coast isolates, six isolates from Bangladesh, three isolates from Mexico, three isolates from Brazil, and single isolates from Chile, Australia, Thailand, and Japan did not contain the seventh pandemic VSP-II and only two strains showed the *V. cholerae* RC385 VSP-II island amplification pattern: one isolated from a sewage sample collected in Brazil in 1978 and a second strain from Mexico. All were negative for VSP-I (Grim *et al.*, 2010; Taviani *et al*., 2010) except for *V. cholerae* V-69, an isolate from a water sample collected off North Point State Park near Baltimore, MD, in the Chesapeake Bay, USA (Colwell *et al*., 1981). The latter was positive for both VSP-I primer pairs, but negative for the seventh pandemic marker (Grim *et al.,* 2010). The only *V. cholerae* O139 environmental isolate present in collection was negative for all three primers pairs (Grim *et al.,* 2010). Further analysis of this strains wasconfirmed by as O139 wb* by PCR and slide agglutination with O139 antisera (Grim *et al.*, 2010; Taviani *et al*., 2010).

V. cholerae non-O1/non-O139

Among *V. cholerae* non-O1/non-O139 isolates, 70% did not harbor VSP-II, 26% contained *V. cholerae* RC385 VSP-II and two the *V. cholerae* TMA21 VSP-II (Table 2), showing these are the most common variants in the non-epidemic *V. cholerae* population. Interestingly, among the *V. cholerae* non-O1/O139 isolates, five strains gave a positive

signal for VC0180-0181 target, suggesting the presence of a form of the VSP-I island. Two of these, *V. cholerae* V-5 and V-63, gave positive signals also for VC0175, indicating the presence of the seventh pandemic VSP-I. Both are *V. cholerae* serogroup O37, V-5 is a strain isolated from Jones Fall, MD, in 1976, and *V. cholerae* V-63 was isolated from Hawkins Point, MD, in 1977 (Grim *et al* 2010). The other three strains were *V. cholerae* VL426 and ATCC14547, both biovar albensis (Spencer, 1955) and *V. cholerae* RC341; all three strains were bioluminescent. Two *V.* mimicus, both isolated from water samples collected near the Louisiana coast, were positive for VC0180-0181 indicating the presence of a VSP-I variant yet not identified (Grim *et al* 2010). Furthermore, among the *V. cholerae* non-O1/O139 isolates, seven carried the RC385 VSP-II (Table 6.6). Among these, two were isolated from Chesapeake Bay, MD, USA, same location as *V. cholerae* RC385 (Grim *et al.*, 2010; Taviani *et al.*, 2010). Of the remainder, one was isolated from a sewage sample collected in Brazil, one was from Czechoslovakia, two were from Japan, and one was from Bangladesh. It should be noted that four of 15 *V. mimicus* strains also were positive for the *V. cholerae* RC385 VSP-II variant. All 47 *V.cholerae* non-O1/O139 from Bangladesh were negative for both VSP-I and VSP-II genomic islands (Grim *et al* 2010; Taviani *et al* 2010).

Conclusions

In conclusion, although all seventh pandemic *V.cholerae* O1 El Tor and O139 of clinical origin contained the seventh pandemic VSP-I, the island is not exclusive to this group, evidenced by its presence in the three Chesapeake Bay isolates.

Comparative genomic analysis of 23 *V. cholerae* strains belonging to different serotypes, originated from diversified geografical locations and isolated over an extended period of time, has led to the discovery of one new variant of the VSP-I and three new variants of the VSP-II genomic islands (Grim *et al*., 2010, Taviani *et al*., 2010). This observation is remarkable, since VSP-I and VSP-II were originally considered to be conserved genetic markers of seventh pandemic *V. cholerae* used in epidemiological investigations (Dziejman *et al*., 2002; O'Shea *et al*., 2004). To date, two other examples of sequence variation whitin *V. cholerae* VSP-II were described (Dziejman *et al.*, 2005; Nusrin *et al.*, 2009). Our analysis adds further insight to the knowledge of this genomic clusters and their evolution in *V. cholerae*.

From the standpoint of genetic comparison, it is clear that the island has undergone significant genetic rearrangement. Two loci, at the 3' end of the VC0498 and VC0511, may represent hot spots for recombination events within the conserved genomic backbone of the island. It appears that the VSP-II cluster has evolved into different variants by acquisition and loss of indels at specific loci within a conserved core.

The VSP-II variant found in *V. cholerae* O1 El Tor CIRS101 has a significant deletion compared to the other two variants presumably circulating among *V. cholerae* O1 El Tor strains: the seventh pandemic and the Peruvian VSP-II. Although its function remains to be elucidated, the CIRS101 VSP-II presence is clearly dominant in recent *V. cholerae* O1 isolates from two cholera endemic sites of Bangladesh, but not in *V. cholerae* O139 isolated from those sites, the latter possessing the prototypical seventh pandemic VSP-II. These data are surprising because these two areas under study, *V. cholerae* O1 and O139 share the same environment and host population, but appear not to

have exchanged this genomic island. This suggests there might be other factors, biological or molecular, influencing horizontal gene transfer.

In Bangladesh, by tracking VSP-II variants, we were able to detect a shift between "old" and "new" pandemic clones of *V. cholerae* O1 El Tor, based on the fact that a 1994 strain (*V. cholerae* O1 MJ1236) carries the prototypical seventh pandemic VSP-II, while those isolated during 2004-2007, carry the new CIRS101 variant. It is of paramount importance to know whether the same shift occurred in clinical *V. cholerae* isolates from Africa or South America to be able to determine if this event is region-specific.

By not being present in non-epidemic isolates of *V. cholerae* non-O1/O139 suggests that the CIRS101 VSP-II confers a selective advantage when in the human host but not when in the aquatic environment. In this regard, it is noteworthy that *V. cholerae* O1 El Tor CIRS101 carries a variant of the CTX cluster found in a group of newly emerged seventh pandemic clones, referred to as El Tor/classical 'hybrid' or 'altered' strains (Nair *et al*., 2006) Therefore, the new *V. cholerae* CIRS101 VSP-II may have arisen in a genomic background positively selected in the human host (hybrid strains appear to produce more cholera toxin), likely becoming dominant among epidemic clones. There has been evidence to suggest a link between the evolutionary success of the two clusters (CTX and VSP-II) is not indicated, based on the presence of a canonical seventh pandemic VSP-II in two other hybrid strains, *V. cholerae* O1 MJ1236 (Bangladesh, 1994) and B33 (Mozambique, 2004).

The VSP-II circulating among *V. cholerae* non-O1/non-O139 and *V. mimicus* is the RC385 VSP-II. Despite different serotype and significant genetic diversity among the

strains, this variant appears to be stable in isolates obtained at different times and geographical locations while TMA21 and MZO-3 VSP-II show only limited distribution. The presence of the new VSP-II variants was not correlated with the presence of a new VSP-I, indicating that the two gene clusters derive from a different history of genetic exchange among *V. cholerae* non-O1/non-O139 and *V. mimicus.*

Publications

Taviani, E., C. J. Grim, J. Choi, J. Chun, B. Haley, N. A. Hasan, A. Huq and R. R. Colwell (2010). "Discovery of novel Vibrio cholera VSP-II genomic islands using comparative genomic analysis." FEMS Microbiology Letters **308**(2): 130-137.

Grim, C. J., J. Choi, J. Chun, Y.-S. Jeon, E. Taviani, N. A. Hasan, B. Haley, A. Huq and R. R. Colwell (2010). "Occurrence of the *Vibrio cholerae* Seventh Pandemic VSP-I Island and a New Variant." OMICS: A Journal of Integrative Biology **14**(1): 1-7.

Chapter 7. Conclusions

Results of this research contributed to our knowledge of the nature of ICEs found in *V. cholerae* and their relationship with other genetic elements, such as pathogenicity islands, and offer a glimpse of a better understanding of the extent to which these elements play a role, not only in epidemic *V. cholerae*, but also in *V. cholerae* non-O1/O139 pathogenesis and ecology.

Overall the comparative analysis of ICEs in recent seventh pandemic clones of *V. cholerae* O1 El Tor, the so-called 'hybrid strains' that are causing cholera in Asian and African countries, further confirmed the conclusions of the entire comparative genomic analysis. *V. cholerae* O1 Matlab hybrid is believed to have emerged in India and moved to Africa, where it became established as the prevalent type. In contrast, in Bangladesh, an altered *V. cholerae* O1 El Tor, typified by *V. cholerae* CIRS101, either emerged from the Matlab clone or independently, and became the prevalent type. Similar insight can be drawn from the distribution of the VSP-II genomic island. In Bangladesh, by tracking VSP-II variants, we were able to detect a shift between “old” and “new” pandemic clones of *V. cholerae* O1 El Tor, based on the fact that a 1994 strain (*V. cholerae* O1 MJ1236) carries the prototypical seventh pandemic VSP-II, while those isolated during 2004-2007, carry the new CIRS101 variant.

From analysis of an ICE-like element found in a clinical isolate of *V. cholerae* O37, it emerged that this element could be associated with the pathogenic potential of

non-O1, non-O139 *V. cholerae* which, even though not clear, is different with respect to toxigenic *V. cholerae*. Moreover, the unusual genetic organization displayed by ICE*Vch*Ban8 and its association with another major pathogenicity island in *V. cholerae,* suggests it belongs either to a novel class of ICE or is an ancient element, offering evidence to the hypothesis that ICEs are progenitors and/or potential dissemination tools for PAIs.

Interestingly, the results reported have shown the presence of ICEs in strains of *V. cholerae* isolated from geographical areas where they had never before been discovered. The element was detected for the first time in *V. cholerae* non-O1/O139 isolated in 1976 and 1977 in the US, long before the first appearance in this species in India, where it had persisted in the aquatic *V. cholerae* non-O1/O139 population for over twenty years suggesting a cryptic functional role of the ICEs in the ecology of these organisms. This hypothesis is further confirmed by the abundance of strains of *V. cholerae* carrying ICEs that were detected in waters around both Northern (Chesapeake Bay) and Southern (Mexico, Peru, Brazil, Chile) regions of this continent, as well as in Iceland, a region that is geographically isolated and where cholera has never been documented. Overall, these results indicate that ICEs are present in the environment where they constitute a genetic reservoir for assemblage of diverse genetic profiles and play a primary role, yet unknown, in the ecology of *V. cholerae.* This is also true for VSP-II, originally considered to be a conserved genetic marker of seventh pandemic *V. cholerae.* It wasdetected in several environmental strains isolated at different chronological times and from different geographical locations.

Our analysis also revealed that ICEs isolated in America had an independent

origin and evolution with respect to that circulating in Asia and Africa. As demonstrated by the presence of ICEs in epidemic *V. cholerae* O1 El Tor isolated in Mexico, including from the first case of cholera, which appeared genetically diverse from any other SXT-related elements found in African and Asian epidemic *V. cholerae*, these elements most likely had been evolutionarily separated from their relatives in the Eastern hemisphere, reflecting separation of their hosts.This result further confirms the hypothesis that *V. cholerae* is a natural inhabitant of the aquatic environment and its local populations have evolved separately and continuously, adjusting to a changing environment via the plasticity of its genome. Furthermore, ICEs are important genetic tools that contribute to fitness of this bacterium in its native habitat.

In summary, our analysis adds further insight to the knowledge of factors that influence horizontal gene transfer and its role in the evolution, pathogenesis, and ecology of *V. cholerae*. From the standpoint of genetic comparison, it is clear that the mobile genetic elements of *V. cholerae* have undergone significant molecular rearrangement. The structural analysis of ICEs and VSP-II revealed a large number of profiles, confirming they are of high genetic plasticity and that their variable content most likely contributes significantly to genomic evolution responsible for adaptation of this bacterial host to new ecological niches.

Bibliography

Ahmed, A., S. Shinoda and T. Shimamoto (2005). "A variant type of *Vibrio cholerae* SXT element in a multidrug-resistant strain of *Vibrio fluvialis*." FEMS Microbiol Ecol **15**: 241-247.

Aldova, E., K. Lazickova, E. Stepankova and J. Lietava (1968). "Isolation of noagglutinable vibrios from an enteritis outbreak in Czechoslovakia." J. Infect. Dis. **118**: 25-31.

Amita, S. R. Chowdhury, M. Thungapathra, T. Ramamurthy, G. B. Nair and A. Ghosh (2003). "Class I integrons and SXT elements in El Tor strains isolated before and after 1992 *Vibrio cholerae* O139 outbreak, Calcutta, India." Emerg Infect Dis **9**(4): 500-502.

Ausubel, F., R. Brent, R. Kingston, D. Moore, J. Seidman, J. Smith and K. Struhl (1990). Current protocols in Molecular Biology. New York, Green Publishing Associates and John Wiley & Sons.

Azaro MA, Landy A (2002). Chapter 7-λ integrase and the λ Int family. In Mobile DNA II, Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) pp 118–148. Washington, DC: ASM Press.

Bagchi, K., P. Echeverria, J. D. Arthur, O. Sethabutr, O. Serichantalergs and C. W. Hoge (1993). "Epidemic of diarrhea caused by *Vibrio cholerae* non-O1 that produced heat-stable toxin among Khmers in a camp in Thailand." Journal of Clinical Microbiology (13): 1315-1317.

Bani, S., P. N. Mastromarino, D. Ceccarelli, A. Le Van, A. M. Salvia, Q. T. Ngo Viet, D. H. Hai, D. Bacciu, P. Cappuccinelli and M. M. Colombo (2007). "Molecular characterization of ICEVchVie0 and its disappearance in *Vibrio cholerae* O1 strains isolated in 2003 in Vietnam." FEMS Microbiol Lett **266**(1): 42-48.

Beaber, J. W., V. Burrus, B. Hochhut and M. K. Waldor (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**(12): 2065-2070.

Beaber, J. W., B. Hochhut and M. K. Waldor (2002). "Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*." J Bacteriol **184**(15): 4259-4269.

Beaber, J. W., B. Hochhut and M. K. Waldor (2004). "SOS response promotes horizontal dissemination of antibiotic resistance genes." Nature **427**(6969): 72-74.

Beaber, J. W. and M. K. Waldor (2004). "Identification of operators and promoters that control SXT conjugative transfer." J Bacteriol **186**(17): 5945-5949.

Beltran, P., G. Delgado, A. Navarro, F. Trujilto, R. K. Selander and A. Cravioto (1999). "Genetic diversity and population structure of *Vibrio cholerae*." J. Clin. Micro. **37**(3): 581-590.

Blakely, G. and D. Sherratt (1996). "Determinants of selectivity in Xer site-specific recombination." Genes & Development **10**(6): 762-773.

Boltner, D., C. MacMahon, J. T. Pembroke, P. Strike and A. M. Osborn (2002). "R391: a conjugative integrating mosaic comprised of phage, plasmid, and transposon elements." J Bacteriol **184**(18): 5158-5169.

Böltner, D. and A. M. Osborn (2004). "Structural comparison of the integrative and conjugative elements R391, pMERPH, R997, and SXT." Plasmid **51**(1): 12-23.

Boyd, E., S. Almagro-Moreno and M. Parent (2009). "Genomic islands are dynamic, ancient integrative elements in bacterial evolution." Trends in Microbiology **17**(2): 47 - 53.

Brussow, H in Hensel, M. and H. Schmidt (2008). Horizontal gene transfer in the evolution of pathogenesis, Cambridge University Press.

Burland, V., Y. Shao, N. T. Perna, G. Plunkett, F. R. Blattner and H. J. Sofia (1998). "The complete DNA sequence and analysis of the large virulence plasmid of Escherichia coli O157:H7." Nucleic Acids Research **26**(18): 4196-4204.

Burrus, V., J. Marrero and M. Waldor (2006). "The current ICE age: biology and evolution of SXT-related integrating conjugative elements." Plasmid **55**(3): 173-183.

Burrus, V., G. Pavlovic, B. Decaris and G. Guédon (2002). "Conjugative transposons: the tip of the iceberg." Molecular Microbiology **46**(3): 601-610.

Burrus, V., R. Quezada-Calvillo, J. Marrero and M. K. Waldor (2006). "SXT-related integrating conjugative element in New World *Vibrio cholerae*." Appl Environ Microbiol **72**(4): 3054-3057.

Burrus, V. and M. Waldor (2004). "Shaping bacterial genomes with integrative and conjugative elements." Res Microbiol **155**: 376-386.

Burrus, V. and M. K. Waldor (2003). "Control of SXT Integration and Excision." J. Bacteriol. **185**(17): 5045-5054.

Burrus, V. and M. K. Waldor (2004). "Formation of SXT tandem arrays and SXT-R391 hybrids." J Bacteriol **186**(9): 2636-2645.

Carver, T. J., K. M. Rutherford, M. Berriman, M.-A. Rajandream, B. G. Barrell and J. Parkhill "ACT: the Artemis comparison tool." Bioinformatics **21**(16): 3422-3423.

Ceccarelli, D., A. Daccord, M. Rene and V. Burrus (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**(15): 5328-5338.

Ceccarelli, D., A. Salvia, J. Sami, P. Cappuccinelli and M. Colombo (2006). "A new cluster of plasmid-located class 1 integrons in *V. cholerae* O1 and *dfrA15* cassette integron in *V. parahaemolyticus* isolated in Angola." Antimicrob Agents Chemother **50**(7): 2493-2499.

Chatterjee, S., K. Ghosh, A. Raychoudhuri, G. Chowdhury, M. K. Bhattacharya, A. K. Mukhopadhyay, T. Ramamurthy, S. K. Bhattacharya, K. E. Klose and R. K. Nandy (2009). "Incidence, Virulence Factors, and Clonality among Clinical Strains of Non-O1, Non-O139 *Vibrio cholerae* Isolates from Hospitalized Diarrheal Patients in Kolkata, India." J. Clin. Microbiol. **47**(4): 1087-1095.

Chatterjee, S. N. and K. Chaudhuri (2003). "Lipopolysaccharides of *Vibrio cholerae*. I. Physical and chemical characterization." Biochim Biophys Acta **1639**(2): 65-79.

Cholera Working Group, I., B (1993). "Large epidemic of cholera-like disease in Bangladesh caused by *Vibrio cholerae* O139 synonym Bengal." Lancet **342**: 387-390.

Chun, J., C. J. Grim, N. A. Hasan, J. H. Lee, S. Y. Choi, B. J. Haley, E. Taviani, Y.-S. Jeon, D. W. Kim, J.-H. Lee, T. S. Brettin, D. C. Bruce, J. F. Challacombe, J. C. Detter, C. S. Han, A. C. Munk, O. Chertkov, L. Meincke, E. Saunders, R. A. Walters, A. Huq, G. B. Nair and R. R. Colwell (2009). "Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*." Proceedings of the National Academy of Sciences **106**(36): 15442-15447.

Coetzee, J. N., N. Datta and R. W. Hedges (1972). "R Factors from Proteus rettgeri." J Gen Microbiol **72**(3): 543-552.

Cohan, F. M. (2002). "What are bacterial species?" Annual Review of Microbiology **56**: 457-488.

Cohan, F. M. (2006). "Towards a conceptual and operational union of bacterial systematics, ecology, and evolution." Philosophical Transactions of the Royal Society B: Biological Sciences **361**(1475): 1985-1996.

Cohan, F. M. and E. B. Perry (2007). "A Systematics for Discovering the Fundamental Units of Bacterial Diversity." Current biology : CB **17**(10): R373-R386.

Colwell, R., R. Seidler, J. Kaper, S. Joseph, S. Garges, H. Lockman, D. Maneval, H. Bradford, N. Roberts, R. Remmers, I. Huq and A. Huq (1981). "Occurrence of *V. cholerae* serotype O1 in Maryland and Louisiana estuaries." Appl Environ Micbrobiol **41**: 555-558.

Colwell, R. R. (1996). "Global climate and infectious disease: the cholera paradigm." Science **274**(5295): 2025-2031.

Cox, M. M. (2001). "RECOMBINATIONAL DNA REPAIR OF DAMAGED REPLICATION FORKS IN ESCHERICHIA COLI: Questions." Annual Review of Genetics **35**(1): 53-82.

Craig, N. (2002). Mobile DNA II, ASM Press.

Dalsgaard, A., M. J. Albert, D. N. Taylor, T. Shimada, R. Meza, O. Serichantalergs and P. Echeverria (1995). "Characterization of Vibrio cgolerae non-O1 serogroups obtained from an outbreak of diarrhea in Lima, Peru." J Clin Microbiol **33**(10): 2715-2722.

Dalsgaard, A., A. Forslund, D. Sandvang, L. Arntzen and K. Keddy (2001). "*Vibrio cholerae* O1 outbreak isolates in Mozambique and South Africa in 1998 are multiple-drug resistant, contain the SXT element and the aadA2 gene located on class 1 integrons." J Antimicrob Chemother **48**(6): 827-838.

Dalsgaard, A., A. Forslund, N. V. Tam, D. X. Vinh and P. D. Cam (1999). "Cholera in Vietnam: changes in genotypes and emergence of class I integrons containing aminoglycoside resistance gene cassettes in *Vibrio cholerae* O1 strains isolated from 1979 to 1996." J Clin Microbiol **37**(3): 734-741.

Davis, B. M. and M. K. Waldor (2000). "CTXphi contains a hybrid genome derived from tandemly integrated elements." Proc Natl Acad Sci U S A **97**(15): 8572-8577.

Davison, J. (1999). "Genetic Exchange between Bacteria in the Environment." Plasmid **42**(2): 73-91.

De la Cruz, F. and J. Davies (2000). "Horizontal gene transfer and the origin of species: lessons from bacteria." Trends in Microbiology **8**(3): 128-133.

Delcher, A. L., S. Kasif, R. D. Fleischmann, J. Peterson, O. White and S. L. Salzberg (1999). "Alignment of whole genomes." Nucleic Acids Res **27**: 2369 - 2376.

Deonier, R. C. and L. Mirels (1977). "Excision of F plasmid sequences by recombination at directly repeated insertion sequence 2 elements: involvement of recA." Proceedings of the National Academy of Sciences of the United States of America **74**(9): 3965-3969.

Dziejman, M., E. Balon, D. Boyd, C. M. Fraser, J. F. Heidelberg and J. J. Mekalanos (2002). "Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease." Proc Natl Acad Sci U S A **99**(3): 1556-1561.

Dziejman, M., D. Serruto, V. C. Tam, D. Sturtevant, P. Diraphat, S. M. Faruque, M. H. Rahman, J. F. Heidelberg, J. Decker, L. Li, K. T. Montgomery, G. Grills, R.

Kucherlapati and J. J. Mekalanos (2005). "Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system." Proc Natl Acad Sci U S A **102**(9): 3465-3470.

Ehara, M., B. M. Nguyen, D. T. Nguyen, C. Toma, N. Higa and M. Iwanaga (2004). "Drug susceptibility and its genetic basis in epidemic *Vibrio cholerae* O1 in Vietnam." Epidemiol Infect **132**(4): 595-600.

Elvers, K. T., G. Wu, N. J. Gilberthorpe, R. K. Poole and S. F. Park (2004). "Role of an Inducible Single-Domain Hemoglobin in Mediating Resistance to Nitric Oxide and Nitrosative Stress in Campylobacter jejuni and Campylobacter coli." J. Bacteriol. **186**(16): 5332-5341.

Esposito, D. and J. J. Scocca (1997). "The integrase family of tyrosine recombinases: evolution of a conserved active site domain." Nucleic Acids Research **25**(18): 3605-3614.

Farina, C., F. Marini, E. Schiaffino, I. Luzzi, A. M. Dionisi, F. Leoni, D. Ottaviani and S. Bordoni (2010). "A fatal *Vibrio cholerae* O37 enteritis." J Med Microbiol: jmm.0.023093-023090.

Faruque, S., M. Albert and J. Mekalanos (1998). "Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*." Microbiol Mol Biol Rev **62**: 1301-1314.

Faruque, S. M. and J. J. Mekalanos (2003). "Pathogenicity islands and phages in *Vibrio cholerae* evolution." Trends Microbiol **11**(11): 505-510.

Faruque, S. M., J. Zhu, Asadulghani, M. Kamruzzaman and J. J. Mekalanos (2003). "Examination of diverse toxin-coregulated pilus-positive *Vibrio cholerae* strains fails to demonstrate evidence for Vibrio pathogenicity island phage." Infect Immun **71**(6): 2993-2999.

Frost, L. S., K. Ippen-Ihler and R. A. Skurray (1994). "Analysis of the sequence and gene products of the transfer region of the F sex factor." Microbiol. Mol. Biol. Rev. **58**(2): 162-210.

Frost, L. S., R. Leplae, A. O. Summers and A. Toussaint (2005). "Mobile genetic elements: the agents of open source evolution." Nat Rev Micro **3**(9): 722-732.

Garriss, G., M. K. Waldor and V. Burrus (2009). "Mobile Antibiotic Resistance Encoding Elements Promote Their Own Diversity." PLoS Genet **5**(12): e1000775.

Gevers, D., F. M. Cohan, J. G. Lawrence, B. G. Spratt, T. Coenye, E. J. Feil, E. Stackebrandt, Y. V. de Peer, P. Vandamme, F. L. Thompson and J. Swings (2005). "Re-evaluating prokaryotic species." Nat Rev Micro **3**(9): 733-739.

Gil, A. I., V. R. Louis, I. N. Rivera, E. Lipp, A. Huq, C. F. Lanata, D. N. Taylor, E. Russek-Cohen, N. Choopun, R. B. Sack and R. R. Colwell (2004). "Occurrence and distribution of *Vibrio cholerae* in the coastal environment of Peru." Environ Microbiol **6**(7): 699-706.

Giron, J. A., M. M. Levine and J. B. Kaper (1994). "Longus: a long pilus ultrastructure produced by human enterotoxigenic Escherichia coli." Mol Microbiol **12**(1): 71-82.

Griffith, F. 1928. The significance of pneumococcal types. J. Hyg. 27.

Gogarten, J. P., W. F. Doolittle and J. G. Lawrence (2002). "Prokaryotic Evolution in Light of Gene Transfer." Molecular Biology and Evolution **19**(12): 2226-2238.

Gogarten, J. P. and J. P. Townsend (2005). "Horizontal gene transfer, genome innovation and evolution." Nat Rev Micro **3**(9): 679-687.

Gopaul, D. N. and G. D. Van Duyne (1999). "Structure and mechanism in site-specific recombination." Current Opinion in Structural Biology **9**(1): 14-20.

Grim, C. J., J. Choi, J. Chun, Y.-S. Jeon, E. Taviani, N. A. Hasan, B. Haley, A. Huq and R. R. Colwell (2010). "Occurrence of the *Vibrio cholerae* Seventh Pandemic VSP-I Island and a New Variant." OMICS: A Journal of Integrative Biology **14**(1): 1-7.

Grys, T. E., M. B. Siegel, W. W. Lathem and R. A. Welch (2005). "The StcE Protease Contributes to Intimate Adherence of Enterohemorrhagic Escherichia coli O157:H7 to Host Cells." Infect. Immun. **73**(3): 1295-1303.

Hacker, J. and J. Kaper (2002). Pathogenicity islands and the evolution of pathogenic microbes, Springer.

Haren, L., B. Ton-Hoang and M. Chandler (1999). "Integrating DNA: Transposases and Retroviral Integrases." Annual Review of Microbiology **53**(1): 245-281.

Hayes, W. (1953). "Observations on a Transmissible Agent Determining Sexual Differentiation in Bacterium coli." J Gen Microbiol **8**(1): 72-88.

Hensel, M. and H. Schmidt (2008). Horizontal gene transfer in the evolution of pathogenesis, Cambridge University Press.

Hochhut, B., J. W. Beaber, R. Woodgate and M. K. Waldor (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**(4): 1124-1132.

Hochhut, B., Y. Lotfi, D. Mazel, S. M. Faruque, R. Woodgate and M. K. Waldor (2001). "Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins." Antimicrob Agents Chemother **45**(11): 2991-3000.

Hochhut, B., J. Marrero and M. Waldor (2000). "Mobilization of plasmids and chromosomal DNA mediated by the SXT element, a Constin found in *Vibrio cholerae* O139." J Bacteriol **182**: 2034-2047.

Hochhut, B., J. Marrero and M. K. Waldor (2000). "Mobilization of plasmids and chromosomal DNA mediated by the SXT element, a constin found in *Vibrio cholerae* O139." J Bacteriol **182**(7): 2043-2047.

Hochhut, B. and M. K. Waldor (1999). "Site-specific integration of the conjugal *Vibrio cholerae* SXT element into prfC." Mol Microbiol **32**(1): 99-110.

Holliday, R. (1964). "A mechanism for gene conversion in fungi." Genetics Research **5**(02): 282-304.

Ichinose, Y., K. Yamamoto, N. Nakasone, M. J. Tanabe, T. Takeda, T. Miwatani and M. Iwanaga (1987). "Enterotoxicity of El Tor-like hemolysin of non-O1 *Vibrio cholerae*." Infect Immun **55**(5): 1090-1093.

Iwanaga, M., C. Toma, T. Miyazato, S. Insisiengmay, N. Nakasone and M. Ehara (2004). "Antibiotic Resistance Conferred by a Class I Integron and SXT Constin

in *Vibrio cholerae* O1 Strains Isolated in Laos." Antimicrob. Agents Chemother. **48**(7): 2364-2369.

Iyer, L., J. Vadivelu and S. D. Puthucheary (2000). "Detection of virulence associated genes, haemolysin and protease amongst *Vibrio cholerae* isolated in Malaysia." Epidemiol Infect **125**(1): 27-34.

Jermyn, W. S. and E. F. Boyd (2002). "Characterization of a novel Vibrio pathogenicity island (VPI-2) encoding neuraminidase (nanH) among toxigenic *Vibrio cholerae* isolates." Microbiology **148**(Pt 11): 3681-3693.

Juiz-Rio, S., C. Osorio, V. de Lorenzo and M. Lemos (2005). "Subtractive hybridization reveals a high genetic diversity in the fish pathogen *Photobacterium damselae subsp. piscicida*: evidence of a SXT-like element." Microbiology **151**: 2659-2669.

Kalendar R, Lee D, Schulman AH (2009). FastPCR Software for PCR Primer and Probe Design and Repeat Search. Genes, Genomes and Genomics, **3**(1): 1-14

Kaper, J., H. Lockman, R. Colwell and S. Joseph (1979). "Ecology, serology, and enterotoxin production of *Vibrio cholerae* in Chesapeake Bay." Appl Environ Microbiol **37**: 91-103.

Kaper, J. B., J. G. J. Morris and M. M. Levine (1995). "Cholera." Clin. Microbiol. Rev. **8**(1): 48-86.

Karaolis, D. K., R. Lan, J. B. Kaper and P. R. Reeves (2001). "Comparison of *Vibrio cholerae* pathogenicity islands in sixth and seventh pandemic strains." Infect Immun **69**(3): 1947-1952.

Konstantinidis, K. T., A. Ramette and J. M. Tiedje (2006). "The bacterial species definition in the genomic era." Philosophical Transactions of the Royal Society B: Biological Sciences **361**(1475): 1929-1940.

Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. D. Lauder and W. M. Rehrauer (1994). "Biochemistry of homologous recombination in Escherichia coli." Microbiol. Mol. Biol. Rev. **58**(3): 401-465.

Kurtz, S., A. Phillippy, A. Delcher, M. Smoot, M. Shumway, C. Antonescu and S. Salzberg (2004). "Versatile and open software for comparing large genomes." Genome Biology **5**(2): R12.

Lan, R. and P. R. Reeves (2000). "Intraspecies variation in bacterial genomes: the need for a species genome concept." Trends in Microbiology **8**(9): 396-401.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Lawley, T. D., W. A. Klimke, M. J. Gubbins and L. S. Frost (2003). "F factor conjugation is a true type IV secretion system." FEMS Microbiology Letters **224**(1): 1-15.

Lawrence, J. G. and H. Hendrickson (2005). "Genome evolution in bacteria: order beneath chaos." Current Opinion in Microbiology **8**(5): 572-578.

Lawrence and Hendrickson in Hensel, M. and H. Schmidt (2008). Horizontal gene transfer in the evolution of pathogenesis, Cambridge University Press.

Lawrence, J. G., Hendrickson Heather (2004). "Chromosome Structure and Constraints on Lateral Gene Transfer." Dynamical Genetics: 319-336.

Lee, C. T., C. Amaro, K. M. Wu, E. Valiente, Y. F. Chang, S. F. Tsai, C. H. Chang and L. I. Hor (2008). "A common virulence plasmid in biotype 2 Vibrio vulnificus and its dissemination aided by a conjugal plasmid." J Bacteriol **190**(5): 1638-1648.

Lim, J. Y., H. J. La, H. Sheng, L. J. Forney and C. J. Hovde (2010). "Influence of Plasmid pO157 on Escherichia coli O157:H7 Sakai Biofilm Formation." Appl. Environ. Microbiol. **76**(3): 963-966.

Lin, T.-L., C.-Z. Lee, P.-F. Hsieh, S.-F. Tsai and J.-T. Wang (2008). "Characterization of Integrative and Conjugative Element ICEKp1-Associated Genomic Heterogeneity in a Klebsiella pneumoniae Strain Isolated from a Primary Liver Abscess." J. Bacteriol. **190**(2): 515-526.

Lorenz, M. G., and W. Wackernagel (1994). Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev. 58:563-602.

Louis, V., E. Russek-Cohen, N. Choopun, I. Rivera, B. Gangle, S. Jiang, A. Rubin, J. Patz, A. Huq and R. Colwell (2003). "Predictability of *Vibrio cholerae* in Chesapeake Bay." Appl Environ Microbiol **69**(5): 2773-2785.

Lukinmaa, S., K. Mattila, V. Lehtinen, M. Hakkinen, M. Koskela and A. Siitonen (2006). "Territorial waters of the Baltic Sea as a source of infections caused by *Vibrio cholerae* non-O1, non-O139: report of 3 hospitalized cases." Diagn Microbiol Infect Dis **54**(1): 1-6.

Mahillon, J. and M. Chandler (1998). "Insertion sequences." Microbiol Mol Biol Rev **62**(3): 725-774.

Marrero, J. and M. K. Waldor (2007). "Determinants of Entry Exclusion within Eex and TraG Are Cytoplasmic." J. Bacteriol. **189**(17): 6469-6473.

Marrero, J. and M. K. Waldor (2007). "The SXT/R391 Family of Integrative Conjugative Elements Is Composed of Two Exclusion Groups." J. Bacteriol. **189**(8): 3302-3305.

Mayr, E. (1982). The growth of biological thought. Cambridge, Massachusetts London England, The Belknap Press of Harvard University Press.

Mazariego-Espinosa, K., A. Cruz, M. A. Ledesma, S. A. Ochoa and J. Xicohtencatl-Cortes (2010). "Longus, a Type IV Pilus of Enterotoxigenic Escherichia coli, Is Involved in Adherence to Intestinal Epithelial Cells." J. Bacteriol. **192**(11): 2791-2800.

McGrath, B. and J. Pembroke (2004). "Detailed analysis of the insertion site of the mobile elements R997, pMERPH, R392, R705 and R391 in *E. coli* K12." FEMS Microbiol Ecol **237**(1): 19-26.

McGrath, B. M., J. A. O'Halloran, A. V. Piterina and J. T. Pembroke (2006). "Molecular tools to detect the IncJ elements: A family of integrating, antibiotic resistant mobile genetic elements." Journal of Microbiological Methods **66**(1): 32-42.

McLeod, S. M., V. Burrus and M. K. Waldor (2006). "Requirement for *Vibrio cholerae* integration host factor in conjugative DNA transfer." J Bacteriol **188**(16): 5704-5711.

Meibom, K. L., M. Blokesch, N. A. Dolganov, C. Y. Wu and G. K. Schoolnik (2005). "Chitin induces natural competence in *Vibrio cholerae*." Science **310**(5755): 1824-1827.

Mohapatra, H., S. S. Mohapatra, C. K. Mantri, R. R. Colwell and D. V. Singh (2008). "*Vibrio cholerae* non-O1, non-O139 strains isolated before 1992 from Varanasi, India are multiple drug resistant, contain intSXT, dfr18 and aadA5 genes." Environ Microbiol **10**(4): 866-873.

Mukhopadhyay, A. K., S. Garg, G. B. Nair, S. Kar, R. K. Ghosh, S. Pajni, A. Ghosh, T. Shimada, T. Takeda and Y. Takeda (1995). "Biotype traits and antibiotic susceptibility of *Vibrio cholerae* serogroup O1 before, during and after the emergence of the O139 serogroup." Epidemiol Infect **115**(3): 427-434.

Murphy, D. B. and J. T. Pembroke (1995). "Transfer of the IncJ plasmid R391 to recombination deficient Escherichia coli K12: Evidence that R391 behaves as a conjugal transposon." FEMS Microbiology Letters **134**(2-3): 153-158.

Murphy, R. A. and E. F. Boyd (2008). "Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates." J Bacteriol **190**(2): 636-647.

Nair, G. B., F. Qadri, J. Holmgren, A. M. Svennerholm, A. Safa, N. A. Bhuiyan, Q. S. Ahmad, S. M. Faruque, A. S. Faruque, Y. Takeda and D. A. Sack (2006). "Cholera due to altered El Tor strains of *Vibrio cholerae* O1 in Bangladesh." J Clin Microbiol **44**(11): 4211-4213.

Nakamura, Y., T. Itoh, H. Matsuda and T. Gojobori (2004). "Biased biological functions of horizontally transferred genes in prokaryotic genomes." Nat Genet **36**(7): 760-766.

Nunes-Düby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger and A. Landy (1998). "Similarities and differences among 105 members of the Int family of site-specific recombinases." Nucleic Acids Research **26**(2): 391-406.

Nusrin, S., A. I. Gil, N. A. Bhuiyan, A. Safa, M. Asakura, C. F. Lanata, E. Hall, H. Miranda, B. Huapaya, C. Vargas G, M. A. Luna, D. A. Sack, S. Yamasaki and G. B. Nair (2009). "Peruvian *Vibrio cholerae* O1 El Tor strains possess a distinct region in the Vibrio seventh pandemic island-II that differentiates them from the prototype seventh pandemic El Tor strains." J Med Microbiol **58**(3): 342-354.

O'Halloran, J. A., B. M. McGrath and J. T. Pembroke (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**(1): 99-105.

O'Shea, Y. A., S. Finnan, F. J. Reen, J. P. Morrissey, F. O'Gara and E. F. Boyd (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**(Pt 12): 4053-4063.

Osorio, C. R., J. Marrero, R. A. F. Wozniak, M. L. Lemos, V. Burrus and M. K. Waldor (2008). "Genomic and Functional Analysis of ICEPdaSpa1, a Fish-Pathogen-Derived SXT-Related Integrating Conjugative Element That Can Mobilize a Virulence Plasmid." J. Bacteriol. **190**(9): 3353-3361.

Pembroke, J. T. and V. P. Anna (2006). "A novel ICE in the genome of Shewanella putrefaciens W3-18-1: comparison with the SXT/R391 ICE-like elements." FEMS Microbiology Letters **264**(1): 80-88.

Pembroke, J. T., C. MacMahon and B. McGrath (2002). "The role of conjugative transposons in the Enterobacteriaceae." Cellular and Molecular Life Sciences **59**(12): 2055-2064.

Pembroke, J. T. and D. B. Murphy (2000). "Isolation and analysis of a circular form of the IncJ conjugative transposon-like elements, R391 and R997: implications for IncJ incompatibility." FEMS Microbiology Letters **187**(2): 133-138.

Pembroke, J. T. and E. Stevens (1984). "The Effect of Plasmid R391 and Other IncJ Plasmids on the Survival of Escherichia coli After UV Irradiation." J Gen Microbiol **130**(7): 1839-1844.

Rajanna, C., J. Wang, D. Zhang, Z. Xu, A. Ali, Y. M. Hou and D. K. Karaolis (2003). "The vibrio pathogenicity island of epidemic *Vibrio cholerae* forms precise extrachromosomal circular excision products." J Bacteriol **185**(23): 6893-6901.

Ramamurthy, T., S. Garg, R. Sharma, S. K. Bhattacharya, G. Balakrish, T. Nair, T. Shimada, Y. Takeda, T. Karasawa, H. Kurazano, A. Pal and Y. Takeda (1993). "Emergence of novel strains of *Vibrio cholerae* with epidemic potential in southern and eastern India." Lancet **341**: 703-704.

Restrepo, D., S. S. Huprikar, K. VanHorn and E. J. Bottone (2006). "O1 and non-O1 *Vibrio cholerae* bacteremia produced by hemolytic strains." Diagn Microbiol Infect Dis **54**(2): 145-148.

Rudra, S., R. Mahajan, M. Mathur, K. Kathuria and V. Talwar (1996). "Cluster of cases of clinical cholera due to *Vibrio cholerae* 010 in east Delhi." Indian J Med Res **103**: 71-73.

Saha, P. K., H. Koley, A. K. Mukhopadhyay, S. K. Bhattacharya, G. B. Nair, B. S. Ramakrishnan, S. Krishnan, T. Takeda and Y. Takeda (1996). "Nontoxigenic *Vibrio cholerae* 01 serotype Inaba biotype El Tor associated with a cluster of cases of cholera in southern India." J Clin Microbiol **34**(5): 1114-1117.

Sandkvist, M. (2001). "Biology of type II secretion." Molecular Microbiology **40**(2): 271-283.

Schubert, S., A. Rakin and J. Heesemann (2004). "The Yersinia high-pathogenicity island (HPI): evolutionary and functional aspects." International Journal of Medical Microbiology **294**(2-3): 83-94.

Scott, J. R., and G. G. Churchward. (1995). Conjugative transposition. Annu. Rev. Microbiol. 49:367-397

Sharma, C., M. Thungapathra, A. Ghosh, A. K. Mukhopadhyay, A. Basu, R. Mitra, I. Basu, S. K. Bhattacharya, T. Shimada, T. Ramamurthy, T. Takeda, S. Yamasaki, Y. Takeda and G. B. Nair (1998). "Molecular analysis of non-O1, non-O139 *Vibrio cholerae* associated with an unusual upsurge in the incidence of cholera-like disease in Calcutta, India." J Clin Microbiol **36**(3): 756-763.

Sharma, N. C., P. K. Mandal, R. Dhillon and M. Jain (2007). "Changing profile of *Vibrio cholerae* O1, O139 in Delhi & its periphery (2003-2005)." Indian J Med Res **125**(5): 633-640.

Smith, M. C. M. and H. M. Thorpe (2002). "Diversity in the serine recombinases." Molecular Microbiology **44**(2): 299-307.

Spampinato, C. and P. Modrich (2000). "The MutL ATPase Is Required for Mismatch Repair." J. Biol. Chem. **275**(13): 9863-9869.

Staley, J. T. (2006). "The bacterial species dilemma and the genomic–phylogenetic species concept." Philosophical Transactions of the Royal Society B: Biological Sciences **361**(1475): 1899-1909.

Tauxe, R., L. Seminario, R. Tapia and M. Libel (1994). The Latin American Epidemic. *Vibrio cholerae* and Cholera: Molecular to Global Perspectives. I. Wachsmuth, P. Blake and O. Olsvik. Washington, D.C., ASM**:** 321-344.

Taviani, E., D. Ceccarelli, M. N. LÃ¡zaro, S. Bani, P. A. Cappuccinelli, R. R. Colwell and M. M. Colombo (2008). Environmental Vibrio spp., isolated in Mozambique, contain a polymorphic group of integrative conjugative elements and class 1 integrons, Blackwell / Wiley.

Taviani, E., C. J. Grim, J. Choi, J. Chun, B. Haley, N. A. Hasan, A. Huq and R. R. Colwell (2010). "Discovery of novel Vibrio cholera VSP-II genomic islands using comparative genomic analysis." FEMS Microbiology Letters **308**(2): 130-137.

Taviani, E., C. J. Grim, J. Chun, A. Huq and R. R. Colwell (2009). "Genomic analysis of a novel integrative conjugative element in *Vibrio cholerae*." FEBS Letters **583**(22): 3630-3636.

Thomas, C. M. (2000). The horizontal gene pool: bacterial plasmids and gene spread. Harwood Academic Publishers, Amsterdam.

Thompson, C. C., F. L. Thompson, K. Vandemeulebroecke, B. Hoste, P. Dawyndt and J. Swings (2004). "Use of recA as an alternative phylogenetic marker in the family Vibrionaceae." Int J Syst Evol Microbiol **54**(Pt 3): 919-924.

Thungapathra, M., Amita, K. K. Sinha, S. R. Chaudhuri, P. Garg, T. Ramamurthy, G. B. Nair and A. Ghosh (2002). "Occurrence of antibiotic resistance gene cassettes aac(6')-Ib, dfrA5, dfrA12, and ereA2 in class I integrons in non-O1, non-O139 *Vibrio cholerae* strains in India." Antimicrob Agents Chemother **46**(9): 2948-2955.

Toma, C., N. Nakasone, T. Song and M. Iwanaga (2005). "*Vibrio cholerae* SXT element, Laos." Emerg Infect Dis **11**(2): 346-347.

Toussaint, A. and C. Merlin (2002). "Mobile Elements as a Combination of Functional Modules." Plasmid **47**(1): 26-35.

Waldor, M., H. Tschape and J. Mekalanos (1996). "A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O130." J Bacteriol **178**: 4157-4165.

Waldor, M. K. and J. J. Mekalanos (1994). "Emergence of a new cholera pandemic: molecular analysis of virulence determinants in *Vibrio cholerae* O139 and development of a live vaccine prototype." J Infect Dis **170**(2): 278-283.

Waldor, M. K. and J. J. Mekalanos (1996). "Lysogenic conversion by a filamentous phage encoding cholera toxin." Science **272**(5270): 1910-1914.

Wang, H., A. P. Roberts, D. Lyras, J. I. Rood, M. Wilks, and P. Mullany. (2000). Characterization of the ends and target sites of the novel coniugative transposon Tn5397 from *Clostridium difficile*: excision and circularization is mediated by the large resolvase TndX. J. Bacteriol. 182:3775-3783.

Whittle, G., N. B. Shoemaker, and A. A. Salyers. (2002). The role of *Bacteroides* conjugative transposons in the dissemination of antibiotic resistance genes. Cell. Mol. Life Sci. 59:2044-2054.

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

Wozniak, R. A. F. and M. K. Waldor (2009). "A Toxinâ€“Antitoxin System Promotes the Maintenance of an Integrative Conjugative Element." PLoS Genet **5**(3): e1000439.

Yam, W. C., M. L. Lung, K. Y. Ng and M. H. Ng (1989). "Molecular epidemiology of *Vibrio cholerae* in Hong Kong." J Clin Microbiol **27**(8): 1900-1902.

Yamamoto, T., G. B. Nair, M. J. Albert, C. C. Parodi and Y. Takeda (1995). "Survey of in vitro susceptibilities of *Vibrio cholerae* O1 and O139 to antimicrobial agents." Antimicrob Agents Chemother **39**(1): 241-244.

Yokota, T. and S. Kuwahara (1977). "Temperature-sensitive R plasmid obtained from naturally isolated drug-resistant *Vibrio cholerae* (biotype El Tor)." Antimicrob Agents Chemother **11**(1): 13-20.