Acquisition and Evolution of SXT-R391 Integrative Conjugative Elements in the Seventh-Pandemic Vibrio cholerae Lineage


Updated information and services can be found at:
http://mbio.asm.org/content/5/4/e01356-14.full.html

http://mbio.asm.org/content/5/4/e01356-14.full.html#SUPPLEMENTAL

This article cites 51 articles, 26 of which can be accessed free at:
http://mbio.asm.org/content/5/4/e01356-14.full.html#ref-list-1

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more>>
Acquisition and Evolution of SXT-R391 Integrative Conjugative Elements in the Seventh-Pandemic *Vibrio cholerae* Lineage

Matteo Spagnoletti,a Daniela Ceccarelli,a Adrien Rieux,a Marco Fondi,c Elisa Taviani,c,d,e Renato Fani,c Mauro M. Colombo,d,e Rita R. Colwell,f,g François Ballouxh

University College London Genetics Institute, University College London, London, United Kingdom; Maryland Pathogen Research Institute, University of Maryland, College Park, Maryland, USA; Department of Biology, University of Florence, Florence, Italy; Department of Biology and Biotechnology Charles Darwin, Sapienza University of Rome, Rome, Italy; Centro de Biotecnologia, Universidade Eduardo Mondlane, Maputo, Mozambique; Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

**ABSTRACT** SXT-R391 Integrative conjugative elements (ICEs) are self-transmissible mobile genetic elements able to confer multidrug resistance and other adaptive features to bacterial hosts, including *Vibrio cholerae*, the causative agent of cholera. ICEs are arranged in a mosaic genetic structure composed of a conserved backbone interspersed with variable DNA clusters located in conserved hot spots. In this study, we investigated ICE acquisition and subsequent microevolution in pandemic *V. cholerae*. Ninety-six ICEs were retrieved from publicly available sequence databases from *V. cholerae* clinical strains and were compared to a set of reference ICEs. Comparative genomics highlighted the existence of five main ICE groups with a distinct genetic makeup, exemplified by ICE*Vch*Ind5, ICE*Vch*SXT, ICE*Vch*Ind6, and ICE*Vch*Ban11. ICE*Vch*Ind5 (the most frequent element, represented by 70 of 96 elements analyzed) displayed no sequence rearrangements and was characterized by 46 single nucleotide polymorphisms (SNPs). SNP analysis revealed that recent inter-ICE homologous recombination between ICE*Vch*Ind5 and other ICEs circulating in gammaproteobacteria generated ICE*Vch*Moz10, ICE*Vch*Ind6, and ICE*Vch*Ban11. Bayesian phylogenetic analyses indicated that ICE*Vch*Ind5 and SXT were independently acquired by the current pandemic *V. cholerae* O1 and O139 lineages, respectively, within a period of only a few years.

**IMPORTANCE** SXT-R391 ICEs have been recognized as key vectors of antibiotic resistance in the seventh-pandemic lineage of *V. cholerae*, which remains a major cause of mortality and morbidity on a global scale. ICEs were acquired only recently in this clade and are acknowledged to be major contributors to horizontal gene transfer and the acquisition of new traits in bacterial species. We have reconstructed the temporal dynamics of SXT-R391 ICE acquisition and spread and have identified subsequent recombination events generating significant diversity in ICEs currently circulating among *V. cholerae* clinical strains. Our results showed that acquisition of SXT-R391 ICEs provided the *V. cholerae* seventh-pandemic lineage not only with a multidrug resistance phenotype but also with a powerful molecular tool for rapidly accessing the pan-genome of a large number of gamma-proteobacteria.

Received 21 May 2014 Accepted 11 July 2014 Published 19 August 2014


Editor Julian Davies, University of British Columbia

Copyright © 2014 Spagnoletti et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Matteo Spagnoletti, m.spagnoletti@ucl.ac.uk.

*Vibrio cholerae*, the causative agent of cholera, is a native inhabitant of the aquatic ecosystem and represents a serious issue for global public health in developing countries. More than 200 serogroups characterized by different somatic O antigens have been described to date, but only serogroups O1 and O139 have been linked to epidemics (1). Seven cholera pandemics have been recorded since 1817 (2). The current seventh pandemic started in 1961 and is caused by *V. cholerae* O1 biotype El Tor, which replaced the Classical biotype responsible for previous pandemics (1). Several sublineages of *V. cholerae* O1 El Tor have emerged over the past 50 years, the most noteworthy being *V. cholerae* O139 Bengal (3, 4). This epidemic serogroup emerged in late 1992 in the Indian subcontinent, and its distribution is currently limited to Asia (3, 5).

Variants of *V. cholerae* O1 showing features of both Classical and El Tor biotypes have been repeatedly isolated in Asia and Africa and are collectively referred to as atypical El Tor (6). In recent years, several studies using comparative genomics have elucidated the evolutionary relationships of these genetically distinct *V. cholerae* variants (5, 7–9). It was demonstrated that they all belong to a single lineage known as the current seventh-pandemic clade (5). Mutreja et al. produced a robust phylogenetic reconstruction for the lineage responsible for the seventh pandemic, suggesting that it spread from the Bay of Bengal in at least three overlapping waves of strains that shared a common ancestor in the 1950s (8). An event at the point of transition between the wave 1 and wave 2 pandemic clones was the acquisition of an integrative conjugative element (ICE) of the SXT-R391 family (8, 10). This transition was dated to approximately 1978 to 1984 and coincides with the dating of the most recent common ancestor (MRCA) of the...
the O1 and O139 serogroups. None of the *V. cholerae* O1 strains belonging to the first wave of pandemic transmission contain an ICE (8).

SXT-R391 ICES are a well-studied family of self-transmissible mobile genetic elements, and recent findings suggest that these elements are present in a wide range of environmental *Gamma-proteobacteria*, occupying different ecological niches in the aquatic environment (11, 12). The prototypical elements, SXT and R391, were isolated from *V. cholerae* O139 in India and *Providencia rettgeri* in South Africa, respectively (13, 14). ICES are able to transfer by conjugation, using a circular intermediate, and integrate into and replicate as part of the host chromosome (15). The backbone of ICES is composed of 52 core genes, of which 25 are required for key functions of integration/excision, conjugative transfer, and regulation (10). This backbone serves as a scaffold for the integration of variable DNA located within conserved sites named variable regions (VR1, VR2, and VR3) and “hotspots” (HS1, HS2, HS3, HS4, and HS5) (10). Core genes are ~50 kb in overall size, and an additional variable genetic cargo of ~30 to ~60 kb can be inserted into the hot spots (10). Variable DNA content can confer ICE-specific features, such as multidrug and heavy-metal resistance, restriction modification systems, and alternative metabolic pathways, to the bacterial host (12, 16–18). After the identification of the SXT element in *V. cholerae* O139 (13), comparative analysis revealed that at least two ICES of the same family, ICE*Vch*Ind5 and ICE*Vch*Moz10, were circulating in *V. cholerae* O1 El Tor strains and that both showed rearrangements in their variable regions compared with SXT (10, 15). ICE*Vch*Ind5 was recognized as the prevalent ICE associated with *V. cholerae* El Tor strains circulating worldwide (19–21), but a wide-range study using genomic data is still lacking.

Recombination appears to play a crucial role in shaping ICE structure and driving its evolution (22, 23). Inter-ICE homologous recombination can take place through the exchange of large DNA fragments between two different ICES redundantly transferred in the same cell (22). Such recombination happens frequently in *in vitro* experiments and depends on host (*recA*) and ICE (*bet* and *exo*) genes (22, 24). Subsequent conjugative transfer of two ICES can occur, as only two exclusion groups (S and R) have been detected within the family (25). ICES of different exclusion groups can temporarily coexist in the same host chromosome with formation of tandem arrays (23). Arrays are excellent sub-

---

**FIG 1** BLAST atlas using ICE*Vch*Ind5 showing the genetic rearrangements in the ICE data set. Contigs from each *de novo* assembly of 96 *V. cholerae* genomes were mapped against the sequence of ICE*Vch*Ind5 (outer circle, with annotations). ICE profiles are shown with different colors according to their variable levels of genetic content within the hot spots. Green, ICE*Vch*Ind5 (74 ICES); orange, ICE*Vch*Moz10 (6 ICES); red, ICE*Vch*Ind6 (13 ICES); purple, ICE*Vch*Ban11 (1 ICE); blue, SXT (2 ICES). Absence of color represents missing genetic material compared to ICE*Vch*Ind5. Gray bars represent deletion regions gap 1 and gap 2 in the antibiotic resistance gene cluster and hotspot 5, respectively. Color gradients are proportional to the BLAST percent identity (90% to 100%).
strates for recombination and promote formation of hybrid ICEs, while backbone synteny is ensured by selection of functional ICEs able to be transferred and maintained in the bacterial population (22). Transposases, ISCR (insertion sequence common region) elements, and other insertion sequences are abundant among variable regions of SXT-R391 ICEs (10, 26); they can promote genetic rearrangements and/or acquisition of new cargo genes in the ICE genetic scaffold (27). A remarkable example of this phenomenon is provided by the organization of the SXT resistance cluster, a ~14-kb region composed of more than seven transposases and ISCR2 elements, which embeds multiple resistance determinants between _rumB_ and _rumA_ (26, 28).

ICE genetic organization, transfer, and regulation have been extensively studied (15, 29), but little is known about the ICE evolutionary dynamics of a single host species. Here we report the prevalence, diversity, and dynamics of recombination within the SXT-R391 family, as well as their appearance and spread in the seventh-pandemic _V. cholerae_ lineage. Comparative and recombination analyses were complemented with an estimate of ICE acquisition in order to reconstruct a comprehensive picture of processes underlying microevolution of this family of mobile genetic elements.

**RESULTS**

**SXT-R391 ICEs in _Vibrio cholerae_ O1 and O139 clinical strains.**

Selection of _V. cholerae_ genomes was carried out based on the year of strain isolation, geographic location, nucleotide sequence quality, and presence of an ICE (see Table S1 in the supplemental material). The overall ICE genetic organization in the data set was investigated to detect structural rearrangements within core and variable regions. All _V. cholerae_ genomes analyzed (see Table S1) were _de novo_ assembled and the assemblies screened against reference elements ICE_VchInd5, ICE_VchMoz10, and SXT, prevalent ICEs associated with _V. cholerae_ O1 and O139 worldwide (8, 10, 19). The presence or absence of variable genes with respect to reference ICEs was determined using BLAST (30), and BLAST atlases were constructed for each of the three ICEs (Fig. 1; see also Fig. S1A and S1B in the supplemental material). The BLAST atlas showing structural rearrangements of all 96 elements against the most prevalent ICE in the data set (ICE_VchInd5) is presented in Fig. 1.

Eighty-two of 96 ICEs had >98% sequence similarity with elements previously detected in clinical _V. cholerae_ isolates as follows: for ICE_VchInd5, profile 1 (see Table S1 in the supplemental material; shown in green in Fig. 1); for ICE_VchMoz10, profile 2 (see Table S1; shown in orange in Fig. 1); and for SXT, profile 5 (see Table S1; shown in blue in Fig. 1). The remaining 14 ICEs carried major rearrangements in the variable regions compared to ICE_VchInd5 (innermost circles in red and purple) (Fig. 1). Thirteen of the 14 elements revealed identical genetic contents and were retrieved from _V. cholerae_ O1 strains isolated in India, Nepal, and Thailand (8, 31) (profile 3 [see Table S1; shown in red in Fig. 1]). An additional ICE, isolated in Bangladesh in 2000, was unique in its variable DNA content (profile 4 [see Table S1; shown in purple in Fig. 1]). These novel ICEs were named ICE_VchInd6 (n = 13) and ICE_VchBan11 (n = 1), respectively.

We detected structural variations within ICE groups due to major deletions. For example, six ICEs belonging to ICE_VchInd5 (_V. cholerae_ O1, Bangladesh 1991, and Tanzania 2009), ICE_VchMoz10 (_V. cholerae_ O1, Mozambique 2005), and SXT (_V. cholerae_ O139, Bangladesh 2002) (see Table S1 in the supplemental material) shared a large deletion in the antibiotic resistance cluster comprising genes _floR_, _strB_, and _sul2_ (gap 1; Fig. 1). A second major rearrangement is represented by gap 2 (Fig. 1). Two ICEs belonging to the ICE_VchInd5 group and isolated in Nepal in 2010 (see Table S1) exhibited a deletion from _rumA_ to _traI_. The deletion comprised backbone genes _s024_, _s025_, and _s026_, as well as the entire genetic cluster located in hotspot 5 and part of _traI_ encoding a putative relaxase. The two ICEs are likely defective elements unable to transfer by conjugation, as a functional _traI_ is required for successful ICE transfer (32).

The major gaps for hotspot 4, hotspot 3, variable region 2, and hotspot 5 observed in the BLAST atlas of ICE_VchInd5 (Fig. 1) are due to the different variable-region content of each ICE. The genetic organization of the five ICE profiles is shown in Fig. 2. Since variable genes encode a large array of functions, only hotspots 3 and 5 are discussed here. Hotspot 3 contains genes that encode diguanylate cyclases in SXT, a class 4 integron carrying the trimethoprim resistance _dfrA1_ gene in ICE_VchInd5, ICE_VchInd6, and ICE_VchMoz10 and two putative genes annotated as exo-nucleases and helicases in ICE_VchBan11. Higher variability is
observed in hotspot 5, which carries a cluster of nine genes of unknown function in ICEVchInd5, a possible restriction modification system in ICEVchInd6, five open reading frames (ORFs) annotated as helicases in ICEVchBan11, and a cluster of eight ORFs, including those encoding a putative ATPase, type II restriction enzymes, and a phage growth limitation factor in ICEVchMoz10 and SXT (Fig. 2) (10).

**Genetic variability and signatures of recombination among ICE groups.** To study genetic diversity within and between the five ICE profiles, single nucleotide polymorphisms (SNPs) were called by comparing reads of all isolates to ICEVchInd5. We discovered a total of 2,967 SNPs among the 96 ICEs. The reads of each ICE belonging to a specific profile were mapped against the reference ICE selected for each group: ICEVchInd5 (n = 70, 46 SNPs), ICEVchMoz10 (n = 6, 12 SNPs), SXT (n = 2, 8 SNPs), and ICEVchInd6 (n = 13, 4 SNPs). Analysis was not done on ICEVchBan11 as it was represented by a single isolate. The vast majority of SNPs occurred between the five ICE groups, and only a limited number of SNPs was detected within each ICE group.

To assess and compare the levels of recombination in core versus noncore ICE regions, we built neighbor-net networks, as implemented by SplitsTree 4 (33). Two separate datasets were considered, one comprising the entire ICE sequence and one with only the core backbone. The five *V. cholerae* ICE profiles were used, together with those of all SXT-R391 ICEs available to date that have been isolated from environmental non-O1/non-O139 *V. cholerae* strains and other Gammaproteobacteria species (i.e., *Providencia rettgeri*, *Providencia alcalifaciens*, *Shewanella putrefaciens*, *Proteus mirabilis*, *Photobacterium damsela*, and *Vibrio fluvialis*) (see Table S1 in the supplemental material). The resultant neighbor-net networks were characterized by extensive reticulation, as shown by analysis of both whole ICE sequences and core regions only (see Fig. S2), and the results of the Phi test for recombination were highly statistically significant (P < 0.0001) in both analyses. The presence of extensive reticulation in the core gene set (see Fig. S2B) points to inter-ICE homologous recombination as an important force shaping ICE structures.

**Evidence of recent inter-ICE homologous recombination in seventh-pandemic *V. cholerae*.** The distribution of SNPs along the 52 conserved core genes of five representative ICEs was further investigated by directly mapping reads from ICEVchBan11, ICEVchInd6, ICEVchMoz10, and SXT isolates against the conserved backbone of ICEVchInd5. To identify possible donors to hybrid ICE formation, reads were also mapped against core genes of the same additional eight ICEs used for network analysis. Potential recent inter-ICE recombination events were identified by detecting dense clusters of SNPs separated by clear-cut boundaries (34).

This analysis indicated that ICEVchBan11 was generated by recombination between ICEVchInd5 and ICEPmnuUSA1, which was isolated from a uropathogenic *Proteus mirabilis* strain (Fig. 3A). Similarly, ICEVchInd6 underwent hybrid rearrangement between ICEVchInd5 and ICEVfInd1 isolated from a clinical *V. fluvialis* strain. Dense clusters of SNPs between *traD* and *s068* were detected, while the rest of the sequence showed no SNPs compared to ICEVchInd5 (Fig. 3B). ICEVchMoz10 showed recombination similar to that seen with ICEVchInd6, sharing large regions of the backbone with ICEVchInd5 and ICEVfInd1 (Fig. 3C). Analysis revealed no SNPs in the first -7 kb (*xis-to-rumB* region) of ICEVchInd5, ICEVchBan11, ICEVchInd6, and ICEVchMoz10 (Fig. 3A to C). Conversely, the distributions of SNPs in SXT and ICEVchInd5 did not point to evidence of recent homologous recombination affecting backbone genes of the two elements (Fig. 3D).

**Acquisition date of ICEVchInd5 and SXT.** Acquisition of SXT-R391 ICEs was dated by Mutreja and colleagues as occurring between 1978 and 1984 by the use of genome-wide SNP data within a Bayesian framework (8). Those authors estimated the lower boundary for acquisition of ICEs as the time to the MRCA (TMRCA) of all *V. cholerae* strains carrying ICEs. In this study, we refined the dating analysis by estimating directly the age of ICEVchInd5 and that of SXT rather than the TMRCA of the strains carrying them. We included ICE sequences from the ICEVchInd5 and SXT groups, but ICEs shown to be recombinants (ICEVchMoz10, ICEVchInd6, and ICEVchBan11) were not considered. Nucleotide alignments were built on ICEVchInd5-like and SXT-like ICEs separately, in order to reconstruct two phylogenies and date the acquisition of each of the two groups of ICEs independently.

Previous analysis supported identification of ICEVchInd5 as the most prevalent ICE, i.e., present in 74 of 96 samples isolated between 1989 and 2010, and likely the oldest ICE in our data set. Inferences were realized using BEAST (35) on an alignment of 71 ICEs (70 ICEs plus ICEVchInd5, excluding four elements with deletions) (see Fig. S3A in the supplemental material). The time of acquisition of ICEVchInd5 (i.e., the TMRCA) was dated to 1985 (95% highest posterior density [HPD] interval, 1980 to 1989) (Fig. 4A). Posterior distributions of substitution rates are given with confidence intervals in Fig. S4. To ensure that our estimate for the time of acquisition of ICEVchInd5 elements was not affected by cryptic recombination, we tested for recombination between ICEVchInd5 elements using five different algorithms implemented in Recombination Detection Program 4 (RDP4) (36) and the Phi test from the PhiPack software (33). While the results of all tests obtained with RDP4 turned out to be statistically nonsignificant, those of the Phi test were borderline significant (P = 0.039). Using SplitsTree4 software and manually inspecting the alignments, we created a reduced alignment by removing nine ICEs that were showing potential recombination (see Table S1). This new reduced alignment of 62 ICEs led to nonsignificant Phi test results (P = 0.302) and a median TMRCA qualitatively similar to that of the whole alignment (results not shown), indicating that the dating analysis was not affected by cryptic recombination between ICEVchInd5 elements.

We repeated similar analyses of four sequences of SXT-like ICEs extracted from *V. cholerae* O139 isolates between 1992, the year of its first appearance, and 2002. For this set of ICEs, acquisition was dated to 1992 (95% HPD, 1990 to 1992) (Fig. 4B).

**DISCUSSION**

The availability of whole-genome sequence data from *V. cholerae* clinical isolates offers a remarkable opportunity to study the microevolution of SXT-R391 ICEs in a single host species. To date, most of the studies have focused on the evolution of the current pandemic lineage of *V. cholerae*, where SXT-R391 ICEs are ubiquitous (5, 8, 9). Both Chun et al. and Mutreja et al. have hinted at the acquisition of SXT-related ICEs as an important milestone in *V. cholerae* genome dynamics and evolution (5, 8). Our report refines this scenario and offers new insights into the dynamics of acquisition and recombination of the SXT-R391 family of ICEs.
In this paper, we demonstrate that both multiple acquisitions and homologous recombination were important in shaping the current ICE diversity in the *V. cholerae* seventh-pandemic lineage. We infer that two distinct elements, ICE*Vch*Ind5 and SXT, were independently acquired by ICE-free *V. cholerae*. After that, homologous recombination with other ICEs, likely acquired by conjugation, played an important role in further shaping the ICE genetic architecture and promoting the formation of new hybrid elements in *V. cholerae* O1 El Tor.

We classified ICEs currently circulating among *V. cholerae* clinical strains into at least five different groups, representing elements with identical genetic cargoes in their variable regions (Fig. 2). Genetic variation within ICE groups is low, with the exception of a few major deletions, mostly located in the antibiotic resistance cluster (Fig. 1). Similar deletions were documented in ICEs from the Haitian epidemic from 2011 and 2012, indicating that this region is prone to frequent rearrangements (37). This is not entirely surprising given the highly recombinogenic nature of the ICE antibiotic resistance cluster encoding Tn3-like and ISCR2 elements recognized as powerful gene capture and movement tools (26, 38, 39).

The phylogenetic methods we employed to test for single versus multiple ICE acquisitions and to estimate the dates of these events rely on the analyzed sequences not having been affected by recent genetic recombination. Previous studies have demonstrated that SXT-R391 ICEs generate their own genetic diversity by mediating recombination between two ICEs arranged in a tandem array (23). ICE-encoded proteins Bet and Exo, working in a RecA-independent homologous recombination system, mediate the formation of hybrid ICEs and are probably responsible for most of the genetic variability observed in the SXT-R391 family (22). However, the operation of this process with identical or sim-
ilar elements is inhibited by the ICE entry exclusion system, which prevents redundant transfer by converting the host cells into poor recipients (25). As a consequence of this constraint, homologous recombination is allowed only between ICEs of different exclusion groups. Moreover, ICE loss from the host chromosome followed by reacquisition of another ICE is a most unlikely event as all the ICEs described in *V. cholerae* encode a toxin-antitoxin system that ensures ICE maintenance (40).

Data indicating the absence of substantial recombination within exclusion groups and secondary loss of ICEs are supported by the temporal patterns we report for distribution of ICEs and their genetic organization in the seventh-pandemic lineage. While we detected extensive recombination post-ICE acquisition by seventh-pandemic *V. cholerae* strains, the backbone genes of ICE*VchInd5* and SXT did not show any recent signature of inter-ICE homologous recombination (Fig. 3D). Our analyses of ICE alignments dated the acquisitions of ICE*VchInd5* and SXT to 1985 and 1992, respectively. Comparison with the *V. cholerae* genotype phylogeny proposed by Mutreja et al. (8) indicates that the first isolates carrying ICE*VchInd5* are represented by wave 2 strains, an outgroup of the wave 3 lineage, where ICE*VchInd5* is dominant. The MRCA of 1981 for O1 and O139 serogroups used as a proxy for the acquisition of the SXT-R391 ICEs by Mutreja et al. (8) predates our estimation for the acquisition of ICE*VchInd5* (1985), although the confidence intervals partially overlap. However, the date proposed by Mutreja et al. does not correspond to the acquisition of SXT and/or the O139 antigen but marks only the split between the El Tor progenitor of the O139 serogroup and the rest of the seventh-pandemic lineage. Interestingly, when we looked at the two O1 strains closest to the O139 sublineage (strains PRL5 and A109 [accession no. ERS013145 and ERS013173, respectively]) (8), we found no ICEs in their genomes. Our estimation for the acquisition date of SXT is set 10 years later and precedes the first outbreak of *V. cholerae* O139 in the Indian subcontinent by just 1 year. While we are confident about the accuracy of the dating of ICE*VchInd5* acquisition, our date for SXT should be treated with caution because it was estimated using the only four SXT element sequences currently available. This small set may not capture the entirety of the genetic diversity of the clade and hence may correspond to a more recent MRCA. Thus, the date of acquisition for SXT presented here is likely to represent an underestimation.

In addition to SXT and ICE*VchInd5*, we detected significant recent inter-ICE homologous recombination in our analysis of other ICE profiles. Network analysis (see Fig. S3 in the supplemental material) and analysis of the distribution of SNPs along 52 conserved ICE core genes (Fig. 3) revealed that ICE*VchInd6*, ICE*VchBan11*, and ICE*VchMoz10* are likely the result of recent inter-ICE homologous recombination events involving ICE*VchInd5* and other ICEs circulating in other *Gammaproteobacteria*, such as ICE*PmniUSA1* and ICE*VflInd1* (Fig. 3) isolated from *P. mirabilis* and *V. fluvialis*, respectively. While it is difficult to identify recombination donors unequivocally because of the mosaic nature of SXT-R391 ICEs, the observed SNP patterns strongly support the hypothesis that the three ICEs were recently generated within *V. cholerae* strains likely previously carrying ICE*VchInd5* (Fig. 3). No doubt, the aquatic environment and its bacterial inhabitants provide SXT-R391 ICEs with a genetic pool from which to draw new genetic material and expand their gene repertoire.

Our data indicate that ICE acquisition by seventh-pandemic *V. cholerae* was not a single event but that SXT and ICE*VchInd5* were independently acquired by ICE-free *V. cholerae* strains. A schematic representation of the acquisition of the SXT-R391 ICEs is presented in Fig. 5. Furthermore, the analyses presented here reveal that ICE acquisition by *V. cholerae* promoted subsequent ICE rearrangements and generated significant variability on the genomic scale. In fact, the ability of SXT-R391 ICEs to mobilize large regions of host DNA and/or ICE-associated genomic islands was previously documented (41). Katz et al. (37) reported a large (~400-kb) inversion around the ICE integration region of a Haitian *V. cholerae* isolate. The recent description of ICE-encoded extended-spectrum cephalosporin resistance (*bla<sub>CMY-2</sub>*) in a clinical *Proteus mirabilis* strain isolated in Japan (42) further highlights the ability of this family of mobile genetic elements to acquire and spread novel drug resistance genes. The impact of ICE acquisition gains noteworthy relevance in light of recent findings showing *V. cholerae* seventh-pandemic clones to be poorly transformable (37).
In summary, these results illustrate how ICE integration promotes both element-associated and chromosomal rearrangement, thus acting as a horizontal gene transfer hot spot for the host strain.

**MATERIALS AND METHODS**

ICE data set creation, *de novo* assembly, and comparative analysis. The NCBI archive and the Sequence Read Archive (SRA) were screened for SXT-R391 ICE sequences. A list of all genome sequences and reference ICEs used in this study is presented in Table S1. All downloaded reads were refiltered and trimmed using a FASTX-Toolkit with a minimum Q20 quality-score cutoff before performing the assembly step. The resultant data set was then assembled *de novo* using Velvet 1/1/06 (43). Contigs related to ICE*Vch*Ind6 and ICE*Vch*Ban11 were extracted from the genomic assemblies of *V. cholerae* 4605 and *V. cholerae* 4672 (accession no. ERS013257 and ERS016137, respectively) and used as references for further analysis. The backbone sequences of ICE*Vch*Ind5, SXT, R391, ICE*Pdu*Spa1, ICE*Vch*Hai2, ICE*Vch*Mex1, ICE*Pnu*P01, ICE*Pal*Ban1, ICE*Pml*SA1, ICE*Vch*Ban11, ICE*Vfl*Ind1, ICE*Vch*Ind6, and ICE*Vch*Moz10 used for recombination analysis were built by concatenating ICE core regions and removing all variable genes in Artemis (44). BLAST atlas maps were built by comparing concatenated *de novo* assemblies with reference sequences of ICE*Vch*Ind5 (GQ463142), ICE*Vch*Moz10 (AIHZ200000000), and SXT (AY055428) using BLASTn with a >90% identity threshold within the CGView Comparison Tool (30, 45). The Artemis Comparison Tool and CGView were used for further comparisons and to visualize the maps (45–47).

**Variant calling and homologous recombination analysis.** Reads were organized in groups reflecting similarity with their reference sequences based on the results of the analysis performed on the *de novo* assembly. Bowtie2 (48) was used to map processed reads belonging to samples of each ICE group against their reference sequences (ICE*Vch*Ind5, ICE*Vch*Ind6, ICE*Vch*Moz10, and SXT) and all 96 samples against ICE*Vch*Ind5. The resultant sam files were converted to bam, coordinate sorted with Picard Tools (http://picard.sourceforge.net), and imported in CLC Genomics Workbench 6.0 for variant calling. Clark Probabilistic Variant Caller was used for this purpose with a variant probability threshold of 90% and with minimum sequence coverage of 10×.

SNP calling for recombination analysis was performed with the same settings described above, mapping representative reads corresponding to ICE*Vch*Ban11 (accession number ERS016137), ICE*Vch*Ind6 (accession number ERS013257), ICE*Vch*Moz10 (accession number ERS013126), and SXT (accession number ERS013124) profiles against the 52 core genes of ICE*Vch*Ind5 and a set of eight additional ICEs found in non-O1/non-O139 *V. cholerae* strains and other Gammaproteobacteria (see Table S1 in the supplemental material). SNPs were filtered to include only synonymous SNPs to avoid potential selection bias introduced by strong selective pressure on specific ICE backbone genes. SNPs were called only on coding regions to avoid length discrepancies in the intergenic regions of the ICE backbone.

The initial search for recombination was conducted by building neighbor-net and split-decomposition networks using the SplitsTree4 software (33). A Phi test for recombination was applied using a permutation test of 100,000 iterations and PhiPack software (49). Additional tests for recombination detection were conducted using Recombination Detection Program 4 (RDP4) v.4.35 (36) and five different methods (RDP, MaxChi, Bootscan, Chimera, and Siscan). All the tests were executed with default parameters, and the multiple-comparison corrected  P-value cutoff value was set to 0.05.

**Bayesian analysis.** ICE consensus sequences for each sample were extracted from the reads mapped against ICE*Vch*Ind5 and SXT using a coverage threshold of >10× and the base-calling quality scores for handling SNP conflicts as implemented by CLC Genomic Workbench. Sequences were then aligned with MUSCLE (50), PartitionFinder (51) was used to estimate the best-fit nucleotide substitution model for each nucleotide subset. Bayesian phylogenetic analyses were performed with BEAST 1.7.4 (35) on both whole elements and various nucleotide partitions (coding positions 1 and 2, coding position 3, noncoding sites, and pseudogenes). Substitution and clock models were unlinked in BEAST, while the tree topologies of the four nucleotide subsets were assumed to be the same. A separate substitution model and molecular clock but a single tree was estimated for each partition. We assumed a lognormal relaxed clock to allow variation in rates among branches. To minimize prior assumptions about demographic history, we adopted a Bayesian skyline plot approach in order to integrate data over different coalescent histories. Rate variations among sites were modeled with a discrete gamma distribution with four rate categories. The tree was calibrated using tip dates only, with sample time spans ranging from 1992 to 2002 and from 1989 to 2010 for SXT-like and ICE*Vch*Ind5-like elements, respectively. BEAST analyses were performed on two datasets. The first contained 71 ICE*Vch*Ind5-like element sequences and the second the 4 SXT-like sequences. A third run was performed on a subset of 62 ICEs of the ICE*Vch*Ind5 group to validate the results. Flat priors (i.e., uniform distributions) were applied for the substitution rate (1.10^{-10} to 1.10^{-3} substitutions/site/year) as well as for the age of any node in the tree, including the height of the root (1500 to 1989 and 1500 to 1992 for the MRCAs of the ICE*Vch*Ind5 and SXT groups, respectively). Posterior distributions of parameters, including divergence times and substitution rates, were estimated by Markov chain Monte Carlo (MCMC) sampling in BEAST. For each analysis that we ran, we combined five independent chains from which samples were drawn.
every 5,000 MCMC steps from a total of 50,000,000 steps, after a burn-in

count of 5,000,000 steps was discarded. Convergence to the stationary
distribution and sufficient sampling were checked by inspection of poste-
rionary samples. The best-supported tree was estimated from the combined
samples using the Maximum clade credibility method implemented in
Tree Annotator after a burn-in level of 10% was discarded.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/

S1, TIF file, 1.1 MB.
S2, TIF file, 0.2 MB.
S3, TIF file, 0.8 MB.
S4, TIF file, 0.2 MB.
Table S1, PDF file, 0.3 MB.
Figure S2, TIF file, 0.2 MB.

ACKNOWLEDGMENTS

This work was supported by European Research Council grant ERC
260801—BIGIDEA (to F.B.). M.S. was supported by a fellowship from the
Institute Pasteur—Fondazione Cenci Bolognetti, Rome, Italy. D.C.
was supported by National Institutes of Health grant no. 2RO1A1039129-
31A2.

We are thankful to Genevieve Garriss (Institute Pasteur, Paris, France)
for insightful comments on the manuscript.

REFERENCES

http://dx.doi.org/10.1073/pnas.0337468100.

lated from the mucus of the coral Fungia echinata from Andaman Sea.


