TITLE: Comprehensive functional analysis of the eighteen Vibrio cholerae N16961 toxin-antitoxin systems substantiates their role in stabilizing the superintegron.

RUNNING TITLE: Analysis of Vibrio cholerae toxin-antitoxin modules

Naeem Iqbal 1, 2, §, Anne-Marie Guérout 1, 2, Evelyne Krin 1, 2, Frédérique Le Roux 3 and Didier Mazel 1, 2, #

1 Institut Pasteur, Unité Plasticité du Génome Bactérien, Département Génomes et Génétique, F-75015 Paris, France
2 CNRS, UMR3525, F-75015 Paris, France
3 Station Biologique, Place Georges Teissier, 29682 Roscoff Cedex, France

#, Corresponding author: mazel@pasteur.fr

§: Present address: Centre de recherche en infectiologie, Centre Hospitalier Universitaire de Québec, Québec G1V 4G2, Canada.
Abstract

The role of chromosomal toxin-antitoxin (TA) systems, ubiquitous within the genomes of free-living bacteria, is still debated. We have scanned the *Vibrio cholerae* N16961 genome for class 2 TA genes and identified 18 gene-pair candidates. Interestingly, all but one are located in the chromosome 2 superintegron (SI). The single TA found outside of the SI is located on chromosome 1 and is related to the well-characterized HipAB family known to play a role in antibiotic persistence. We investigated this clustering within the SI and its possible biological consequences by performing a comprehensive functional analysis on all of the putative TA systems. We demonstrate that the 18 TAs identified encode functional toxins and that their cognate antitoxins are able to neutralize their deleterious effects when expressed in *Escherichia coli*. In addition, we reveal that the seventeen predicted TA systems of the SI are transcribed and expressed in their native context from their own promoters, a situation rarely found in integron cassettes. We tested the possibility of interactions between non-cognate pairs of all toxins and antitoxins and found no cross-interaction between any of the different TAs. Although these observations do not exclude other roles, they clearly strengthen the role of TA systems in stabilizing the massive SI cassette array of *V. cholerae*.

Importance

The chromosomal toxin-antitoxin systems have been shown to play various, sometimes contradictory roles, ranging from genomic stabilization to bacterial survival via persistence. Determining the interactions between TA systems hosted within the same bacteria is essential to understand the hierarchy between these different roles. We identify here the full set of class 2 TAs carried in the *Vibrio cholerae* N16961 genome and found they are all, with a single exception, located in the chromosome 2 superintegron. Their characterization, in terms of functionality, expression and possible cross-interactions, support their main role as being the stabilization of the 176 cassette-long array of the superintegron, but does not
exclude dual roles, such as stress response elements, persistence, and bacteriophage defense through abortive infection mechanisms.
Introduction

Toxin-antitoxin (TA) systems were discovered in 1983 on plasmid F of Escherichia coli (1) and were shown to be involved in stable plasmid maintenance by post-segregational killing, a mechanism distinct from replication and partition. These systems typically consist of an operon of two genes which encode a toxin that targets an essential cellular function and an antitoxin that binds to and inhibits the toxin. Toxin activity is regulated through differential stability of the stable toxin and the labile antitoxin; loss of a TA system by the progeny during cell division results in cell death by the action of the stable toxin (for reviews, see (2, 3)). The antitoxin, in most cases, also acts as a transcriptional autorepressor of the operon and the degradation of the antitoxin results in the transcriptional activation of the TA operon.

Currently there are three types of TA systems, which are classified according to the nature of their antitoxin (4). For the type 1 system, such as tisA/tisB (5) and hok/sok (6), toxins are hydrophobic proteins and antitoxins are sRNAs. Type 2 TAs are the most represented among TA systems and both their toxins and antitoxins are present as proteins (7). The third type of TA system is currently only represented by the toxIN locus, which was discovered in 2009 on a cryptic plasmid of Erwinia carotovora (8). It consists of a protein-RNA toxin-antitoxin pair, where RNA antitoxin ToxI inhibits the toxicity of toxin ToxN by interfering with its biochemical activity (9). Type 2 TA systems are ubiquitous in prokaryotic genomes and have more recently been identified as genuine components of the chromosomes of most free-living bacteria (7, 10, 11) with up to 88 predicted TAs in the Mycobacterium tuberculosis genome (12). Several hypotheses have been proposed for the biological roles of chromosomal TAs, but these are not yet clear and remain under debate. TAs have been proposed to be physiological or developmental regulators of processes such as programmed cell death (PCD), growth and/or development and persister cell formation. PCD refers to cell death mediated by an intracellular death program through a cascade of signalization in response to an external stress (13), however, this hypothesis is controversial, as several groups have failed to reproduce the original observations (see (4)). Gerdes and coworkers
proposed that TA loci function as stress response elements and that they help the cells cope with nutrient starvation by modulating their growth (14). In the context of their role for survival, TA systems have been proposed to induce a dormant cellular condition termed persistence (15-17). Persisters are non-growing, dormant cells exhibiting low levels of translation and this state could allow them to overcome stress conditions, such as antibiotic treatment (reviewed in (18)). TAs have also been involved in bacteriophage defense through abortive infection mechanisms, where bacteria commit suicide to prevent phage spreading (19-21). Two roles for TAs regarding genetic stability have also been proposed. The first is as an anti-addiction module, in which TAs can protect the host genome from colonization by an incoming mobile element or from a plasmid carrying a TA from the same functional family by allowing its harmless loss through neutralization of the invading toxin by the chromosomal antitoxin (22). This hypothesis is supported by the fact that some chromosomal TA systems can cross-interact with their plasmidic homologs, i.e. a chromosomally encoded toxin is neutralized by a plasmid encoded antitoxin or vice versa (for examples, see (22-24).

Recently, we have shown that the chromosomally encoded \textit{phd/doc} system of \textit{V. cholerae} can interact with the homologous \textit{phd/doc} system encoded by phage P1 (25). These observations emphasize that chromosomal TA systems may act as "exclusion" systems to protect the bacteria from an excess of invading mobile genetic elements (plasmids, transposons, phages) carrying identical TA systems (22, 24). In addition, chromosomal TA systems, by interfering with their homologs encoded by mobile genetic elements, could drive the evolution of the latter ones to select the systems having toxins that are no longer recognized by the anti-addiction modules. Finally, similar to their role in plasmid stabilization, TAs have been proposed to stabilize chromosomal regions by preventing accidental deletions, especially when located in unstable segments such as mobile genetic elements (26-28). The \textit{mosAT} and \textit{tad-ata} systems have been shown to be responsible for the stabilization of two different integrative conjugative elements (ICEs) by post-segregational killing as in plasmids (28, 29). The insertion of two systems (\textit{RelBE1} and \textit{ParDE1}) from \textit{Vibrio vulnificus} stabilized a 165 kb non-essential chromosomal region in \textit{E. coli} (27).
Strikingly, TAs are very common in cassettes of chromosomal integrons, especially those TA systems known to be involved in plasmid stabilization (26, 27, 30), and are present often in high numbers, especially in superintegrons (SIs) (31).

Bacterial chromosomes are subject to lateral DNA exchange through genomic parasites such as integrons, conjugative transposons and bacteriophages. Among these, integrons are very efficient in capturing and rearranging the ORFs embedded in gene cassettes and converting them to functional genes by ensuring their correct expression. Vibrio SIs gather hundreds of highly diverse gene cassettes (32), with a large majority of ORFs of unknown functions. The analysis of available Vibrionales genomes in 2007 led to the identification of 1677 gene cassettes (33). Cassettes are generally promoterless and are thought to constitute a silent reservoir of adaptive functions (26, 33). Expression of these cassettes relies on their relative position from the Pc promoter and any silent cassette can be called on for expression through recombination to an attI or attI-proximal location (for a review see (31)). Integron integrase expression, which is commonly controlled by the SOS response (34, 35), governs cassette recombination. Thus cassette recombination is triggered in times of stress, while in non-stressful conditions, the cassette arrays stay steady and the selective pressure exerted on distal silent cassettes is assumed to be low if not non-existent. Previous analysis of the V. cholerae N16961 genome had led to the identification of 13 type 2 TA systems, all localized in the SI (7, 26). Interestingly, these potential TA cassettes are among the few cassettes that seem to carry their own promoter (25, 31). The fact that these cassettes are most likely expressed led us to propose that they could play a stabilizing role by preventing cassette(s) loss through homologous recombination between identical attC sites (the cassette recombination site) or between copies of repeated identical cassettes (26, 27). Chromosomal regions such as SIs are not generally essential for bacterial survival but rather encode for functions that aid the bacteria in rapidly adapting to a changing environment. In this way, TA systems contribute to maintain the potential genes that can confer a major selective advantage to their host under stress, and these genes may disappear rapidly in steady growth conditions in the absence of TA modules. Previous studies of the V. cholerae TA
cassettes of the HigBA, ParDE and Phd/Doc families gave further support to this role (25, 30, 138, 36). However, these properties do not restrain cassette microevolution and the role of TA systems as anti-addiction modules. Here, we re-analyzed the N16961 genome and identified 5 potential new TA systems in addition to the 13 previously listed. Strikingly, four candidates were also embedded within integron cassettes, while the fifth is related to HipAB and does not seem to be embedded within a mobile genetic element. Among the previously described potential TA cassettes present in the *V. cholerae* N16961 SI, 3 parDE, 2 higBA and the phd-doc TA cassettes had already been shown to be functional TA systems (25, 30, 36, 37). In this study, we established the functionality of all 11 remaining candidates, as well as the sedentary HipAB homologue we identified. Thus, we demonstrate for the first time that all of these 17 SI cassette systems are functional, and that they are all expressed in their native context, demonstrating that they all carry their own promoter, a rare property for integron cassettes. Importantly, we show that these systems are very specific. Indeed, we found that even for systems belonging to the same TA family, there is no cross-interaction between noncognate toxins and antitoxins. The roles that these multiple TA cassettes can play in this genomic structure are discussed in light of their lack of cross-interactions and specific pattern of transcription.
Results

156 TA System identification in the *V. cholerae* N16961 genome.

157 We had previously identified and characterized several TA cassettes carried in different chromosomal integrons of various *Vibrio* genomes (25-27), and suspected their presence in chromosomal integrons of other genera such as *Treponema denticola* (31). A previous analysis of the *V. cholerae* N16961 genome had identified a total of 13 TA candidates, all located within the SI region (7), a unique property among *Vibrios*. Indeed, analysis of the other available *Vibrio* genomes showed that TA candidate distribution was not restricted to the integron cassette array, and TAs could be found in other chromosomal contexts (see (31)). This led us to re-analyze the genome of N16961 for the presence of potential TAs.

166 Starting from the data obtained using RASTA (38) on the N16961 genome listed in the toxin-antitoxin database TADB (39), we carefully checked the predicted list of 19 potential TA systems, where 16 were carried in integron cassettes, 1 on chromosome 2 but outside of the SI, and 2 on chromosome 1. We found that 2 of the 3 TAs which were not in the SI cassette array were erroneous predictions. The third TA candidate carries gene VC0815, whose 453 amino acid long product is related to HipA, and is associated with VC0814, a 160 codon gene with no known homologues, but which contains an HTH domain and whose translation start site overlaps the final codons of VC0815, an organization commonly found in TA operons.

176 In parallel, we searched all SI cassettes carrying two ORFs, and recovered 17 cassettes, which included the 16 cassettes listed in TADB (Fig. 1) and the *phd-doc* cassette that we previously characterized. Ten of the seventeen TA systems belong to four well-characterized families (7) and three cassettes were found to be hybrid TA systems (ie. a TA system whose toxin and antitoxin do not belong to the same family according to the TA search software described in (11)). These hybrid systems correspond to a *phd-parE* gene pair (VCA0422-VCA0423), a *stbD-relE* (VCA0477-VCA0478), and a *relB-parE* (VCA0488-VCA0489). The last four predicted TA cassettes have at most one gene having poor homology to a known
toxin or antitoxin. The identical VCA0318-VCA0319 and VCA0481-VCA0482 cassettes encode proteins having respective homology to N-acetyltransferases and ribbon-helix-helix (RHH) transcription factors; the first protein was found to be a toxin and the second its antitoxin (see below). Cassette VCA0486-VCA0487 also encodes an RHH transcription factor-like protein, which was also found to be an AT for the VCA0487 toxic product (also carrying an N-acetyltransferase motif). Lastly, the VCA0332-VCA0333 cassette was found to express an RHH transcription factor-like protein acting as an AT for the VCA0332 toxin which belongs to an uncharacterized family of conserved proteins, COG2929.

Functionality of the sedentary HipAB candidate identified in chromosome 1 of *V. cholerae*.

Both candidate HipAB genes, VC0814 encoding the putative antitoxin and VC0815 encoding the HipA-like protein, were independently cloned in compatible plasmids under the control of the P\textsubscript{lac} promoter for VC0814 in pSU18, and under the arabinose-inducible P\textsubscript{BAD} promoter of the low-copy-number vector pBAD43 (a pSC101 derivative, (40)) for VC0815. Both types of plasmids were introduced into *E. coli* DH5\textalpha cells. We found that cells carrying the VC0814 plasmid were able to grow in the presence of IPTG, showing that this gene was not toxic for *E. coli*. Transformants carrying the VC0815 plasmid were selected in the presence of 1% glucose to repress the P\textsubscript{BAD} promoter. Plating these cells on LB agar + 0.2% arabinose resulted in cell death, showing that VC0815 was indeed encoding a powerful toxin. We then co-transferred the two plasmids into DH5\textalpha cells, repeated the arabinose induction test, and found that in the presence of expressed VC0814, VC0815 was no longer toxic. Thus VC0814-VC0815 forms a genuine TA system.

Functionality of the toxins carried in the superintegron cassettes.

Among these 17 potential TA systems, 3 cassettes were found in two identical copies. These correspond to 2 parDE loci (parDE1 and parDE3), 2 relBE loci (relBE2 and relBE4) and to the new TA encoded by VCA0318-VCA0319 and VCA0481-VCA0482, thus downsizing the
number of different systems to 14 (Fig. 1). To test the functionality of the toxins of *V. cholerae* N16961 chromosomal TA modules, each of its 14 putative toxin genes was cloned into pBAD43 (Table S1), as described above for VC0815, and introduced into *E. coli* DH5α cells in the presence of glucose. The toxin activity of each candidate was first tested in an *E. coli* background to avoid potential interference resulting from the presence of resident chromosomal toxin and/or antitoxin genes. Under glucose repression, *E. coli* recombinants carrying each of the fourteen toxin genes were able to grow on LB-agar plates. When replicated on plates containing 0.2% arabinose, the cells did not grow for any of the tested toxins, except those expressing VCA0423. The gene coding for this putative toxin ends in the middle of the VCR (the *V. cholerae* cassette specific attC site). This rare configuration generally reflects a frame shift mutation or a fusion between unrelated cassettes. Comparison with other *V. cholerae* genomes revealed the origin of this odd feature. Indeed, in strains O395 / 569B, we found that the VCA0423 equivalent (VCO395-0843, hereafter called VCA0423c) has a different and shorter C-terminal end, with a stop codon properly located upstream of the VCR. The N16961 VCA0423 odd C-terminal corresponds to the sequence of a cassette (VCO395-0861) located 18 cassettes downstream of VCO395-0843. We then tested the functionality of VCA0423c, and found that it was toxic. Thus, it is very likely that an accidental fusion between these two cassettes has occurred in an ancestor of N16961, inactivating the toxin gene while keeping the promoter and antitoxin gene operational. The rest of the study was made with the O395 VCA0423c toxin gene.

Hence, all the 14 putative toxin genes of *V. cholerae* TA modules are active, and thus toxic, in *E. coli* DH5α, confirming previous characterizations of the HigBA, ParDE and Phd-Doc subset (25, 30, 36, 37, 41).

**Functionality of the antitoxins of the *V. cholerae* TA cassettes**

Each of the fourteen putative antitoxin genes was cloned into a high-copy-number expression vector pUC18 under the control of the *P* _{lac} promoter and introduced into *E. coli* DH5α cells (see materials and methods, Table S1). The functionality of the 14 antitoxins was...
tested through the fact that expression of a functional antitoxin should prevent cell death due
to toxin expression triggered in the presence of arabinose. *E. coli* cells containing each
antitoxin cloned into pUC18 were transformed with the cognate toxin compatible plasmid and
grown in the presence of glucose. Co-transformants were then replicated on medium
containing either arabinose or glucose. Growth in these two conditions was found to be
similar for the 14 tested TA modules. This showed that all of the TA cassettes found in the *V.
cholerae* SI encode functional toxins and antitoxins.

**Functionality of the toxins of TA modules in *V. cholerae***

Each of the 14 toxins found to be toxic in *E. coli*, was introduced by transformation into its
natural host *V. cholerae* N16961, in order to establish if the TA systems were expressed from
their original SI location, and whether expression of the antitoxin could prevent the toxin’s
deleterious activity. The functionality of the P_BAD promoter in *V. cholerae* has already been
established (see for example (30, 36, 41, 42) and pBAD43 derivatives carrying the toxin
genes were introduced by transformation and selection on glucose-containing media. Then,
the expression of the toxins was induced by replica plating on a 0.2% arabinose-containing
medium and cell viability was determined. We found that the chromosomally encoded
antitoxins counteracted the toxicity of the overexpressed toxins and that the cells showed a
similar growth on arabinose as on glucose for all the toxins tested. We used the ccdAB TA
system previously characterized in the *V. fischeri* SI (26), as a control of our setup. Indeed,
N16961 does not carry a ccdAB locus, and we found that once transformed with a
pBAD43::ccdB plasmid, transfer from glucose- to arabinose-containing media triggered cell
death.

**Expression of the TA cassettes in the native SI context.**

In order to formally demonstrate the expression of these TA cassettes and confirm the
results obtained in the N16961 transformation experiments described above, we performed a
semi-quantitative RT-PCR, with or without AMV reverse transcriptase, for each of those TA,
with the exception of phd-doc, as its expression in the same conditions has previously been published (25). Amplification of the expected products was only obtained after reverse transcription (Fig. S1), demonstrating their efficient transcription. Furthermore, their expression in cells harvested at mid-log or early stationary phases was analyzed by real-time quantitative RT-PCR on total RNA. We found that apart from higBA2 and VCA0486-0487, all cassettes have a higher expression during exponential growth compared to stationary phase, the highest ratio being obtained for the duplicated parDE1/3 cassette (Table 1).

**Toxin activity assays in E. coli.**

In order to better understand the activity of these toxins, as we could not test their toxicity in *Vibrio cholerae*, we investigated the activity of eleven of these toxins in *E. coli* under conditions where toxins were induced in a liquid medium for specific periods of time (60-180 min), with the cells then being plated on a glucose-containing medium and incubated to grow with the toxins repressed. We found that all toxins induced at least a 3-log drop in CFUs in the first hour of induction, but in all cases a small fraction of cells was able to survive at least as long as 3 hours following induction of the toxin (Fig. S2). We verified the presence of active toxin in these cells, as when these surviving cells were replicated on LB-agar plates containing either glucose or arabinose, in all cases more than 90% of cells were unable to grow on arabinose (data not shown). This demonstrates that the toxins were still toxic for these cells and they were not spontaneously inactivated mutants. The consistent blockage of the toxicity by the expression of each protein’s cognate antitoxin suggests that this toxicity is due to a discrete activity that can be neutralized, and is not a nonspecific consequence of protein overexpression.

**Cross-interactions between all toxins and antitoxins carried in the seventeen superintegron cassette**

Among these 17 TA systems, 10 belong to four different, well-characterized families i.e. Hig, Rel, Par and Phd-Doc. The HigB and RelE toxins belong to the RelE superfamily, but their
antitoxins are unrelated (11). The Hig family contains two remotely related higBA loci. The Rel family has four relBE loci, where two relBE loci (relBE2 and relBE4) have identical ORFs and regulatory sequences. The Par family consists of 3 parDE systems and this family also has two identical parDE loci (parDE1 and parDE3), while there is only 1 phd-doc locus. To gain insight into the degree of specificity of these TA modules, we explored whether or not an antitoxin can counteract the activity of a noncognate toxin. All toxins were expressed from the same vector, which is compatible with the vector used to express all antitoxins. We thus transformed all strains carrying the different antitoxins with each of the plasmids carrying the toxins, selected transformants on glucose and subsequently grew the cells in arabinose-containing media to determine if growth could occur in toxin expression conditions. The cells were unable to grow when the expression of a toxin was induced in the presence of a noncognate antitoxin (Table 2). No cross-interaction was observed among all 87 crosses between antitoxins and noncognate toxins showing that interactions are highly specific, even between TA systems of the same family (Table 2).

Occurrence of these TA cassettes in Vibrio species

We first explored the conservation of the 14 unique TA cassettes of N16961 among the other Vibrio cholerae genomes that have been entirely or partially sequenced. We found that these cassettes are highly conserved among the V. cholerae strains and isolates with fully assembled genomes (Table S2), while their distribution in strains with partially sequenced genomes is extremely variable. However, this does not reflect the actual TA cassette content of the SI in these strains.

We also carried out a study of the distribution of these TAs in the genomes of other Vibrio species. We observed that these TAs are variably present in other Vibrio genomes, and sometimes not embedded in cassettes (Table 3). Three TAs, HigBA2, VCA0477-478 and Phd-Doc, were uniquely found in V. cholerae (Table 3), but it is possible that this apparent lack of cassette exchange reflects a bias in the sampling, as we previously showed that the Phd-Doc cassette was also present in the Vibrio metschnikovii SI (26).
Discussion

In a previous analysis in 2005, Pandey and Gerdes identified 13 TA operons in the V. cholerae N16961 genome. Strikingly, their analysis revealed that all of these were embedded within cassettes of the superintegron (7). Since, TA systems have gained even more attention, as they have been the subject of multiple analyses and many new systems have been described (11). This, together with the fact that the role of these genetic modules is highly debated, led us to undertake a re-analysis of the toxin-antitoxin system content of V. cholerae genomes. We uncovered 5 novel TA modules in N16961, 4 of which were carried within integron cassettes, giving a total of 17 TA cassettes (Fig. 1). These cassettes are highly conserved among V. cholerae strains and isolates (Table S2). The single TA system that we identified and characterized, which is not carried in the SI encodes a toxin of the HipA family and its antidote. This toxin has been shown to inactivate the glutamyl-tRNA synthetase through phosphorylation and was identified through its major role in persister formation in E. coli (15, 16).

We have demonstrated the functionality of the 17 V. cholerae TA cassettes and explored their cross-interactions. Among these, ten correspond to well-defined TA families, and six of these cassettes have been previously characterized to various extents. These are the three ParDE, the two HigBA and the single Phd-Doc cassettes (25, 30, 36, 41). Two of the three ParDE family cassettes correspond to a duplication of the same cassette (parDE1 and parDE3) separated by 43 kb in N16961. The genes of both ParDE cassettes, parDE1/3 and parDE2 are in opposite orientation compared to the majority of cassettes whose expression relies on the Pc promoter (Fig. 1) (43). The ParE toxin is a gyrase inhibitor that induces DNA degradation and activates the SOS response to DNA damage, resulting in inhibition of cell division and reducing cell viability. In agreement with the recent findings of Yuan and coworkers (36, 41), our results confirm that parDE1/3 and parDE2 encode functional toxins whose toxicity was completely blocked by their cognate antitoxins. We further confirmed no cross interaction between the three members of Par family. (36, 41). As in C. crescentus,
which carries three ParDE systems (44), transcription of the two functional paralogous \textit{parDE} loci is differentially regulated, as \textit{parDE}1/3 is highly induced during exponential growth phase, while \textit{parDE}2 is only induced 1.6 fold (Table 1), supporting the idea that these paralogous systems act as distinct functional units (44). The two HigBA cassettes, \textit{higBA}1 and \textit{higBA}2, carry operons oriented as classical cassette genes, i.e., in the same direction as transcription from the \textit{Pc} promoter (Fig. 1). Both HigB toxins have been shown to inhibit the global rate-of-translation and preferentially cleave translated mRNA \textit{in vivo} (30). Our results are similar to those published previously (30, 37), and confirm that both of the \textit{V. cholerae} \textit{higBA} loci encode functional toxins and that their toxicity is only neutralized by expression of their cognate antitoxins. As for ParDE, we found that the two paralogs are differently regulated during growth (Table 1).

We characterized in this study, both in terms of functionality and expression, the remaining 11 TA cassettes. The \textit{V. cholerae N16961} SI contains 4 \textit{relBE} homologous cassettes, two of which (\textit{relBE}2 and \textit{relBE}4) are identical. As with ParDE cassettes, the \textit{relBE} genes are in opposite orientation compared to canonical integron cassettes. RelE toxins are translation inhibitor cytotoxins that cleave mRNAs from their 5' ends, leading to a rapid translational shutdown that is lethal or inhibitory to the host cells (45). We show here that all the \textit{relBE} loci of \textit{V. cholerae} N16961 encode functional toxins and that the activity of these toxins is completely counteracted by their cognate antitoxins. We also observed that no cross interactions exist between the non-cognate RelE and RelB proteins. In contrast to ParDE and HigBA TA systems, the four paralogs RelBE are similarly regulated during growth (Table 1).

The seven other TA systems in the superintegron are either hybrid or novel systems. There is one \textit{phd-parE}-like (VCA0422/0423), one \textit{stbD-relE}-like (VCA0477/0478), one \textit{relB-parE}-like (VCA0488/0489) according to the TA search software described in (11), and four systems with ribbon-helix-helix (RHH) transcription factor-like antitoxins (VCA0319 (its duplication in VCA0482), VCA0333, VCA0486 (Table S3)), associated with toxins that are either N-acetyl-transferases (VCA0318/0481, VCA0332 and VCA0487) or conserved proteins of unknown function (VCA0332). We showed that all are expressed and that their...
transcription, except for VCA0486-0487, is induced during exponential growth (Table 1). We demonstrate that all but one of these TA loci encode functional toxin-antitoxins. The non-functional cassette of N16961, VCA0422-0423, carries a toxin (VCA0423) that has been inactivated through an illegitimate recombination with another cassette, which exchanged the 3' part of the carried gene leading to a new C-terminal part for the original VCA0423. By chance, we found that this was a recent event as several *V. cholerae* strains, including 569B and O395, carry the original VCA0422-0423c cassette, for which we demonstrated toxin functionality.

We tested all possible cross-interactions between the different antitoxins and toxins that could lead to toxin inhibition and found that all the antitoxins only block their cognate toxin activity, even when TAs belonged to the same family (Table 2). The 17 identified TA cassettes are scattered throughout the entire 176 cassette-long array of the N16961 superintegron (Fig. 1). The first TA cassette corresponds to cassette 18 in the array and the last is located in position 174, 2 cassettes before the end cassette (see the annotated *V. cholerae* N16961 genome of the MicroScope platform at the French Genoscope for a precise mapping [https://www.genoscope.cnrs.fr/agc/microscope/mage/]). As mentioned above, these cassettes are highly conserved among *V. cholerae* strains and isolates (Table S2).

As mentioned above, we have shown that all of these cassettes are expressed and carry their own promoter, an extremely rare situation for integron cassettes. This rare property guarantees that they are transcribed independently of their location inside the cassette array. Indeed, integron cassettes seldom carry a promoter, and their expression occurs when they are recombined at, or close to, the *attI* site which carries a strong promoter, Pc (43, 46). This dynamic expression model, which allows the expression of adaptive genes only when needed is considered to be central for integron the evolutionary success (31). In the *V. cholerae* N16961 SI, as for other integrons, this constraint impacts the orientation of the cassette ORFs; among the 188 consecutive ORFs of its SI, only 27 are in opposite orientation relative to the Pc promoter. Strikingly, among these 18 correspond to genes carried in 9 of the 17 TA cassettes. It is interesting to note that in this configuration, even if
transcription extends beyond the cassette boundary, it will not express the surrounding genes as they are coded on the opposite, complementary, strand. The SI TA expression was found to be higher during exponential growth than during entry into stationary phase, except for two systems (Table 1). It was previously suggested that chromosomal TA loci function as stress response elements as they help the cells cope with nutrient starvation (14). Indeed, transcription of TA loci was found to be strongly increased during nutrient starvation in *Escherichia coli* (47-49), and during heat shock in *Sulfolobus solfataricus* (50). This has also been observed for the two *higBA* cassettes, whose transcription is induced during amino acid starvation (30). All but two antitoxins are predicted to carry a structural domain known to mediate interactions with DNA (Table S3), suggesting that they can regulate the expression of the TA operons in certain conditions. Our results advocate for a major role of these 17 TA cassettes is the stabilization of the SI cassette arrays, as previously proposed when we characterized the first two TA cassettes (26, 27). Chromosome-encoded TA loci may act to stabilize their neighboring DNA, mobile genetic elements and large genomic islands because the loss of any given TA locus (e.g. by slippage during replication or illegitimate recombination) will result in a cessation of de novo antitoxin production, toxin activation and arrest of cell growth and/or viability. This hypothesis is reinforced by the absence of cross talk between the SI TA we observed (Table 2).

This possible role better explains the presence of 17 functional TA loci grouped in the SI of *V. cholerae* (Fig. 1), as the *V. cholerae* SI is tightly packed with genes (one gene per 600 bp and at least 176 *attC* sites) (26). The *V. cholerae* *attC* sites, called VCR, are 123 bp long specific sequences which show more than 90% identity in 149 out of the 176 VCRs (26). Thus, any recombination between repeated copies of cassettes that can result in a large-scale deletion will be counterselected by one of the 14 unique TA loci, as far as the TA is located in that fragment. The *V. cholerae* TA loci are likely to constrain the emergence of daughter cells carrying deletions within the SI, which may otherwise survive if there were no functional TA loci and thus contribute to the stable perpetuation of the *V. cholerae* genome.

The facts that all these TA cassettes are expressed from their own promoter and that 15 out
17 are more expressed during exponential growth, which is when the replication is active, support this stabilization role. Indeed, recombination and slippage, which would lead to accidental cassette loss, are more frequent during active replication than when bacteria reach stationary phase and stop replication. However, it is likely that the TA cassettes could also have additional roles. These cassettes can protect their host genome from colonization by incoming mobile elements or plasmids carrying TAs from the same functional families by allowing their harmless loss through neutralization of the invading toxin by the chromosomal antitoxin (22, 25). These TAs can also play a role in bacteriophage defense through abortive infection mechanisms, as already demonstrated in other bacteria (19-21). Vibrio cholerae ecology is intricately linked to phage infection. Indeed, phage predation has been reported to be a factor that influences seasonal epidemics of cholera in Bangladesh (51). Furthermore, phage infections have been shown to occur in the human host, leading to both a decreased load of bacteria (52) and to an alteration of the virulence potential of V. cholerae shed from cholera patients (53).

We observed that when ectopically expressed in E. coli, the efficiency of the different tested toxins on cell viability vary to a large extent. However, all lead to the formation of a fraction of cells that, if not dividing, survive toxin activity and are in a state very similar to persistence (Fig. S2). Persisters are non-growing, dormant cells having low levels of translation and their state may be related to viable-but-non-culturable conditions (VBNC) which has been commonly observed in Vibrio species, including V. cholerae (for a review, see (54). The molecular mechanisms underlying this phenomenon have not yet been identified, but those TAs may have similar properties in V. cholerae and play a key role in persister formation, even when carried in SI cassettes, as found by the groups of Lewis and Hill for E. coli TAs systems (16, 17). Indeed, in E. coli, the over-expression of different TA locus enriched for persister cells and ectopic expression of TA loci dramatically increased the number of cells exhibiting multidrug resistance (16-18).

The quasi-absence of TA loci from the V. cholerae core genome is puzzling, especially as we identified TAs in other vibrios both inside and outside their integrons in their genome (31).
(see Table 3). The presence of expressed TAs within the integron cassettes may have allowed the deletion of TAs located outside of the SI due to TA cross-interactions, given the possibility that these cassettes could fulfill the physiological role(s) played by chromosomal TAs. This hypothesis now needs to be tested.
Materials and methods

Bacterial strains and media

*E. coli* DH5α (lab collection) and wild type *V. cholerae* (55) were used to generate recombinant strains carrying various plasmid constructs, including toxins and antitoxins listed in Table 1. Bacterial strains (Table S4) were grown in Luria-Bertani (LB) medium at 37°C. Antibiotics were used at the following concentrations: Ampicillin, (100µg/ml); Spectinomycin, (100µg/ml for *V. cholerae* or 50µg/ml for *E. coli*) and Chloramphenicol (25µg/ml for *E. coli* or 5µg/ml for *V. cholerae*). Glucose (1%) or arabinose (0.2%) was used to repress or induce the P<sub>BAD</sub> promoter, respectively.

Construction of expression vectors

All genes were amplified by PCR using the primers listed in Table S5 and *V. cholerae* N16961 chromosomal DNA as a template. Each of the toxin genes was cloned into the expression vector pBAD43, carrying the spectinomycin resistance gene under the control of the arabinose inducible P<sub>BAD</sub> promoter. Each of the antitoxin genes was cloned into the expression vector pUC18 (Invitrogen), carrying the ampicillin resistance gene, under the control of the promoter P<sub>LAC</sub>. Each construction was confirmed by Sanger sequencing.

RNA preparation and RT-PCR. Total RNA was purified from LB cultures harvested in the middle and at the end of the exponential phase as previously described (25). RT-PCR was performed on 500 ng *V. cholerae* N16961 total RNA with QF and QR primers (Table S5) using the Access RT-PCR system (Promega) with or without AMV reverse transcriptase. Real-time quantitative RT-PCR were performed with the same primers and analyzed as previously described (25). The expression level of tested genes was normalized using the 16S rRNA of *V. cholerae* and previously described primers 16SQF and 16SQR, (25). The experiments were repeated twice independently.
Toxicity assays

All *E. coli* and *V. cholerae* strains were grown on LB-agar with appropriate antibiotics and 1% glucose at 37°C overnight and replicated on LB-agar with antibiotics and either 0.2% arabinose or 1% glucose, and then grown at 37°C.

Toxin activity assays in liquid medium

*E. coli* strains containing either only a toxin or a toxin-antitoxin pair were grown in LB with appropriate antibiotics and 1% glucose to an optical density at 600nm (OD$_{600}$) of ~0.3 to 0.5, washed with LB, re-suspended in LB with antibiotics and either 1% glucose or 0.2% arabinose, and then grown at 37°C. At successive time points, aliquots were removed, the OD$_{600}$ was measured, and the bacteria were plated onto selective plates containing 1% glucose to enumerate CFU (colony-forming unit). The functionality of the toxin expressing plasmids in surviving clones was verified after transformation in a naïve *E. coli*.

Acknowledgments.

We thank Jason Bland and Sebastian Aguilar Pierlé for helpful reading of the manuscript. Work in the Mazel laboratory is funded by the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS-UMR 3525), the French National Research Agency (ANR-12-BSV3-0015), the French Government's Investissement d'Avenir program, Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant ANR-10-LABX-62-IBEID) and the European Union Seventh Framework Programme (FP7-HEALTH-2011-single-stage) "Evolution and Transfer of Antibiotic Resistance" (EvoTAR). NI was supported by a Franco-Pakistani doctoral fellowship.


Figure 1. Distribution of the different Toxin-Antitoxin cassettes in the *V. cholerae* N16961 superintegron. The external boundaries of the SI cassette array are indicated by striped genes *rplt-intIA* on one side, and VCA0510 on the other. Genes carried in cassettes are represented by plain pentagons and *attC* sites by black triangles. TA genes are represented by pentagons framed in green for the antitoxins, and in red for the toxins; the orientation of the pentagon and its location above or below the black line indicate the relative orientation of the genes. TAs belonging to well-defined families are indicated by the following colored pentagons: blue for RelBE, green for ParDE, dark red for HigBA, light green for Phd-Doc; light orange indicates unknown TA families. Identical TA cassettes are circled with a specific color. The fact that not all 176 SI cassettes are represented is indicated by dashes in the black line.
Table 1. Relative expression of the TA cassettes according to the growth phase.

<table>
<thead>
<tr>
<th>TA cassette</th>
<th>hgb1</th>
<th>hgb2</th>
<th>parD1-3</th>
<th>parD2</th>
<th>phd</th>
<th>relB1</th>
<th>relB2-4</th>
<th>relB3</th>
<th>319-482</th>
<th>332</th>
<th>422</th>
<th>477</th>
<th>486</th>
<th>488</th>
</tr>
</thead>
<tbody>
<tr>
<td>exponential/stationary</td>
<td>2.4</td>
<td>0.9</td>
<td>7.1</td>
<td>1.6</td>
<td>3</td>
<td>2.5</td>
<td>2.8</td>
<td>2.1</td>
<td>2.8</td>
<td>3.3</td>
<td>1.7</td>
<td>2.5</td>
<td>0.9</td>
<td>2.8</td>
</tr>
</tbody>
</table>

1. Quantitative RT-PCR were performed on total RNA from exponential and stationary growth phase culture.
Table 2. Cross interactions between toxin and antitoxins carried in the N16961 TA cassettes.

<table>
<thead>
<tr>
<th>Toxins</th>
<th>HigB 1</th>
<th>HigB 2</th>
<th>ParE 1-3</th>
<th>ParE 2</th>
<th>RelE 1</th>
<th>RelE 2-4</th>
<th>RelE 3</th>
<th>Doc 318-481</th>
<th>332</th>
<th>423c</th>
<th>478</th>
<th>487</th>
<th>489</th>
</tr>
</thead>
<tbody>
<tr>
<td>HigA 1</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HigA 2</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ParD 1-3</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ParD 2</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RelB 1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RelB 2-4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RelB 3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Phd</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>319-482</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>333</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>422</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>477</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>486</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>488</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 3. Distribution of *V. cholerae* cassettes related TA systems in other *Vibrio* species genomes and their location (•, cassette; ■, chromosomal not in cassette; ▲, plasmid).

<table>
<thead>
<tr>
<th>Bacterial strain</th>
<th>HigBA1</th>
<th>ParDE1</th>
<th>RelBE1</th>
<th>Phd/Doc1</th>
<th>VCA0318-3191</th>
<th>VCA0332-3331</th>
<th>VCA0422-4231</th>
<th>VCA0477-4781</th>
<th>VCA0486-4871</th>
<th>VCA0488-4891</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1-3</td>
<td>2</td>
<td>1-3</td>
<td>1-3</td>
<td>1-3</td>
<td>1-3</td>
<td>1-3</td>
<td>1-3</td>
</tr>
<tr>
<td><em>Vibrio cholerae</em> N16961</td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><em>Vibrio fischeri</em> MJ11</td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><em>Vibrio fischeri</em> ES114</td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><em>Vibrio harveyi</em> ATCC BAA-1116</td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><em>Vibrio parahaemolyticus</em> RIMD</td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><em>Vibrio splendidus</em> 12B01</td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><em>Vibrio splendidus</em> LGP32</td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Aliivibrio salmonicida</td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><em>Vibrio vulnificus</em> CMCP6</td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><em>Vibrio vulnificus</em> YJ016</td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>

1. HigBA1 (N16961 genome): VCA0391-392; HigBA2: VCA0468-469.
2. ParDE1 (N16961 genome): VCA0311-312; ParDE2: VCA0359-VIBCH20336; ParDE3: VCA0385-386
3 RelBE1 (N16961 genome): VCA0324-325; RelBE2: VCA0348-349; RelBE3: VCA0444-445; RelBE4: VCA0503-504
4 Phd/Doc (N16961 genome): VCACDS418614D-VCA0475
5 Seven hybrid and new systems of N16961 genome
6 Embedded in cassette but mobilized away from SI
7 Cassette repeated twice