Genetic drivers of chromosomal integron stability

Egill Richard[1,2], Baptiste Darracq[1,2], Eloi Littner[2,3,4], Claire Vit[1,2], Clémence Whiteway[1], Julia Bos[1], Florian Fournes[1], Geneviève Garriss[1], Valentin Conte[1], Delphine Lapaillerie[5,6], Vincent Parissi[5,6], François Rousset[7], Ole Skovgaard[8], David Bikard[7], Eduardo P. C. Rocha[3], Didier Mazel[1#] and Céline Loot[1#]

[1]Institut Pasteur, Unité Plasticité du Génome Bactérien, CNRS UMR3525, 75724 Paris, France.
[2]Sorbonne Université, ED515, F-75005, Paris, France.
[3]Institut Pasteur, Université Paris Cité, CNRS UMR3525, Microbial Evolutionary Genomics, Paris, France.
[4]DGA CBRN Defence, 91710 Vert-le-Petit, France.
[5]University of Bordeaux, Fundamental Microbiology and Pathogenicity Laboratory, CNRS, UMR 5234, SFR TransBioMed, Bordeaux, France.
[6] Viral DNA Integration and Chromatin Dynamics Network (DyNAVir), France.
[7]Synthetic Biology, Department of Microbiology, Institut Pasteur, Paris, France.
[8]Department of Science, Systems and Models, Roskilde University, Roskilde DK-4000, Denmark.

[#]Correspondent footnote: celine.loot@pasteur.fr and didier.mazel@pasteur.fr
Tel: +33 1 4061 3287

Running title: A new role of toxin–antitoxin systems in Sedentary Chromosomal Integrons.

Key words: bacterial evolution and plasticity, genome organization, genetic stability, site-specific recombination, mobile elements, sedentary chromosomal integron, toxin–antitoxin cassettes.

Abstract

The integron is a bacterial recombination system that allows acquisition, stockpiling and expression of promoterless genes embedded in cassettes. Some integrons, like the one found in the second chromosome of *Vibrio cholerae*, can be particularly massive, gathering hundreds of non-expressed cassettes. It is unclear how such genetic structures can be stabilized in bacterial genomes. Here, we reveal that the orientation of integrons toward replication within bacterial chromosomes is essential to their stability. Indeed, we show that upon inversion of the *V. cholerae* chromosomal integron, the cassette excision rate of its cassettes is dramatically increased. This correlates with a strong growth defect which we show is mostly due to the excision of a remarkable type of cassettes bearing their own promoter and encoding toxin–antitoxin systems. This "abortive excision" of toxin–antitoxin systems can prevent the inversion of chromosomal integrons and the associated extensive loss of cassettes. Our analysis of the available sedentary chromosomal integrons in genome database show an over-representation of toxin–antitoxin cassettes in large integrons. This study thus provides a striking example of the relationship between genome organization, genome stability and an emerging property of toxin–antitoxin systems.

Introduction

Integrons are bacterial recombination systems that act as a platform capable of capturing, stockpiling, excising and reordering mobile elements called cassettes. They were originally found to be the genetic systems responsible for the gathering of resistance determinants in some mobile elements [1]. Integrons play broader roles in bacterial adaptation, especially the numerous, large and yet poorly understood sedentary integrons [2–5]. We distinguish two types of integrons: Mobile Integrons (MI) and Sedentary Chromosomal Integrons (SCI). Both MIs and SCIs share however the same general organization: a stable platform and a variable array of cassettes (Fig 1a). The stable platform of the integron contains i) the integrase gene (*intI*) under the control of its promoter P_{int}, ii) the *attI* integration site and iii) the cassette P_C promoter driving the expression of the genes encoded in the cassette array located downstream. The variable part consists in an array of cassettes, each of which are generally composed of a promoterless gene associated to a recombination site called *attC*. Only the first few cassettes (those closest to the P_C promoter) can be expressed, while the rest represents a low-cost memory of valuable functions for the cell [6]. Upon expression of the integrase, the cassettes can be excised and then

re-integrated at the *attI* integration site and therefore become expressed [7]. The shuffling of integron cassettes allows bacteria to screen for the set of functions that would optimize its survival in a given environment. Remarkably, we previously found that the integrase, which catalyze the cassette shuffling, is part of the SOS regulon [8]. For this reason, integrons are described as "on demand" adaptation systems [9,10]. The integron integrase (IntI) has a very singular place among the broad family of the tyrosine recombinases [11]. While it recognizes the *attI* site under its double stranded (ds) form through its primary sequence, the *attC* sites are recombined as a single stranded (ss) form [12,13]. Indeed, these sites show a very limited conservation in sequence, but they all share an imperfect palindromic organization that favorizes the formation of a secondary structure that is recognized and recombined by IntI (Fig 1b). Although both the bottom and the top strands of an *attC* site can form a secondary structure, the recombination of the bottom strand (*bs*) is about 10^3 more efficient and the unpaired structural features of the site play a central role in this strand selectivity [13,14]. Selectivity for the *bs* of the structured *attC* site is essential to the correct orientation of the cassette upon integration at the *attI* site, allowing its expression by the P_C promoter [14].

In addition to their sedentary nature, the main feature that distinguishes SCIs from MIs is the size of their cassette array. While MIs can store up to 10 cassettes, SCIs can easily store dozens of them (up to 301 in *Vibrio vulnificus*), encompassing a substantial fraction of their host genomes [3]. For this reason, it was argued that the high genetic capacitance of SCIs allows them to serve as cassette reservoir for MIs which can compile from it the most relevant cassette repertoire in a given environment [10,15]. However, it is not clear how such massive and silent structure as SCIs can be stabilized in bacterial genomes. Our recent *in silico* analyses revealed that most of SCIs (83%) have a specific orientation relative to the chromosome replication, such that the recombinogenic *bs* of *attC* sites in the cassette array is carried by the replication leading strand template [15]. We previously showed that the recombinogenic single strand of *attC* sites is less accessible to the integrase on this strand than when located on the lagging strand template, where discontinuous replication leaves single-strand gaps that could facilitate folding (Fig 1c) [16]. In the latter orientation, the cassette excision rate would expectedly be higher, while the integration rate would not be changed, likely leading to a massive loss of cassettes in SCIs. Under these settings, such large arrays might not be achievable, and the SCI role as a genetic reservoir would be limited. Direct evidence supporting this model, however, is still lacking.

Here we investigate how the orientation of SCIs impacts their stability. We used as a model system the SCI of *Vibrio cholerae* (127 kb long and containing 179 cassettes), also called Super Integron (SI) [5]. Using phage recombinases, we were able to invert the *V. cholerae* SI and show

that, while the cassette integration rate is not impacted, the cassette excision rate is dramatically increased. We also show that the inversion of the SI is associated to a strong growth defect correlating with a high cell mortality rate in presence of a functional integrase only. We find that toxin–antitoxin (TA) modules, encoded by some cassettes carrying their own promoter (as opposed to the rest of the cassettes), are in a large part responsible for the decreased fitness induced by the increased excision rate in the SI Inverted (Inv) strain. As it has already been proposed, the presence of TA modules within SCI arrays likely prevents cassette deletion through large chromosomal rearrangements [17,18]. However, we propose here that when the rate of cassette excision is highly increased, as in the case of SI inversion, the TA cassettes would also efficiently excise as single cassettes, killing the host. Our results shed light a new role of TA cassettes in constraining the orientation of SCIs through "abortive excision" which optimizes the genetic capacitance of SCIs.

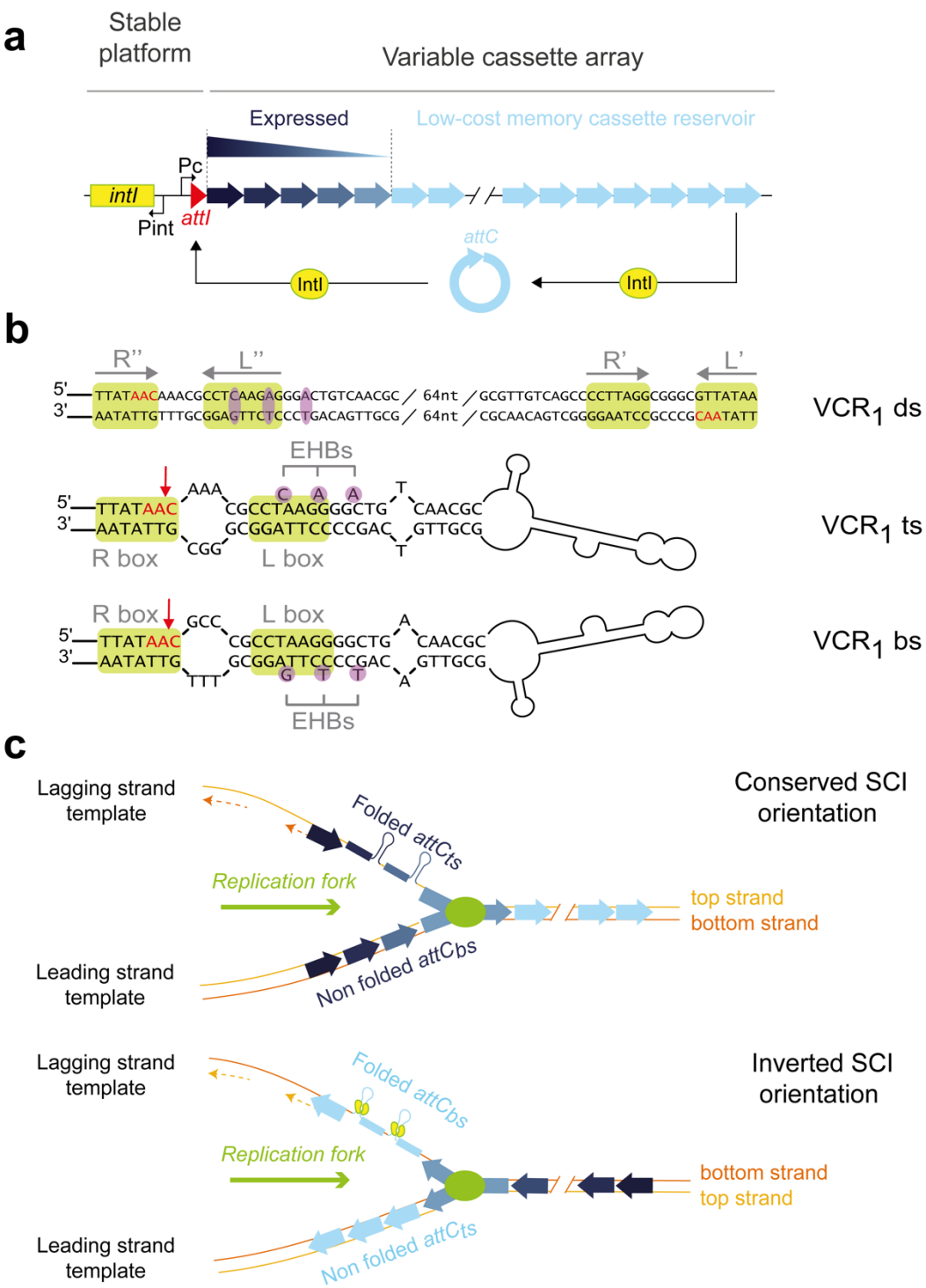

Figure 1: The integron system and the importance of its orientation towards replication. **a.** Schematic representation of the integron. The stable platform consists of a gene coding for the integrase (*intI*, in yellow) and its promoter P_{int}, the cassette insertion site *attI* (in red) and the cassette promoter P_C driving the expression of the downstream cassettes along an expression gradient. Cassettes are displayed as arrows where the tip represent the *attC* site. Their color intensity represents their level of expression from dark blue (highly expressed) to light blue

(not expressed). Non-expressed cassettes can be excised through an intramolecular *attC* × *attC* reaction and be reintegrated in 1st position near the P_C promoter through an intermolecular *attC* × *attI* reaction and become expressed. The combination of excision and integration is referred to as cassette shuffling.

b. Sequence of the VCR associated with the first cassette of the SI of *V. cholerae* N16961 (VCA0292), that we call VCR_1. In its ds (double stranded) form, the R'/R'' and L'/L'' boxes are highlighted in green. The extra helical bases (EHB) are highlighted in purple. The two different strands are also shown as structured hairpins. The CAA motif that is recognized and cleaved (between the C and the A) by the integrase is shown in red. This cleavage point is on the same side of the EHBs in the top strand (ts) while they are apart from one another on the bottom strand (bs).

c. Mechanistic insight on the issue of integron orientation. An array of cassettes is represented while it is replicated. In the lagging strand, Okazaki fragments are represented as dotted lines, leaving stretches of ssDNA on the correspondent template. In their conserved orientation, SCIs recombinogenic $attC_{bs}$ are carried by the continuously replicated leading strand template which supposedly limits their structuration. The structured $attC_{ts}$ are not recombined by the integrase. In the inverted orientation their more frequent structuration is expected to lead to increased binding of the integrase and higher recombination frequencies.

Results

The inversion of the Super Integron in *V. cholerae* dramatically increases the cassette excision events in the array.

The Super Integron in *V. cholerae* is naturally oriented so that the *bs* of the VCRs, the *V. cholerae attC* site historical name, are carried by the leading strand template [15] (Fig 1c). To investigate the importance of this orientation in the cassette array dynamics, we inverted in-place the entire SI of *V. cholerae* N16961 strain (see Materials and Methods, Fig S1). From these SI inverted (SI Inv) strain, we re-inverted back the SI to its original orientation (SI Reinv strain). We first monitored the effect of SI inversion on cassette shuffling by cultivating overnight *V. cholerae* with the SI in both orientations, in presence or absence of IntIA (Fig 2a). A PCR was then performed on independent clones with primers located in the *attIA* integration site and in the first cassette of the array (VCA0292) with an expected size of 275 bp in absence of shuffling (Fig 2a). Very few shuffling events could be detected in the SI Reinv clones (3/48), whereas we observed large amplicons in 46/48 of the SI Inv clones, testifying to the insertion

of multiple cassettes at the *attIA* site (Fig 2b). Performing our classical conjugation assay to deliver cassettes in *V. cholerae* recipient strains [7], we obtained a constant integration rate of these cassettes at the *attIA* site regardless of the SI orientation (Fig S2). We therefore conclude that the increase in shuffling in our assay strongly suggests a significant increase in the cassette excision rate in the SI Inv strain. To monitor the cassette excision rate at the population level in strains with SI in both orientations, we designed an SI-wide cassette excision assay. Like the previous shuffling assay, we grew *V. cholerae* strains overnight in the absence or presence of integrase, but this time in the absence of the *attIA* site to focus solely on excision events. To access cassette rearrangements along the array, we used long-read sequencing (Materials and Methods, Fig 2c).

For each population, a total of ~80000 reads with an average length of ~6.4 kb was produced and the coverage within the SI was about 134X leading to a detection limit of about 7×10^{-3} for single cassette excision events or 4×10^{-5} when considering the whole SI (179 cassettes). Each read captured the state of a segment of the SI in a single cell and allowed an easy detection of cassette deletion on this segment. We could identify 478 cassette excision events in the population of *V. cholerae* with an inverted SI and expressing the integrase, encompassing 120 cassettes where excision could be observed at least once. Not a single event was detected in the native orientation of the SI or in absence of the integrase. Thus, we conclude that in these strains the excision rate is below 4×10^{-5} (i.e., the limit of detection), that is at least 1250 times less than in the SI Inv in presence of IntIA where the average excision rate is of 5×10^{-2} (when excluding the 59 cassettes where excision could not be detected). This striking augmentation of the excision rate in the SI Inv is arguably the direct consequence of the increased recombinogenic potential of the VCRs. These results allowed to establish a direct link between the inversion of the SI and a highly increased cassette excision frequency.

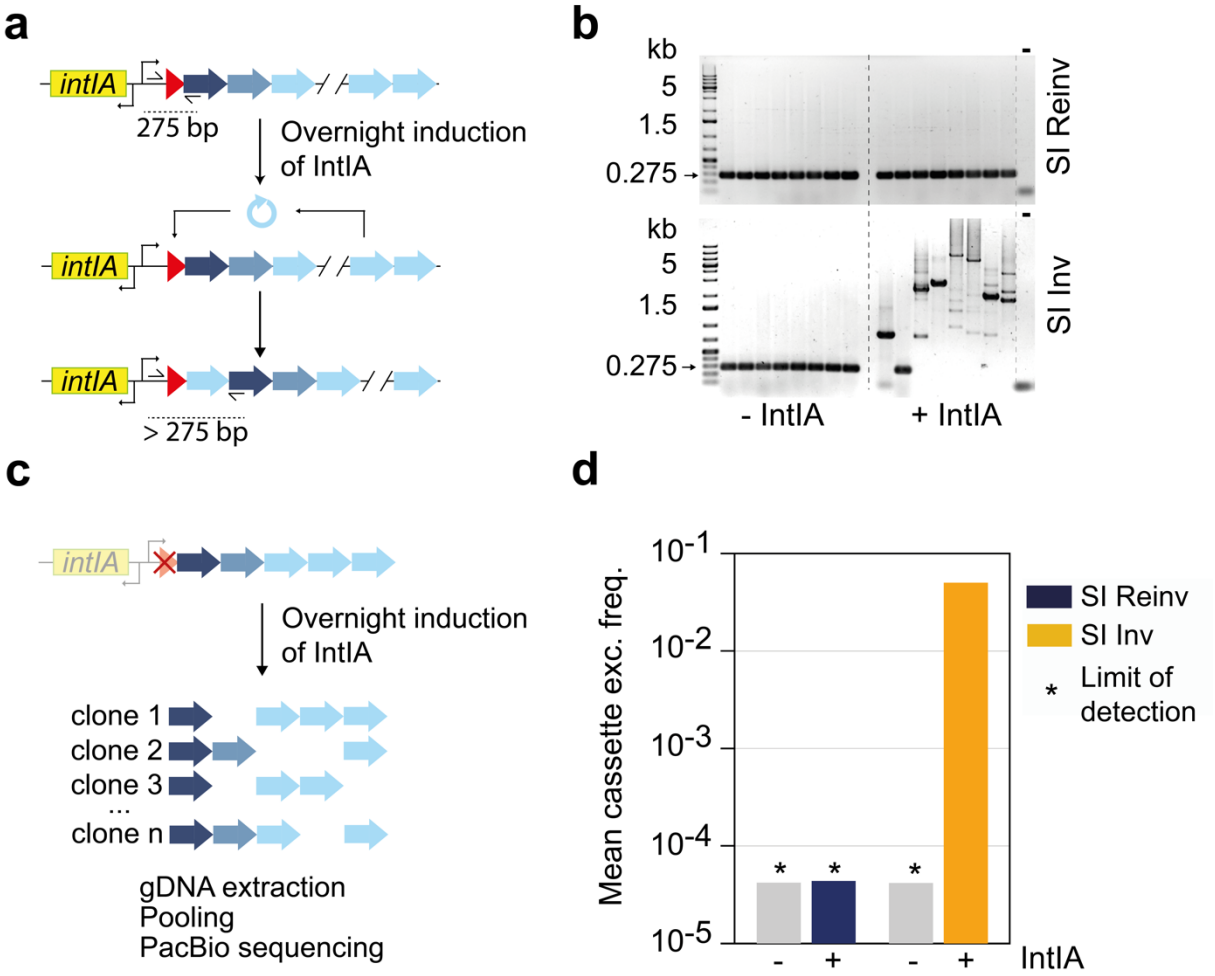

Figure 2: Cassette shuffling and excision assays in the Super Integron inverted and reinverted *V. cholerae* strains

a. Schematic representation of the assay designed to detect cassette shuffling.

b. PCR analysis of the cassettes shuffled in 1st position in individual clones in presence (+IntIA) or in absence of the integrase (-IntIA) for one overnight. The expected PCR product in absence of shuffling is 275 bp. Longer bands are evidence shuffling events. Multiple bands may come from the random slippage of the DNA polymerase in presence of highly repeated sequences (multiple VCRs in one amplicon).

c. Schematic representation of the assay designed to detect cassette excision. A strain of *V. cholerae* Δ*attIA* is used to initiate a culture composed of clones experiencing various cassette excision events. Clones from this heterogeneous population are represented (from 1 to n) each with different cassettes missing.

d. Mean cassette excision frequency of all the cassettes in the array of the SI. The cassette excision frequency of the SI Inv strain in presence of integrase represents the mean of the frequencies that could be calculated. No excision event could be detected in the other

conditions; hence the limit of detection is represented instead (~4 × 10^{-5}). The grey bars represent the frequencies obtained in absence of integrase in the SI Inv and SI Reinv strains.

The inversion of the Super Integron in *V. cholerae* leads to a growth defect in presence of a functional integrase.

Genome rearrangements are not very rare [19,20], and if a single inversion was sufficient to empty a SCI of all its cassettes in a few generations, then the massive and silent structures that are SCIs would probably not exist. Hence, we reasoned that there might be a high selective pressure on the orientation of the SI, preventing any spontaneous inversion. To test this, we set out to measure growth of the SI Inv strain (Fig 3a). While we did not observe any growth defect in the SI Inv strain in absence of integrase, we found an important growth defect upon expression of the integrase in this strain compared to the SI Reinv strain (Fig 3a). Growth rate is ~35% lower in the SI Inv strain expressing the integrase compared to the controls (Fig 3b). We observed the same growth defect while inducing the endogenous integrase through the SOS response using sub-lethal concentrations of ciprofloxacin (Fig 3c), confirming our previous conclusions in this more natural setting. To further evaluate the cost of the inversion of the SI, we set out a competition experiment where SI Inv and Reinv strains were co-cultivated for 24h starting with an initial 1:1 ratio (Fig 3d). While no substantial deviation to the original ratio between SI Inv and Reinv strains could be observed in absence of integrase, the expression of the integrase in those strains quickly led to a strong disadvantage of the SI Inv strain compared to the SI Reinv. This confirms the high fitness cost of the inversion of the SI of *V. cholerae* in presence of the integrase. Expressing a catalytically inactive integrase that binds to the $attC_{bs}$ (Fig S3) without cleaving it did not produce any growth defect (Fig 3e), ruling out a scenario where the mere binding of more accessible $attC_{bs}$ in the SI Inv strain could explain the growth defect. Hence, the cleavage activity of the integrase was necessary to induce a growth defect in the SI Inv. Yet, the integrase can cleave both *attIA* and VCR so to determine if it was the increased shuffling or the mere cassette excision that imposed such a strong cost of SI inversion, we performed growth curve in cells lacking the *attIA* site (Fig 3f). We still observed a growth defect, hence confirming that it is the increased cassette excision rate that explained the growth defect of the SI Inv strain.

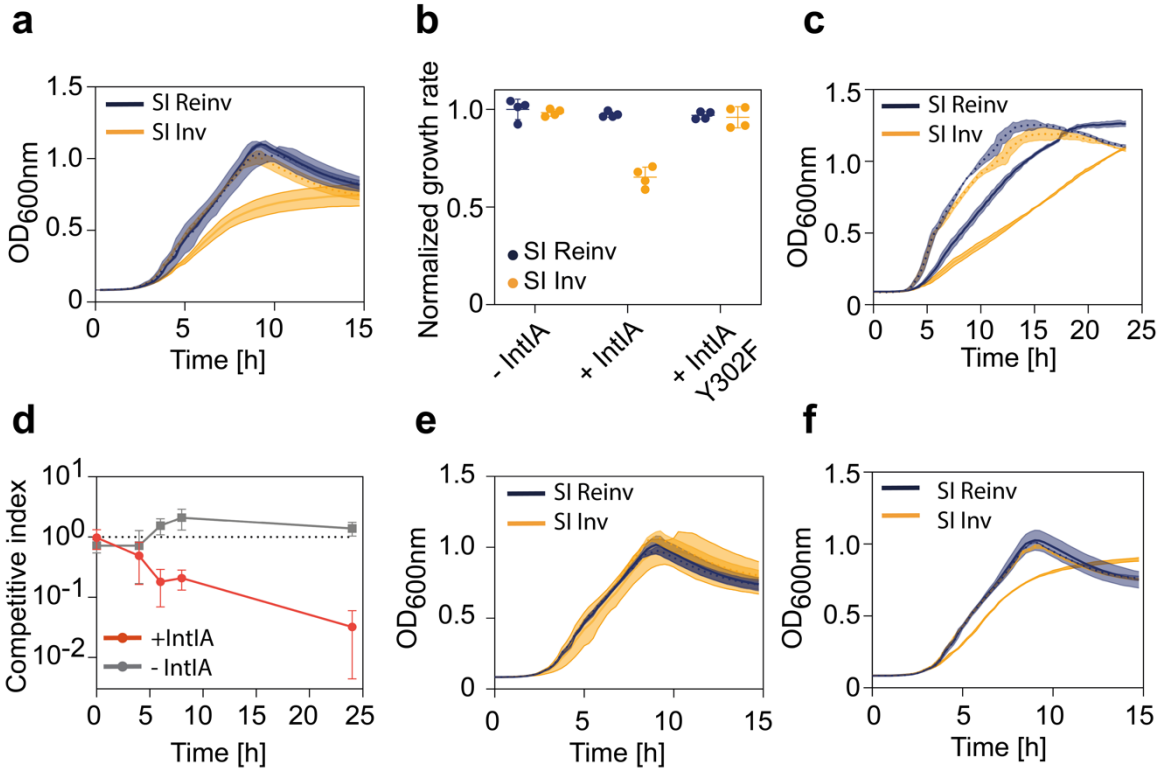

Figure 3: Growth of the Super Integron inverted and reinverted *V. cholerae* strains

a. Growth curve of SI Inv (orange) or SI Reinv (blue) *V. cholerae* strains in presence (full lines) or in absence of IntIA (dotted lines). Curve corresponds to the mean of four biological replicates, each with an average of at least two technical replicates, and the shade corresponds to the standard errors at each timepoint (same for all growth curves).

b. Growth rates of the SI Inv and SI Reinv strains normalized by the mean growth rate of the SI Reinv strain in absence of IntIA. The normalized growth rate of each strain respectively in absence of integrase (- IntIA) and, in presence of the WT (+ IntIA) or the catalytically inactive integrase (+ IntIA Y302F) are represented.

c. Growth curve of the SI Inv and Reinv strains in presence (full lines) or in absence (dotted lines) of a sub-inhibitory concentration of ciprofloxacin (10 ng. mL^{-1}) inducing the endogenous IntIA expression.

d. Competitive index of the SI Inv strain compared to the Reinv strain in absence (gray) or in presence of IntIA (red) in function of time (24h co-cultures). An index of 1 represents a ratio of 1:1 of the two strains in the mix. The lower the index, the lower the ratio of SI Inv compared to SI Reinv. Index was calculated with three biological replicates for each time point and the means and standard errors are represented.

e. Growth curve of the SI Inv and Reinv strains in presence of the catalytically inactive IntIA (full lines) or in absence of IntIA (dotted lines).

f. Growth curve of the SI Inv and Reinv strains in presence (full lines) or in absence (dotted lines) of IntIA, this time in a strain lacking an *attIA* site.

The growth defect in the Super Integron Inverted strain in presence of integrase is associated with increased cell death.

The growth defect previously observed in the integrase-expressing SI Inv strain could result from a slower division rate of each individual cell or a higher mortality rate at each generation. To discriminate between these two possibilities, we needed information at the single cell level. We therefore set out to observe in widefield microscopy live cells of our different strains growing on an agar pad mimicking the liquid medium used for generating the growth curves. An exponentially growing culture of our different strains was diluted in such a way that the growth of microcolonies stemming from a dozen cells per condition could be followed for 180 min (Fig 4a, Movie S1 and S2). In both the SI Inv or Reinv strains expressing the integrase, we tracked the development of microcolonies that could all behave differently, with some colonies growing fast and some much slower. However, we observed that this heterogeneity of development of the microcolonies was more pronounced in the SI Inv strain expressing the integrase resulting in a broader distribution of microcolony sizes after 100 min of culture (Fig 4b, Movie S2). This reflects the fact that, whereas the SI Reinv colonies generally follow an expected developmental trajectory, in the SI Inv strain, many microcolonies stop growing after a few divisions or do not grow at all. Some events of cell lysis, characterized by sudden release by some cells of diffracting matter [21], could also be observed specifically in the SI Inv strain (Fig 4c, Movie S3). These observations favor a model where the expression of the integrase in the SI Inv strain induces cell death in a part of the population.

We tested this model in a live and dead assay where a differential coloration of the cells allows discrimination between living (green) and dead (red) cells in a population (Fig 4d). We observed a clear increase of the proportion of dead cells in culture of SI Inv strain expressing the integrase compared to the control conditions. Automated cytometry allowed us to quantify cell viability in our control strains (Fig 4e), where a basal mortality rate of ~3-5% was measured, compared to ~16% in the SI Inv strain expressing the integrase. Interestingly, expression of a catalytically inactive integrase in the SI Inv strain resulted to a mortality rate similar to the basal level (Fig 4e), consistent with the observation that was made in the previous growth measurement (Fig 3b and 3e). These data indicated that inversion of the SI led to a high mortality rate that was dependent on the activity of the integrase.

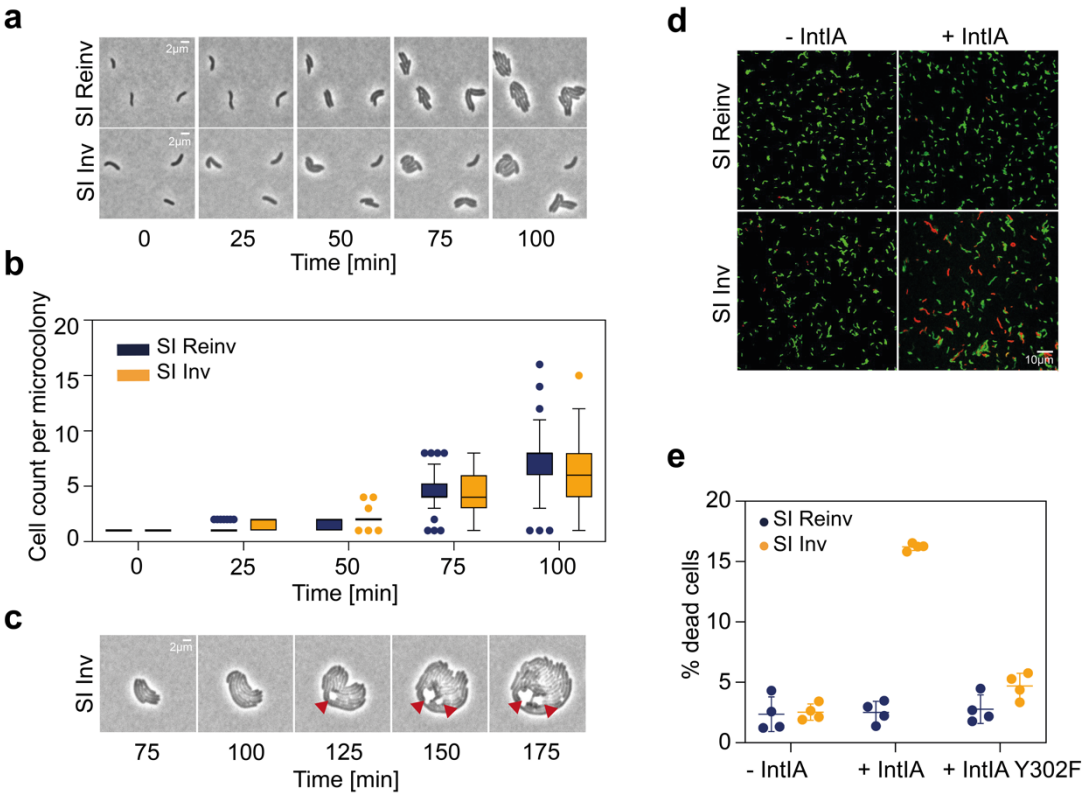

Figure 4: Cell viability of the Super Integron inverted and reinverted *V. cholerae* strains

a. Bright-field microscopy images taken from 180 min time-lapse series of live cells growing on a MOPS agar pad. Typically captions have to be kept rather short and to the point. A representative sample of 3 microcolonies of SI Inv or SI Reinv strains is shown. Scale bar is 2 µm.

b. Microcolony growth of the SI Inv and Reinv strains on MOPS agar pads. Boxplots represent the median and 95% C.I of the number of cells in 10 microcolonies tracked during 180 min.

c. Representative example of cell lysis events (red arrows) captured during the time-lapse experiments of live SI Inv cells growing on MOPS agar pads and expressing the integrase. Lysis events were not observed in the SI Reinv strain expressing the integrase. Scale bar is 2 µm.

d. Fluorescence microscopy images resulting from the Live and Dead assay on the SI Inv and Reinv strains expressing or not the integrase. Green fluorescence indicates live cells (SYTO-9) whereas red fluorescence (Propidium iodide (PI)) indicates dead cells. Scale bar is 10 µm.

e. Quantification of cell death by cytometry as a measure of cells positively stained with PI over the total number of cells counted (10,000 cells per replicate). Four biological replicates, means and standard deviation are represented.

Fitness defect in the SI Inverted strain is largely explained by the excision of toxin–antitoxin systems.

The fact that SI inversion results in increased cassette excision and that this is associated with increased growth arrest and cell death dependent on integrase catalytic activity led us to hypothesize that the increased excision of some cassettes could be deleterious for the cell. While most cassettes do not carry a promoter and their loss through excision should be harmless for the cell, a significant subset of cassettes encodes toxin–antitoxin (TA) systems that do carry a promoter [18,22,23]. TA systems are operons that encode a stable toxic protein and a corresponding unstable antitoxin capable of neutralizing its cognate toxin activity or expression, so that the co-expression of the toxin and antitoxin is harmless for the cell. When the expression of the gene pair is stopped, the antitoxin is degraded, resulting in cell death or growth arrest [24]. In *V. cholerae* N16961, out of the 20 TA systems that have been identified, 19 are concentrated within the SI [23,25] (Fig 5a). We hypothesized that the increased excision of the cassettes encoding TA systems, alongside all the non-expressed cassettes, could be at the origin of the increased growth arrest and cell death observed in the SI Inv strain. Indeed, cassettes encoding TA systems are supposedly as likely to be excised as "regular" cassettes, as shown by the comparison of the pfold values (a measure of the ability of an *attC* site to form a recombinogenic secondary structure, see Materials and Methods) of their respective *attC* sites (Fig S4a). In addition, we observe excision of the TA-encoding cassettes when they are duplicated (Fig S4b), suggesting that the non-duplicated TA cassettes can also be excised but result in such a large fitness defect that we cannot observe their excision in our data set. By combining the approaches proposed by Banno and coll. [26] and Reiss and coll. [27], we designed a highly multiplexed CRISPR base-editing tool to effectively inactivate all the numerous TA cassettes of the array (see Materials and Methods, Fig S5). After a tool validation step on a single target (i.e., the *lacZ* gene, Fig S5b and c), 12 guides were designed to target 15 toxins (3 of which are duplicated) such that a STOP codon was introduced, resulting in at least a one-third truncation of each toxin (Fig S5d). The remaining 4 toxins, that could not be inactivated through base-editing, were inactivated by introducing a stop codon through the more classical allelic exchange method [28]. It resulted in *V. cholerae* strains where all the TA within the SI were inactivated (TAi). We therefore assessed the effect this could have on growth of both SI Inv and Reinv TAi strains. In absence of integrase (Fig 5b), we do not observe substantial differences between all the strains whether the SI is inverted or not. In presence of integrase, the growth defect characteristic of the SI Inv strain appears to be largely rescued upon inactivation of the TA systems (Fig 5c). Performing a competition assay, we found again, in

presence of integrase, the SI Inv to be severely out competed by the SI Reinv strain but this was alleviated in absence of TA systems (Fig 5d). Similarly, the apparent fitness defect of TAi strains could also be confirmed by competition assay. While the SI Inv TAi still outcompetes the SI Inv strain by more than a factor three in presence of integrase, the SI Reinv TAi strain has an impacted fitness compared to the SI Reinv strain carrying TA systems (Fig 5e). A cell viability assay showed that the improved fitness of the SI Inv TAi strain compared with the SI Inv strain was associated with decreased cell death (Fig S6).

Finally, although the restoration of growth is important in absence of TA systems in the SI Inv strain, the phenotypic rescue is not perfect. We reasoned that the intense recombination activity in the SI Inv strain in presence of integrase could also affect the replication process explaining this gap. There are examples of macromolecular conflicts between DNA binding complex and the replication fork, sometimes leading to important impact on growth [29,30]. To address this, we performed Marker Frequency Analysis (MFA), a technique that gives access to the replication pattern by deep sequencing of gDNA extracted from an exponentially growing bacterial culture [30,31] and previously used in *V. cholerae* [32]. In absence of integrase, the MFA profile seems normal for both the SI Inv and Reinv strains (Fig S7). However, in presence of integrase, in the SI Inv strain, a severe drop of coverage can be observed at the genomic location corresponding to the SI specifically. This means that there is an important replication slow-down within the SI in this condition. It was possible that the lack of constraint on the cassette array in absence of TA systems could lead to a quick and massive loss of cassettes and the associated replication defect. For this reason, we sequenced single clones stemming from the culture of 4 SI Inv TAi independent strains to assess their cassette content after one overnight culture in presence of the integrase. Although two clones had effectively lost respectively 83% and 52% of their cassette content (*i.e* respectively 104 and 94 cassettes), the other two had only lost 9% and 11% of their cassette content (*i.e* respectively 16 and 20 cassettes). We conclude that the increased fitness of the SI inv TAi strain is not due to a massive loss of cassettes but rather to the absence of the detrimental effect of losing TA systems as single cassettes.

Therefore, we conclude that the growth defect is mostly due to the increased excision of TA cassettes together with non-expressed cassettes even if we cannot exclude a slight effect of a disruption of the replication process by the intense cassette dynamics.

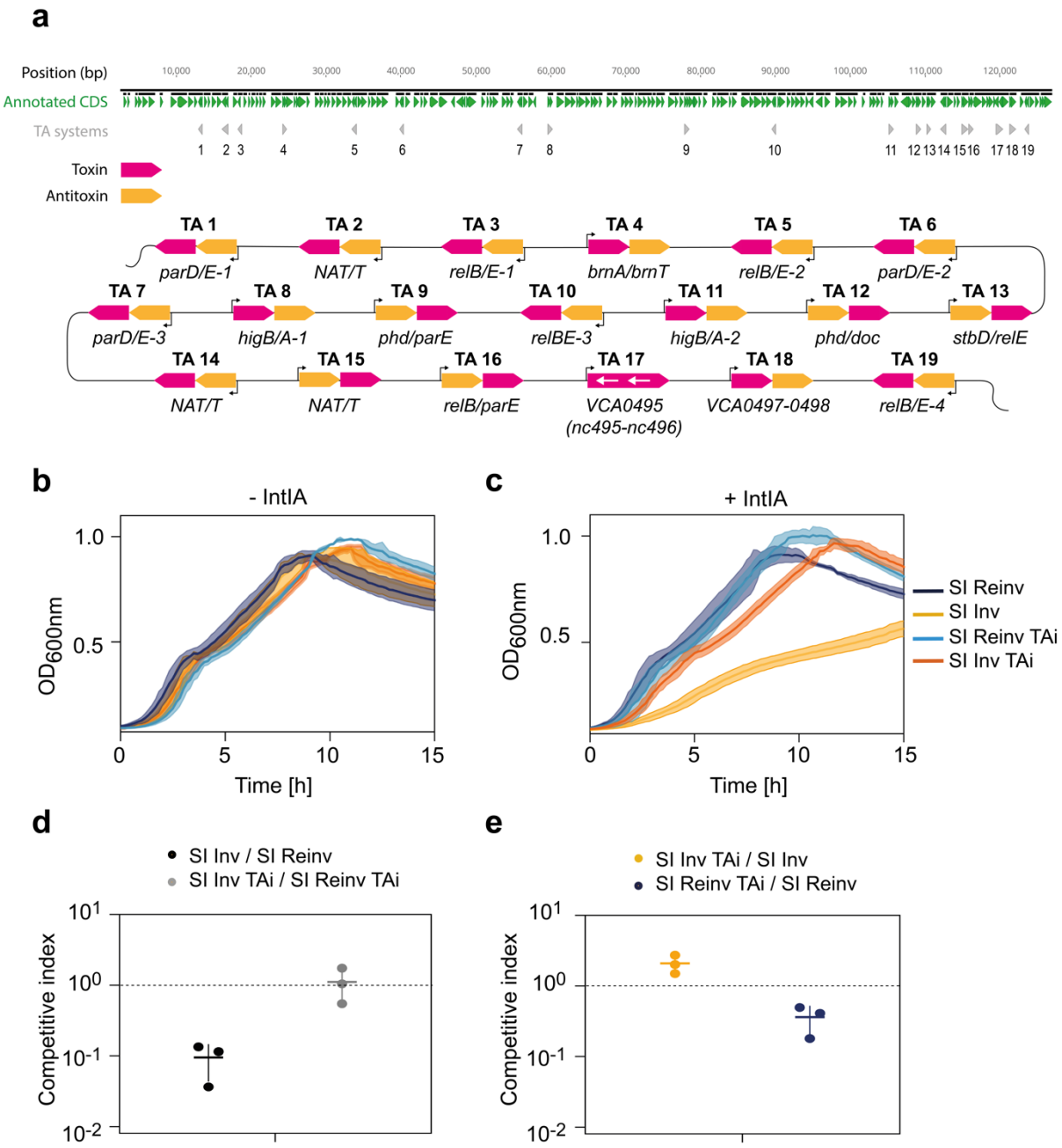

Figure 5: Large rescue of the growth defect in the TA inactivated SI Inverted strains of *V. cholerae*.

a. Representation of the SI of *V. cholerae* and the repartition of the various TA systems numbered from 1 to 19. A detail of the TA systems (name, orientation in the array, order of toxin/antitoxin) is represented below. NAT/T means N-acetyltransferase/transcription factor.

b, c. Growth curve of the strains of *V. cholerae* with its SI Inv or Reinv and the TA WT or inactivated (TAi) b, in absence (dotted lines) or c, in presence of IntIA (full lines).

d. Competitive index of the SI Inv strain compared to the SI Reinv strain (black) and of the SI Inv TAi strain compared to the SI Reinv TAi strain (grey).

e. Competitive index of the SI Inv strain compared to the SI Inv TAi strain (orange) and of the SI Reinv strain compared to the SI Reinv TAi strain (blue) after 24h of co-culture. For both d and e, an index of 1 represents a ratio of 1:1 of the two strains in the mix (figured in dotted lines).

TA cassettes are consistently over-represented in large Sedentary Chromosomal Integrons.

To assess if a correlation could exist between the number of TAs and the size of the SCI cassette arrays, we used IntegronFinder 2.0 [4] to detect all of the sedentary integrons in complete genomes of the RefSeq NCBI database. We identified 398 genomes with at least one SCI (defined as an integron comprising strictly more than 10 cassettes, as suggested in [4]) or one large CALIN (BigCALIN, defined as a SCI lacking a functional integrase). In each of these genomes, we screened TA systems with TASmania [33], a predictor with very low false negative rate. Our goal was to retrieve even unusual TA systems that would be rarely found outside SCIs and hence missed by more conservative predictors. Following a validation step relying on the well annotated genome of the *V. cholerae* N16961 strain, we decided to consider as a toxin (respectively antitoxin) hit any protein matching a TASmania toxin (respectively antitoxin) profile with an e-value lower than 10^{-3}. Then, we compared the number of toxins and antitoxins detected in the elements (SCI/BigCALIN) with the total number of cassettes in each of the 398 genomes. We found a positive correlation between the two, indicating that larger SCIs have more TA cassettes (Fig 6a). To assess the specific enrichment of TAs in SCI/BigCALIN when compared to the rest of the genome, we performed in each isolate a contingency table analysis. Out of the 398 isolates, 246 verified the hypothesis of a TA enrichment within SCI/BigCALIN ($p<0.05$, Fisher test and Benjamini-Hochberg correction). There are 87 genomes with SCI/BigCALIN devoid of genes coding for toxins or antitoxins, and 65 encode at least one of them but do not display a significant enrichment.

Most of the genomes with a significant over-representation of TAs in SCI/BigCALIN were in *Vibrio*. This may reflect a particularly strong effect in this genus, but we cannot exclude that it simply results from other elements having too few cassettes to provide a statistically significant signal of over-representation. Focusing on its 327 available genomes confirms a clear association between TA frequency and SCI/BigCALIN size (Fig 6b) and reveals that TA systems tend to be over-represented within integrons across 80% of the 280 genomes containing at least one SCI/BigCALIN. We studied the patterns of this over-representation in function of the phylogeny of Vibrio. We observe that the enrichment is scattered across the tree,

highlighting the important role of TAs in a broad range of Vibrio species (Fig 6c). This association holds true with the very stringent Bonferroni correction, although with a less sparse distribution on the phylogenetic tree, centered around a clade comprising *V. cholerae* (Fig S8). Interestingly, in this species, some TA systems (from the parDE family) are conserved in most isolates (82/86). This conservation may suggest a long period of co-evolution of integrons and TA systems, at least in *V. cholerae*. Although it does not exclude the regular turnover of some other TA systems, TA cassettes may have contributed to the evolution of SCI to a large extent.

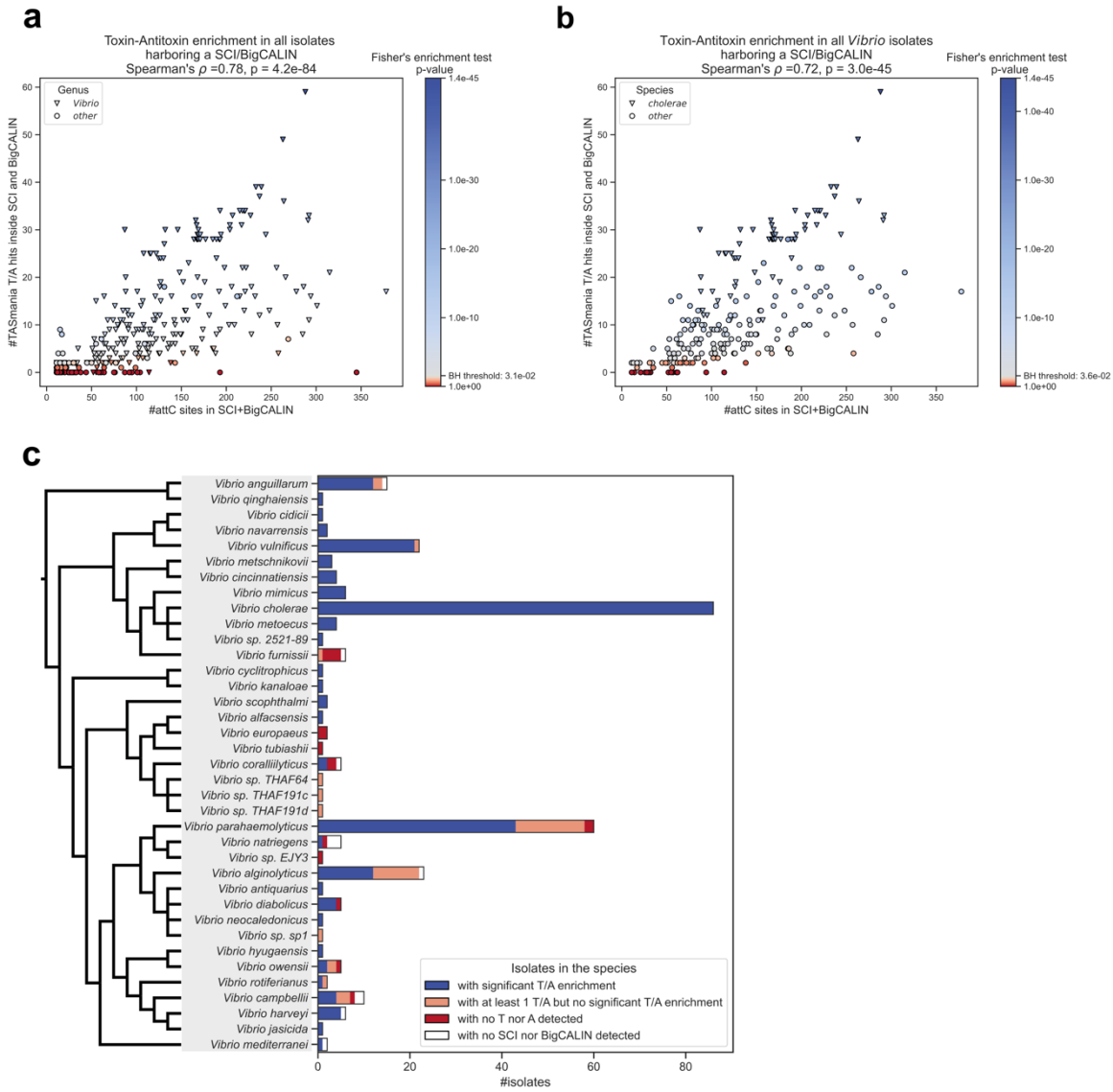

Figure 6: Toxin-Antitoxin landscape in SCIs determined by comparative genomics.

a, b. Each point represents a single NCBI RefSeq isolate harboring at least one SCI or BigCALIN (>10 *attC* sites). The x-axis displays the total number of *attC* sites in the (potentially multiple) SCI and BigCALIN of the isolate. The y-axis exhibits the total number of TASmania toxin and antitoxin HMM hits in the SCI and BigCALIN of the isolate (e-value<10^{-3}). The

Spearman nonparametric rank correlation coefficient is shown at the top of the graph, together with its significance p-value. Point hue indicates the Benjamini-Hochberg adjusted p-value of Fisher's test for TASmania hits enrichment in SCI+BigCALIN compared to the rest of the genome (scale on the right of the graph, significant enrichment in blue). **a.** All of the isolates harboring at least one SCI or BigCALIN are displayed. The shape of each point on the graph represents whether the isolate belongs to the *Vibrio* genus (triangles) according to the NCBI taxonomy or to another genus (circles). **b.** Same graph as a, but only with isolates belonging to the *Vibrio* genus. The shape of each point on the graph represents whether the isolate belongs to the *Vibrio cholerae* species (triangle) or to another species (round).
c. Distribution of *Vibrio* isolates by species according to the toxin/antitoxin enrichment inside the SCI/BigCALIN they harbor. The cladogram used to group species was adapted from the phylogeny reconstructed by Sawabe and coll. [34].
T: toxin, A: antitoxin

Discussion

We previously noticed a strong bias in the orientation of the large SCIs toward the replicon replication [15]. To understand the selective constraints associated with this specific orientation, we tested the effect of the inversion of the paradigmatic SCI of *V. cholerae*. We show that the inversion of this SCI leads to a dramatic increase of the cassette excision rate. Indeed, *attC* sites recombine in their single-stranded form and were expected to have a different recombination potential depending on the orientation of the cassette array, due to the integrase preference for one of the structured strands ($attC_{bs}$). Upon inversion of the cassette array, the recombinogenic $attC_{bs}$ now carried by the lagging strand template are much more recombined by the integrase due to the discontinuous replication on this strand, which facilitated the structuration of the $attC_{bs}$, leading to at least a 1250-fold increase of cassette excision. Our results show, for the first time in an SCI, that this specificity of the integrase allows to regulate the recombination potential of cassettes depending on the availability of structured $attC_{bs}$. Therefore, the quasi-systematic presence of the $attC_{bs}$ on the leading strand template [15] greatly increases the stability of SCIs and must be essential to be able to maintain a large pool of cassettes within their array, ensuring their role as a reservoir of cassettes. Nevertheless, genome rearrangements are not rare events [19,20], and if a single inversion could result in the loss of a non-transcribed region of

several dozens of kilobase pairs in just a few generations, one would expect SCI arrays to be much smaller than observed. Therefore, the existence of large SCIs implied the existence of an efficient selective pressure on their orientation. Strikingly, the inversion of the SI was also associated with a strong fitness defect in presence of the integrase. This altered growth of the SI Inv strain makes a lot of sense with respect to integron evolution. The growth defect did not translate into a consistently slower division rate at the single cell level, but rather to a higher heterogeneity in the outcome of growth for each cell, with a higher mortality rate and possibly more frequent cell growth arrest. While the mere binding of the integrase on the structured *attC* sites could have been enough to cause this growth defect by impairing DNA replication, we showed that the cleavage activity of the integrase was necessary to impact the fitness of the SI Inv strain. The systematic inactivation of the 19 TA cassettes present in the SI allowed us to make a direct link between the increased cassette excision in the SI Inv strain and its decreased viability in presence of a functional integrase.

Indeed, we observed an almost complete rescue of the growth defect in the SI Inv strain devoid of functional TA cassettes. Therefore, this suggests that the increased cassette excision frequency in the SI Inv leads to the increased loss of TA cassettes and to the subsequent cell growth inhibition or cell death. A residual part of the growth defect in the SI Inv strain could be attributable to the disruption of replication by the recombination reactions occurring within the array, as suggested by the MFA profile of the SI Inv strain in presence of integrase.

TA systems are regularly found in mobile genetic elements and contribute to their stability. In the case of *V. cholerae*, their presence all along the SI cassette array was proposed to avoid large scale rearrangements of the cassette array leading to the simultaneous loss of dozens of cassettes [9,18]. Indeed, the high prevalence of repeated sequences within SCIs might lead to massive loss of cassettes by homologous recombination in absence of regularly interspaced TA modules within the array. But this stabilizing role is independent of the integrase and is relevant regardless of the orientation of the SI. Here, we propose a secondary stabilizing activity of TA modules encoded by integron cassettes. This activity arises from the direct association between TA systems with *attC* sites, which strikingly interconnects cassette excision frequency with cell viability (Fig 7). While the excision of most cassettes of the array is completely harmless for the cell because they do not contain a promoter and are therefore not expressed, the excision of TA modules immediately imposes a prohibiting fitness cost. Since TA cassettes have supposedly the same chance to be excised than any other, this means that any excision event can potentially be lethal for the cell or at least lead to growth arrest. Although it may seem surprising that TA systems are associated with recombination sites, making them easy to lose

and to kill its host, we argue that this could be beneficial at the population level, analogously to their role in abortive infection (Abi) in the context of phage infection [24]. The very strong fitness cost of cassette excision imposed by the random chance of excising a TA must translate into counter-selection of any condition in which the excision rate is highly increased. In particular we showed here that the presence of TA cassettes heavily penalizes the inversion of the SI of *V. cholerae* which highly increases the cassette excision rate. We call this process "abortive cassette excision", where cassette excision leads to the suicide of a fraction of the population that is proportional to the overall cassette excision rate (Fig 7). The strength of this process, is that the sensor (the *attC* site) and the effector (TA system) are precisely combined into one piece (a TA cassette). This process might be an emergent property of TA systems once associated with an *attC* site which proves to be a very effective way to ensure the stability of large SCIs by driving the evolution of cassette excision rate towards low values. The fact that the TA cassettes all carry their own promoter strongly support that their role is different from the other cassettes carrying other adaptive functions. Strikingly, among the 1423 complete mobile integrons carried by plasmids, only one TA was found (specifically in the MI of the 248 kb plasmid from the bacterium *Comamonas testosteroni* (GCF_014076475.1)) [35]. This very low percentage of TA cassettes within MIs perhaps means that their recruitment can occur but lead to the abortive excision of the TA cassettes due to a too high cassette excision frequency in MIs compared to SCIs [15]. This process further underlines the importance of TA cassettes within the arrays of chromosomal integrons. It fits the observations made previously that there are many TAs in SCI [18], and the current identification of their significant and perhaps ancient over-representation in the SCI of many *Vibrio* species compared to the rest of the genome. This long co-evolution of *Vibrio* species, particularly *V. cholerae*, with their integron-encoded TA systems may have enabled SCI to have access to a vast repertoire of cassettes and thus expand their genetic capacitance.

The recent and striking evolutionary success of the integron system to annihilate antibiotic treatment of Gram-negative pathogens, relies on the complementarity between MIs and SCIs. The capacity of bacterial adaptation "on demand" is mediated by MIs, while the maintenance of a large reservoir of cassettes is ensured by SCIs. The functioning of MIs has been extensively studied over the past decades, whereas that of SCIs and especially their maintenance has always remained enigmatic. Here, we revealed that the maintenance of SCIs is controlled by the TA cassettes. In addition to their known role in plasmid stability or viral defense [24], TAs associated with integron recombination sites ensure the interaction between host genome organization and plasticity through abortive cassette excision.

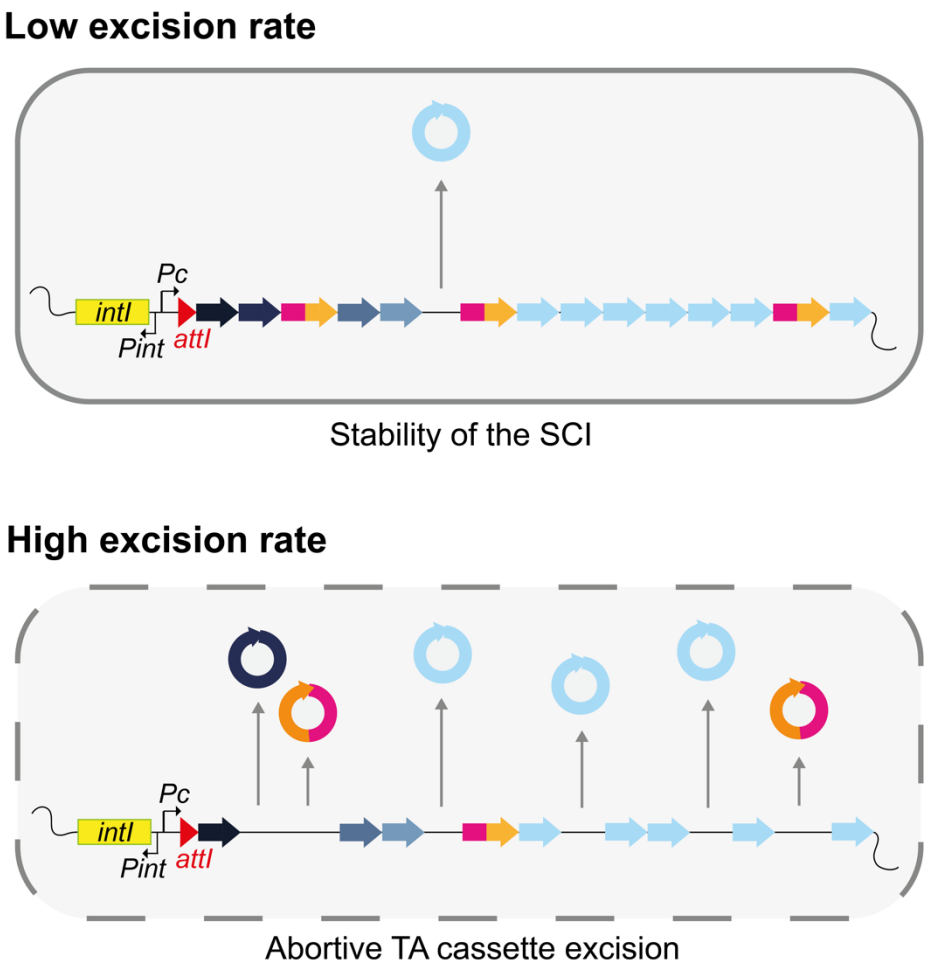

Figure 7: Model of the "Abortive TA cassette excision" process in an integron array containing TA cassettes.

Model of the emerging action of TA cassettes (in pink and orange) within the SI to stabilize the array of cassettes. The stability of the SCI (up) in conditions where the excision rate is low, and the proposed "abortive excision of TA cassettes" model killing the cell when the excision rate becomes too high are represented.

Materials and methods

Bacterial strains, plasmids and primers

The different bacterial strains, plasmids and primers that were used in this study are described respectively in Table S1, S2 and S3.

Media

Vibrio cholerae and *Escherichia coli* strains were grown in Luria Bertani (LB) at 37°C. *V. cholerae* strains containing a plasmid with a thermo-sensitive origin of replication were grown at 30°C. Thymidine (dT) and diaminopimelic acid (DAP) were supplemented, when necessary, to a final concentration of 300 µM. Glucose (Glu), L-arabinose (Ara) and Fructose were added respectively at final concentrations of 10g/L, 2g/L and 10g/L. X-Gal was used at a final concentration of 100 µM. DAPG was used at a final concentration of 50 µM. Antibiotics were respectively used at the following concentrations (for *V. cholerae* and *E. coli* respectively): carbenicillin (100 µg/ml; 100 µg/ml), chloramphenicol (5 µg/ml; 25 µg/ml), kanamycin (25 µg/ml; 25 µg/ml), rifampicin (1 µg/ml; 150 µg/ml), spectinomycin (100 µg/ml or 200 µg/ml in presence of glucose; 50 µg/ml), zeomycin (50 µg/ml; 50 µg/ml). In order to avoid catabolic repression during various tests using arabinose as inducer for the expression of the integrase, cells were grown in a synthetic rich medium: MOPS Rich supplemented with Fructose (1%) as a carbon source and arabinose (0.2%) as an inducer.

SI inverted and reinverted strain constructions

The inversion of the SI was performed using a genetic tool developed in the lab and designed to target the relocation of chromosomal DNA with bacteriophage attachment sites (HK in this case). Two DNA fragments were inserted respectively upstream and downstream of the SI in the N16961 hapR$^+$ strain (8637). The upstream fragment contained a kanamycin resistance gene and the 5' end of a carbenicillin resistance gene associated with the $attR_{HK}$ site. The downstream fragment contained a zeocin resistance gene and the partner 3' end of the carbenicillin resistance gene associated with the $attL_{HK}$ site (Fig S1). The strain containing the upstream and downstream fragments prior to inversion is referred to as the parental strain (Par). The expression of the HK_{022} integrase and excisionase (as a directional factor, pA401) in the parental strain led to the $attR_{HK} \times attL_{HK}$ reaction resulting in the inversion of the SI and to the reconstitution of the full copy of the *bla* gene, making it possible to select on carbenicillin for clones with an inverted SI (Fig S1). As a control for the following experiments, the SI was re-

inverted to its original orientation using solely the HK_{022} integrase (p8507) to perform the $attP_{HK} \times attB_{HK}$ reaction in the SI inv strain.

Parental strain:

Successive natural transformations of 8637 strain with PCR fragments produced from pF384 (fragment Kan^R) and pF850 (fragment Zeo^R) were performed. We used o4286 and o4302 pairwise primers to produce the Km^R fragment and o4657 and o4662 to produce the Zeo^R fragment.

SI inverted strain:

Transformation of I857-859 strains by the pA401 plasmid were performed at 30°C and in presence of glucose 1% to repress the HK integrase excisionase expression. Transformants were selected on spectinomycin (the marker resistance carried by the plasmid) and glucose containing plates at 30°C. The protocol of transformation is described above. SI inversion was performed by cultivating the obtained transformant clones during 12H (overnight) at 30°C in presence of spectinomycin and arabinose 0.2 % (to induce the HK integrase excisionase expression). SI inverted clones are selected by plating the resulting culture on plates containing carbenicillin and at 37°C (to favor the loss of the pA401 thermosensitive replication plasmid).

SI reinverted strain:

Transformation of the SI inverted strain by the p8507 plasmid was transformed at 30°C (to repress the thermoinducible HK integrase promoter). Transformants were selected on Spectinomycin (the marker resistance carried by the plasmid) containing plates at 30°C. The protocol of transformation is described above. SI re-inversion was performed by cultivating the obtained transformant clones up to OD_{600} ~0.3 at 30°C and by shifting the temperature to 37°C during 90 min. SI reinverted clones are selected by plating the resulting culture on plate without carbenicillin at 42°C (to favor the loss of the p8507 thermosensitive replication plasmid). The reinversion of the SI was checked by confirming the carbenicillin sensitivity of several obtained clones (by plating them on carbenicillin containing plates).

Automated growth curve measurements.

Overnight (ON) cultures of the indicated strain were diluted 1/1,000 and then grown for 20h in the indicated medium. Bacterial preparations were distributed by triplicate in 96-well microplates. Growth-curve experiments were performed using a TECAN Infinite microplate reader, with absorbance measurements (600 nm) taken at 10-min intervals. Slopes during exponential phase were directly obtained using the “GrowthRates” R package.

Toxin–antitoxins inactivation by allelic exchange

We performed allelic exchange to construct N16961 lacking the TA systems that could not be targeted using the Base-editing tool. To this purpose, we constructed and used different variants of the pMP7 vector, respectively pP897, pP898, pP900 and pP902. We followed the same protocols as previously described [28]. Briefly, the suicide vector pMP7 contains a R6K origin of replication and its replication is then dependent on the presence of the Π protein in the host cell. The Π3813 cell, a *pir*+ CcdB resistant *E. coli* strain, was used for cloning the different pMP7 plasmids. Once constructed, these vectors were transformed into the β3914 donor strain in order to deliver by conjugation the pMP7 vector into the desired recipient strain. Homology regions corresponding to the genomic DNA from the recipient strain have been cloned in the different pMP7 vector to allow the integration of the plasmid by homologous recombination. The only way for pMP7 vector to replicate into recipient strains is then to integrate into the host genome after a first crossover. After conjugation, integration of the entire pMP7 vector were then selected by plating cells on Cm plates lacking DAP. Next, cells were grown in presence of L-arabinose (0.2%) in order to express the CcdB toxin.

The expression of this toxin allows to kill cells in which the second crossover that leads to the excision of pMP7 backbone did not take place. This method allows us to obtain mutants of *V. cholerae* that are devoid of any antibiotic resistance marker. Since we did not performed deletions, the verification of the gene replacement by PCR only was prohibited. For this reason, the STOP containing version of the toxin was also associated to a BamHI restriction site to allow an easier screening of the correct gene replacement by digesting the appropriate PCR and looking for a restriction profile. After that, the PCR products containing the BamHI sites were checked by sequencing. The primers used for PCR screening and sequencing are described Table S3.

Golden Gate Assembly

By default, the vector expressing dCas9 carries a "random" guide that does not target any locus in *E. coli* nor in *V. cholerae* but that do contain two BsaI restriction sites. To change the guide, we perform a "Golden Gate Assembly" as such: two oligonucleotides have to be designed in the form 5'–TAGTNNNNNNNNNNNNNNNNNNNN–3' and 5'-TTTGNNNNNNNNNNNNNNNNNNNN-3' where the 20 "N" is the targeted sequence. For the annealing, 6uL of each oligo is mixed with 2uL of T4 DNA Ligase Buffer and 0.4 µL of T4

PNK for a total volume of 20 µL and then incubated for 30 min at 37°C. We then add 1 µL of NaCl (1M), incubate 5 min at 95°C and finally let slowly cool down to RT for at least 4h.
To insert the desired guide into the dCas9 expressing vector, we prepare the following mix: 2 µL of the dCas9 plasmid, 2 µL of annealed oligos, 1 µL of Cutsmart buffer, 1 µL of BsaI enzyme, 1 µL of ATP (1mM), 1 µL of ligase, 2 µL H_20. We then incubate the mix in a thermocycler with the following steps: 3 min at 37°C (for digestion), 4 min at 16°C (for ligation) and alternate between those steps 25 times. We finish with one cycle of 5 min at 50°C and 5 min at 80°C for enzymes inactivation. At this step, the mix is ready to be transformed in a cloning *E. coli* strain and the successful sgRNA insertions are screened by PCR using one primer used at the annealing step together with the o2406 primer.

Base editing

The tool used for base editing is derived from the catalytically dead Cas9 (dCas9) that was described in [26] for CRISPRi. The construction consists in a gene fusion between dCas9 and CDA1, a cytidine deaminase that catalyzes the deamination of cytidine, resulting in a uridine base which will later be replicated as a thymine (C → U→ T). A linker of 100 AA separated the dCas9 and the CDA1. A Uracil Glycosylase Inhibitor (UGI) was also fused to CDA1 to increase the base editing rate as described by Banno and coll. as well as a LVA degradation tag to decrease the toxicity the construct [26]. Due to a low efficiency transformation rate of this construct in *Vibrio cholerae*, the construct was delivered by conjugation. Hence, the appropriate construct was first transformed in a β3914 donor strain. Then both the donor and receptor strains were cultivated to exponential phase, mixed 1:1 in a 2 mL tube and centrifugated (6000 rpm for 6 min).
The pellet was spread on a membrane on a plate containing DAP to sustain growth of the donor strain and DAPG for induction of base editing, and the plate was incubated for 3h at 37°C. After incubation, the membrane was resuspended in 5mL LB and vortexed for 30s to resuspend the conjugated cells which were then plated on appropriate media: Cm_5 + Glc1% + Xgal to obtain both the CFU/mL after conjugation and the base editing efficiency or Rif_1 + Cm_5 + Glc1% to obtain the rate of apparition rifampicin resistant clones as a proxy for global mutation rate. Blue colonies had a functional *lacZ* gene while the white colonies were synonym for a mutated *lacZ* gene. All targeted sites were checked by PCR and sequencing (see Table S3). Then the whole genomes of the TAi strains were sequenced (Illumina).

Super Integron cassette shuffling assay

A single clone of each strain was isolated on plate and then inoculated for an overnight culture in LB + Spec + Glc 1%. The next day, the cells were inoculated at 1/50th for 2h in MOPS Rich + Spec + Fructose 1% + Arabinose 0.2% as a pre-induction step until they were at exponential phase. They were then inoculated at 1/1000th and grew for 20h until stat phase. Cultures where streaked on 1% agarose plates containing MH+ Spec + Glc 1% and single clones were isolated. A total of 48 PCR per condition was performed using an oligo (o5778, see Table S3) hybridizing in the promoter of *intIA* (stable part) and one (o1401) hybridizing in the first cassette of the array (variable part). A representative sample of the PCR profile for each condition was migrated in 1% agarose for visualization.

Suicide conjugation assay

This assay has been previously described by Vit and coll. 2021 [7]. We used the suicide conjugative vector pSW23T (pD060) that allows the delivery of the bottom strand of the $attC_{aadA7}$ recombination site. For each condition assay, at least 16 recombinant clones were isolated on appropriate plates and analyzed by PCR. To determine precisely if the pSW23T vector has been inserted into the *attIA* site of the SCI we used 5778 and SWend primers. These primers hybridize respectively in a sequence upstream of *attIA* in *V. cholerae* chromosome 2 or downstream of $attC_{aadA7}$ in the pSW23T vector. For each condition assay, at least three PCR reactions were purified and sequenced to confirm the insertion point.

Super Integron cassette excision assay

A single clone of each strain was isolated on plate and then inoculated for an overnight culture in LB + Spec + Glc 1%. The next day, the cells were inoculated at 1/50th for 2h in MOPS Rich + Spec + Fructose 1% + Arabinose 0.2% as a pre-induction step until they were at exponential phase. They were then inoculated at 1/1000th and grew for 20h until stat phase. Finally, 1mL of this culture was used to inoculate a 100 mL culture in LB + Spec + Glc1 % in order to filter the dead cells of the latter culture. The resulting culture constitutes a mixed population where each individual might have experienced a diverse set of cassette excision events. Bulk DNA was extracted from these mixed population (Qiagen genomic DNA extraction kit) and sequenced using the Pacific Bioscience (PacBio) long-read sequencing technology. Mapping was performed using minimap2 [36] and the cassette deletions were detected using a homemade R program. For each cassette, the number of deletion events was divided by the average sequencing depth for that cassette to determine the cassette excision rate. In order not to confuse

the excision rate with the random chance of an excision event to be propagated, the cultures of the different strains were replicated 5 times independently. 60 cassettes could not be detected as excised in our assay. The mean excision rate across the SI is the mean value of the 119 excision rates that could be calculated.

Protein purification

Bacterial strain BL21(DE3)plysS was transformed with a pET-derived plasmids expressing the His-IntI integrase (pM413 and pM415). The protein purification was carried out as described by Demarre and coll. [37] .

Electromobility shift assay

Each reaction contained 500 ng Poly[d(I-C)], 12 mM Hepes–NaOHpH 7.7, 12% glycerol, 4mM Tris–HCl pH 8.0, 60 mM KCl, 1 mM EDTA, 0.06 mg/ml BSA, 1 mM DTT, 10% Tween20, 0.01 nM specified 32P-labeled DNA oligonucleotide (Table S3) and the specified quantities of purified His-tagged IntIA and IntIAY302F, in a final volume of 20 µl. The samples were incubated at 30 ºC for 10 min without the probe followed by 20 min with the probe, then loaded to a 5% native polyacrylamide gels with 0.5x TBE as running buffer. The gels were visualized using a Molecular Dynamics intensification screen and a Typhoon FLA 9500 laser scanner

Competition assay

The SI Inv strains are resistant to Carb, which is not the case for the SI Reinv strains (see above). We used this difference to perform a competition assay. One clone of each strain was isolated on plate and used to inoculate an overnight culture. Before to start the competition assay, the cells were precultured at $1/50^{th}$ in LB + Spec + Glc 1% until exponential phase to be not biased by the dead cells from the overnight culture. Cells were then mixed at 1:1 ratio at $1/200^{th}$ in MOPS Rich + Fructose 1% + Spec + Arabinose 0.2% for induction of the integrase. The co-culture was plated on petri dishes containing MH + Spec + Glc 1% to get the total CFU/mL and in parallel on MH + Carb + Spec + Glc 1% plates to get the number of SI Inv CFU/mL in that same culture. We plated at t_0, $t_0 + 2h$, $t_0 + 4h$, $t_0 + 6h$, $t_0 + 8h$, $t_0 + 24h$. The competitive index (I) was calculated as such : $I = CFU_{SI\ Inv} / CFU_{SI\ Reinv} - CFU_{SI\ Inv}$ so that an index of 1 represents a situation where the ratio of SI Inv and SI Reinv is 1:1 (Neutrality). A high index indicates that SI Inv has a competitive advantage compare to its Reinv counterpart, and *vice-versa*.

Live and dead assay

Overnight cultures were performed in LB + Spec + Glucose 1%. Then day cultures were inoculated at 1:1000 in MOPS Rich + Spec + Fructose 1% + Arabinose 0.2% medium to OD600nm~ 0.8. Then 50 µL of the latter culture was dyed using the using the LIVE/DEAD™ *Bac*Light™ Bacterial Viability Kit, that is a mix of two dyes (Propidium iodide, or PI and SYTO-9). PI specifically stains the dead cells in red. SYTO-9 specifically stains viable cells in green. Cell viability was then assessed by microscopy and flow cytometry. For microscopy, dyed cells were placed of 1.4% agarose pad containing MOPS Rich + Spec + Fructose 1% + Arabinose 0.2% and observed using the EVOS M7000 microscope (100X objective, Texas Red filter to observe red fluorescence and FITC filter to observe green fluorescence). For flow cytometry, 20 µL of dyed cells were added to 200 µL of PBS and red fluorescence was assessed for 50 000 cells. The result is displayed as the proportion of cells that did display red fluorescence (dead cells).

Microscopy setup for live imaging

Bacterial cells were grown to mid-exponential phase in liquid MOPS Rich + Spec + Fructose 1% + Arabinose 0.2% medium and then transferred to 1.4% agarose-padded slides containing MOPS Rich + Spec + Fructose 1% + Arabinose 0.2%. A coverslip was placed on top of the agarose pad and sealed with a vaseline:lanolin:paraffin mix (ratio 1:1:1) to prevent pad evaporation. Slides were incubated at 37°C under an inverted wide-field microscope (Zeiss ApoTome) for time-lapse video recording. Video frames were taken at 1 min interval time for a total duration of 180 min, using a Plan Apo 63× objective (numerical aperture = 1.4) using a Hamamatsu sCMOS ORCA-Flash 4.0 v3 and autofocus module (Definite focus, Zeiss) (Pasteur Institute Imaging Facility Imagopole). A total of 8 movies were recorded and analyzed across all strains.

Marker Frequency Analysis

MFA was performed as described in [32]. Briefly, cells were grown in MOPS + Spec + Arabinose 0.2% + Fructose 1% at 37°C under agitation. Genomic DNA was prepared using the DNeasy® Tissue Kit (Qiagen) from 30ml of exponentially growing cells (OD_{450} ~0.15). The remaining culture was washed twice and was kept under agitation in MOPS + Spec + Glucose 1% at 37°C

to repress expression of IntIA and to prepare gDNA from cells in stationary phase (24h of culture). Libraries were prepared for Illumina sequencing (150 bp, paired end). The resulting reads were mapped with bowtie2 [38]. The genomic coverage of the sequencing data stemming from the exponentially growing cultures of each strain was normalized with the corresponding data in stationary phase. The normalized coverage is represented in function of the relative position to the origin of replication of each chromosome as colored dots for 1 kb bins and red dots for 10 kb bins. Dots are not shown when at least one position has a coverage of zero within the bin (typically due to repetitive sequences that prevents correct reads mapping). The gradient of coverage from *ori* to *ter* serves as a proxy for the replication dynamics and the slope translates the local speed of the replication fork [31].

Comparative genomics analysis of TA systems within SCI

Genomes dataset and SCI/BigCALIN annotation

To study the enrichment of TA systems within SCI across bacterial genomes, we used the dataset and integron predictions of IntegronFinder 2.0 [4]. Briefly, this dataset consists of 21,105 complete genomes retrieved from NCBI RefSeq on 30 March 2021. We defined SCIs as complete integrons with at least 11 predicted *attC* sites. Several isolates of *Vibrio cholerae* appeared to be devoid of any SCI, carrying instead a large (>10 *attC* sites) CALIN (cluster of *attC* sites lacking an integrase), that we called BigCALIN. A closer investigation showed that these BigCALIN were preceded by a pseudogenized integrase, which makes them similar to the SI found in *Vibrio cholerae*. These elements could have been recently inactivated, or the putative pseudogenes could result from sequencing artifacts. Hence, we decided to include in our analysis all the isolates carrying at least one SCI or BigCALIN.

Toxin and antitoxin prediction

In each genome harboring at least one SCI or BigCALIN, we predicted toxins and antitoxins with TASmania [33]. More precisely, we downloaded toxin and antitoxin HMM models from TASmania's website (accessed on 25 November 2021) and screened for them with the hmmscan command of HMMER 3.3 [39] (Nov 2019; http://hmmer.org/). As the command was run with default parameters, no threshold was used to filter hits in the first screening.
To reduce the number of false positives, we evaluated TASmania's performance on the genome of *Vibrio cholerae* N16961's secondary chromosome, known to harbor 19 TA systems [23,25], all

located within the SI. We observed that applying an e-value threshold of 10^{-3} allowed to keep all the hits corresponding to characterized TA systems (as described by Iqbal and coll. [23]), while eliminating most of the other hits. Hence, we selected this value to filter TASmania hits in all of the genomes comprised in our analysis.

Contingency analysis of TA enrichment within SCI/BigCALIN

For each genome, we performed a contingency table analysis to test for an enrichment of TA systems within SCI/BigCALIN. The contingency table was built as following: in each genome, each protein was classified as, on one side, belonging or not to an SCI/BigCALIN, and, on the other side, containing a TASmania hit or none. A Fisher one-sided test for statistical enrichment significance was performed on each contingency table. The resulting p-values were adjusted with the Benjamini-Hochberg and the Bonferroni corrections. An isolate was claimed to harbor a SCI/BigCALIN significantly enriched in TA when the adjusted p-value was lower than 0.05. The Benjamini-Hochberg and Bonferroni corrections were re-computed in the analysis focusing on *Vibrio* isolates.

Vibrio *genus cladogram*

To identify a potential evolutionary trend explaining the TA enrichment within SCI/BigCALIN across the *Vibrio* genus, we reconstructed a cladogram grouping *Vibrio* species together. More precisely, we took as basis the phylogeny published by Sawabe and coll. [34] and removed the species that did not harbor any SCI/BigCALIN. The species comprised in our dataset that were absent of the phylogeny were then added as a sister branch of its closest relative species.

Acknowledgements

We thank Jun Teramoto and Akihiko Kondo for sharing their construct for base-editing. We thank Clarisabel Garcia Rodriguez for her help with microscopy. We thank Marc Monot, Laurence Ma, Juliana Pipoli Da Fonseca and Thomas Cocklaer from the Biomics platform, C2RT, Institut Pasteur, Paris, France, supported by France Génomique (ANR-10-INBS-09) and IBISA.

Funding

This work was supported by the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS-UMR 3525), the Fondation pour la Recherche Médicale (FRM Grant No.

EQU202103012569), the ANR Chromintevol (ANR-21-CE12-0002-01), the French Government's Investissement d'Avenir program Laboratoire d'Excellence 'Integrative Biology of Emerging Infectious Diseases' [ANR-10-LABX-62-IBEID], the Ministère de l'Enseignement Supérieur et de la Recherche and the Direction Générale de l'Armement (DGA).

Conflict of interest statement

None declared.

Bibliography

1. Stokes, H.W., and Hall, R.M. (1989). A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. Mol. Microbiol. *3*, 1669–1683. 10.1111/j.1365-2958.1989.tb00153.x.
2. Cury, J., Jové, T., Touchon, M., Néron, B., and Rocha, E.P. (2016). Identification and analysis of integrons and cassette arrays in bacterial genomes. Nucleic Acids Res. *44*, 4539–4550. 10.1093/nar/gkw319.
3. Cambray, G., Guerout, A., and Mazel, D. (2010). Integrons. 10.1146/annurev-genet-102209-163504.
4. Néron, B., Littner, E., Haudiquet, M., Perrin, A., Cury, J., and Rocha, E.P.C. (2022). IntegronFinder 2.0: Identification and Analysis of Integrons across Bacteria, with a Focus on Antibiotic Resistance in Klebsiella. Microorganisms *10*. 10.3390/microorganisms10040700.
5. Mazel, D., Dychinco, B., Webb, V.A., and Davies, J. (1998). A distinctive class of integron in the Vibrio cholerae genome. Science *280*, 605–608. 10.1126/science.280.5363.605.
6. Collis, C.M., and Hall, R.M. (1995). Expression of antibiotic resistance genes in the integrated cassettes of integrons. Antimicrob. Agents Chemother. *39*, 155–162. 10.1128/aac.39.1.155.
7. Vit, C., Richard, E., Fournes, F., Whiteway, C., Eyer, X., Lapaillerie, D., Parissi, V., Mazel, D., and Loot, C. (2021). Cassette recruitment in the chromosomal Integron of Vibrio cholerae . Nucleic Acids Res., 1–18. 10.1093/nar/gkab412.
8. Guerin, É., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Re, S. Da,

Gonzalez-Zorn, B., Barbé, J., Ploy, M.C., and Mazel, D. (2009). The SOS response controls integron recombination. Science (80-.). *324*, 1034. 10.1126/science.1172914.

9. Escudero, J.-A., Loot, C., Nivina, A., and Mazel, D. (2014). The Integron: Adaptation On Demand. Microbiol. Spectr. *79*, 277–306. 10.1007/978-0-387-77863-1_14.
10. Richard, E., Darracq, B., Loot, C., and Mazel, D. (2022). Unbridled Integrons : A Matter of Host Factors. Cells *11*, 1–22.
11. Escudero, J.A., Loot, C., Parissi, V., Nivina, A., Bouchier, C., and Mazel, D. (2016). Unmasking the ancestral activity of integron integrases reveals a smooth evolutionary transition during functional innovation. Nat. Commun. *7*, 10937. 10.1038/ncomms10937.
12. Johansson, C., Kamali-Moghaddam, M., and Sundström, L. (2004). Integron integrase binds to bulged hairpin DNA. Nucleic Acids Res. *32*, 4033–4043. 10.1093/nar/gkh730.
13. Bouvier, M., Demarre, G., and Mazel, D. (2005). Integron cassette insertion : a recombination process involving a folded single strand substrate. EMBO J. *24*, 4356–4367. 10.1038/sj.emboj.7600898.
14. Nivina, A., Escudero, J.A., Vit, C., Mazel, D., and Loot, C. (2016). Efficiency of integron cassette insertion in correct orientation is ensured by the interplay of the three unpaired features of attC recombination sites. Nucleic Acids Res. *44*, 7792–7803. 10.1093/nar/gkw646.
15. Loot, C., Nivina, A., Cury, J., Escudero, J.A., Ducos-Galand, M., Bikard, D., Roch, E.P.C., and Mazel, D. (2017). Differences in integron cassette excision dynamics shape a trade-off between evolvability and genetic capacitance. MBio *8*. 10.1128/mBio.02296-16.
16. Loot, C., Bikard, D., Rachlin, A., and Mazel, D. (2010). Cellular pathways controlling integron cassette site folding. EMBO J. *29*, 2623–2634. 10.1038/emboj.2010.151.
17. Rowe-magnus, D. a, Guerout, A., Biskri, L., Bouige, P., and Mazel, D. (2003). Comparative Analysis of Superintegrons : Engineering Extensive Genetic Diversity in the Vibrionaceae Comparative Analysis of Superintegrons : Engineering Extensive Genetic Diversity in the Vibrionaceae. Genome Res., 428–442. 10.1101/gr.617103.
18. Szekeres, S., Dauti, M., Wilde, C., Mazel, D., and Rowe-Magnus, D.A. (2007). Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. Mol. Microbiol. *63*, 1588–1605. 10.1111/j.1365-2958.2007.05613.x.
19. Belda, E., Maya, A., and Silva, F.J. (2005). Genome rearrangement distances and gene

order phylogeny in γ-proteobacteria. Mol. Biol. Evol. *22*, 1456–1467. 10.1093/molbev/msi134.

20. Val, M., Kennedy, S.P., Soler-bistué, A.J., Barbe, V., Bouchier, C., Ducos-galand, M., Skovgaard, O., Mazel, D., and Paris, F.- (2014). Fuse or die : how to survive the loss of Dam in Vibrio cholerae. *91*, 665–678. 10.1111/mmi.12483.
21. Wong, F., and Amir, A. (2019). Mechanics and Dynamics of Bacterial Cell Lysis. Biophys. J. *116*, 2378–2389. 10.1016/j.bpj.2019.04.040.
22. Rowe-Magnus, D.A., Guerout, A.M., Biskri, L., Bouige, P., and Mazel, D. (2003). Comparative analysis of superintegrons: Engineering extensive genetic diversity in the vibrionaceae. Genome Res. *13*, 428–442. 10.1101/gr.617103.
23. Iqbal, N., Guérout, A.M., Krin, E., Le Roux, F., and Mazel, D. (2015). Comprehensive functional analysis of the 18 Vibrio cholerae N16961 toxin-antitoxin systems substantiates their role in stabilizing the superintegron. J. Bacteriol. *197*, 2150–2159. 10.1128/JB.00108-15.
24. Jurėnas, D., Fraikin, N., Goormaghtigh, F., and Van Melderen, L. (2022). Biology and evolution of bacterial toxin-antitoxin systems. Nat. Rev. Microbiol. *0123456789*. 10.1038/s41579-021-00661-1.
25. Krin, E., Baharoglu, Z., Sismeiro, O., Varet, H., and Mazel, D. (2022). Systematic transcriptome analysis allows the identification of new type I and type II Toxin / Antitoxin systems located in the superintegron of Vibrio cholerae. Res. Microbiol. *In Press*.
26. Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H., and Kondo, A. (2018). Deaminase-mediated multiplex genome editing in Escherichia coli. Nat. Microbiol. *3*, 423–429. 10.1038/s41564-017-0102-6.
27. Reis, A.C., Halper, S.M., Vezeau, G.E., Cetnar, D.P., Hossain, A., Clauer, P.R., and Salis, H.M. (2019). Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. Nat. Biotechnol. *37*, 1294–1301. 10.1038/s41587-019-0286-9.
28. Le Roux, F., Binesse, J., Saulnier, D., and Mazel, D. (2007). Construction of a Vibrio splendidus mutant lacking the metalloprotease gene vsm by use of a novel counterselectable suicide vector. Appl. Environ. Microbiol. *73*, 777–784. 10.1128/AEM.02147-06.
29. Lang, K.S., and Merrikh, H. (2018). The Clash of Macromolecular Titans: Replication-Transcription Conflicts in Bacteria. Annu. Rev. Microbiol. *72*, 71–88.

10.1146/annurev-micro-090817-062514.

30. Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., Woodward, J.J., Dreifus, J.E., and Merrikh, H. (2017). Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell *170*, 787-799.e18. 10.1016/j.cell.2017.07.044.
31. Skovgaard, O., Bak, M., Løbner-Olesen, A., and Tommerup, N. (2011). Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. Genome Res. *21*, 1388–1393. 10.1101/gr.117416.110.
32. Val, M.E., Marbouty, M., De Lemos Martins, F., Kennedy, S.P., Kemble, H., Bland, M.J., Possoz, C., Koszul, R., Skovgaard, O., and Mazel, D. (2016). A checkpoint control orchestrates the replication of the two chromosomes of Vibrio cholerae. Sci. Adv. *2*. 10.1126/sciadv.1501914.
33. Akarsu, H., Bordes, P., Mansour, M., Bigot, D.J., Genevaux, P., and Falquet, L. (2019). TASmania: A bacterial toxin-antitoxin systems database. PLoS Comput. Biol. *15*, 1–28. 10.1371/journal.pcbi.1006946.
34. Sawabe, T., Ogura, Y., Matsumura, Y., Feng, G., Rohul Amin, A.K.M., Mino, S., Nakagawa, S., Sawabe, T., Kumar, R., Fukui, Y., et al. (2013). Updating the Vibrio clades defined by multilocus sequence phylogeny: Proposal of eight new clades, and the description of Vibrio tritonius sp. nov. Front. Microbiol. *4*, 1–14. 10.3389/fmicb.2013.00414.
35. Huyan, J., Tian, Z., Zhang, Y., Zhang, H., Shi, Y., Gillings, M.R., and Yang, M. (2020). Dynamics of class 1 integrons in aerobic biofilm reactors spiked with antibiotics. Environ. Int. *140*, 105816. 10.1016/j.envint.2020.105816.
36. Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. 10.1093/bioinformatics/bty191.
37. Demarre, G., Frumerie, C., Gopaul, D.N., and Mazel, D. (2007). Identification of key structural determinants of the IntI1 integron integrase that influence attC × attI1 recombination efficiency. Nucleic Acids Res. *35*, 6475–6489. 10.1093/nar/gkm709.
38. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. 10.1038/nmeth.1923.
39. Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics *14*, 755–763. 10.1093/bioinformatics/14.9.755.