Genetic drivers of chromosomal integron stability

Egill Richard, Baptiste Darracq, Eloi Littner, Claire Vit, Clémence Whiteway, Julia Bos, Florian Fournes, Geneviève Garriss, Valentin Conte, Delphine Lapaillerie, et al.

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 and contain 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, its plasticity is dramatically increased. This correlates with a strong growth defect which we show is mostly due to the excision of a particular type of cassettes bearing their own promoter and encoding toxin-antitoxin systems. This so called "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 a robust correlation between the size of the cassette array and the number of toxin-antitoxin cassettes. 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 may play broader roles in bacterial adaptation, especially the numerous, large and poorly understood sedentary integrons [2–5]. Hence, 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. The

shuffling of integron cassettes allows bacteria to screen for the set of functions that optimize its survival in a given environment. Remarkably, we previously found that the integrase allowing the cassette shuffling is part of the SOS regulon since the P_{int} contains a lexA box [7]. Numerous stresses can trigger the SOS response in bacteria, among which the action of some antibiotics [8]. Therefore, the very stress that induces the expression of the integrase allows the screening for a set of cassettes that may in turn enable the cell to escape from it. For this reason, integrons are described as "on demand" adaptation systems [9,10].

The integron integrase has a very singular place among the broad family of the tyrosine recombinases. 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 [11,12]. Indeed, those sites show a very limited conservation in sequence, but they all share an imperfect palindromic organization that allows the formation of a secondary structure that is recognized and recombined by IntI (Fig 1b). The imperfections in the palindromic organization of the *attC* sites reveal 3 types of unpaired structural features upon folding : the Unpaired Central Spacer (UCS), the Extra Helical Bases (EHB) and the Variable Terminal Structure (VTS) [13]. Although both the bottom and the top strand 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 [12,14]. 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]. The two main recombination reactions that take place in integrons are the *attI* × *attC* and the *attC* × *attC* reactions. The *attI* × *attC* reaction, that leads to cassette insertion, involves both dsDNA (*attI*) and ssDNA (*attC*) which results to the formation of an asymmetric Holiday junction after the first strand exchange and that is resolved through replication [15]. For its part, the *attC* × *attC* reaction is associated to cassette excision and was found less efficient than the *attI* × *attC* reaction since it requires the coincidental structuration of two *attC* sites instead of one [9].

In addition to their sedentary nature, the main feature that distinguishes SCIs from MIs is the capacity of their cassette array. While MIs can store up to 10 cassettes , SCIs can easily store dozens (up to 301 in *Vibrio vulnificus*), encompassing a substantial fraction of their hosts' 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 [10,16]. 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 replication, such that the

recombinogenic *bs* of *attC* sites in the cassette array is carried by the replication leading strand template [16].

We previously showed that the recombinogenic single strand of *attC* sites is less accessible to the integrase when on this strand than if located on the lagging strand template, where discontinuous replication leaves single-strand gaps that could facilitate folding (Fig 1c) [17]. In the latter orientation, the cassette excision rate would expectedly be higher, while the integration rate would not be changed, leading to a decrease of available functions found in SCIs. Under these settings, such big arrays might not be achievable and the SCI role as a genetic reservoir would be limited, directly impacting the versatility of MIs that benefit from it. Direct evidence supporting this model, however, is still lacking.

Here we address the constraints in the orientation of SCIs using that of *Vibrio cholerae*, called the Super Integron (SI), that we inverted. We show that, while the integration is not impacted, the cassette excision rate is dramatically increased within the SI upon its inversion, therefore impairing its genetic capacitance. 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. This high fitness cost associated with SI inversion likely prevents spontaneous inversion events from occurring and may explain the observed rarity of SCIs carrying $attC_{bs}$ on the lagging strand template [16]. We find that toxin-antitoxin (TA) modules encoded by some of the cassettes and that carry 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 SCIs arrays likely prevents cassette deletion through large chromosomal rearrangements [18,19]. However, here we propose that they could also be excised as single cassettes, killing the host when the cassette excision rate is highly increased like in the case of the inversion of the SI. 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.

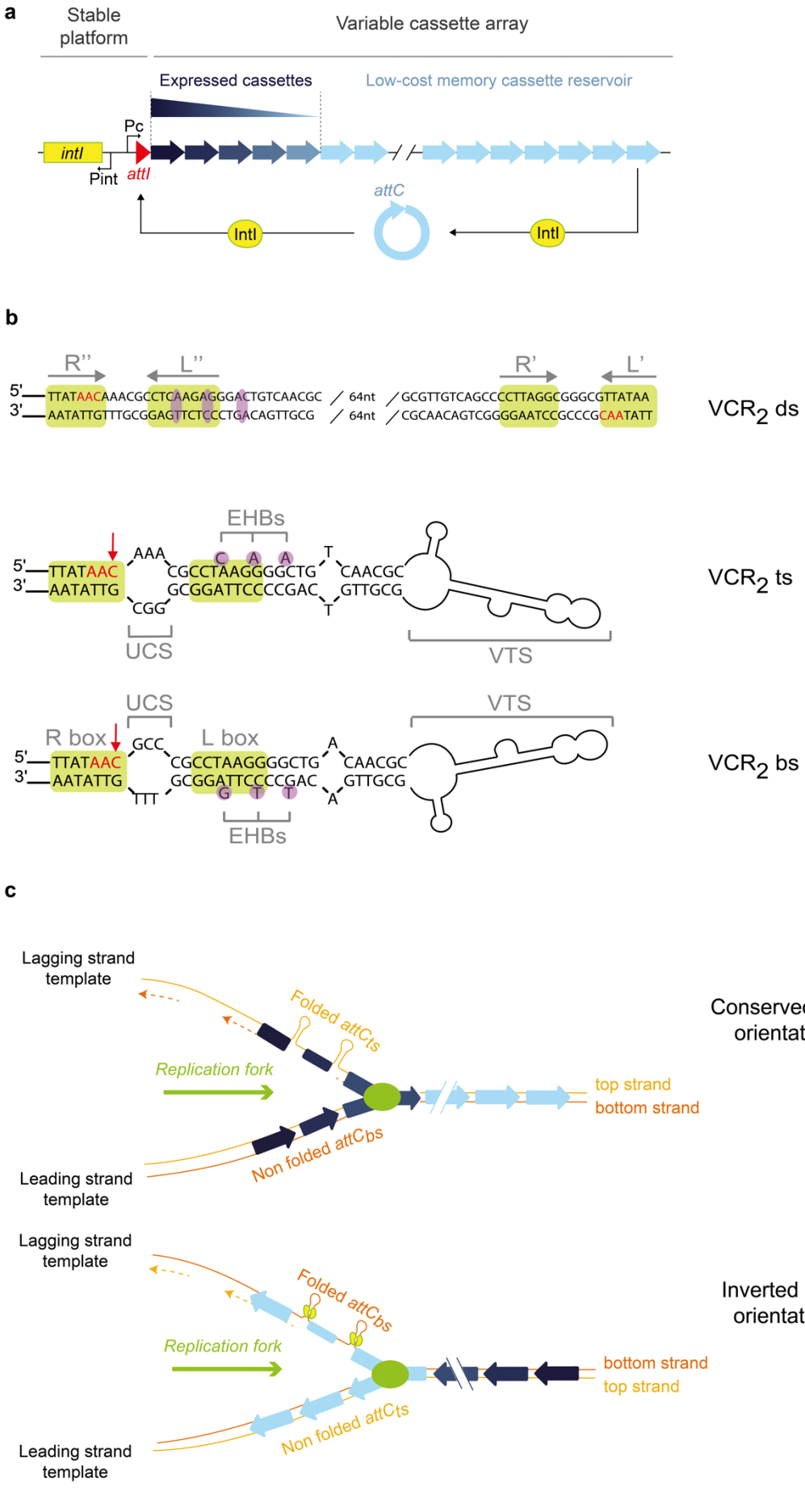

Figure 1: The integron system and the importance of its orientation towards replication.
a. Schematic representation of the integron. The stable platform consists of a gene coding for the integrase (*intI*, in yellow) and its promoter P_{int}, the cassette insertion site *attI* (in red) and the cassette 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_2. 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 $attC_{bs}$ (dark orange) are carried by the continuously replicated leading strand template which supposedly limits their structuration. The structured $attC_{ts}$ (light orange) 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* does not affect the cassette recruitment capability of the *attIA* site

The SI is a 127kb long structure containing 179 cassettes whose *attC* sites are called *Vibrio cholerae* Repeats (VCR). It is naturally oriented so that the *bs* of the VCRs are carried by the

leading strand template [16] (Fig 1c). To investigate the importance of this orientation in the cassette array dynamics, we set out to invert the entire SI of *V. cholerae* N16961 strain. To do this, we used a phage-based site-specific recombination system previously developed in the lab and designed to target the relocation of chromosomal pieces of DNA (see Material and Methods and Fig S1) [20]. As a result, we obtained a strain in which the SI is inverted (SI Inv) and a strain in which the SI is re-inverted back to its original orientation (SI Reinv). Each strain was generated as three biological replicates.

We tested the cassette recruitment capability of the *attIA* site in both SI orientations using a suicidal conjugation assay previously developed in the lab [12]. In this assay, an *attC*-containing plasmid mimicking an integron cassette is delivered in a single-stranded form through conjugation to the SI Inv and SI Reinv *V. cholerae* strains containing a vector (pSC101, low copy number plasmid) expressing the IntIA integrase or an empty version as a negative control. Once delivered in the recipient strain, this *attC*-containing plasmid cannot replicate autonomously and the only way for it to be maintained is to insert itself into the host by recombination with an *att* site carried by the SI (Fig 2a). In presence of IntIA, the cassette integration frequency in *attIA* site is around 10^{-2} for both the SI Inv and Reinv strains (Fig 2b). This is consistent with our recent work that showed that the high efficiency of the *attIA* × *attC* reaction in *V. cholerae* allows a strong cassette recruitment capacity of the SI [21]. However, no significant difference in cassette integration frequency was observed between the SI Inv and Reinv strains. This can be explained by the fact that the *attIA* site recombines in its *ds* form and thus its recruitment capacity should not be affected by its orientation towards replication. Only rare events of integration (less than 10^{-6}) are observed in those strains carrying the empty plasmid. This is consistent with our recent results showing that the induction of SOS through conjugation led to a sufficient expression of the endogenous integrase to integrate incoming cassettes at a low frequency [21].

A junction PCR between the *attIA* site and the incoming synthetic cassette was performed to confirm the integration at the *attIA* site in both strains (Fig 2c). In the SI Reinv recombinant clones, 100% of the PCR performed (96/96) were positive and the amplicons were of the expected size (240 bp) for an integration at the *attIA* site. In the SI Inv recombinant clones, the same PCR was positive in almost all cases (95/96) but yielded amplicons larger than expected in 23% of the clones. Sequencing of the latter amplicons revealed that, in these clones, one or several cassettes from the SI had been excised and reintegrated at the *attIA* site (Fig S2). These data suggested that, while expressing the integrase to allow the integration of the synthetic cassette, we also induced the cassette shuffling which is largely increased in the SI Inv strain.

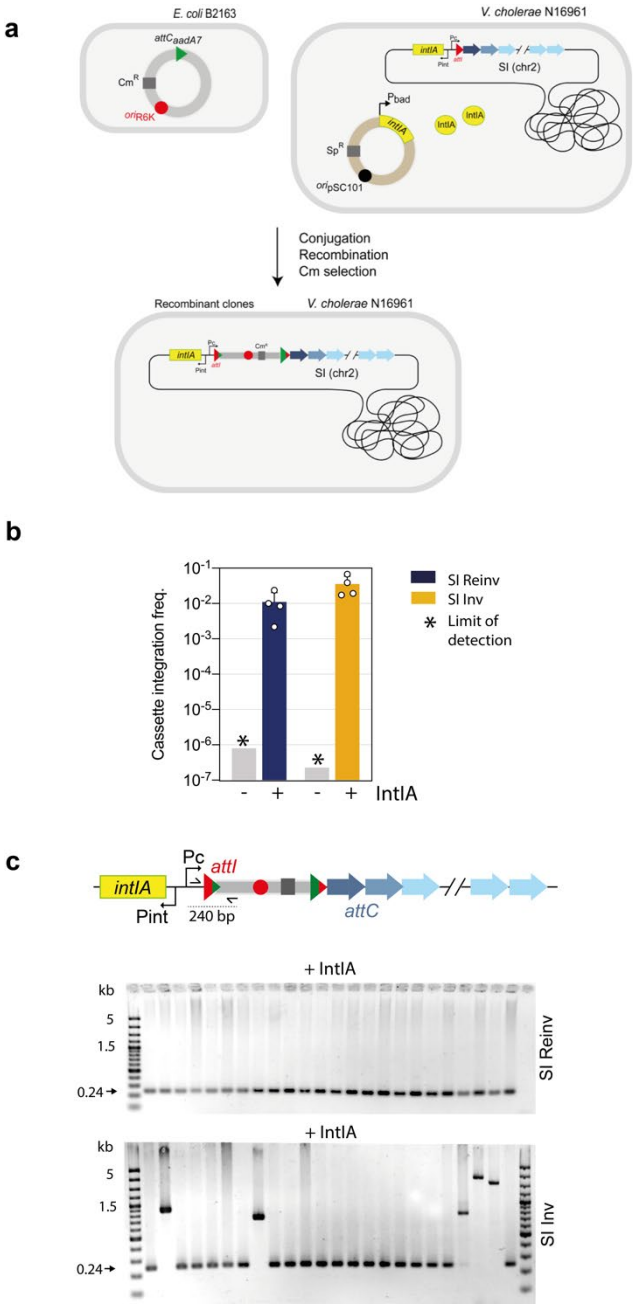

Figure 2: Cassette integration assay in the Super Integron of *V. cholerae*.

a. Schematic representation of the suicidal conjugation assay used to evaluate cassette integration frequency. The suicidal plasmid confers resistance to chloramphenicol (Cm^R) and has an R6K origin of replication that is dependent of the presence of the Π protein to replicate, that is present in the donor strain but not in the recipient *V. cholerae* strain. The incoming cassette is expected to integrate in the *attIA* site.

b. Cassette integration frequencies. Insertion of a synthetic cassette delivered by conjugation in a strain of *V. cholerae* with its SI Inverted (Inv, orange bar) or Reinverted (Reinv, blue bar). The grey bars represent the frequencies obtained in absence of integrase.

c. PCR analysis of the integration events of the synthetic cassette in strains containing the *attIA* site. The PCR amplifies the junction between the *attIA* site and the incoming cassette. A band of 240 bp is expected so that a longer band indicates a shuffling event occurring after the integration of the synthetic cassette.

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

In the SI Inv strain, the *bs* of the VCRs are carried by the lagging strand template where 1-2kb stretches of ssDNA are formed upon the passage of the replication fork [22], which should increase the recombinogenic potential of the VCRs as we previously showed that Int1 could chase SSB (Fig 1c) [23,24]. We monitored the shuffling of the cassettes within the SI by cultivating overnight *V. cholerae* with the SI in both orientations, carrying either the plasmid expressing IntIA or the empty one (Fig 3a). 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 3a). Very few shuffling events could be detected in SI Reinv clones (3/48), while we quasi-systematically observed large sized amplicons in the SI Inv clones (46/48) translating the insertion of multiple cassettes at the *attIA* site (Fig 3b). This further confirmed the increase of cassette shuffling in the SI Inv strain, in accordance with the integration assay results (see above). Furthermore, considering the virtually constant integration rate at the *attIA* site in both orientations of the SI, this increase in shuffling strongly suggested an important increase of the cassette excision rate in the SI Inv strain

To monitor cassette excision in both oriented SI strains, we designed a SI-wide cassette excision assay (Fig 3c). Similarly, to the previous shuffling test, single clones of the *V. cholerae* strains with the SI in both orientations were used to inoculate overnight cultures in absence or in presence of the integrase, but this time in absence of the *attIA* site to focus on the excision events only. To filter the cells that died following an excision event, the overnight cultures were then used to inoculate a second culture in conditions where the expression of the integrase was turned off (see Material and Methods). Even though all the cultures initially derived from a single clone of the respective strains, the second culture consisted in an heterogenous population where each of the individuals may have experienced a different set of excision events that could be revealed by sequencing. With this approach, a cassette regularly found to be missing in the final culture could be either lost early in the culture or excised at a high frequency. To avoid the confusion between the timing and the frequency of excision events,

each strain was cultivated 8 times on different days starting from independent clones, then the genomic DNA (gDNA) of each culture was extracted and pooled in a balanced manner. This way the amplification of rare excision events randomly occurring early in the assay was significantly diluted. Finally, as most SI cassettes are not expressed, a cassette found consistently lost in our assay after pooling of multiple independent replicates could be interpreted as a cassette that excised at a high frequency rather than a cassette that negatively impacted the fitness of its host.

Since we focused on excision events, that is structural variants, we sequenced the resulting mixed gDNA samples on a PacBio platform. For each population, a total of ~25000 reads with an average length of ~7kb were produced and the coverage within the SI was about 50X leading to a detection limit of about 2×10^{-2} for single cassette excision events or 10^{-4} when considering the whole SI (179 cassettes). The read mapping was performed using minimap2 and only a dozen of reads out of ~100000 failed to be mapped. Each molecule sequenced from the mix of gDNA captioned the state of a part of the cassette array in a single individual from an heterogenous population of the correspondent strain. The cassette deletions could be recovered from the mapping data and 475 cassette excision events could be recorded in the population of *V. cholerae* with an inverted SI and expressing the integrase. Not a single event was detected in the other populations, that is in absence of integrase, or in presence of integrase in the native orientation of the SI. Thus, we conclude that in these strains the excision rate is below 10^{-4} (i.e., the limit of detection).

In the SI Inv, 60 cassettes were not found to be excised in our assay. These may be cassettes with a low excision frequency or cassettes whose excision was deleterious for the cell, knowing that these included the 13 cassettes coding for non-duplicated TA modules. The other 120 cassettes excised with an average frequency of 7×10^{-2} with frequencies ranging from 1.2×10^{-2} to 4.4×10^{-1} (Fig 3e). When comparing these data with the detection limit SI-wide, we concluded that the inversion of the SI leads at least to a 2500-fold increase of the cassette excision rate upon expression of the integrase. This striking augmentation of the excision rate 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.

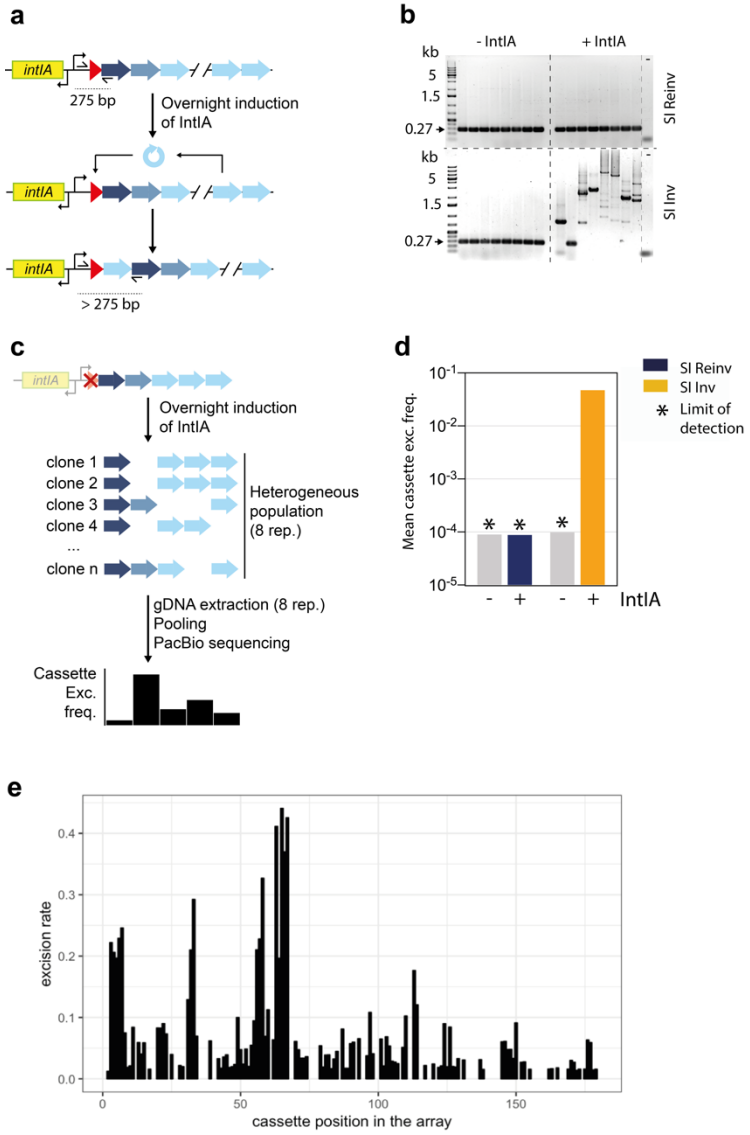

Figure 3: Cassette shuffling assay in the Super Integron of *V. cholerae* and effect of SI inversion on its plasticity.

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

b. PCR analysis of the cassettes shuffled in 1st position in individual clones exposed or not to the integrase 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 single clone 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. The final result that is cassette excision frequency for each cassette is figuratively represented at the bottom.

d. Average 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 (120 cassettes with at least 1 excision event recorded). No excision event could be detected in the other conditions; hence the limit of detection is represented instead (~10^{-4}). The grey bars represent the frequencies obtained in absence of integrase in the SI Inv and SI Reinv strains.

e. Cassette excision frequency for each of the 180 cassettes of the SI, defined as the frequency at which a read is found with a deletion corresponding to a cassette. Minimum value: 0.012; maximum value: 0.44. Mean (excluding the values of 0): 0.07.

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

The increase in the frequency of cassette excision upon inversion of the SI was such that we reasoned that there must be a limiting factor preventing such an event to occur in natural settings. Indeed, genome rearrangements are not very rare [25], and if a single inversion was sufficient to empty a SCI of all its cassettes in a few generations, then the massive structures of SCIs would not exist. Hence, we thought 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 4a). 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. We evaluate the growth rate to be ~35% lower in the SI Inv strain expressing the integrase compared to the controls (Fig4b). This confirms the existence of a strong selective pressure on the orientation of the SI. This severe phenotype was observed by overexpressing the integrase on a low copy number pSC101 plasmid, which led us to investigate the outcome of growing the SI Inv strain while inducing the endogenous integrase. The integrase is part of the SOS regulon [7,9,26], which can be triggered by the action of DNA damaging agents such as ciprofloxacin [8]. We cultivated the same strains, devoid of any plasmid and in presence of sub-lethal concentrations of ciprofloxacin (Fig 4c). We did observe a growth defect in all strains but with a stronger impairment in the SI Inv strain, 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 a 1:1 ratio (Fig 4d). In this experiment, we used the presence of the *bla* gene conferring resistance to carbenicillin specifically in the SI Inv strain to discriminate it from the SI Reinv strain. The co-cultures were plated on *ad hoc* plates and the CFU measured at different time points. While no substantial deviation to the original ratio between SI Inv and Reinv strains could be observed when those strains contained an empty plasmid, the expression of the integrase in those strains led to a strong disadvantage of the SI Inv strain compared to the SI Reinv. Indeed, 6h after the initiation of the culture (beginning of the stationary phase), the competition index of the SI Inv strains was of ~ 0.1 and continued its decrease to reach ~0.05 after 24h of co-culture (Fig 4d). This confirms the high fitness cost of the inversion of the SI of *V. cholerae* in presence of the integrase.

The absence of any phenotype on the growth of the SI Inv strain in the absence of integrase indicated that this protein had a key role in the growth defect observed when the SI is inverted. We set out to test if that could be explained by the mere increased binding of $attC_{bs}$ by the integrase and the subsequent disruption of replication as it can be observed with many DNA binding proteins [27]. In that scenario, the actual cleavage activity of the integrase would not be necessary to cause a growth defect upon inversion of the integrase. The integrase is a tyrosine-recombinase, thus mutation of the catalytic tyrosine 302 into a phenylalanine (Y302F) leads to a catalytically inactive integrase [28]. Through electrophoretic mobility shift assay (EMSA), we showed that a variant of IntIA mutated in the catalytic tyrosine (Y302F) had a binding capacity comparable to that of the WT IntIA (Fig S3). Hence, we tested the effect of the expression of that Y302F mutated integrase that binds to the $attC_{bs}$ without cleaving it, in the context of an inverted SI (Fig 4e). We did not observe any growth defect in that setting, ruling out a scenario where the mere binding of more accessible $attC_{bs}$ in the SI Inv strain could explain the growth defect upon expression of a WT integrase in that strain. 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 4f). We still observed a growth defect, indicating that it is the increased cassette excision rate that explained the growth defect of the SI Inv strain.

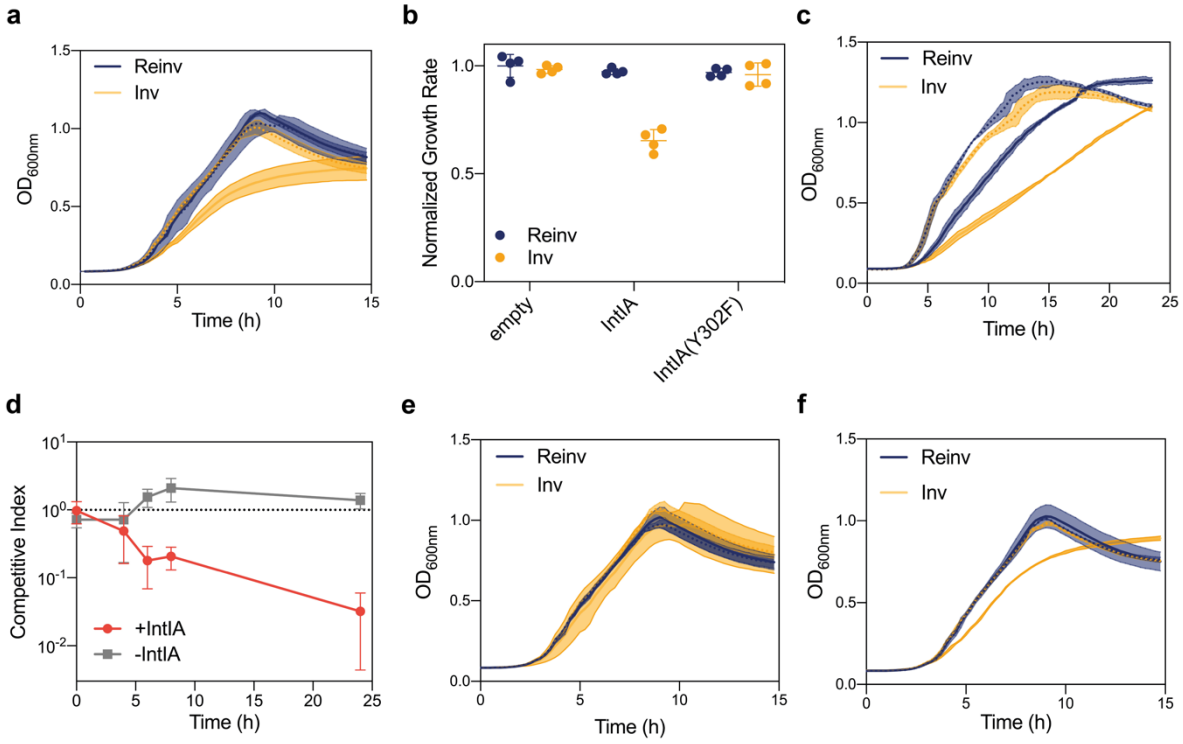

Figure 4: Growth of the Super Integron inverted and reinverted *V. cholerae* strains

a. Growth curve of the strains of *V. cholerae* with its SI Inv or Reinv in presence of an empty plasmid (dotted line) or of a plasmid expressing IntIA (full line). 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 (orange) and SI Reinv (blue) strains normalized by the mean growth rate of the SI Reinv strain in absence of integrase. The normalized growth rate of each strain expressing respectively no integrase, the WT IntIA and the catalytically inactive integrase (IntIA Y302F) are represented.

c. Growth curve of the SI Inv and Reinv strains without any plasmid but in presence (full lines) or in absence (dotted lines) of a sub-inhibitory concentration of ciprofloxacin (10 $ng.mL^{-1}$).

d. Competitive index of the SI Inv strain compared to the Reinv strain in presence of an empty plasmid (gray) or in presence of a plasmid containing the integrase (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 an empty plasmid (dotted line) or of a plasmid expressing the catalytically inactive IntIA (Y302F mutation, full line)

f. Growth curve of the SI Inv and Reinv strains in presence of an empty plasmid (dotted line) or of a plasmid expressing 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

Growth curves provide information about the behavior of a population of bacterial cells. Therefore, the growth defect previously observed in the integrase-expressing SI Inv strain could just as easily result from a slower division rate of each individual cell as it could be explained by 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 100 min (Fig 5a). In both the SI Inv or Reinv strains expressing the integrase, we could observe 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 5b). This reflects the fact that, whereas in the SI Reinv strain, 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 [29], could also be observed specifically in the SI Inv strain (Fig 5c). 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 5d). 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 5e), where a basal mortality rate of ~3-5% was measured, compared to ~16% in the SI Inv strain expressing a WT integrase. Interestingly, expression of a catalytically inactive integrase in the SI Inv strain resulted to a mortality rate similar to the basal level (Fig 5e), consistent with the observation that was made in the previous growth measurement. These data indicated that inversion of the SI led to a high mortality rate that was dependent on the activity of the integrase.

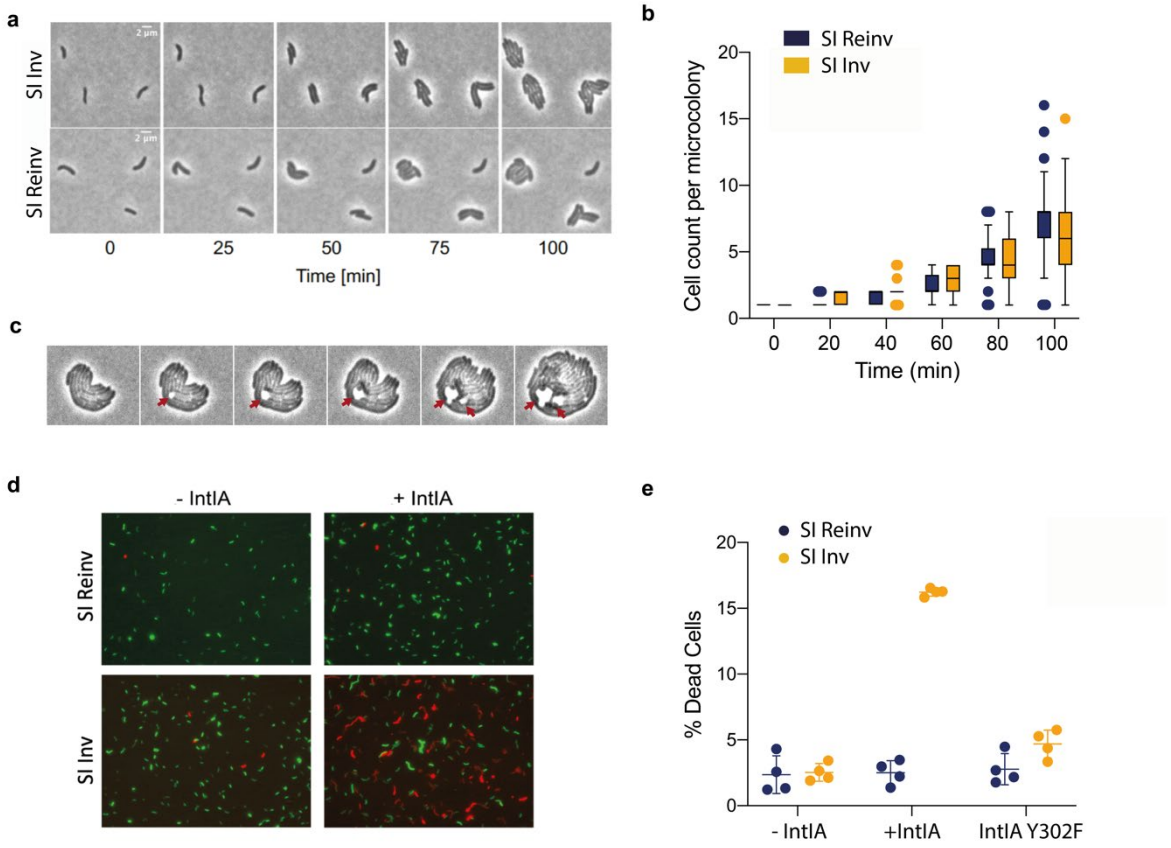

Figure 5: Cell viability of the Super Integron inverted and reinverted *V. cholerae* strains

a. Bright-field microscopy images taken from 100 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 100 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.

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.

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 modules

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. Indeed, while most cassettes do not carry a promoter and their loss through excision should be harmless for the cell, a significant subset of cassettes encode Toxin-Antitoxin (TA) modules that do carry a promoter [19,30,31]. TA modules 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 [32]. In *V. cholerae*, 20 TAs have been identified, 19 of which are concentrated within the SI [31] (Fig 6a). 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 (Fig 5). 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 Material and Methods) of their respective *attC* sites (Fig S4). To test this hypothesis, we set out to knock-out all the TA modules simultaneously in order to check if that would restore a normal growth in the SI Inv strain in presence of the integrase. Due to the large number of TA modules, we designed a multiplex approach to effectively inactivate them. While the classical CRISPR-Cas approaches have already been used for multiplex editing in bacteria [33,34], the number of targets (19 toxins) was a limiting factor in our case, which led us to use base editing that had been successfully used in *E. coli* to edit 6 loci simultaneously [35]. The base editing tool that we used was based on a dCas9 vector extensively used in *E. coli* [36] and that we showed to enable efficient CRISPR interference in *V. cholerae* (Fig S5a). Following the approach proposed by Banno and coll. [35], dCas9 was fused to the *Petromyzon marinus* cytosine deaminase (PmCDA1), an Uracil Glycosylase Inhibitor (UGI) and a degradation signal peptide (LVA) (Fig S5b). The efficiency of the base editor was assessed by introducing a non-sense mutation in the *lacZ* gene to get an easy read-out of the mutation (Fig S5c). The introduction of a non-sense mutation through base-editing implies to find a codon that can be easily switched to a STOP codon near a PAM sequence, a

constraint that limits the available targets. One of the guides satisfying these conditions in the *lacZ* gene was tested and led to a base editing efficiency close to 100% (Fig S5d), allowing to extend this approach to target TA cassettes. Due to the constraints of the base editing gene inactivation strategy, 4 toxins did not contain any satisfying target and had to be inactivated through a classical gene replacement approach (see Material and Methods). 12 guides were designed to target the remaining 15 toxins (3 of which are duplicated) in such a way that the STOP introduced resulted at least to a truncation of a 1/3rd of the toxin (Fig S5e). The co-expression of the array of guides and the base-editor led to the very efficient simultaneous inactivation of all the targeted toxins (Fig 6b). It resulted in a *V. cholerae* strain where all the TA within the SI were inactivated (TAi). Multiple clones of SI Reinv and SI Inv strains with mutated TA were generated in this way.

Once the TA were inactivated in both SI Inv and Reinv strains, we assessed the effect this could have on growth. In absence of integrase (Fig 6c, left panel), apart from a slight growth defect of the SI Reinv TAi strain compared to the SI Reinv strain, we do not observe significant difference between all the strains whether the SI is inverted or not. In presence of integrase, we find again the growth defect characteristic of the SI Inv strain, but this appears to be substantially rescued upon inactivation of the TA modules (Fig 6c, right panel). Importantly, because inactivation of the TA systems seems to also have affected the SI Reinv strain, the difference between SI Inv and Reinv in TAi context appears now negligible. This was further confirmed by a competition assay (Fig 6d), where we found again the SI Inv to be severely out competed by the SI Reinv strain but which was completely alleviated in absence of TA systems (Fig 6d). 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 6e). 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).

In absence of TA cassettes, the lack of constraint on the cassette array made it possible that increased fitness of the SI Inv TAi expressing IntIA could simply be due to an early loss of its array and its possible associated cost. For this reason, we sequenced one clone stemming from the culture of 4 SI Inv TAi independent strains to assess its 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) which could not

explain the increased growth rate in the hypothesis that the growth defect of the SI Inv was only due to an increased cost of the cassette array.

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 asked what could be the missing factor explaining this gap. We reasoned that the intense recombination activity accompanying the inversion of the SI in presence of integrase could also affect the replication process. There are examples of macromolecular conflicts between DNA binding complex and the replication fork, sometimes leading to important impact on growth [27,37]. 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 [37,38] and previously used in *V. cholerae* [39]. Indeed, in such settings, the ratio between loci near the *ori* are enriched compare to those near the *ter* region and this translates into a gradient of coverage from *ori* to *ter* upon mapping of the sequencing data. Importantly, the MFA profile is obtained by normalizing the coverage data of the exponentially growing bacterial population with the coverage data of the same population at stationary phase. This way, a drop of coverage cannot be interpreted as a mapping issue, but rather to a local disruption of replication. In absence of integrase, the MFA profile seems normal for both the SI Inv and Reinv strains (Fig S8), but in presence of integrase, a severe drop of coverage can be observed at the genomic location corresponding to the SI specifically in the SI inv strain. This suggests that the activity of the integrase might also have a role to play in the decreased fitness of the SI Inv strain.

Therefore, we show that, if the growth defect of the SI Inv strain is effectively due in large part to increased excision of TA systems, it may also due, albeit with a more limited effect on fitness, to a possible disruption of the replication process by the high cassette dynamics in the SI Inv.

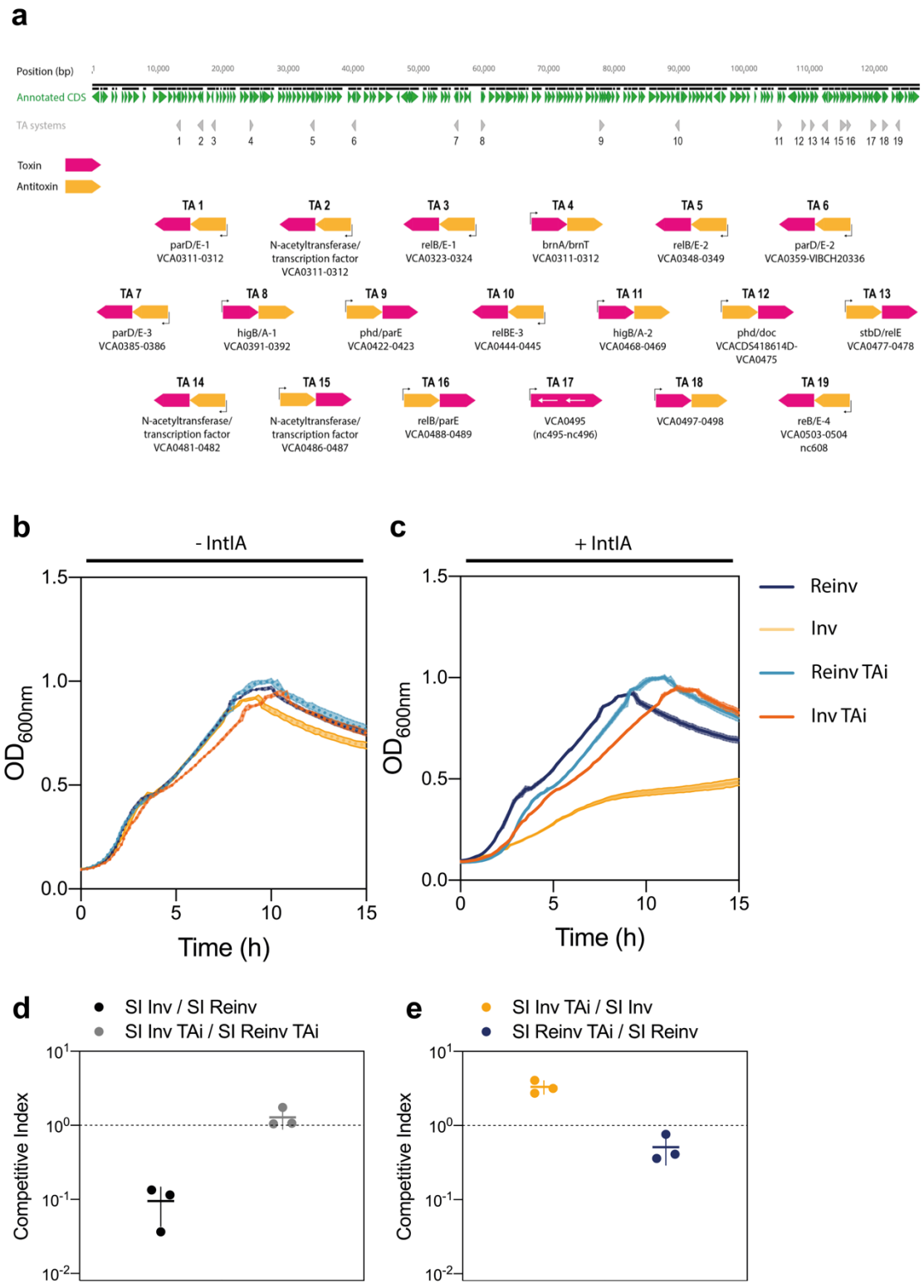

Figure 6: Partial rescue of the growth defect in the TA inactivated SI Inverted strains of *V. cholerae*.

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.

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 presence of an empty plasmid (dotted line) or c, in presence of a plasmid expressing IntIA (full line).

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). 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).

Toxin-Antitoxin landscape in Sedentary Chromosomal Integrons determined by comparative genomics

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 [40], 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 TAs (Fig 7a). 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 toxin/antitoxin 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 7b), 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 7c). 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 S7). 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.

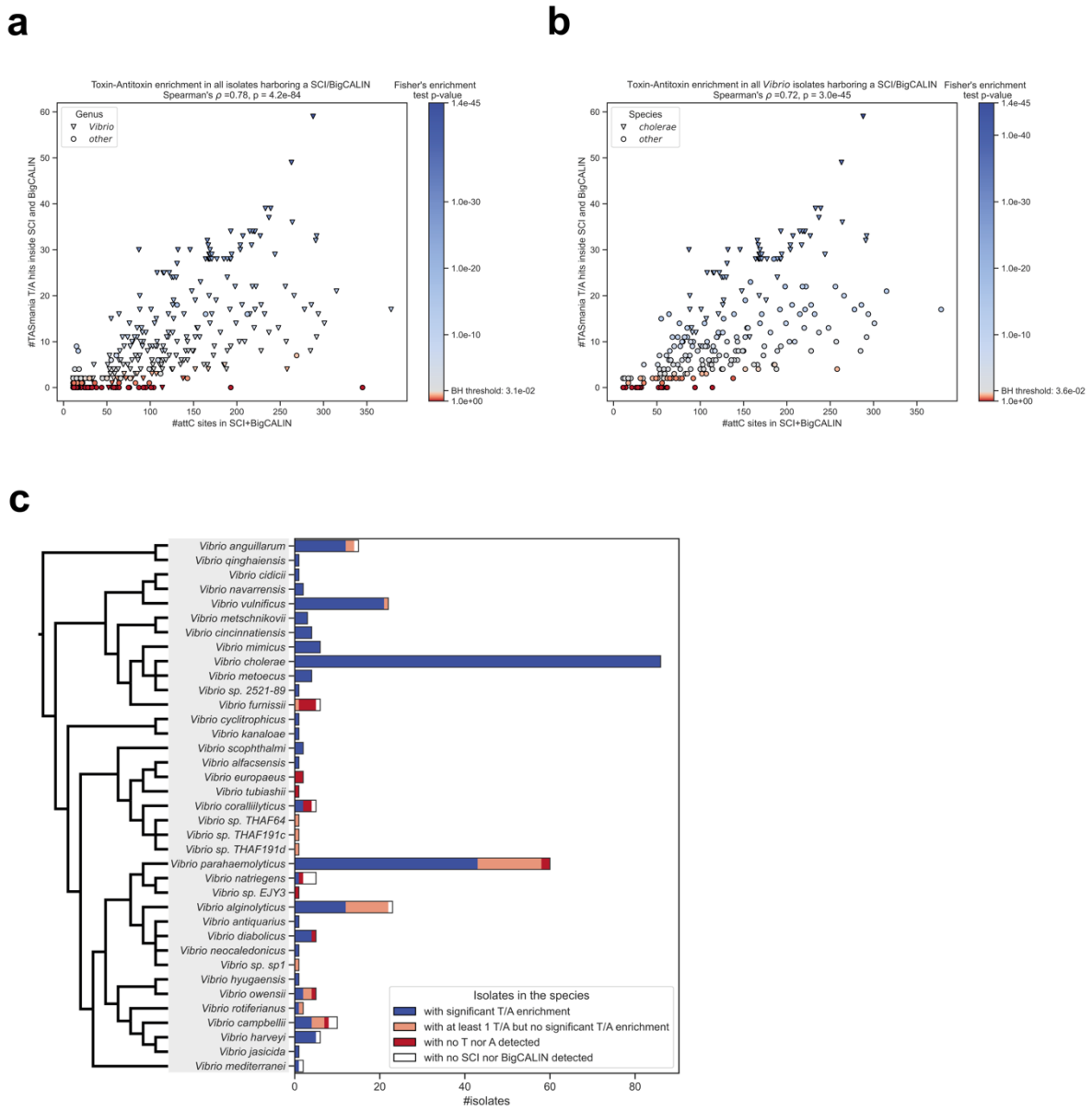

Figure 7: Toxin-Antitoxin landscape in SCIs determined by comparative genomics.

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. [41].

T: toxin, A: antitoxin

Discussion

We previously noticed a strong bias in the orientation of the large SCIs toward the replicon replication [16]. 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 plasticity of its cassette array. While cassette integration is not impacted by the inversion, the cassette excision frequency is highly increased. This is in accordance with the recombination mechanisms of the integron and the nature of its substrates. Indeed, the *attI* site recombines as *ds*DNA so that it was expected that its orientation within a genome should not affect its cassette recruitment capacity. In contrast, *attC* sites that recombine in their single-stranded form 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 that strand that facilitated the structuration of the $attC_{bs}$, leading to at least a 100-fold increase of cassette excision. The integrase strong preference for the $attC_{bs}$ as opposed to the $attC_{ts}$ is necessary for the proper function of the integron by allowing the integration of the cassettes in the correct orientation at the *attI* site. Our results show, for the first time in an SCI, that this specificity of the integrase allow to regulate the recombination potential of cassettes by acting on the availability of structured $attC_{bs}$. Therefore, the quasi-systematic presence of the $attC_{bs}$ on the leading strand template [16] 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.

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. Indeed, while the presence of *attC*$_{bs}$ on the leading strand template greatly increases the stability of SCIs, this fact alone could not explain the existence of such massive and mostly silent structures. Genome rearrangements are not rare events, and if a single inversion could result in the loss of a non-transcribed region of several dozen kb 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 a selective pressure on their orientation. 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 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. 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. Most importantly, the discrepancy of growth between SI Inv and Reinv was nearly completely abolished in absence of functional TA cassettes within the array. This suggests that the increased cassette excision frequency in the SI Inv leads to the increased loss of TA modules and to the subsequent cell growth inhibition or cell death even if 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.

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 was proposed to avoid large scale rearrangements of the cassette array leading to the simultaneous loss of dozens of cassettes [9,19]. 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 presence or absence of 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 that arises from their direct association with an *attC* site (Fig 8). This unique association interconnects cassette excision frequency with cell viability. 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, 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. [32]. 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 proportional to the overall cassette excision rate (Fig 8). 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). We do not claim that this process has been positively selected for, but rather see it as an emergent property of TA systems once associated with an *attC* site, and one that is proving to be a very effective way to ensure the stability of large SCIs. 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)) [42]. This very low percentage of TA cassettes within MIs perhaps means that their recruitment can occur but lead to abortive excision of the TA cassette due to too high a cassette excision frequency in MIs compared to SCIs [16].

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 [19], 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 access a vast repertoire of cassettes and thus expand their genetic capacitance.

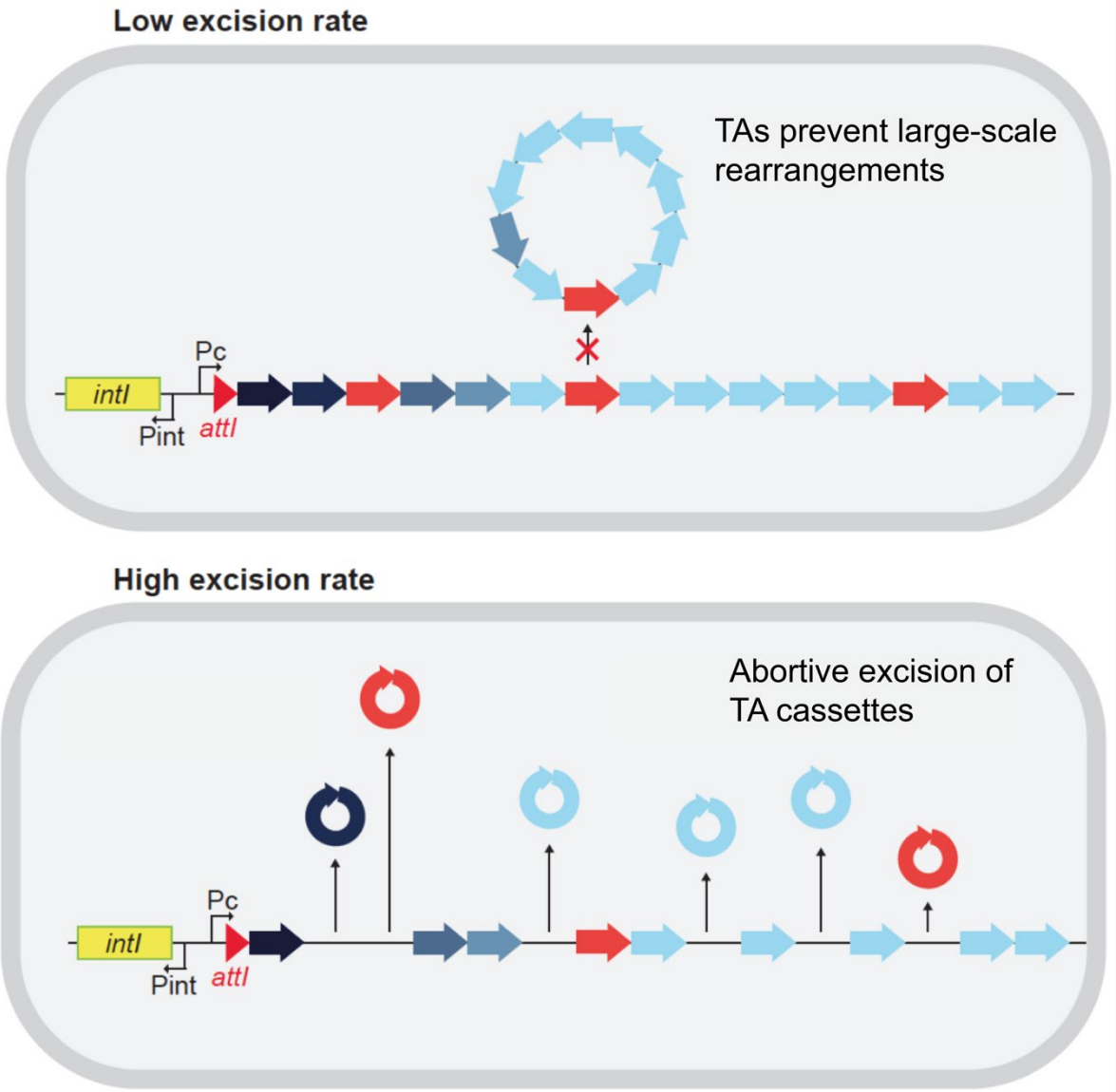

Figure 8: Model of the "abortive cassette excision" process in an integron array containing TA cassettes.

Model of the double action of TA cassettes (in red) within the SI to stabilize the array of cassettes. The classical activity preventing large scale rearrangements (up), and the proposed "abortive cassette excision" process driving low excision rates by killing the cell when the excision rate becomes too high.

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 strain constructions

Two DNA fragments were inserted respectively upstream and downstream of the SI in a N16961 hapR$^+$ strain (WT). 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. 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) 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. As a control for the following experiments, the SI was re-inverted to its original orientation using the HK_{022} integrase to perform the $attP_{HK} \times attB_{HK}$ reaction in the SI inv strain.

The high frequency of that recombination reaction allowed an easy screening of the re-inversion events by PCR. We used a genetic tool developed in the lab and designed to target the relocation of chromosomal DNA using bacteriophage attachment sites. Prophage excision from the host chromosome relies on site-specific recombination between to sequences flanking the phage,

termed attachment sites *attL* and *attR*. This recombination is carried out by phage-specific recombinases "Int" and its cognate RDF (recombination directionality factor), or excisionase, Xis. The tool developed in the lab uses the attachment sites from two different bacteriophages: HK and λ. Each attachment site the *attL/attR* pairs was associated with part of a marker that is rendered fully functional when the two sites recombine: HK sites were associated with a part of a *bla* gene (ampicillin resistance) and the λ sites with part of a *lacZ* gene. Co-expression of the HK and λ integrases/excisionases triggers the recombination of the partner *attL/attR* sites, which allows for the reconstitution of fully functional *bla* and *lacZ* markers and the selection of clones in which the relocation has occurred. This genetic tool has been used to construct the re-located SI strains used in this study. We used also this tool in this study to perform the SI inversion. In this case, we only used the HK bacteriophage properties. Indeed, we inserted the attL and attR HK sites on each side of the SI. The $attR_{HK}$ site is associated with the km resistance conferring gene and the $attL_{HK}$ site with the zeocin resistance conferring gene. Each *att* sites were associated with a part of a *bla* gene (ampicillin resistance). Expression of the HK excisionase triggers the recombination of the partner *att* sites, which allows for the reconstitution of fully functional *bla* marker and the selection of clones in which the SI inversion has occurred.

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 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 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 re-inverted 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 pB203, pK590, pK584 and p6780. We followed the same protocols as previously described [43]. 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.

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 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.

CRISPR interference assays

In order to silence gene expression, we use CRISPR dead Cas9 (dCas9), allowing to target a specific locus, designated by a sgRNA, and to disrupt transcription (CRISPR interference or CRISPRi) without causing dsDNA breaks like with a normal CRISPR Cas9. The plasmid used to express dCas9 is a medium copy number plasmid carrying a p15A origin and the guide(s) is (are) on that same plasmid. The plasmid carries a chloramphenicol resistance gene (cat). While the guide(s) is (are) expressed in a constitutive way, the dCas9 gene is under the control of the pHlf promoter inducible by the 2,4-Diacetylphloroglucinol (DAPG). Induction was performed using a final concentration of 50uM. The guides were inserted using the Golden Gate Assembly approach (see below).

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'–TAGTNNNNNNNNNNNNNNNNNNN–3' and 5'-TTTGNNNNNNNNNNNNNNNNNNN-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 oligo.

Base editing

The tool used for base editing is almost the same that the one used for interference, except that we use a modified version of dCas9 as described in [35].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. 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.

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 100 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.

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 [40]. 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 [44] (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 17 TA systems [31], 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. [31]), 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 was 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. [41] 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 Marc Monot, Laurence Ma, Juliana Pipoli Da Fonseca and Thomas Cocklaer from the Biomics platform, C2RT, Institut Pasteur, Paris, France.

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.; Hall, R.M. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol. Microbiol.* **1989**, *3*, 1669–1683, doi:10.1111/j.1365-2958.1989.tb00153.x.
2. Cury, J.; Jové, T.; Touchon, M.; Néron, B.; Rocha, E.P. Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res.* **2016**, *44*, 4539–4550, doi:10.1093/nar/gkw319.
3. Cambray, G.; Guerout, A.; Mazel, D. Integrons. **2010**, doi:10.1146/annurev-genet-102209-163504.
4. Néron, B.; Littner, E.; Haudiquet, M.; Perrin, A.; Cury, J.; Rocha, E.P.C. IntegronFinder 2.0: Identification and Analysis of Integrons across Bacteria, with a Focus on Antibiotic Resistance in Klebsiella. *Microorganisms* **2022**, *10*, doi:10.3390/microorganisms10040700.
5. Mazel, D.; Dychinco, B.; Webb, V.A.; Davies, J. A distinctive class of integron in the Vibrio cholerae genome. *Science* **1998**, *280*, 605–8, doi:10.1126/science.280.5363.605.
6. Collis, C.M.; Hall, R.M. Expression of antibiotic resistance genes in the integrated cassettes of integrons. *Antimicrob. Agents Chemother.* **1995**, *39*, 155–162, doi:10.1128/aac.39.1.155.
7. Guerin, É.; Cambray, G.; Sanchez-Alberola, N.; Campoy, S.; Erill, I.; Re, S. Da; Gonzalez-Zorn, B.; Barbé, J.; Ploy, M.C.; Mazel, D. The SOS response controls integron recombination. *Science (80-.).* **2009**, *324*, 1034, doi:10.1126/science.1172914.
8. Baharoglu, Z.; Mazel, D. SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiol. Rev.* **2014**, *38*, 1126–1145, doi:10.1111/1574-6976.12077.

9. Escudero, J.-A.; Loot, C.; Nivina, A.; Mazel, D. The Integron: Adaptation On Demand. *Microbiol. Spectr.* **2014**, *79*, 277–306, doi:10.1007/978-0-387-77863-1_14.
10. Richard, E.; Darracq, B.; Loot, C.; Mazel, D. Unbridled Integrons : A Matter of Host Factors. *Cells* **2022**, *11*, 1–22.
11. Johansson, C.; Kamali-Moghaddam, M.; Sundström, L. Integron integrase binds to bulged hairpin DNA. *Nucleic Acids Res.* **2004**, *32*, 4033–4043, doi:10.1093/nar/gkh730.
12. Bouvier, M.; Demarre, G.; Mazel, D. Integron cassette insertion : a recombination process involving a folded single strand substrate. *EMBO J.* **2005**, *24*, 4356–4367, doi:10.1038/sj.emboj.7600898.
13. Bouvier, M.; Ducos-Galand, M.; Loot, C.; Bikard, D.; Mazel, D. Structural features of single-stranded integron cassette attC sites and their role in strand selection. *PLoS Genet.* **2009**, *5*, doi:10.1371/journal.pgen.1000632.
14. Nivina, A.; Escudero, J.A.; Vit, C.; Mazel, D.; Loot, C. 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.* **2016**, *44*, 7792–7803, doi:10.1093/nar/gkw646.
15. Loot, C.; Ducos-Galand, M.; Escudero, J.A.; Bouvier, M.; Mazel, D. Replicative resolution of integron cassette insertion. *Nucleic Acids Res.* **2012**, *40*, 8361–8370, doi:10.1093/nar/gks620.
16. Loot, C.; Nivina, A.; Cury, J.; Escudero, J.A.; Ducos-Galand, M.; Bikard, D.; Roch, E.P.C.; Mazela, D. Differences in integron cassette excision dynamics shape a trade-off between evolvability and genetic capacitance. *MBio* **2017**, *8*, doi:10.1128/mBio.02296-16.
17. Loot, C.; Bikard, D.; Rachlin, A.; Mazel, D. Cellular pathways controlling integron cassette site folding. *EMBO J.* **2010**, *29*, 2623–2634, doi:10.1038/emboj.2010.151.
18. Rowe-magnus, D. a; Guerout, A.; Biskri, L.; Bouige, P.; Mazel, D. Comparative Analysis of Superintegrons : Engineering Extensive Genetic Diversity in the Vibrionaceae Comparative Analysis of Superintegrons : Engineering Extensive Genetic Diversity in the Vibrionaceae. *Genome Res.* **2003**, 428–442, doi:10.1101/gr.617103.
19. Szekeres, S.; Dauti, M.; Wilde, C.; Mazel, D.; Rowe-Magnus, D.A. Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Mol. Microbiol.* **2007**, *63*, 1588–1605, doi:10.1111/j.1365-2958.2007.05613.x.

20. Bland, M.J.; Ducos-Galand, M.; Val, M.E.; Mazel, D. An att site-based recombination reporter system for genome engineering and synthetic DNA assembly. *BMC Biotechnol.* **2017**, *17*, 1–10, doi:10.1186/s12896-017-0382-1.
21. Vit, C.; Richard, E.; Fournes, F.; Whiteway, C.; Eyer, X.; Lapaillerie, D.; Parissi, V.; Mazel, D.; Loot, C. Cassette recruitment in the chromosomal Integron of Vibrio cholerae. *Nucleic Acids Res.* **2021**, *49*, 5654–5670, doi:10.1093/nar/gkab412.
22. Okazaki, R.; Okazaki, T.; Sakabe, K.; Sugimoto, K.; Sugino, A. Mechanism of DNA chain growth, possible discontinuity and unusual secondary structure of newly synthesized chains. *PNAS* **1967**, *95*, 63–70, doi:10.1016/0022-2836(75)90335-6.
23. Loot, C.; Parissi, V.; Escudero, J.A.; Amarir-Bouhram, J.; Bikard, D.; Mazela, D. The integron integrase efficiently prevents the melting effect of escherichia coli single-stranded dna-binding protein on folded attC sites. *J. Bacteriol.* **2014**, *196*, 762–771, doi:10.1128/JB.01109-13.
24. Mukhortava, A.; Poge, M.; Grieb, M.S.; Nivina, A.; Loot, C.; Mazel, D.; Schlierf, M. Structural heterogeneity of attC integron recombination sites revealed by optical tweezers. *Nucleic Acids Res.* **2019**, *47*, 1861–1870, doi:10.1093/nar/gky1258.
25. Belda, E.; Maya, A.; Silva, F.J. Genome rearrangement distances and gene order phylogeny in γ-proteobacteria. *Mol. Biol. Evol.* **2005**, *22*, 1456–1467, doi:10.1093/molbev/msi134.
26. Cambray, G.; Sanchez-Alberola, N.; Campoy, S.; Guerin, É.; Da Re, S.; González-Zorn, B.; Ploy, M.C.; Barbé, J.; Mazel, D.; Erill, I. Prevalence of SOS-mediated control of integron integrase expression as an adaptive trait of chromosomal and mobile integrons. *Mob. DNA* **2011**, *2*, 1–15, doi:10.1186/1759-8753-2-6.
27. Lang, K.S.; Merrikh, H. The Clash of Macromolecular Titans: Replication-Transcription Conflicts in Bacteria. *Annu. Rev. Microbiol.* **2018**, *72*, 71–88, doi:10.1146/annurev-micro-090817-062514.
28. Butler, M.I. Tyrosine Recombinase Retrotransposons and Transposons. *Mob. DNA III* **2015**, 1271–1291, doi:10.1128/microbiolspec.mdna3-0036-2014.
29. Wong, F.; Amir, A. Mechanics and Dynamics of Bacterial Cell Lysis. *Biophys. J.* **2019**, *116*, 2378–2389, doi:10.1016/j.bpj.2019.04.040.
30. Rowe-Magnus, D.A.; Guerout, A.M.; Biskri, L.; Bouige, P.; Mazel, D. Comparative analysis of superintegrons: Engineering extensive genetic diversity in the vibrionaceae. *Genome Res.* **2003**, *13*, 428–442, doi:10.1101/gr.617103.
31. Iqbal, N.; Guérout, A.M.; Krin, E.; Le Roux, F.; Mazel, D. Comprehensive functional

analysis of the 18 Vibrio cholerae N16961 toxin-antitoxin systems substantiates their role in stabilizing the superintegron. *J. Bacteriol.* **2015**, *197*, 2150–2159, doi:10.1128/JB.00108-15.

32. Jurėnas, D.; Fraikin, N.; Goormaghtigh, F.; Van Melderen, L. Biology and evolution of bacterial toxin-antitoxin systems. *Nat. Rev. Microbiol.* **2022**, *0123456789*, doi:10.1038/s41579-021-00661-1.
33. Jiang, W.; Bikard, D.; Cox, D.; Zhang, F.; Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **2013**, *31*, 233–239, doi:10.1038/nbt.2508.
34. Ao, X.; Yao, Y.; Li, T.; Yang, T.T.; Dong, X.; Zheng, Z.T.; Chen, G.Q.; Wu, Q.; Guo, Y. A multiplex genome editing method for Escherichia coli based on CRISPR-Cas12a. *Front. Microbiol.* **2018**, *9*, 1–13, doi:10.3389/fmicb.2018.02307.
35. Banno, S.; Nishida, K.; Arazoe, T.; Mitsunobu, H.; Kondo, A. Deaminase-mediated multiplex genome editing in Escherichia coli. *Nat. Microbiol.* **2018**, *3*, 423–429, doi:10.1038/s41564-017-0102-6.
36. Rousset, F.; Cabezas-Caballero, J.; Piastra-Facon, F.; Fernández-Rodríguez, J.; Clermont, O.; Denamur, E.; Rocha, E.P.C.; Bikard, D. The impact of genetic diversity on gene essentiality within the Escherichia coli species. *Nat. Microbiol.* **2021**, *6*, 301–312, doi:10.1038/s41564-020-00839-y.
37. Lang, K.S.; Hall, A.N.; Merrikh, C.N.; Ragheb, M.; Tabakh, H.; Pollock, A.J.; Woodward, J.J.; Dreifus, J.E.; Merrikh, H. Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. *Cell* **2017**, *170*, 787-799.e18, doi:10.1016/j.cell.2017.07.044.
38. Skovgaard, O.; Bak, M.; Løbner-Olesen, A.; Tommerup, N. Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Res.* **2011**, *21*, 1388–1393, doi:10.1101/gr.117416.110.
39. Val, M.E.; Marbouty, M.; De Lemos Martins, F.; Kennedy, S.P.; Kemble, H.; Bland, M.J.; Possoz, C.; Koszul, R.; Skovgaard, O.; Mazel, D. A checkpoint control orchestrates the replication of the two chromosomes of Vibrio cholerae. *Sci. Adv.* **2016**, *2*, doi:10.1126/sciadv.1501914.
40. Akarsu, H.; Bordes, P.; Mansour, M.; Bigot, D.J.; Genevaux, P.; Falquet, L. TASmania: A bacterial toxin-antitoxin systems database. *PLoS Comput. Biol.* **2019**, *15*, 1–28, doi:10.1371/journal.pcbi.1006946.
41. 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. Updating the Vibrio clades defined by multilocus sequence phylogeny: Proposal of eight new clades, and the description of Vibrio tritonius sp. nov. *Front. Microbiol.* **2013**, *4*, 1–14, doi:10.3389/fmicb.2013.00414.

42. Huyan, J.; Tian, Z.; Zhang, Y.; Zhang, H.; Shi, Y.; Gillings, M.R.; Yang, M. Dynamics of class 1 integrons in aerobic biofilm reactors spiked with antibiotics. *Environ. Int.* **2020**, *140*, 105816, doi:10.1016/j.envint.2020.105816.
43. Le Roux, F.; Binesse, J.; Saulnier, D.; Mazel, D. Construction of a Vibrio splendidus mutant lacking the metalloprotease gene vsm by use of a novel counterselectable suicide vector. *Appl. Environ. Microbiol.* **2007**, *73*, 777–784, doi:10.1128/AEM.02147-06.
44. Eddy, S.R. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755–763, doi:10.1093/bioinformatics/14.9.755.