Cassette recombination dynamics within chromosomal integrons are regulated by toxin–antitoxin systems.

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]* Céline Loot[1]*

[1]Institut Pasteur, Université Paris Cité, CNRS UMR3525, Unité Plasticité du Génome Bactérien, 75015 Paris, France.
[2]Sorbonne Université, ED515, F-75005, Paris, France.
[3]Institut Pasteur, Université Paris Cité, CNRS UMR3525, Microbial Evolutionary Genomics, 75015 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]Institut Pasteur, Université Paris Cité, CNRS UMR6047, Synthetic Biology, 75015 Paris, France.
[8]Department of Science, Systems and Models, Roskilde University, Roskilde DK-4000, Denmark.

*Corresponding authors: celine.loot@pasteur.fr and didier.mazel@pasteur.fr
Tel: +33 1 40 61 32 87

Short title: New role of toxin–antitoxin cassettes in integrons

Abstract

Integrons are adaptive bacterial devices that rearrange promoter less gene cassettes into variable ordered arrays under stress conditions, to sample combinatorial phenotypic diversity. Chromosomal integrons often carry hundreds of silent gene cassettes, with integrase-mediated recombination leading to rampant DNA excision and integration, posing a potential threat to genome integrity. How this activity is regulated and controlled, particularly through selective pressures, to maintain such large cassette arrays is unknown. Here we show a key role of promoter-containing toxin–antitoxin (TA) cassettes as abortive systems that kill the cell when the overall cassette excision rate is too high. These results highlight the importance of TA cassettes regulating

the cassette recombination dynamics and provide insight into the evolution and success of integrons in bacterial genomes.

Teaser

The accumulation of cassette functions in integrons is ensured by toxin–antitoxin systems which kill the cell when the cassette excision rate is too high.

Introduction

Integrons are bacterial recombination systems that are capable of capturing, stockpiling, and reordering mobile elements called cassettes, to control their expression. They were originally found to be the genetic systems responsible for the acquisition of antimicrobial resistance determinants in some mobile elements (*1*). Integrons contain a stable platform and a variable cassette array (Fig 1a). The stable platform is defined by 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. The cassettes in the array are generally composed of a promoter less 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 represent silent and valuable functions for the cell (*2*). These cassettes can be excised and then re-integrated at the *attI* integration site and therefore become expressed (*3*). This cassette shuffling ensures a combinatorial phenotypic diversity which 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 catalyzes the cassette shuffling, is only expressed in stress conditions (*4*). For this reason, integrons are described as "on demand" adaptation systems (*5, 6*).

The integron integrase (IntI) has a very singular place among the broad family of the tyrosine recombinases (*7*). 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 (*8, 9*). 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 than that of the top (*9*). This ensures a correct orientation of the cassette upon integration at the *attI* site, allowing its expression by the P_C promoter (*10*).

We distinguish two types of integrons: Mobile Integrons (MIs) carried by conjugative plasmids and mobilizable among bacteria, and Sedentary Chromosomal Integrons (SCIs). In addition to their sedentary nature, the main feature that distinguishes SCIs from MIs is the size of their cassette array. While MIs do not typically store more than 10 cassettes, SCIs often contain

dozens of cassettes — with the largest SCI from *Vibrio vulnificus* encoding 301 — thus encompassing a substantial fraction of their host genomes (*11*). SCI cassettes constitute an almost infinite variety of functions extending far beyond antibiotic resistance and linked to bacterial key adaptive functions (*12*). The high genetic capacitance of SCIs allows them to serve as a repository of cassettes for MIs to compile the most relevant repertoire of cassettes in a specific environment (*6, 13*). The remarkable adaptive success of the integron system is partly due to this intricate connection between SCIs and MIs. Despite all that, twenty-five years after their discovery, SCIs remain enigmatic genetic structures. We still ignore how such structures can accumulate so many silent cassettes and be maintained in bacteria to provide such an adaptive device for future stresses. Among the cassettes of SCIs, a distinct type is notable, encoding toxin–antitoxin (TA) systems and containing their own promoter (*14-17*). TA systems encode a stable toxin that inhibits an essential cellular process and an unstable antitoxin that counteracts its cognate toxin. When the expression of the gene pair is stopped, the antitoxin is degraded, resulting in cell death or growth arrest (*18*). The presence of toxin–antitoxin systems within SCI arrays likely prevents cassette deletion through large chromosomal rearrangements thus maintaining the array (*14, 15*). Nevertheless, this sole factor is not sufficient to understand how SCIs are kept so massive. Indeed, in stress conditions, when the integrase is expressed, the cassette excision rate must be tightly regulated. To ensure integron cassette accumulation, the cassette excision rate must be kept below the integration rate, and to maintain the large integrons, the two rates must be at least similar. An imbalance in favor of the excision rate would lead *in fine* to the loss of most silent cassettes. Here, directly linked to their cassette nature, we demonstrated that the TAs play an additional role regulating the cassette recombination dynamics. We used as model the *Vibrio cholerae* N16961 strain responsible for the ongoing 7th cholera pandemic and which contain a particularly massive SCI (~130 kb and 179 cassettes (*19*)). By developing a genetic assay to measure the SCI cassette dynamics at the population level, we demonstrate that, the cassette excision rate is dramatically increased all along the SCI when we "artificially" inverted it on-site. We show that the inversion of the SCI is also associated to a strong growth defect correlated with a high cell mortality rate in presence of a functional integrase only. By successfully inactivating all the 19 TA cassettes within the inverted SCI (*16, 17*) (Fig 1b), using highly multiplexed CRISPR base editing, we demonstrate that these cassettes are in a large part responsible for the observed mortality phenotype.

Our results highlight a new role for TA cassettes via "abortive cassette excision" depending on the integrase. Under conditions where the cassette excision rate is high and could compromise the maintenance of large SCIs, TA cassettes are also excised at a high rate as individual cassettes, thus killing cells. Through this, TAs exert a selective pressure that drives integrons to adopt a

configuration in which the excision rate of their cassettes is kept low. This has probably led to the preferential orientation of SCI with respect to the replication fork that we previously observed (*13*), and which limits the rate of excision. The discovery of this new type of regulation in integrons adds to the already known regulatory network. It enables silent cassettes to be accumulated and maintained, making them available to produce combinatorial phenotypic diversity and respond to future stresses.

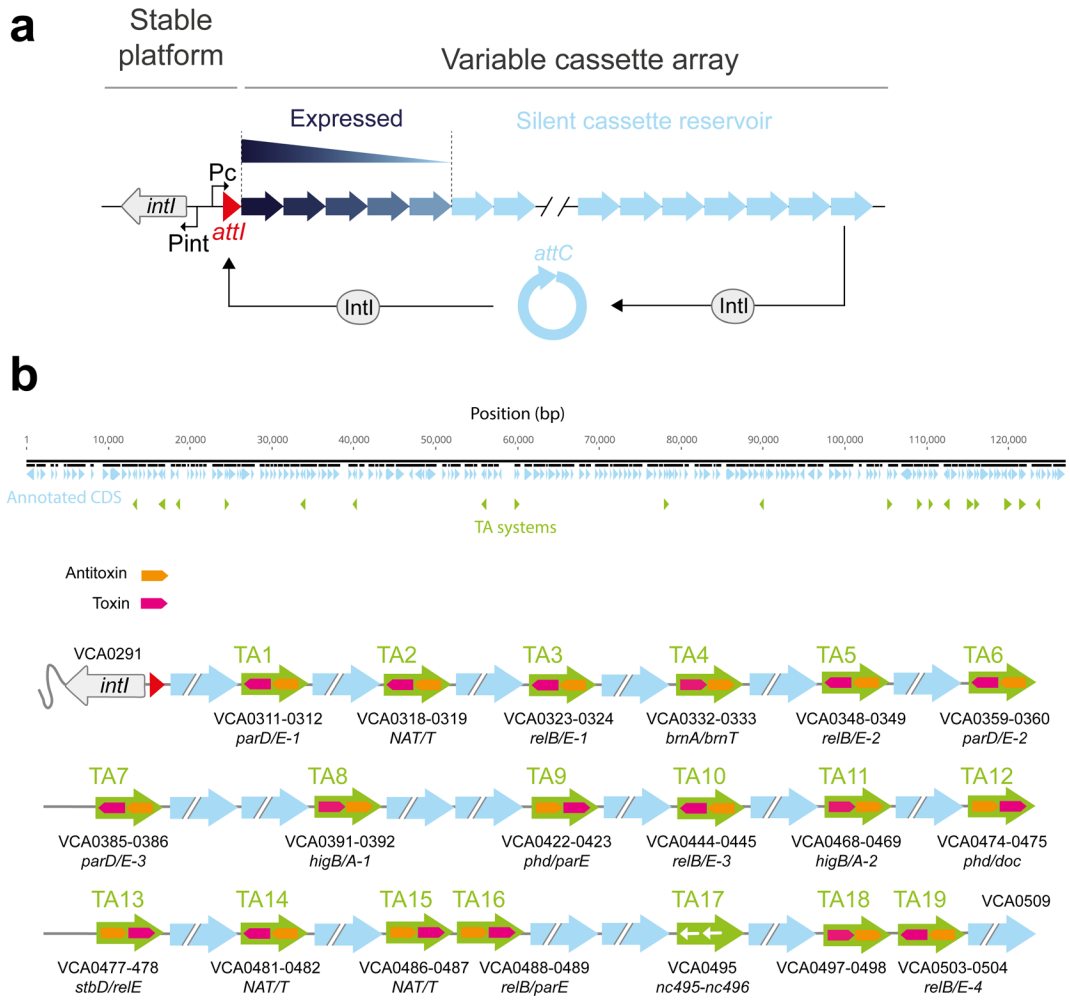
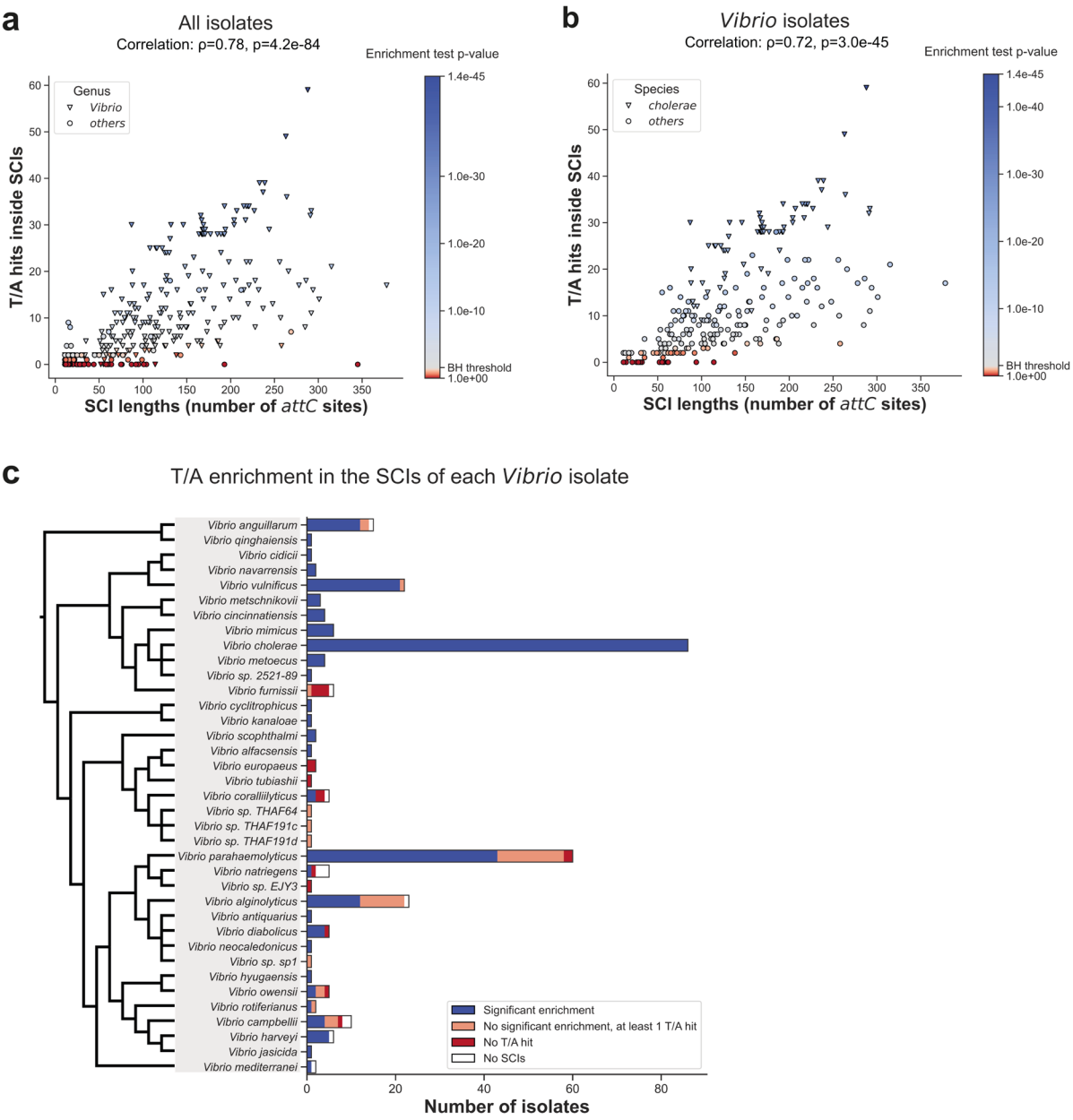


Figure 1: The integron system and the toxin–antitoxin cassettes.

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

122 promoter P_C driving the expression of the downstream cassettes along an expression gradient.
123 Cassettes are displayed as arrows where the tip represent the *attC* site. Their color intensity
124 represents their level of expression from dark blue (highly expressed) to light blue (not expressed).
125 Non-expressed cassettes can be excised through an intramolecular *attC* × *attC* reaction and be
126 reintegrated in 1st position near the P_C promoter through an intermolecular *attC* × *attI* reaction and
127 become expressed. The combination of excision and integration is referred to as cassette shuffling.
128 **b.** Representation of the SCI of *V. cholerae* and the repartition of the various toxin–antitoxin (TA)
129 systems numbered from 1 to 19. A detail of the TA systems (name, orientation in the array, order
130 of toxin/antitoxin) is represented below. NAT/T means N-acetyltransferase/transcription factor.
131
132 **Results**
133
134 **SCIs are specifically enriched in toxin–antitoxin systems**
135 To better understand the role of the TAs as cassettes in SCI arrays and to assess whether SCI size
136 and TA number are correlated, we systematically explored all TA carried by SCIs. We used
137 IntegronFinder 2.0 (*20*) to detect all of the sedentary integrons in complete genomes of the RefSeq
138 NCBI database. We identified 398 genomes with at least one SCI (defined as an integron
139 comprising strictly more than 10 cassettes, as suggested in (*20*)). In each of these genomes, we
140 screened TA systems with TASmania (*21*). Then, we compared the number of toxins and antitoxins
141 detected in the SCIs with the total number of cassettes in each of the 398 genomes. We found a
142 positive correlation between the two, indicating that larger SCIs have more TA cassettes (Fig 2a).
143 To assess the specific enrichment of TAs in SCIs when compared to the rest of the genome, we
144 performed in each isolate a contingency table analysis. Out of the 398 isolates, 246 verified the
145 hypothesis of a TA enrichment within SCI ($p<0.05$, Fisher test with Benjamini-Hochberg
146 correction). Of the remaining isolates, 87 genomes lack genes encoding toxins or antitoxins and 65
147 encode at least one of them but do not show significant enrichment. As most of the genomes with
148 a significant over-representation of TAs in SCIs belonged to the *Vibrio* genus (Fig 2a), we decided
149 to analyze them thoroughly. Among the 280 *Vibrio* genomes, 225 verified the hypothesis of a TA
150 enrichment within SCI (Fig 2b). This may reflect a particularly strong effect in this genus, but we
151 cannot exclude that it simply results from other SCIs having too few cassettes to provide a
152 statistically significant signal of over-representation. We studied the patterns of this over-
153 representation in function of the phylogeny of *Vibrio*. We observe that the enrichment signal is
154 scattered across the tree, highlighting the important role of TAs in a broad range of *Vibrio* species
155 (Fig 2c). This association holds true with the very stringent Bonferroni correction, although with a
156 less sparse distribution on the phylogenetic tree, centered around a clade comprising *V. cholerae*

157 (Fig S1). Interestingly, in this species, some TA systems (in the parDE family) are conserved in
158 most isolates (82/86). This conservation may suggest a long period of co-evolution of integrons and
159 TA systems, at least in *V. cholerae*. Although it does not exclude the regular turnover of some other
160 TA systems, TA cassettes may have contributed to the evolution of SCIs to a large extent.



161

162

163 **Figure 2: Toxin–antitoxin landscape in SCIs**

164 **a, b.** Each point represents a single NCBI RefSeq isolate harboring at least one SCI (>10 *attC* sites).
165 The x-axis displays the total number of *attC* sites in the (potentially multiple) SCI of the isolate.
166 The y-axis exhibits the total number of TASmania toxin and antitoxin HMM hits in the SCI of the
167 isolate (e-value<10^{-3}). The Spearman nonparametric rank correlation coefficient is shown at the top

168 of the graph, together with its significance p-value. Point hue indicates the p-value of Fisher's test
169 for TASmania hits enrichment in SCI compared to the rest of the genome (scale on the right of the
170 graph, significant enrichment in blue). The Benjamini-Hochberg correction for multiple testing was
171 applied (see below for the corresponding computed thresholds) **a.** All the isolates harboring at least
172 one SCI are displayed. The shape of each point on the graph represents whether the isolate belongs
173 to the *Vibrio* genus (triangles) according to the NCBI taxonomy or to another genus (circles).
174 Benjamini-Hochberg correction threshold: 0.031. **b.** Same graph as a, but only with isolates
175 belonging to the *Vibrio* genus. The shape of each point on the graph represents whether the isolate
176 belongs to the *Vibrio cholerae* species (triangle) or to another species (round). Benjamini-Hochberg
177 correction threshold: 0.036.
178 **c.** Distribution of *Vibrio* isolates by species according to the toxin/antitoxin enrichment inside the
179 SCI they harbor. The cladogram used to group species was adapted from the phylogeny
180 reconstructed by Sawabe and coll (*22*).
181 T: toxin, A: antitoxin
182 BH: Benjamini-Hochberg correction
183
184 **The inversion of the SCI in *V. cholerae* dramatically increases the cassette excision rate in the**
185 **array.**
186
187 Evolutionary factors have enabled the integron cassette accumulation and maintenance to
188 lead to the emergence of massive and stable chromosomal integrons. To specify these factors and
189 determine if TA cassettes have a role in this evolution process, we decided to render the cassette
190 array of large integrons unstable and determine how these "genetically modified" integron
191 structures can be counter-selected. To do this, we inverted the *V. cholerae* SCI on place (SCI Inv
192 strain) and re-inverted it back to its original orientation (SCI Reinv strain) (Fig 3a, Fig S2 and
193 Materials and Methods). Indeed, we previously establish that when the recombinogenic *bs* of *attC*
194 sites are located on the lagging strand template, where discontinuous replication leaves single-
195 strand gaps, this favors *attC* site folding and cassette excision (*13, 23*). We therefore hypothesized
196 that the natural SCI orientation, where the *bs* of *attC* sites are located on the leading strand template,
197 disfavors cassette excision events and that the SCI inversion would be a way to make the cassette
198 array unstable.
199 We first monitored the effect of SCI inversion on cassette shuffling by cultivating overnight
200 *V. cholerae* with the SCI in both orientations, in presence or absence of IntIA (Fig 3b). A PCR was
201 then performed on independent clones with primers located in the *attIA* integration site and in the

202 first cassette of the array (VCA0292) with an expected size of 275 bp in absence of shuffling (Fig
203 3b). Very few shuffling events could be detected in the SCI Reinv clones (3/48), whereas we
204 observed large amplicons in 46/48 of the SCI Inv clones, testifying to the insertion of multiple
205 cassettes at the *attIA* site (Fig 3c). Performing our classical conjugation assay to deliver cassettes
206 in *V. cholerae* recipient strains (*3*), we obtained a constant integration rate of these cassettes at the
207 *attIA* site regardless of the SCI orientation (Fig S3). Our results therefore demonstrate that SCI
208 inversion results in a substantial increase in cassette excision while the integration rate remains
209 stable.

210 To monitor the cassette excision rate at the population level in strains with SCI in both
211 orientations, we designed an SCI-wide cassette excision assay. Like the previous shuffling assay,
212 we grew *V. cholerae* strains overnight in the absence or presence of integrase, but this time in the
213 absence of the *attIA* site to focus solely on excision events. To assess cassette rearrangements along
214 the array, we used long-read sequencing (Materials and Methods, Fig 3d). Each read captured the
215 state of a segment of the SCI in a single cell and allowed an easy detection of cassette deletion on
216 this segment. For each population, a total of ~80000 reads with an average length of ~6.4 kb was
217 produced. For the *V. cholerae* population corresponding to the SCI inverted strain expressing
218 integrase, we identified 478 cassette excision events, encompassing 120 cassettes where excision
219 was observed at least once leading to a cassette excision rate of 3×10^{-2} (Fig 3e). For the populations
220 of strains with the SCI in the native orientation with and without integrase and, in the strain with
221 an inverted SCI in absence of the integrase, the excision rates are below 7×10^{-5} (*i.e.*, the limit of
222 detection, Fig 3e and Materials and methods). These results reveal a direct link between the
223 inversion of the SCI and a highly increased cassette excision frequency.
224

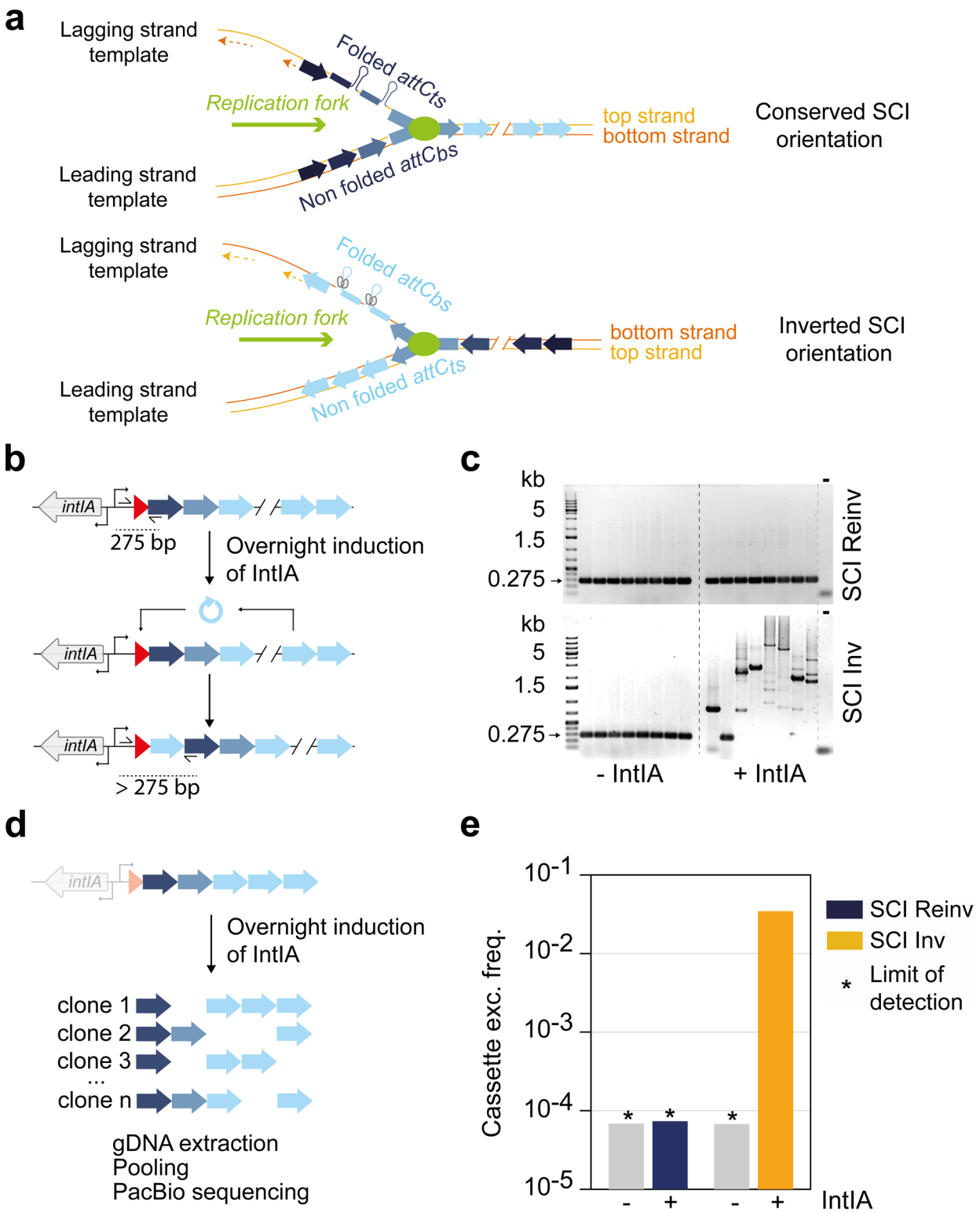
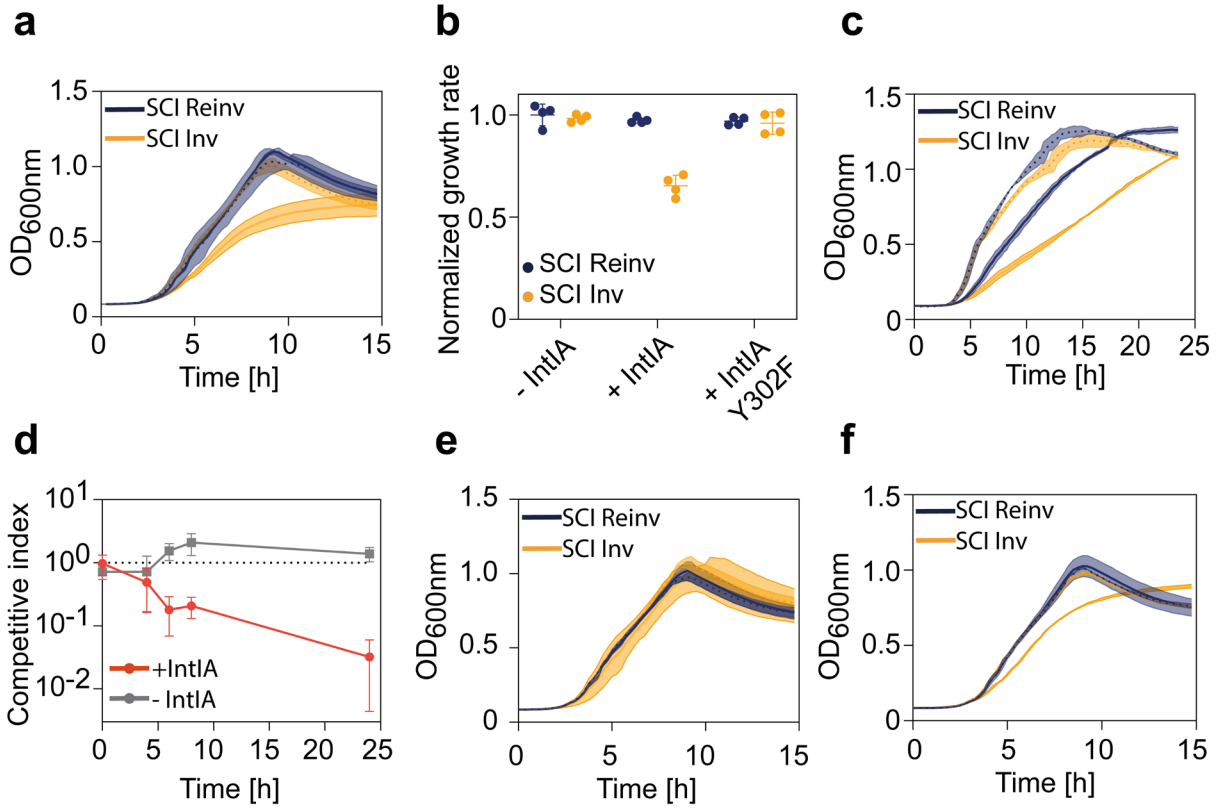


Figure 3: Cassette shuffling and excision assays in the SCI inverted and reinverted *V. cholerae* strains

229 **a.** Mechanistic insight on the issue of integron orientation. An array of cassettes is represented while
230 it is replicated. In the lagging strand, Okazaki fragments are represented as dotted lines, leaving
231 stretches of ssDNA on the correspondent template. In their conserved orientation, SCIs
232 recombinogenic $attC_{bs}$ are carried by the continuously replicated leading strand template which
233 supposedly limits their structuration. The structured $attC_{ts}$ are not recombined by the integrase. In
234 the inverted orientation their more frequent structuration is expected to lead to increased binding of
235 the integrase and higher recombination frequencies.
236 **b.** Schematic representation of the assay designed to detect cassette shuffling.
237 **c.** PCR analysis of the cassettes shuffled in 1st position in individual clones in presence (+IntIA) or
238 in absence of the integrase (-IntIA) for one overnight. The expected PCR product in absence of
239 shuffling is 275 bp. Longer bands are evidence shuffling events. Multiple bands may come from
240 the random slippage of the DNA polymerase in presence of highly repeated sequences (multiple
241 *attC* sites in one amplicon).
242 **d.** Schematic representation of the assay designed to detect cassette excision. A strain of *V. cholerae*
243 Δ*attIA* is used to initiate a culture composed of clones experiencing various cassette excision events.
244 Clones from this heterogeneous population are represented (from 1 to n) each with different
245 cassettes missing.
246 **e.** Cassette excision frequency in the array of the SCI. The cassette excision frequency is calculated
247 for the SCI inverted strain expressing the integrase (see Materials and Methods). No excision event
248 could be detected in the other conditions; hence the limit of detection (*) is represented instead.
249 The grey bars represent the frequencies obtained in absence of integrase in the SCI Inv and SCI
250 Reinv strains.
251
252 **The inversion of the SCI in *V. cholerae* leads to a growth defect in presence of a functional**
253 **integrase.**
254
255 Genome rearrangements are not very rare (*24, 25*), and if a single inversion is sufficient to increase
256 the excision cassette rate and empty a SCI of all its cassettes in a few generations, then the massive
257 and silent structures that are SCIs would probably not exist. We reasoned that there may be counter-
258 selective forces in situations in which the cassette array is unstable, preventing, for example, a
259 spontaneous inversion of the SCI. We set out to measure growth of the SCI Inv strain (Fig 4a).
260 While we did not observe any growth defect in the SCI Inv strain in absence of integrase, we found
261 an important growth defect upon expression of the integrase in this strain compared to the SCI
262 Reinv strain (Fig 4a). Growth rate is ~35% lower in the SCI Inv strain expressing the integrase

compared to the controls (Fig 4b). We observed the same growth defect while inducing the endogenous integrase through the SOS response using sub-lethal concentrations of ciprofloxacin (Fig 4c), confirming our previous conclusions in this more natural setting. To further evaluate the cost of the inversion of the SCI, we set out a competition experiment where SCI Inv and Reinv strains were co-cultivated for 24h starting with an initial 1:1 ratio (Fig 4d). While no substantial deviation to the original ratio between SCI 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 SCI Inv strain compared to the SCI Reinv. This confirms the high fitness cost of the inversion of the SCI of *V. cholerae* in presence of the integrase. Expressing a catalytically inactive integrase that binds to the $attC_{bs}$ (Fig S4) without cleaving it did not produce any growth defect (Fig 4e), ruling out a scenario where the mere binding of more accessible $attC_{bs}$ in the SCI Inv strain could explain the growth defect. Hence, the cleavage activity of the integrase was necessary to induce a growth defect in the SCI Inv. Yet, the integrase can cleave both *attIA* and *attC* sites so to determine if it was the increased shuffling or the mere cassette excision that imposed such a strong cost of SCI inversion, we performed growth curve in cells lacking the *attIA* site (Fig 4f). We still observed a growth defect, hence confirming that it is the increased cassette excision rate that explained the growth defect of the SCI Inv strain.



282

283 **Figure 4: Growth of the SCI inverted and reinverted *V. cholerae* strains**

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

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

292 **c.** Growth curve of the SCI Inv and Reinv strains in presence (full lines) or in absence (dotted lines)
293 of a sub-inhibitory concentration of ciprofloxacin (10 ng/ml) inducing the endogenous IntIA
294 expression.

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

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

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

304

305 **The growth defect in the SCI Inverted strain in presence of integrase is associated with**
306 **increased cell death.**

307

308 The growth defect previously observed in the integrase-expressing SCI Inv strain could
309 result from a slower division rate of each individual cell or a higher mortality rate at each
310 generation. To discriminate between these two possibilities, we needed information at the single
311 cell level. We therefore set out to observe in widefield microscopy live cells of our different strains
312 growing on an agarose pad mimicking the liquid medium used for generating the growth curves.
313 An exponentially growing culture of the SCI Inv and Reinv strains was diluted to display individual
314 cells onto the agarose pads. We recorded the development of microcolonies for 170 min (Fig 5a,
315 Movie S1a, S1b, S2a and S2b) and the number of cells per microcolony in the SCI Inv and Reinv

316 strains expressing the integrase was determined as a measure of growth. We observed that the SCI
317 Inv strain displayed increased heterogeneity in growth across microcolonies resulting in a broader
318 distribution of microcolony sizes at the end of the run (from t=120 min, Fig 5b, Movie S2a and
319 S2b). This reflects the fact that, whereas the SCI Reinv colonies generally follow an expected
320 developmental trajectory, in the SCI Inv strain, many microcolonies stop growing after a few
321 divisions or do not grow at all. Some events of cell lysis, characterized by sudden release by some
322 cells of diffracting matter (*26*), could also be observed specifically in the SCI Inv strain (Fig 5c,
323 Movie S2c). These observations favor a model where the expression of the integrase in the SCI Inv
324 strain induces cell death in a part of the population.

325 We tested this model in a live and dead assay where a differential coloration of the cells
326 allows discrimination between living (green) and dead (red) cells in a population (Fig 5d). We
327 observed a clear increase of the proportion of dead cells in culture of SCI Inv strain expressing the
328 integrase compared to the control conditions. We quantified the cell mortality and observed a high
329 percentage of dead cells (~16%) in the SCI Inv strain expressing the integrase and a basal mortality
330 rate (~3-5%) in the other strains (Fig 5e). Interestingly, expression of a catalytically inactive
331 integrase in the SCI Inv strain resulted to a mortality rate similar to the basal level (Fig 5e),
332 consistent with the observation that was made in the previous growth measurement (Fig 4b and 4e).
333 These data indicated that inversion of the SCI led to a high mortality rate that was dependent on the
334 activity of the integrase.
335
336

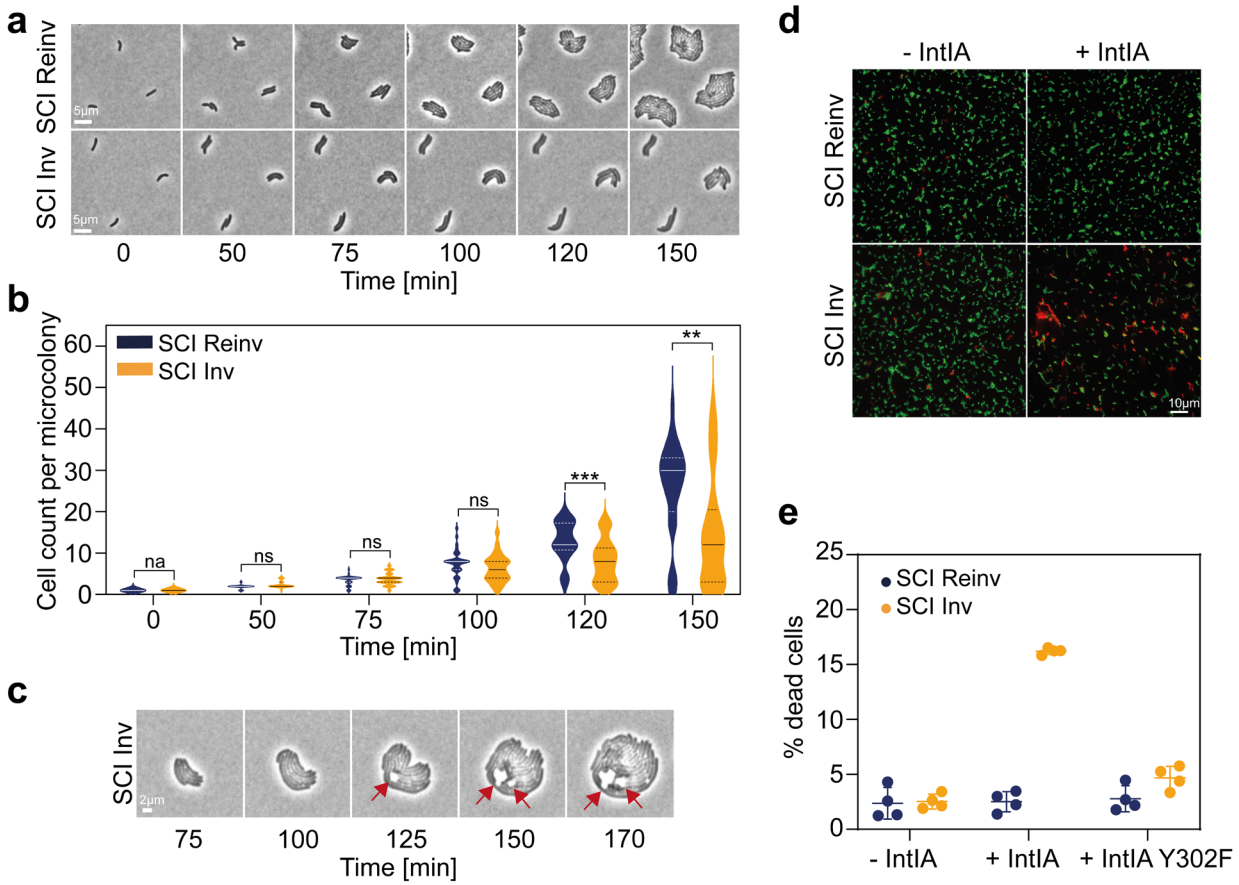


Figure 5: Cell viability of the SCI inverted and reinverted *V. cholerae* strains

a. Bright-field microscopy images taken from 170 min time-lapse series of live cells growing onto a MOPS agarose pad. A representative sample of 3 microcolonies of SCI Inv or SCI Reinv strains is shown. Scale bar is 5 µm.

b. Microcolony growth of the SCI Inv and Reinv strains onto MOPS agarose pads. Violin plots show the distribution of the number of cells per microcolony (27 for SI Inv and 34 for SCI Reinv strains) tracked during 170 min. Median (full line) and quartiles (dotted line) are shown. Statistical comparisons (Mann-Whitney test) are as follow: (na. (not applicable), ns (not significant), *** *Pvalue* < *0.001*, ** < *0.01*).

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

d. Fluorescence microscopy images resulting from the Live and Dead assay on the SCI 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.

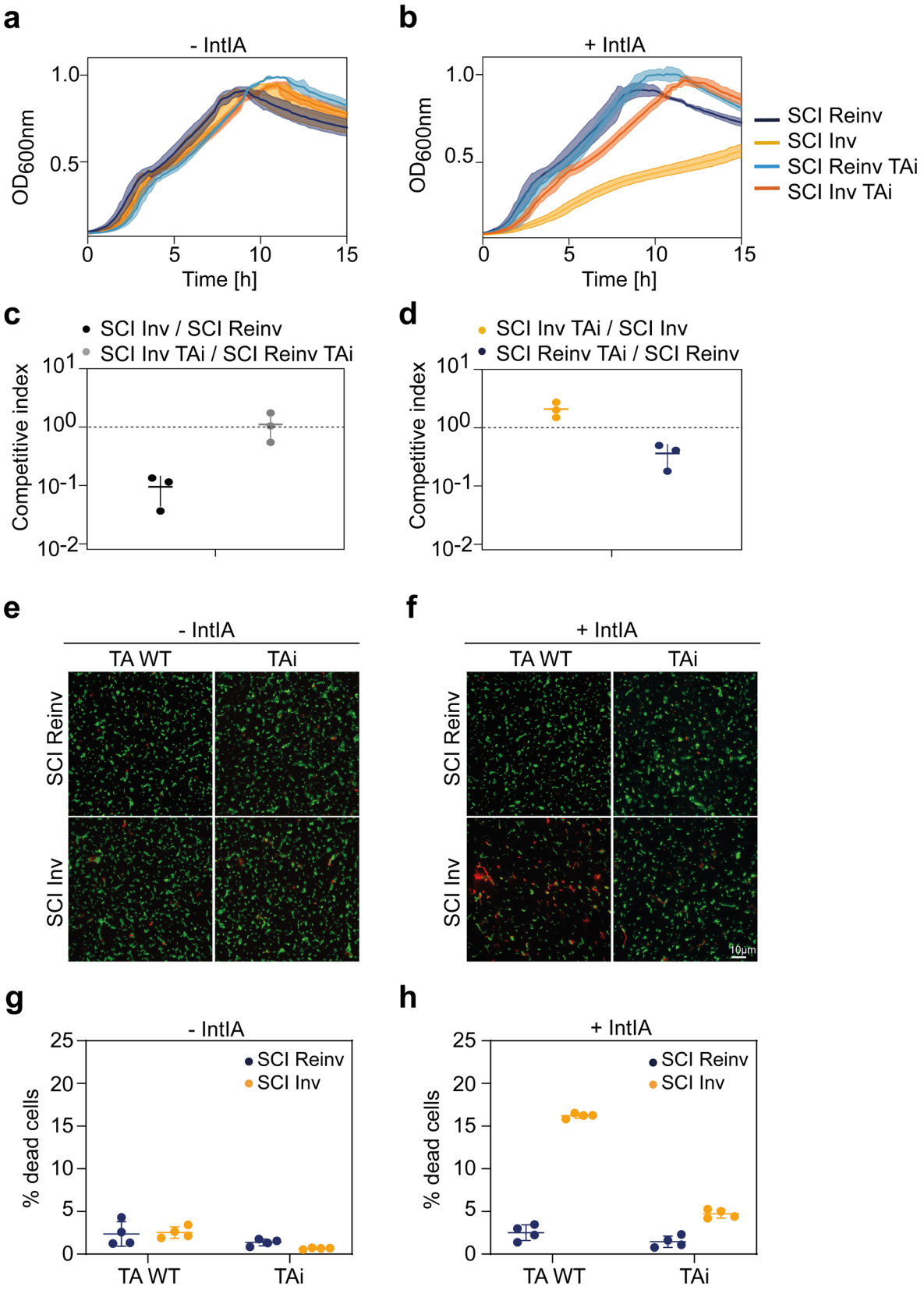
353 **e.** Quantification of cell death as a measure of cells positively stained with PI over the total number
354 of cells counted (~ 1000 cells per replicate). Four biological replicates, means and standard
355 deviation are represented.
356
357 **The fitness defect in the SCI Inverted strain is largely explained by the excision of toxin–**
358 **antitoxin systems.**
359 The fact that SCI inversion results in increased cassette excision and that this is associated
360 with increased growth arrest and cell death dependent on integrase catalytic activity led us to
361 hypothesize that the increased excision of some cassettes could be deleterious for the cell. We
362 hypothesized that the increased excision of the cassettes encoding TA systems, alongside all the
363 non-expressed cassettes, could be at the origin of the increased growth arrest and cell death
364 observed in the SCI Inv strain. Indeed, analysis of *attC* sequences indicate that TA systems are as
365 likely to be excised as other cassettes (Materials and methods, Fig S5a). In addition, we observe
366 excision of the TA-encoding cassettes when they are duplicated (Fig S5b), suggesting that the non-
367 duplicated TA cassettes can also be excised but result in such a large fitness defect that we cannot
368 observe their excision in our data set. By combining prior approaches (*27, 28*), we designed a
369 CRISPR-based cytosine base editors which allowed us to inactivate almost all the TA cassettes of
370 the array in one go. 12 guides were designed to target 15 toxins (3 of which are duplicated) such
371 that a STOP codon was introduced, resulting in at least a one-third truncation of each toxin
372 (Materials and methods, Fig S6). The remaining 4 toxins, for which we were unable to design a
373 suitable guide, were inactivated by introducing a stop codon through the more classical allelic
374 exchange method (*29*). It resulted in *V. cholerae* strains where all the TA within the SCI were
375 inactivated (TAi).
376 We next assessed the effect of TA inactivation on growth of both SCI Inv and Reinv strains.
377 In absence of integrase (Fig 6a), we do not observe substantial differences between all the strains
378 whether the SCI is inverted or not. In presence of integrase, the growth defect characteristic of the
379 SCI Inv strain appears to be largely rescued upon inactivation of the TA systems (Fig 6b).
380 Performing a competition assay, we found again, in presence of integrase, the SCI Inv to be severely
381 out competed by the SCI Reinv strain but this was alleviated in absence of functional TA systems
382 (Fig 6c). Moreover, the SCI Inv TAi outcompetes the SCI Inv strain by more than a factor of three
383 in presence of integrase (Fig 6d). Note that both TAi strains, compared to the *wt* ones, present a
384 slightly impacted fitness (Fig 6d), probably due to off-target mutations consequently to the TA
385 inactivation process. A cell viability assay and a quantification of dead cells showed that the
386 improved fitness of the SCI Inv TAi strain in presence of integrase compared with the SCI Inv

387 strain was associated with decreased cell death (Fig 6e, 6f, 6g and 6h). For more clarity, we reported
388 in these figures, the previous results presented in the figure 5 (i.e. TA WT).

389 Finally, although the restoration of growth is important in absence of TA systems in the SCI
390 Inv strain, the phenotypic rescue is not complete. We reasoned that the intense recombination
391 activity in the SCI Inv strain in presence of integrase could also affect the replication process
392 explaining this gap. To address this, we performed Marker Frequency Analysis (MFA), a technique
393 that gives access to the replication pattern by deep sequencing of gDNA extracted from an
394 exponentially growing bacterial culture (*30-32*). In absence of integrase, the MFA profile seems
395 normal for both the SCI Inv and Reinv strains (Fig S7). However, in presence of integrase, in the
396 SCI Inv strain, a severe drop of coverage can be observed at the genomic location corresponding to
397 the SCI specifically. This means that there is an important replication slow-down within the SCI in
398 this condition. Therefore, we cannot exclude a slight effect of a disruption of the replication process
399 by the intense cassette dynamics in the SCI inverted strains in addition to the TA effect.

400 It was possible that the lack of constraint on the cassette array in absence of TA systems
401 could lead to a quick and massive loss of both cassettes and the associated replication defect. For
402 this reason, we sequenced single clones stemming from the culture of 4 SCI Inv TAi independent
403 strains to assess their cassette content after one overnight culture in presence of the integrase.
404 Although two clones had effectively lost 83% and 52% of their cassette content (*i.e* respectively
405 104 and 94 cassettes), the other two had only lost 9% and 11% of their cassette content (*i.e*
406 respectively 16 and 20 cassettes). Thus, the increased fitness of the SCI inv TAi strain is not due to
407 a massive loss of cassettes but rather to the absence of the detrimental effect of losing TA systems
408 as single cassettes. We conclude that the growth defect in the SCI inverted strains is mostly due to
409 the increased excision of TA cassettes acting as "abortive system" killing the cell. Nonetheless, we
410 cannot completely exclude a slight effect of a disruption of the replication process by the intense
411 cassette dynamics.

412


a
- IntIA
OD600nm
1.0
0.5
0
5
10
15
Time [h]
b
+ IntIA
OD600nm
1.0
0.5
0
5
10
15
Time [h]
SCI Reinv
SCI Inv
SCI Reinv TAi
SCI Inv TAi
c
SCI Inv / SCI Reinv
SCI Inv TAi / SCI Reinv TAi
Competitive index
10^1
10^0
10^{-1}
10^{-2}
d
SCI Inv TAi / SCI Inv
SCI Reinv TAi / SCI Reinv
Competitive index
10^1
10^0
10^{-1}
10^{-2}
e
- IntIA
TA WT
TAi
SCI Reinv
SCI Inv
f
+ IntIA
TA WT
TAi
SCI Reinv
SCI Inv
10µm
g
- IntIA
SCI Reinv
SCI Inv
% dead cells
25
20
15
10
5
0
TA WT
TAi
h
+ IntIA
SCI Reinv
SCI Inv
% dead cells
25
20
15
10
5
0
TA WT
TAi

413

414

415

416 **Figure 6: Impact of the toxin–antitoxin cassette inactivation on *V. cholerae* cell growth and**
417 **viability**
418
419 **a, b.** Growth curve of the strains of *V. cholerae* with its SCI Inv or Reinv and the TA WT or
420 inactivated (TAi), in absence (a) or in presence of the IntIA integrase (b).
421 **c.** Competitive index of the SCI Inv strain compared to the SCI Reinv strain (black) and of the SCI
422 Inv TAi strain compared to the SCI Reinv TAi strain (grey) after 24h of co-culture.
423 **d.** Competitive index of the SCI Inv TAi strain compared to the SCI Inv strain (orange) and of the
424 SCI Reinv TAi strain compared to the SCI Reinv strain (blue) after 24h of co-culture. For both c
425 and d, an index of 1 represents a ratio of 1:1 of the two strains in the mix (figured in dotted lines).
426 **e, f.** Fluorescence microscopy images resulting from the Live and Dead assay on the SCI Inv and
427 Reinv strains carrying TA cassettes in their WT form (TA WT) or carrying the inactivated TA
428 cassettes (TAi), in absence (e) or in presence of the IntIA integrase (f). Green fluorescence indicates
429 live cells (SYTO-9) whereas red fluorescence (Propidium iodide (PI)) indicates dead cells. Scale
430 bar is 10 µm.
431 **g, h.** Quantification of cell death as a measure of cells positively stained with PI over the total
432 number of cells counted (~ 1000 cells per replicate) in absence (g) or in presence of the IntIA
433 integrase (h). Four biological replicates, means and standard deviation are represented.
434 T: toxin, A: antitoxin
435

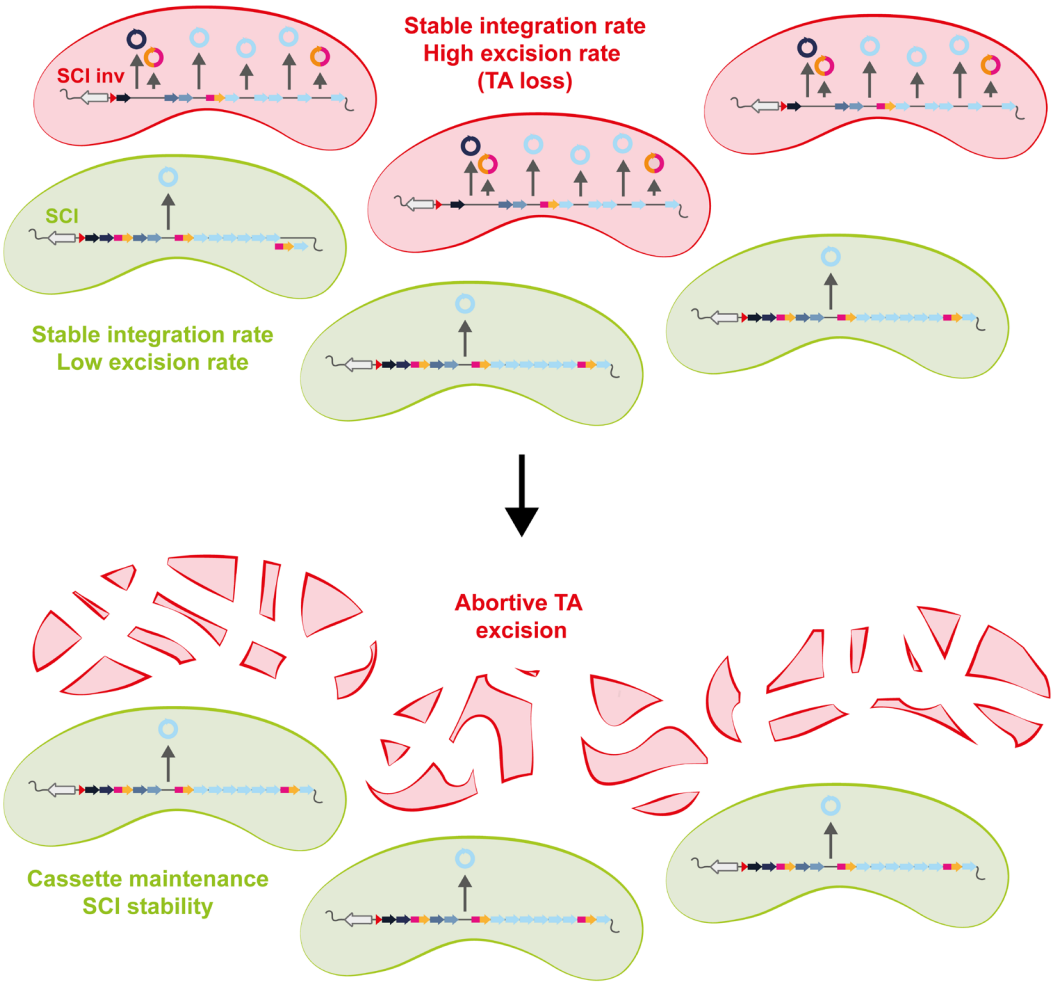
436 ## Discussion

437 The recent and striking evolutionary success of the integron system to annihilate antibiotic
438 treatment of Gram-negative pathogens, relies on the complementarity and interplay between MIs
439 and SCIs. There is ample evidence that the large and sedentary chromosomal integrons are the
440 source of the antibiotic resistance cassettes disseminated by the small and mobile ones (*33, 34*).
441 The SCI arrays may carry up to more than 300 mostly silent cassettes and represent several percent
442 of the bacterial genomes (i.e. 3% in *Vibrio cholerae*) (*5, 6*). How the carriage of hundreds of silent
443 genes, which may become useful at one time, might be selected, remains unknown. Here, we
444 demonstrate that TA cassettes in SCIs regulate the cassette recombination dynamics and stabilize
445 massive SCIs.

446 TA systems are regularly found in mobile genetic elements and contribute to their stability.
447 In the case of *V. cholerae*, their presence all along the SCI cassette array was proposed to avoid
448 large scale rearrangements of the cassette array leading to the simultaneous loss of dozens of
449 cassettes (*5, 14*). Indeed, the high prevalence of repeated sequences within SCIs might lead to

450 massive loss of cassettes by homologous recombination in absence of regularly interspaced TA
451 modules within the array. Strangely, all TA systems described so far in SCIs were found
452 systematically associated with their own *attC* sites, turning them into cassettes (*5, 16, 17, 35*). This
453 is puzzling since they all contain a promoter and they do not need to be integrated into *attI* sites to
454 be expressed. Here, we propose an additional stabilizing activity of TA modules directly linked to
455 their association with *attC* sites and which strikingly interconnects cassette excision frequency with
456 cell viability (Fig 7). While the excision of most cassettes of the array is completely harmless for
457 the cell because they do not contain a promoter and are therefore not expressed, the excision of TA
458 cassettes immediately imposes a prohibiting fitness cost. Since TA cassettes have supposedly the
459 same chance to be excised than any other, this means that any excision event can potentially be
460 lethal for the cell or at least lead to growth arrest. Although it may seem surprising that TA systems
461 are associated with recombination sites, making them easy to lose and to kill its host, we argue that
462 this could be beneficial at the population level, analogously to their role in abortive infection (Abi)
463 in the context of phage infection (*18*). We call this process "abortive cassette excision", where
464 cassette excision leads to the suicide of a fraction of the population that is proportional to the overall
465 cassette excision rate (Fig 7). The strength of this process is that the sensor (the *attC* site) and the
466 effector (TA system) are precisely combined into one piece (a TA cassette). This process represents
467 an emergent property of TA systems once associated with an *attC* site, which proves to be a very
468 effective way to ensure the stability of integron by driving the evolution of cassette excision rate
469 towards low values. The very strong fitness cost of cassette excision imposed by the random chance
470 of excising a TA must translate into counter-selection of any condition in which the excision rate
471 is highly increased. Here we showed that the presence of TA cassettes heavily penalizes the
472 inversion of the SCI of *V. cholerae* which highly increases the cassette excision rate. This makes a
473 lot of sense with respect to integron evolution. This indeed explains our recent *in silico* analyses
474 which reveal that most of SCIs have a specific orientation relative to the chromosome replication
475 (*13*). It fits the observations made previously that there are many TAs in large SCIs (*14*), and the
476 current identification of their significant and perhaps ancient over-representation in the SCI of many
477 *Vibrio* species compared to the rest of the genome. This long co-evolution of *Vibrio* species,
478 particularly *V. cholerae*, with their integron-encoded TA systems may have enabled SCI to have
479 access to a vast repertoire of cassettes and thus expand their genetic capacitance. Strikingly, among
480 the 1423 complete mobile integrons carried by plasmids, only one TA cassette was found
481 (specifically in the MI of the 248 kb plasmid from the bacterium *Comamonas testosteroni*
482 (GCF_014076475.1)) (*36*). This very low percentage of TA cassettes within MIs perhaps means

483 that their recruitment can occur but lead to the abortive excision of the TA cassettes due to a too
484 high cassette excision frequency in MIs compared to SCIs (*13*).
485
486 Finally, in addition to their known role in plasmid stability or viral defense (*18*), TAs
487 associated with integron recombination sites represent special cassettes largely widespread in the
488 large SCIs and acting as selective pressure to ensure the interplay between host genome
489 organization and genome plasticity through that we called "abortive cassette excision".
490



509

510 **Figure 7: Model of the "Abortive cassette excision" process in an integron array containing**
511 **toxin–antitoxin cassettes.**
512 In conditions for which the excision cassette rate is low, the TA cassettes (in pink and orange) are
513 seldom lost. If the excision rate is high, the TA cassettes are highly excised and kill cells. This
514 "abortive model" of TA cassette excision ensures the SCI cassette maintenance.
515 T: toxin, A: antitoxin
516

517 # Materials and Methods
518
519 **Bacterial strains, plasmids and primers**
520 The different bacterial strains, plasmids and primers that were used in this study are described
521 respectively in Table S1, S2 and S3.
522
523 **Media**
524 *Vibrio cholerae* and *Escherichia coli* strains were grown in Luria Bertani (LB) at 37°C. *V. cholerae*
525 strains containing a plasmid with a thermo-sensitive origin of replication were grown at 30°C.
526 Thymidine (dT) and diaminopimelic acid (DAP) were supplemented, when necessary, to a final
527 concentration of 300 µM. Glucose (Glu), L-arabinose (Ara) and Fructose were added respectively
528 at final concentrations of 10g/L, 2g/L and 10g/L. X-Gal was used at a final concentration of 100
529 µM. DAPG was used at a final concentration of 50 µM. Antibiotics were respectively used at the
530 following concentrations (for *V. cholerae* and *E. coli* respectively): carbenicillin (100 µg/ml; 100
531 µg/ml), chloramphenicol (5 µg/ml; 25 µg/ml), kanamycin (25 µg/ml; 25 µg/ml), rifampicin (1
532 µg/ml; 150 µg/ml), spectinomycin (100 µg/ml or 200 µg/ml in presence of glucose; 50 µg/ml),
533 zeocin (50 µg/ml; 50 µg/ml). To avoid catabolic repression during various tests using arabinose as
534 inducer for the expression of the integrase, cells were grown in a synthetic rich medium: MOPS
535 Rich supplemented with Fructose (1%) as a carbon source and arabinose (0.2%) as an inducer.
536
537 **SCI inverted and reinverted strain constructions**
538 The inversion of the SCI was performed using a genetic tool developed in the lab (*37*) and designed
539 to target the relocation of chromosomal DNA with bacteriophage attachment sites (HK in this case).
540 Two DNA fragments were inserted respectively upstream and downstream of the SCI in the
541 N16961 hapR$^+$ strain (8637). The upstream fragment contained a kanamycin resistance gene and
542 the 5' end of a carbenicillin resistance gene associated with the $attR_{HK}$ site. The downstream
543 fragment contained a zeocin resistance gene and the partner 3' end of the carbenicillin resistance
544 gene associated with the $attL_{HK}$ site (Fig S2). The strain containing the upstream and downstream
545 fragments prior to inversion is referred to as the parental strain (Par). The expression of the HK_{022}
546 integrase and excisionase (as a directional factor, pA401) in the parental strain led to the $attR_{HK}$ ×
547 $attL_{HK}$ reaction resulting in the inversion of the SCI and to the reconstitution of the full copy of the
548 *bla* gene, making it possible to select on carbenicillin for clones with an inverted SCI (Fig S2). As
549 a control for the following experiments, the SCI was re-inverted to its original orientation using
550 solely the HK_{022} integrase (p8507) to perform the $attP_{HK}$ × $attB_{HK}$ reaction in the SCI inv strain.

551

552 *Parental strain:*

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

556 *SCI inverted strain:*

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

565 *SCI reinverted strain:*

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

575

576 **Automated growth curve measurements.**

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

582

583 **Toxin–antitoxins inactivation by allelic exchange**

584 We performed allelic exchange to construct N16961 lacking the TA systems that could not be
585 targeted using the Base-editing tool. To this purpose, we constructed and used different variants of
586 the pMP7 vector, respectively pP897, pP898, pP900 and pP902. We followed the same protocols
587 as previously described (*29*). Briefly, the suicide vector pMP7 contains a R6K origin of replication
588 and its replication is then dependent on the presence of the Π protein in the host cell. The Π3813
589 cell, a *pir*+ CcdB resistant *E. coli* strain, was used for cloning the different pMP7 plasmids. Once
590 constructed, these vectors were transformed into the β3914 donor strain to deliver by conjugation
591 the pMP7 vector into the desired recipient strain. Homology regions corresponding to the genomic
592 DNA from the recipient strain have been cloned in the different pMP7 vector to allow the
593 integration of the plasmid by homologous recombination. The only way for pMP7 vector to
594 replicate into recipient strains is then to integrate into the host genome after a first crossover. After
595 conjugation, integration of the entire pMP7 vector were then selected by plating cells on Cm plates
596 lacking DAP. Next, cells were grown in presence of L-arabinose (0.2%) to express the CcdB toxin.
597 The expression of this toxin allows to kill cells in which the second crossover that leads to the
598 excision of pMP7 backbone did not take place. This method allows us to obtain mutants of *V.*
599 *cholerae* that are devoid of any antibiotic resistance marker. Since we did not performed deletions,
600 the verification of the gene replacement by PCR only was prohibited. For this reason, the STOP
601 containing version of the toxin was also associated to a BamHI restriction site to allow an easier
602 screening of the correct gene replacement by digesting the appropriate PCR and looking for a
603 restriction profile. After that, the PCR products containing the BamHI sites were checked by
604 sequencing. The primers used for PCR screening and sequencing are described Table S3.
605

606 **Golden Gate Assembly**

607 By default, the vector expressing dCas9 carries a "random" guide that does not target any locus in
608 *E. coli* nor in *V. cholerae* but that do contain two BsaI restriction sites (*38*). To change the guide,
609 we perform a "Golden Gate Assembly" as such: two oligonucleotides have to be designed in the
610 form 5'–TAGTNNNNNNNNNNNNNNNNNNNN–3' and 5'-
611 TTTGNNNNNNNNNNNNNNNNNNNN-3' where the 20 "N" is the targeted sequence. For the
612 annealing, 6uL of each oligo is mixed with 2uL of T4 DNA Ligase Buffer and 0.4 µL of T4 PNK
613 for a total volume of 20 µL and then incubated for 30 min at 37°C. We then add 1 µL of NaCl (1M),
614 incubate 5 min at 95°C and finally let slowly cool down to RT for at least 4h.
615 To insert the desired guide into the dCas9 expressing vector, we prepare the following mix: 2 µL
616 of the dCas9 plasmid, 2 µL of annealed oligos, 1 µL of Cutsmart buffer, 1 µL of BsaI enzyme, 1
617 µL of ATP (1mM), 1 µL of ligase, 2 µL H_20. We then incubate the mix in a thermocycler with the

618 following steps: 3 min at 37°C (for digestion), 4 min at 16°C (for ligation) and alternate between
619 those steps 25 times. We finish with one cycle of 5 min at 50°C and 5 min at 80°C for enzymes
620 inactivation. At this step, the mix is ready to be transformed in a cloning *E. coli* strain and the
621 successful sgRNA insertions are screened by PCR using the o2406 and o3812 primers.
622
623 **Base editing**
624 The tool used for base editing is derived from the catalytically dead Cas9 (dCas9) previously
625 described (*27*). The construction consists in a gene fusion between dCas9 and CDA1, a cytidine
626 deaminase that catalyzes the deamination of cytidine, resulting in a uridine base which will later be
627 replicated as a thymine (C → U→ T). A linker of 100 AA separated the dCas9 and the CDA1. An
628 Uracil Glycosylase Inhibitor (UGI) was also fused to CDA1 to increase the base editing rate as
629 described by Banno and coll as well as a LVA degradation tag to decrease the toxicity the construct
630 (*27*). Due to a low efficiency transformation rate of this construct in *Vibrio cholerae*, the construct
631 was delivered by conjugation. Hence, the appropriate construct was first transformed in a β3914
632 donor strain. Then both the donor and receptor strains were cultivated to exponential phase, mixed
633 1:1 in a 2 mL tube and centrifugated (6000 rpm for 6 min).
634 The pellet was spread on a membrane on a plate containing DAP to sustain growth of the donor
635 strain and DAPG for induction of base editing, and the plate was incubated for 3h at 37°C. After
636 incubation, the membrane was resuspended in 5mL LB and vortexed for 30s to resuspend the
637 conjugated cells which were then plated on appropriate media: Cm_5 + Glc1% + Xgal to obtain both
638 the CFU/mL after conjugation and the base editing efficiency or Rif_1 + Cm_5 + Glc1% to obtain the
639 rate of apparition rifampicin resistant clones as a proxy for global mutation rate. Blue colonies had
640 a functional *lacZ* gene while the white colonies were synonym for a mutated *lacZ* gene. All targeted
641 sites were checked by PCR and sequencing (see Table S3). Then the whole genomes of the TAi
642 strains were sequenced (Illumina).
643
644 **SCI cassette shuffling assay**
645 A single clone of each strain was isolated on plate and then inoculated for an overnight culture in
646 LB + Spec + Glc 1%. The next day, the cells were inoculated at $1/50^{th}$ for 2h in MOPS Rich + Spec
647 + Fructose 1% + Arabinose 0.2% as a pre-induction step until they were at exponential phase. They
648 were then inoculated at $1/1000^{th}$ and grew for 20h until stat phase. Cultures were streaked on 1%
649 agarose plates containing MH+ Spec + Glc 1% and single clones were isolated. A total of 48 PCR
650 per condition was performed using an oligo (o5778, see Table S3) hybridizing in the promoter of
651 *intIA* (stable part) and one (o1401) hybridizing in the first cassette of the array (variable part). A

652 representative sample of the PCR profile for each condition was migrated in 1% agarose for
653 visualization.
654
655 **Suicide conjugation assay**
656 This assay has been previously described by Vit and coll. 2021 (*3*). We used the suicide conjugative
657 vector pSW23T (*39*) that allows the delivery of the bottom strand of the $attC_{aadA7}$ recombination
658 site (pD060). For each condition assay, at least 16 recombinant clones were isolated on appropriate
659 plates and analyzed by PCR. To determine precisely if the pSW23T vector has been inserted into
660 the *attIA* site of the SCI we used 5778 and SWend primers. These primers hybridize respectively
661 in a sequence upstream of *attIA* in *V. cholerae* chromosome 2 or downstream of $attC_{aadA7}$ in the
662 pSW23T vector. For each condition assay, at least three PCR reactions were purified and sequenced
663 to confirm the insertion point.
664
665 **SCI cassette excision assay**
666 A single clone of each strain was isolated on plate and then inoculated for an overnight culture in
667 LB + Spec + Glc 1%. The next day, the cells were inoculated at $1/50^{th}$ for 2h in MOPS Rich + Spec
668 + Fructose 1% + Arabinose 0.2% as a pre-induction step until they were at exponential phase. They
669 were then inoculated at $1/1000^{th}$ and grew for 20h until stat phase. Finally, 1mL of this culture was
670 used to inoculate a 100 mL culture in LB + Spec + Glc1 % to filter the dead cells of the latter
671 culture. The resulting culture constitutes a mixed population where each individual might have
672 experienced a diverse set of cassette excision events. Bulk DNA was extracted from these mixed
673 population (Qiagen genomic DNA extraction kit) and sequenced using the Pacific Bioscience
674 (PacBio) long-read sequencing technology. Mapping was performed using minimap2 (*40*) and the
675 cassette deletions were detected using a homemade R program. In order not to confuse the excision
676 rate with the random chance of an excision event to be propagated, the cultures of the different
677 strains were replicated eight times independently.
678 The cassette excision frequency corresponds to the total number of cassette excision events divided
679 by the sequencing depth across the SCI. For the SCI inverted strain expressing integrase, we
680 detected 478 excision events encompassing the entire SCI. The calculated frequency corresponds
681 to: N_{exc} / (SCI_{cov} + N_{exc}); where N_{exc} is the number of excision events and SCI_{cov} is the mean
682 coverage for each cassette multiplied by the number of cassettes in the SCI (i.e. 179). We therefore
683 obtained a cassette excision frequency of 478 / (13269.16 + 478) = 3.48e-02.
684 For the SCI inverted strain without integrase and the SCI reinverted strains with and without
685 integrase, for which we did not detect any excision event, we calculated the limit of detection. It

686 corresponds to: 1/ SCI_{cov}. Note that the SCI read coverage obtained for each condition is
687 respectively 13518.08, 14737.07 and 14550.91 corresponding to detection limits of respectively
688 7.40e-05, 6.79e-05 and 6.87e-05.
689
690
691 **Protein purification**
692 IntIA proteins were expressed in BL21 bacteria (*41*). Cells were lysed in 25 mM Tris pH 8, 1M
693 NaCl, 1 mM PMSF, lysozyme and protease inhibitors containing buffer. Proteins were purified by
694 nickel affinity chromatography with a 200 mM imidazole elution. His-tag was removed by
695 incubating proteins at 4°C overnight with TEV protease (1 mg per 15 mg of total proteins containing
696 a 5 mM final concentration of β mercapto-ethanol). After dilution to 150 mM NaCl, proteins were
697 purified on an SP column equilibrated with 25 mM Tris pH8 and 150 mM NaCl. A gradient of 150
698 mM to 1M NaCl was performed. Proteins were then brought to a final concentration of 1 mM DTT
699 and 10% glycerol.
700
701 **Electromobility shift assay**
702 Each reaction contained 500 ng Poly[d(I-C)], 12 mM Hepes–NaOH pH 7.7, 12% glycerol, 4mM
703 Tris–HCl pH 8.0, 60 mM KCl, 1 mM EDTA, 0.06 mg/ml BSA, 1 mM DTT, 10% Tween20, 0.01
704 nM specified 32P-labeled DNA oligonucleotide (Table S3) and the specified quantities of purified
705 His-tagged IntIA and IntIAY302F, in a final volume of 20 µl. The samples were incubated at 30 ºC
706 for 10 min without the probe followed by 20 min with the probe, then loaded to a 5% native
707 polyacrylamide gels with 0.5x TBE as running buffer. The gels were visualized using a Molecular
708 Dynamics intensification screen and a Typhoon FLA 9500 laser scanner
709
710 **Competition assay**
711 The SCI Inv strains are resistant to Carb, which is not the case for the SCI Reinv strains (see above).
712 We used this difference to perform a competition assay. One clone of each strain was isolated on
713 plate and used to inoculate an overnight culture. Before to start the competition assay, the cells were
714 precultured at $1/50^{th}$ in LB + Spec + Glc 1% until exponential phase to be not biased by the dead
715 cells from the overnight culture. Cells were then mixed at 1:1 ratio at $1/200^{th}$ in MOPS Rich +
716 Fructose 1% + Spec + Arabinose 0.2% for induction of the integrase. The co-culture was plated on
717 petri dishes containing MH + Spec + Glc 1% to get the total CFU/mL and in parallel on MH + Carb
718 + Spec + Glc 1% plates to get the number of SCI Inv CFU/mL in that same culture. We plated at
719 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 =

720 $CFU_{SCI\ Inv}$ / $CFU_{SCI\ Reinv}$ – $CFU_{SCI\ Inv}$ so that an index of 1 represents a situation where the ratio of
721 SCI Inv and SCI Reinv is 1:1 (Neutrality). A high index indicates that SCI Inv has a competitive
722 advantage compare to its Reinv counterpart, and *vice-versa*.
723
724 **Live and dead assay**

725 Overnight cultures were performed in LB + Spec + Glucose 1%. Then day cultures were inoculated
726 at 1:1000 inMOPS Rich + Spec + Fructose 1% + Arabinose 0.2% medium to OD600nm~ 0.8. Then
727 50 µL of the latter culture was dyed using the LIVE/DEAD™ *Bac*Light™ Bacterial Viability Kit,
728 that is a mix of Propidium Iodide (PI) and SYTO-9 dyes. PI specifically stains the dead cells in red.
729 SYTO-9 specifically stains viable cells in green. Cell viability was then assessed by microscopy.
730 Stained cells were placed of 1.4% agarose pad containing MOPS Rich and observed using a Zeiss
731 ApoTome inverted wide-field microscope. Snapshot images were taken with a 40-ms and 50-
732 ms exposure time (FITC and DsRED channel respectively), using a Plan Apo 63× objective
733 (numerical aperture = 1.4, +optovar 1.6×) using a Hamamatsu sCMOS ORCA-Flash 4.0 v3 (Institut
734 Pasteur imaging facility imagopole). The result is presented as the proportion of dead cells (red) in
735 the population.
736
737 **Microscopy setup for live imaging**

738 Bacterial cells were grown to mid-exponential phase in liquid MOPS Rich + Spec + Fructose 1%
739 + Arabinose 0.2% medium and then transferred to 1.4% agarose-padded slides containing MOPS
740 Rich + Spec + Fructose 1% + Arabinose 0.2%. A coverslip was placed on top of the agarose pad
741 and sealed with a vaseline:lanolin:paraffin mix (ratio 1:1:1) to prevent pad evaporation. Slides were
742 incubated at 37°C under an inverted wide-field microscope (Zeiss ApoTome) for time-lapse video
743 recording. Video frames were taken at 1 min interval time for a total duration of 170 min, using a
744 Plan Apo 63× objective (numerical aperture = 1.4) using a Hamamatsu sCMOS ORCA-Flash 4.0
745 v3 and autofocus module (Definite focus, Zeiss) (Pasteur Institute Imaging Facility Imagopole). A
746 total of 8 movies were recorded and analyzed across all strains.
747
748 **Marker Frequency Analysis**
749
750 MFA was performed as described in (*32*). Briefly, cells were grown in MOPS + Spec + Arabinose
751 0.2% + Fructose 1% at 37°C under agitation. Genomic DNA was prepared using the DNeasy®
752 Tissue Kit (Qiagen) from 30ml of exponentially growing cells (OD_{450} ~0.15). The remaining culture

753 was washed twice and was kept under agitation in MOPS + Spec + Glucose 1% at 37°C to repress
754 expression of IntIA and to prepare gDNA from cells in stationary phase (24h of culture). Libraries
755 were prepared for Illumina sequencing (150 bp, paired end). The resulting reads were mapped with
756 bowtie2 (*42*). The genomic coverage of the sequencing data stemming from the exponentially
757 growing cultures of each strain was normalized with the corresponding data in stationary phase.
758 The normalized coverage is represented in function of the relative position to the origin of
759 replication of each chromosome as colored dots for 1 kb bins and red dots for 10 kb bins. Dots are
760 not shown when at least one position has a coverage of zero within the bin (typically due to
761 repetitive sequences that prevents correct reads mapping). The gradient of coverage from *ori* to *ter*
762 serves as a proxy for the replication dynamics and the slope translates the local speed of the
763 replication fork (*31*).
764

765 **Comparative genomics analysis of toxin–antitoxin systems within SCI**

766 *Genome dataset and SCI annotation*

767 To study the enrichment of TA systems within SCI across bacterial genomes, we used the dataset
768 and integron predictions of IntegronFinder 2.0 (*20*). Briefly, this dataset consists of 21,105
769 complete genomes retrieved from NCBI RefSeq on 30 March 2021. We defined SCIs as complete
770 integrons with at least 11 predicted *attC* sites. Several isolates of *Vibrio cholerae* appeared to carry
771 large (>10 *attC* sites) CALINs (cluster of *attC* sites lacking an integrase) (*20, 43*). A closer
772 investigation showed that these CALINs were preceded by a pseudogenized integrase, which relates
773 them with SCIs. These elements could have been recently inactivated, or the putative pseudogenes
774 could result from sequencing artifacts. Hence, we included in our analysis all the isolates carrying
775 these CALINs. To simplify the results, we grouped together SCIs and CALINs, without distinction,
776 under the global term SCIs.

777 *Toxin and antitoxin prediction*

778 In each genome harboring at least one SCI, we predicted toxins and antitoxins with TASmania, a
779 predictor with very low false negative rate (*21*). Our goal was to retrieve even unusual TA systems
780 that would be rarely found outside SCIs and hence missed by more conservative predictors. More
781 precisely, we downloaded toxin and antitoxin HMM models from TASmania's website (accessed
782 on 25 November 2021) and screened for them with the hmmscan command of HMMER 3.3 (*44*)
783 (Nov 2019; http://hmmer.org/). As the command was run with default parameters, no threshold was

784 used to filter hits in the first screening. To reduce the number of false positives, we evaluated
785 TASmania's performance on the genome of *Vibrio cholerae* N16961's secondary chromosome,
786 known to harbor 19 TA systems (*16, 17*), all located within the SCI. We observed that applying an
787 e-value threshold of 10^{-3} allowed to keep all the hits corresponding to characterized TA systems (as
788 described by Iqbal and coll. (*16*)), while eliminating most of the other hits. Hence, we selected this
789 value to filter TASmania hits in all the genomes comprised in our analysis and we decided to
790 consider as a toxin (respectively antitoxin) hit any protein matching a TASmania toxin (respectively
791 antitoxin) profile with an e-value lower than 10^{-3}.
792
793 *Contingency analysis of TA enrichment within SCI*

794 For each genome, we performed a contingency table analysis to test for an enrichment of TA
795 systems within SCIs. The contingency table was built as following: in each genome, each protein
796 was classified as, on one side, belonging or not to an SCI, and, on the other side, containing a
797 TASmania hit or none. A Fisher one-sided test for statistical enrichment significance was performed
798 on each contingency table. To correct for multiple testing, the resulting p-values were adjusted
799 either with the Benjamini-Hochberg (main analysis) or the Bonferroni (complementary analysis)
800 corrections. An isolate was claimed to harbor a SCI significantly enriched in TA when the adjusted
801 p-value was lower than 0.05. The Benjamini-Hochberg and Bonferroni corrections were re-
802 computed in the analysis focusing on *Vibrio* isolates.

803 Vibrio *genus cladogram*

804 To identify a potential evolutionary trend explaining the TA enrichment within SCI across the
805 *Vibrio* genus, we reconstructed a cladogram grouping *Vibrio* species together. More precisely, we
806 took as basis the phylogeny published by Sawabe and coll. (*22*) and removed the species that did
807 not harbor any SCI. The species comprised in our dataset that were absent of the phylogeny were
808 then added as a sister branch of the closest relative species.

809
810 **References**
811

812 1. H. W. Stokes, R. M. Hall, A novel family of potentially mobile DNA elements encoding
813 site-specific gene-integration functions: integrons. *Molecular Microbiology* **3**, 1669-1683
814 (1989).
815 2. C. M. Collis, R. M. Hall, Expression of antibiotic resistance genes in the integrated cassettes
816 of integrons. *Antimicrob Agents Chemother* **39**, 155-162 (1995).

817 3. C. Vit *et al.*, Cassette recruitment in the chromosomal Integron of Vibrio cholerae. *Nucleic*
818 *Acids Res* **49**, 5654-5670 (2021).
819 4. E. Guerin *et al.*, The SOS response controls integron recombination. *Science* **324**, 1034
820 (2009).
821 5. J. A. Escudero, C. Loot, A. Nivina, D. Mazel, The Integron: Adaptation On Demand.
822 *Microbiology spectrum* **3**, MDNA3-0019-2014 (2015).
823 6. E. Richard, B. Darracq, C. Loot, D. Mazel, Unbridled Integrons: A Matter of Host Factors.
824 *Cells* **11**, (2022).
825 7. J. A. Escudero *et al.*, Unmasking the ancestral activity of integron integrases reveals a
826 smooth evolutionary transition during functional innovation. *Nature communications* **7**,
827 10937 (2016).
828 8. C. Johansson, M. Kamali-Moghaddam, L. Sundstrom, Integron integrase binds to bulged
829 hairpin DNA. *Nucleic Acids Res* **32**, 4033-4043 (2004).
830 9. M. Bouvier, G. Demarre, D. Mazel, Integron cassette insertion: a recombination process
831 involving a folded single strand substrate. *Embo J* **24**, 4356-4367 (2005).
832 10. A. Nivina, J. A. Escudero, C. Vit, D. Mazel, C. Loot, Efficiency of integron cassette
833 insertion in correct orientation is ensured by the interplay of the three unpaired features of
834 attC recombination sites. *Nucleic Acids Res* **44**, 7792-7803 (2016).
835 11. G. Cambray, A. M. Guerout, D. Mazel, Integrons. *Annu Rev Genet* **44**, 141-166 (2010).
836 12. M. Buongermino Pereira *et al.*, A comprehensive survey of integron-associated genes
837 present in metagenomes. *BMC Genomics* **21**, 495 (2020).
838 13. C. Loot *et al.*, Differences in Integron Cassette Excision Dynamics Shape a Trade-Off
839 between Evolvability and Genetic Capacitance. *mBio* **8**, (2017).
840 14. S. Szekeres, M. Dauti, C. Wilde, D. Mazel, D. A. Rowe-Magnus, Chromosomal toxin-
841 antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Mol*
842 *Microbiol* **63**, 1588-1605 (2007).
843 15. D. A. Rowe-Magnus, A. M. Guerout, L. Biskri, P. Bouige, D. Mazel, Comparative analysis
844 of superintegrons: engineering extensive genetic diversity in the vibrionaceae. *Genome Res*
845 **13**, 428-442. (2003).
846 16. N. Iqbal, A. M. Guerout, E. Krin, F. Le Roux, D. Mazel, Comprehensive Functional
847 Analysis of the 18 Vibrio cholerae N16961 Toxin-Antitoxin Systems Substantiates Their
848 Role in Stabilizing the Superintegron. *J Bacteriol* **197**, 2150-2159 (2015).
849 17. E. Krin *et al.*, Systematic transcriptome analysis allows the identification of new type I and
850 type II Toxin/Antitoxin systems located in the superintegron of Vibrio cholerae. *Res*
851 *Microbiol*, 103997 (2022).
852 18. D. Jurenas, N. Fraikin, F. Goormaghtigh, L. Van Melderen, Biology and evolution of
853 bacterial toxin-antitoxin systems. *Nat Rev Microbiol* **20**, 335-350 (2022).
854 19. D. Mazel, B. Dychinco, V. A. Webb, J. Davies, A distinctive class of integron in the *Vibrio*
855 *cholerae* genome. *Science* **280**, 605-608 (1998).
856 20. B. Neron *et al.*, IntegronFinder 2.0: Identification and Analysis of Integrons across Bacteria,
857 with a Focus on Antibiotic Resistance in Klebsiella. *Microorganisms* **10**, (2022).
858 21. H. Akarsu *et al.*, TASmania: A bacterial Toxin-Antitoxin Systems database. *PLoS Comput*
859 *Biol* **15**, e1006946 (2019).
860 22. T. Sawabe *et al.*, Updating the Vibrio clades defined by multilocus sequence phylogeny:
861 proposal of eight new clades, and the description of Vibrio tritonius sp. nov. *Frontiers in*
862 *microbiology* **4**, 414 (2013).
863 23. C. Loot, C. Turlan, P. Rousseau, B. Ton-Hoang, M. Chandler, A target specificity switch in
864 IS911 transposition: the role of the OrfA protein. *EMBO J* **21**, 4172-4182 (2002).
865 24. E. Belda, A. Moya, F. J. Silva, Genome rearrangement distances and gene order phylogeny
866 in gamma-Proteobacteria. *Mol Biol Evol* **22**, 1456-1467 (2005).

25. M. E. Val *et al.*, Fuse or die: how to survive the loss of Dam in Vibrio cholerae. *Mol Microbiol* **91**, 665-678 (2014).
26. F. Wong, A. Amir, Mechanics and Dynamics of Bacterial Cell Lysis. *Biophys J* **116**, 2378-2389 (2019).
27. S. Banno, K. Nishida, T. Arazoe, H. Mitsunobu, A. Kondo, Deaminase-mediated multiplex genome editing in Escherichia coli. *Nat Microbiol* **3**, 423-429 (2018).
28. A. C. Reis *et al.*, Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. *Nat Biotechnol* **37**, 1294-1301 (2019).
29. F. Le Roux, J. Binesse, D. Saulnier, D. Mazel, 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 (2007).
30. K. S. Lang *et al.*, Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. *Cell* **170**, 787-799 e718 (2017).
31. O. Skovgaard, M. Bak, A. Lobner-Olesen, N. Tommerup, Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Res* **21**, 1388-1393 (2011).
32. M. E. Val *et al.*, A checkpoint control orchestrates the replication of the two chromosomes of Vibrio cholerae. *Sci Adv* **2**, e1501914 (2016).
33. D. A. Rowe-Magnus, A. M. Guerout, D. Mazel, Bacterial resistance evolution by recruitment of super-integron gene cassettes. *Mol Microbiol* **43**, 1657-1669. (2002).
34. F. Le Roux *et al.*, Genome sequence of Vibrio splendidus: an abundant planctonic marine species with a large genotypic diversity. *Environmental microbiology* **11**, 1959-1970 (2009).
35. A. M. Guerout *et al.*, Characterization of the phd-doc and ccd toxin-antitoxin cassettes from Vibrio superintegrons. *J Bacteriol* **195**, 2270-2283 (2013).
36. J. Huyan *et al.*, Dynamics of class 1 integrons in aerobic biofilm reactors spiked with antibiotics. *Environ Int* **140**, 105816 (2020).
37. M. E. Val, O. Skovgaard, M. Ducos-Galand, M. J. Bland, D. Mazel, Genome engineering in Vibrio cholerae: a feasible approach to address biological issues. *PLoS Genet* **8**, e1002472 (2012).
38. F. Rousset *et al.*, The impact of genetic diversity on gene essentiality within the Escherichia coli species. *Nat Microbiol* **6**, 301-312 (2021).
39. G. Demarre *et al.*, A new family of mobilizable suicide plasmids based on the broad host range R388 plasmid (IncW) or RP4 plasmid (IncPα) conjugative machineries and their cognate *E.coli* host strains. *Research in Microbiology* **156**, 245-255 (2005).
40. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
41. D. MacDonald, G. Demarre, M. Bouvier, D. Mazel, D. N. Gopaul, Structural basis for broad DNA specificity in integron recombination. *Nature* **440**, 1157-1162 (2006).
42. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
43. J. Cury, T. Jove, M. Touchon, B. Neron, E. P. Rocha, Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res* **44**, 4539-4550 (2016).
44. S. R. Eddy, Profile hidden Markov models. *Bioinformatics* **14**, 755-763 (1998).

Acknowledgments

We thank Samuel H. Sternberg for critical reading of the manuscript. 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 Cokelaer from the Biomics platform, C2RT, Institut Pasteur, Paris, France, supported by France Génomique (ANR-10-INBS-09) and IBISA. We thank Clarisabel Garcia Rodriguez for her help with microscopy.

Funding:

Institut Pasteur
Centre National de la Recherche Scientifique (CNRS-UMR 3525)
Fondation pour la Recherche Médicale (FRM Grant No. EQU202103012569)
ANR Chromintevol (ANR-21-CE12-0002-01)
French Government's Investissement d'Avenir program Laboratoire d'Excellence 'Integrative Biology of Emerging Infectious Diseases' [ANR-10-LABX-62-IBEID]
Ministère de l'Enseignement Supérieur et de la Recherche and the Direction Générale de l'Armement (DGA)

Author contributions:

Conceptualization: ER, CL, DM, DB, EPCR
Methodology: ER, CL
Investigation: ER, BD, EL, CV, CW, JB, FF, GG, VC, DL, VP, FR, OS, CL
Supervision: CL, DM, DB, EPCR
Writing—original draft: ER, CL, DM
Writing—review & editing:

Competing interests: Authors declare that they have no competing interests.

Data and materials availability:

All data are available in the main text or the supplementary materials. MFA and PacBio sequences are publicly available in The European Nucleotide Archive (ENA). For the MFA analysis, the accession no. are ERR11475429, ERR11475428, ERR11475427, ERR11475426, ERR11475425, ERR11475424, ERR11475423 and ERR11475422. For the PacBio sequencing, the accession no. are ERR11475405, ERR11475404, ERR11475403 and ERR11475402. All other raw data are available upon request to the corresponding author.