Genomic islands targeting dusA in Vibrio species are distantly related to Salmonella Genomic Island 1 and mobilizable by IncC conjugative plasmids — Source link ↗

Romain Durand, Florence Deschênes, Vincent Burrus

Institutions: Université de Sherbrooke

Related papers:

- Genomic islands targeting dusA in Vibrio species are distantly related to Salmonella Genomic Island 1 and mobilizable by IncC conjugative plasmids.

- Genomic islands related to Salmonella genomic island 1; integrative mobilisable elements in trmE mobilised in trans by A/C plasmids

- A novel family of integrases associated with prophages and genomic islands integrated within the tRNA-dihydrouridine synthase A (dusA) gene

- Stability, Entrapment and Variant Formation of Salmonella Genomic Island 1

- Identification and characterization of novel Salmonella mobile elements involved in the dissemination of genes linked to virulence and transmission

Share this paper: 📘 🐦 💼 ✉

View more about this paper here: https://typeset.io/papers/genomic-islands-targeting-dusa-in-vibrio-species-are-4ky93lhbxt

1 **Genomic islands targeting *dusA* in *Vibrio* species are distantly**

2 **related to *Salmonella* Genomic Island 1 and mobilizable by IncC**

3 **conjugative plasmids**

4

5 Romain Durand[1,#a], Florence Deschênes[1] and Vincent Burrus[1]*

6

7 [1] Département de biologie, Université de Sherbrooke, Sherbrooke, Québec,

8 Canada

9 [#a]Current Address: Institut de Biologie Intégrative et des Systèmes, Université

10 Laval, Québec, Canada

11

12 * Corresponding author

13 E-mail: vincent.burrus@usherbrooke.ca

14

15 Short title: *dusA*-specific genomic islands mobilizable by IncC conjugative

16 plasmids

Abstract

Salmonella Genomic Island 1 (SGI1) and its variants are significant contributors to the spread of antibiotic resistance among *Gammaproteobacteria*. All known SGI1 variants integrate at the 3' end of *trmE*, a gene coding for a tRNA modification enzyme. SGI1 variants are mobilized specifically by conjugative plasmids of the incompatibility groups A and C (IncA and IncC). Using a comparative genomics approach based on genes conserved among members of the SGI1 group, we identified diverse integrative elements distantly related to SGI1 in several species of *Vibrio*, *Aeromonas*, *Salmonella*, *Pokkaliibacter*, and *Escherichia*. Unlike SGI1, these elements target two alternative chromosomal loci, the 5' end of *dusA* and the 3' end of *yicC*. Although they share many features with SGI1, they lack antibiotic resistance genes and carry alternative integration/excision modules. Functional characterization of IME*Vch*USA3, a *dusA*-specific integrative element, revealed promoters that respond to AcaCD, the master activator of IncC plasmid transfer genes. Quantitative PCR and mating assays confirmed that IME*Vch*USA3 excises from the chromosome and is mobilized by an IncC helper plasmid from *Vibrio cholerae* to *Escherichia coli*. IME*Vch*USA3 encodes the AcaC homolog SgaC that associates with AcaD to form a hybrid activator complex AcaD/SgaC essential for its excision and mobilization. We identified the *dusA*-specific recombination directionality factor RdfN required for the integrase-mediated excision of *dusA*-specific elements from the chromosome. Like *xis* in SGI1, *rdfN* is under the control of an AcaCD-responsive promoter. Although the integration of IME*Vch*USA3 disrupts *dusA*, it

40   provides a new promoter sequence and restores the reading frame of *dusA* for

41   proper expression of the tRNA-dihydrouridine synthase A. Phylogenetic analysis

42   of the conserved proteins encoded by SGI1-like elements targeting *dusA*, *yicC*,

43   and *trmE* gives a fresh perspective on the possible origin of SGI1 and its

44   variants.

45

46   **Author summary**

47   We identified integrative elements distantly related to the *Salmonella* Genomic

48   Island 1 (SGI1), a key vector of antibiotic resistance genes in

49   *Gammaproteobacteria*. SGI1 and its variants reside at the 3' end of *trmE*, share a

50   large, highly conserved core of genes, and carry a complex integron that confers

51   multidrug resistance phenotypes to their hosts. Unlike members of the SGI1

52   group, these novel genomic islands target the 5' end *dusA* or the 3' end of *yicC*,

53   lack multidrug resistance genes, and seem much more diverse. We showed here

54   that, like SGI1, these elements are mobilized by conjugative plasmids of the IncC

55   group. Based on comparative genomics and functional analyses, we propose a

56   hypothetical model of the evolution of SGI1 and its siblings from the progenitor of

57   IncA and IncC conjugative plasmids via an intermediate *dusA*-specific integrative

58   element through gene losses and gain of alternative integration/excision

59   modules.

3

Introduction

Integrative and mobilizable elements (IMEs) are discrete, mobile chromosomal

regions that can excise from the chromosome and borrow the mating apparatus

of helper conjugative elements to transfer to a new bacterial host [1,2]. IMEs are

usually composed of two main functional modules. The site-specific

recombination module contains genes and *cis*-acting sequences that mediate the

integration of the IMEs into and their excision from the chromosome. The

mobilization module includes the *cis*-acting origin of transfer (*oriT*) and usually

encodes mobilization proteins required to initiate the conjugative transfer at *oriT*

[1]. In its simplest form, the mobilization module only consists of an *oriT* locus

mimicking the *oriT* of the helper element [3–5]. The excision of IMEs is elicited by

conjugative plasmids or integrative and conjugative elements (ICEs). These

helper elements encode the type IV secretion system (T4SS) that translocates

the IME DNA into the recipient cell [1].

Several distinct families of IMEs have been described to date. Most encode

beneficial traits for their host, such as resistance to antibiotics and heavy metals

or bacteriocin synthesis [1,6]. *Salmonella* Genomic Island 1 (SGI1) is certainly

one of the most studied IMEs. Though first described 20 years ago, SGI1 and its

siblings have only recently gained a lot of attention due to their prevalence and

prominent role in the spread of multidrug resistance [7,8]. The canonical 43-kb

SGI1 resides at the 3' end of *trmE* (also known as *mnmE* or *thdF*) in *Salmonella*

enterica serovar Typhimurium DT104 [9]. *trmE* encodes the 5-

carboxymethylaminomethyluridine-tRNA synthase GTPase subunit. SGI1

4

83   variants have been reported in a wide array of *Gammaproteobacteria*, including

84   *Proteus mirabilis* (PGI1), *Acinetobacter baumannii* (AGI1), *Morganella*,

85   *Providencia*, *Enterobacter*, *Escherichia coli*, *Vibrio cholerae* (GI-15), and

86   *Klebsiella pneumoniae* [7,10,11]. Most variants carry a class I integron

87   structurally similar to the In104 integron of SGI1. In104 confers resistance to

88   ampicillin, chloramphenicol/florfenicol, streptomycin/spectinomycin,

89   sulfamethoxazole, and tetracycline [8,12]. SGI1 and its variants are an

90   epidemiological threat exacerbated by their specific mobilization by conjugative

91   plasmids of the incompatibility groups A (IncA) and C (IncC) [13,14]. IncC

92   plasmids contribute to the global circulation of multidrug resistance genes,

93   including NDM metallo-β-lactamase and carbapenemase genes, among a broad

94   range of *Gammaproteobacteria* [15,16]. The transcriptional activator AcaCD

95   encoded by IncC plasmids triggers the excision and mobilization of SGI1 [17,18].

96   SGI1 and most variants share a conserved core set of 28 genes, representing

97   27.4 kb, disrupted by insertion sequences and the class 1 integron inserted at

98   diverse positions (Fig 1, top) [7,9,12]. Thus far, the function of a few conserved

99   genes has been characterized. Together with the *cis*-acting recombination

100   site *attP,* the genes *int* and *xis* form the recombination module of SGI1

101   [13]. *int* encodes the site-specific tyrosine recombinase (integrase) that targets

102   the 3' end of *trmE. xis* encodes the recombination directionality factor (RDF or

103   excisionase) that enhances the excision reaction catalyzed by Int. The

104   mobilization module includes the mobilization genes *mpsAB* and the *oriT* located

105   upstream of *mpsA* [19]. *mpsA* encodes an atypical relaxase distantly related to

5

106 tyrosine recombinases. Unlike most characterized IMEs, SGI1 carries a replicon

107 composed of an iteron-based origin of replication (*oriV*) and the replication

108 initiator gene *rep* [20,21]. SgaCD, a transcriptional activator complex expressed

109 by SGI1 in response to a coresident IncC plasmid, controls *rep* expression

110 [21,22]. The excised replicative form of SGI1 destabilizes the helper plasmid by

111 an unknown process, and is further stabilized by its *sgiAT* addiction module

112 [20,22–24]. Finally, SGI1 encodes three mating pore subunits, TraN$_S$, TraH$_S$, and

113 TraG$_S$, that actively replace their counterparts in the T4SS encoded by the IncC

114 plasmid [25]. The substitution of TraG allows SGI1 to bypass the IncC-encoded

115 entry exclusion mechanism and transfer between cells carrying conjugative

116 plasmids belonging to the same entry exclusion group [26].

117 Given the high similarity between SGI1 variants integrated at *trmE*, we undertook

118 a search for distant SGI1-like IMEs in bacterial genomes using MpsA, TraG$_S$,

119 SgaC, and TraN$_S$ as baits. Here, we report the existence of distantly related IMEs

120 integrated at the 5' end of *dusA* in several species of *Vibrionaceae* and the 3' end

121 of *yicC* in several species of *Gammaproteobacteria*. We have examined the

122 interactions between an IncC plasmid and IME*Vch*USA3, a *dusA*-specific

123 representative IME from an environmental *V. cholerae* strain. The genetic

124 determinants required for the excision of IME*Vch*USA3 and its mobilization by

125 IncC plasmids were characterized. Finally, we took a fresh look at the emergence

126 and evolution of SGI1 and its siblings by conducting phylogenetic analyses and

127 proposed a hypothetical evolutionary pathway of putative IMEs resembling SGI1.

Results

128

Novel integrative elements (IEs) distantly related to SGI1 are inserted in

129

dusA and *yicC* in various *Gammaproteobacteria*

130

131 To find novel SGI1-like elements, we searched the Refseq database using blastp

132 and the primary sequences of MpsA, TraG$_S$, SgaC, and TraN$_S$. Considering the

133 substitution of integration modules can change the integration site [27–29], the

134 integrase Int$_{trmE}$ was excluded from the analysis. We identified 24 distinct

135 integrative elements encoding homologs of the four bait proteins in 36 different

136 bacterial strains (Fig 1, Table 1 and S1 Table). 21 of these IEs are integrated into

137 the 5' end of *dusA* in diverse *Vibrio* species from various origins. The remaining

138 three are located at the 3' end of *yicC* in *E. coli*, *Aeromonas veronii*, *P. mirabilis*,

139 *S. enterica* serovar Kentucky, and *Pokkaliibacter plantistimulans*. The size of the

140 IEs varies from 22.8 kb to 37.1 kb. The conserved genes *mpsA* (together with

141 *mpsB*), *traG*, *traN*, and *sgaC* remain in a syntenic order, though sporadically

142 separated by variable DNA (Fig 1).

143 **Figure 1. Schematic representations of SGI1-related IEs.** The position and

144 orientation of open reading frames (ORFs) are indicated by arrowed boxes.

145 Colors depict the function deduced from functional analyses and BLAST

146 comparisons. Potential AcaCD binding sites are represented by green angled

147 arrows. Each island is flanked by the *attL* and *attR* (vertical grey lines)

148 attachment sites when integrated into the 3' end of *trmE* (light blue), the 5' end of

149 *dusA* (light green), or the 3' end of *yicC* (pink). The annotation of *attL* and *attR*

150 relative to *int* is based on SGI1 (*trmE*) [9], IE*Aba*D1279779 of *Acinetobacter*

151 *baumannii* D1279779 (*dusA*) [30] and MGI*Vfl*Ind1 (*yicC*) [3]. Details regarding

152 ORFs are shown in S1 Dataset.

153 Consistent with the change of integration site, the respective *int* genes of SGI1

154 and the *dusA*- and *yicC*-specific IEs do not share any sequence similarity.

155 Furthermore, unlike SGI1, these novel IEs lack *xis* downstream of *int* (Fig 1).

156 Instead, *yicC*-specific IEs carry two small open reading frames (ORF) upstream

157 of the *attR* site. The putative translation product of the second one shares 35%

158 identity over 65% coverage with the excisionase RdfM of MGI*Vfl*Ind1 [31].

159 Although *dusA*-specific IEs lack *xis* and *rdfM*, all carry an ORF predicted to

160 encode a 76-aminoacyl residue protein containing the pyocin activator protein

161 PrtN domain (Pfam PF11112). Based on its size, position, predicted DNA-binding

162 function, conservation, and evidence presented below, we named this ORF *rdfN*.

163 Phylogenetic analysis of Int$_{yicC}$ proteins of *yicC*-specific SGI1-like IEs form a

164 cluster distinct from the integrases of IMEs mobilizable by IncC plasmids through

165 a MobI protein (Pfam PF19456), such as MGI*Vmi*1, and IMEs that mimic the *oriT*

166 of SXT/R391 ICEs, such as MGI*Vfl*Ind1 [3,17,32] (Fig 2A).

167 **Figure 2. Integrases encoded by the *yicC*- and *dusA*-specific IEs**. Maximum

168 likelihood phylogenetic analyses of Int$_{yicC}$ (A) and Int$_{dusA}$ (B). The trees are drawn

169 to scale, with branch lengths measured in the number of substitutions per site

170 over 400 and 359 amino acid positions for Int$_{yicC}$, and Int$_{dusA}$, respectively. The

171 helper elements and mechanism of mobilization are indicated for each lineage

172 according to the keys shown in the legend box of panel A. The inset of panel B

173   shows logo sequences of the repeats in *attL* and *attR* attachment sites. The

174   arrows indicate the island termini experimentally determined for IE*Aba*D1279779

175   by Farrugia *et al.* [30]. (C) Heatmap showing blastp identity percentages of

176   pairwise comparison of Int$_{dusA}$ representative proteins. Proteins accession

177   numbers are provided in S2 Dataset, except for IE*Aba*D1279779

178   (WP_000534871.1), IE*Ppr*Pf-5 of *Pseudomonas protegens* Pf-5

179   (WP_011060295.1), and IEs of *Burkholderia gladioli* BSR3 (WP_013697845.1),

180   *Bradyrhizobium* sp. BTAi1 (WP_012043559.1), *Agrobacterium* sp. H13-3

181   (WP_013636109.1), and *Neisseria gonorrhoeae* FA 1090 (EFF39980.1).

182   Phylogenetic analysis of Int$_{dusA}$ proteins confirmed that the integrases of these

183   IEs form a monophyletic group exclusive to the *Vibrionaceae* and distinct from

184   those encoded by other *dusA*-specific IEs found in other taxa, including

185   GI*Aca*Bra1 from *Aeromonas caviae* that is likely mobilizable by IncC plasmids via

186   a MobI protein [32] (Fig 2B). Int$_{dusA}$ proteins of the IEs identified here share at

187   least 75% identity, while identities drop below 60% with the non-*Vibrio* Int$_{dusA}$

188   proteins (Fig 2C). Sequence logos built using alignments of the *attL* and *attR*

189   chromosomal junctions revealed a 21-bp imperfect repeat at the extremities of

190   each IE (Fig 2B). This repeat is similar to the one reported for *dusA*-specific IEs

191   found in a broader range of species [30].

192   **Three types of *dusA*-integrated SGI1-related elements**

193   Blastn and blastp analyses using SGI1ΔIn104 as the reference confirmed that

194   the identified *dusA*-specific IEs share limited sequence similarities with SGI1

195   (S1A Fig). Besides the genes encoding MpsA, TraG, SgaC, and TraN, all carry

9

196   the auxiliary mobilization factor gene *mpsB* and the *oriT* sequence (Fig 1).

197   Secondary structure prediction of the aligned *oriT* sequences located upstream

198   of *mpsA* using RNAalifold revealed that despite the sequence divergence, the

199   structure of *oriT* with three stem-loops was strictly conserved (S2B Fig). In

200   contrast, *sgaD* is not strictly conserved and highly divergent from *sgaD* of SGI1

201   when present (Fig 1 and S1A Fig).

202   Comparison using IE*Vch*USA2 as the reference suggests that *dusA*-specific IEs

203   cluster into three distinct types as confirmed by the phylogenetic analysis of

204   concatenated MpsA-TraG-SgaC-TraN (Fig 3A, S1B and S3 Fig). Type 1 *dusA*-

205   specific IEs such as IME*Vch*USA3 are mainly found in *V. cholerae* and lack both

206   *traH* and *sgaD*. Type 2 IEs such as IE*Vch*USA2 lack *sgaD* but carry *traH* (Fig 1

207   and 3A). This lineage only includes two *dusA*-specific IEs of *V. cholerae* but also

208   closely related *yicC*-specific IEs such as IE*Eco*MOD1 and the *trmE*-specific

209   GI*Vch*O27-1. Finally, type 3 IEs such as GI*Vch*USA5 are the most distant from

210   the two other types and SGI1 (Fig 3A). Type 3 IEs carry both *traH* and *sgaD* and

211   reside in diverse *Vibrio* species. With the exception of a few outliers encoded by

212   IEs such as IE*Vch*N2817, IE*Vch*N2708 or IE*Ppl*Ind1, the proteins MpsA, TraG,

213   SgaC and TraN encoded by members of the same type typically share more than

214   95% identity (Fig 3B and S3). MpsA remains the least divergent protein between

215   the three types, sharing at least 65% identity between type 1 and type 3, and

216   from 64% to 93% with SGI1. In contrast, TraG and TraN are the most divergent

217   between types, ranging from 46% to 59% for TraG and from 46% to 76% for

218   TraN.

10

219   Worthy of note, these three distinct lineages of *dusA*-specific IEs are supported

220   by the phylogeny of the *oriT* sequences (S2A Fig). Again, *oriT* loci of type 3 IEs

221   strongly diverge from those of types 1 and 2, as well as from the *oriT* loci of the

222   highly homogenous SGI1 group.

223   **Figure 3. Conserved genes support three main lineages of *dusA*-specific**

224   **SGI1-like IEs.** (A) Maximum likelihood phylogenetic analysis of concatenated

225   MpsA-TraG-SgaC-TraN. The tree is drawn to scale, with branch lengths

226   measured in the number of substitutions per site over 2,637 amino acid positions.

227   Taxa corresponding to IEs targeting *trmE* and *yicC* are indicated by a light blue

228   circle and a red circle, respectively. All other taxa correspond to *dusA*-specific

229   IEs. Phylogenetic relationships of MpsA, TraG, SgaC and TraN proteins are

230   shown separately in S2 Fig. (B) Heatmaps showing blastp identity percentages of

231   pairwise protein comparisons for representatives of MpsA, TraG, SgaC, and

232   TraN. Proteins accession numbers and clusters are provided in S1 Table and S2

233   Dataset.

234   **Variable features found in the *dusA*- and *yicC*-specific IEs**

235   Most variable genes in the identified IEs encode proteins of unknown function. A

236   search for antibiotic resistance determinants using the Resistance Gene Identifier

237   server failed to reveal any known resistance gene. Several IEs encode putative

238   functions altering host processes and virulence, including the transport of ions

239   and small molecules (*ktrAB*, *trkH*, and *kdpD* for potassium uptake and *rcnAR* for

240   nickel/cobalt efflux in IE*Vch*Hai10, sulfite export in IE*Vch*N2817 and IE*Vch*Swe1),

241   c-di-GMP degradation (IE*Vch*Ban1), and fimbriae (IE*Vch*Ban1) (S1 Dataset).

11

242 None of the reported IEs carries the same replication module (*S004-rep-oriV*) as

243 canonical SGI1. Instead, five *dusA*-specific IEs belonging to the type 3 lineage

244 (IE*Vch*USA5, IE*Vch*Bra2, IE*Vpa*Chn1, IE*Vpa*Chn2, and IE*Vpa*Ban1a) encode a

245 putative replication initiator protein with the IncFII_repA domain (Pfam PF02387)

246 (Fig 1, S1 Dataset). IE*Vvu*USA1 encodes a putative helicase with an UvrD_C_2

247 domain (Pfam PF13538), whereas IE*Vch*Hai10, IE*Vch*USA2 and IE*Vch*Swe1

248 encode a putative ATP-dependent helicase (PcrA) and a putative ATP-

249 dependent endonuclease (YbjD). In addition, IE*Vch*N2786, IE*Sen*USA1 and

250 IE*Eco*MOD1 encode a predicted DEAD/DEAH box helicase (Pfam PF00270 and

251 PF00271). The three *yicC*-specific IEs encode a homolog of TrfA (Pfam

252 PF07042), the replication initiator protein of broad-host-range IncP plasmids [33].

253 No replicative functions could be ascribed with confidence to any gene carried by

254 the other *dusA*-specific IEs. Several IEs also encode toxin-antitoxin systems,

255 such as *sgiAT* and *higAB*, which likely enhance their stability (Fig 1). In the type 3

256 IEs IE*Vch*Bra2, IE*Vch*N2708, IE*Vpa*Chn1 and IE*Vpa*Ban1a, *sgiAT* is associated

257 with a gene coding for a putative abortive infection bacteriophage resistance

258 factor (Abi_2, Pfam PF07751). Likewise, IE*Vvu*USA1 carries a gene coding for a

259 different putative abortive infection bacteriophage resistance factor (AbiEii toxin,

260 Pfam PF13304).

261 Finally, IE*Vpa*Ban1a is integrated at *dusA* adjacent to a distinct IE, IE*Vpa*Ban1b,

262 in a tandem fashion. GI*Vpa*Ban1b codes for two predicted integrases sharing

263 44% and 27% identity with Int$_{dusA}$ of IE*Vpa*Ban1a. GI*Vpa*Ban1b encodes a

264   putative type I restriction-modification system, a MobA-like relaxase (MOB$_{P1}$), the

265   mobilization auxiliary factor MobC, and an RdfN homolog (Fig 1).

266   **Non-canonical SGI1-like IEs carry AcaCD-responsive genes**

267   Considering the divergence of the 24 new IEs from prototypical SGI1, we

268   wondered whether an IncC plasmid could mobilize them like SGI1. The hallmark

269   of IncC-dependent mobilization is the presence of AcaCD-responsive promoters

270   in IncC-mobilizable IEs. Hence, we searched for putative AcaCD-binding sites in

271   the sequences of *trmE*-specific IEs (prototypical SGI1 was used as the positive

272   control) and the *yicC*- and *dusA*-specific IEs. In these IEs, an AcaCD-binding

273   motif was predicted upstream of *traN*, *traHG* (or *traG*), *S018*, and *xis* (or *rdfM* or

274   *rdfN*) (Fig 1 and S4 Fig). Moreover, an AcaCD-binding motif was also predicted

275   upstream of *trfA* in the *yicC*-specific IEs.

276   We cloned the promoter sequences of *int*, *traN*, *traG*, *S018*, and *rdfN* of

277   IME*Vch*USA3 upstream of a promoterless *lacZ* reporter gene and monitored the

278   β-galactosidase activity with or without AcaCD. The promoter P_{int} was active

279   regardless of the presence of AcaCD (Fig 4A). In contrast, the four other

280   promoters exhibited weak activity in the absence of AcaCD. Upon induction of

281   *acaDC* expression, P_{traN} and P_{S018} remained unresponsive, while the activities of

282   P_{traG} and P_{rdfN} increased 40 and 400 times, respectively (Fig 4B). The inertia of

283   P_{traN} and P_{S018} toward AcaCD could result from single nucleotide substitutions in

284   the AcaCD binding site previously shown to be essential for recruiting the

285   activator [22]: CCS**AA**AWW instead of CCS**C**AAWW in P_{traN} and CCC**C**AAAA instead of

286   CCC**A**AAAA in P_{S018} (S4 Fig).

13

287   **Figure 4. β-galactosidase activities of the promoters P_{int}, P_{traN}, P_{traG}, P_{S018}**

288   **and P_{rdfN} of IMEVchUSA3 transcriptionally fused to $lacZ$.** (A) Colonies were

289   grown on LB agar with or without arabinose to induce $acaDC$ expression from

290   pBAD-$acaDC$. (B) Induction levels of the same promoters in response to AcaCD.

291   β-galactosidase assays were carried out using the strains of panel A. Ratios

292   between the enzymatic activities in Miller units for the arabinose-induced versus

293   non-induced strains containing pBAD-$acaDC$ are shown. The bars represent the

294   mean and standard error of the mean of three independent experiments.

295   Hence, despite their divergence and different integration sites, these IEs share

296   with SGI1 a common activation mechanism elicited by the presence of an IncC

297   plasmid.

298   **IncC plasmids induce the excision and mobilization of IMEVchUSA3**

299   Next, we tested whether a coresident IncC plasmid could trigger the excision of

300   IMEVchUSA3 from $dusA$ in its original host, $V.$ $cholerae$ OY6PG08. The

301   derepressed IncC plasmid pVCR94[Kn] Δ$acr2$ [34] was introduced into OY6PG08

302   by conjugation from $E.$ $coli$ KH40. The Δ$acr2$ mutation improves the efficiency of

303   interspecific transfer of the plasmid [35]. OY6PG08 Kn[R] transconjugants were

304   tested by PCR to amplify the $attL$ and $attR$ chromosomal junctions, as well as the

305   $attB$ and $attP$ sites resulting from the excision of IMEVchUSA3 (S5A Fig).

306   IMEVchUSA3 was rarely retained in the transconjugants compared to the control

307   IncC-free OY6PG08 clones, suggesting it was unstable and rapidly lost in IncC[+]

308   cells (S5B and S5C Figs).

309   To test the interspecific mobilization of IME*Vch*USA3 from *V. cholerae*

310   OY6PG08, we inserted a selection marker upstream of *traG* and used pVCR94^{Kn}

311   Δ*acr2* as the helper plasmid. IME*Vch*USA3^{Cm} transferred to *E. coli* CAG18439 at

312   a frequency of 7.01×10^{-5} transconjugant/donor CFUs. Amplification of *attL* and

313   *attR* using *E. coli*-specific primers confirmed that IME*Vch*USA3 integrates at

314   *dusA* in *E. coli* (S5D Fig).

315   **Excision of *dusA*-specific IEs depends on *rdfN***

316   To further characterize the biology of IME*Vch*USA3, we measured its excision

317   rate and copy number by qPCR, with and without coresident pVCR94^{Sp}. We also

318   monitored its intraspecific transfer (*E. coli* to *E. coli*) in the same context.

319   Spontaneous excision of the island rarely occurred (<0.001% of the cells) (Fig

320   5A). In contrast, in the presence of the helper plasmid, the free *attB* site was

321   detected in more than 67% of the cells confirming that the IncC plasmid elicits

322   the excision of IME*Vch*USA3^{Kn}. Likewise, the presence of the plasmid resulted in

323   a ~3-fold increase of the copy number of IME*Vch*USA3^{Kn} (Fig 5B), suggesting

324   that the excised form of the island undergoes replication. The frequency of

325   transfer of IME*Vch*USA3^{Kn} was comparable to that of the helper plasmid

326   (~3.5×10^{-2} transconjugants/donor), while the frequency of cotransfer was more

327   than two logs lower (Fig 5C).

328   **Figure 5. Effect of *acaDC* and *rdfN* on the IncC-dependent excision and**

329   **mobilization of IME*Vch*USA3.** (A) IME*Vch*USA3^{Kn} excision rate corresponds to

330   the *attB*/chromosome ratio. (B) IME*Vch*USA3^{Kn} copy number corresponds to the

331   *higA*/chromosome ratio. For panels A and B, all ratios were normalized using the

15

332 control set to 1 and displayed in white. (C) Impact of *acaC*, *acaDC*, *sgaC* and

333 *rdfN* deletions on the mobilization of IME*Vch*USA3. Conjugation assays were

334 performed with CAG18439 (Tc) containing the specified elements as donor

335 strains and VB112 (Rf) as the recipient strain. The bars represent the mean and

336 standard error of the mean obtained from a biological triplicate. ¤ indicates that

337 the excision rate or transfer frequency was below the detection limit. Statistical

338 analyses were performed (on the logarithm of the values for panels A and C)

339 using a one-way ANOVA with Dunnett's multiple comparison test. For panels A

340 and B, statistical significance indicates comparisons to the normalization control.

341 Statistical significance is indicated as follows: ****, $P < 0.0001$; ***, $P < 0.001$; **,

342 $P < 0.01$; *, $P < 0.05$; ns, not significant. (D) Schematic representation of mini-IE

343 inserted at the 5' end of *dusA*. (E) RdfN acts as a recombination directionality

344 factor. Detection of *attB*, *attP*, *attL* and *attR* sites by PCR in colonies of *E. coli*

345 EC100 *dusA*::mini-IE in the presence or absence of *rdfN*. L, 1Kb Plus DNA

346 ladder (Transgen Biotech).

347 Thus far, the factors required to catalyze the excision of *dusA*-specific IEs have

348 not been examined [30]. Whereas all *dusA*-specific IEs lack *xis* downstream of

349 *int*, they carry a small ORF, here named *rdfN*, coding for a putative PrtN homolog

350 (Fig 1) [30]. The deletion of *rdfN* abolished the excision and replication of

351 IME*Vch*USA3[Kn]. Complementation by ectopic expression of *rdfN* from the

352 arabinose-inducible promoter P_{BAD} restored the wild-type excision level but not

353 the replication (Fig 5A and 5B). Likewise, deletion of *rdfN* abolished the

16

354   mobilization of IME*Vch*USA3^{Kn} but had no impact on the transfer of the helper

355   plasmid (Fig 5C), confirming the specific role of *rdfN* in the IE's mobility.

356   To confirm that *rdfN* encodes the sole and only RDF of IME*Vch*USA3, we

357   constructed mini-IE, a minimal version of IME*Vch*USA3 that only contains *int* and

358   a spectinomycin-resistance marker. mini-IE is flanked by *attL* and *attR* and is

359   integrated at *dusA* in *E. coli* EC100 (Fig 5D). Using mini-IE, *attB* and *attP* were

360   detected only upon ectopic expression of *rdfN* from pBAD-*rdfN*, confirming that

361   no other IME*Vch*USA3-encoded protein besides Int and RdfN is required for the

362   excision of the element (Fig 5E). *rdfN* is the essential RDF gene that favors the

363   excision of IME*Vch*USA3 and, most likely, all *dusA*-specific IEs.

364   **A SgaC/AcaD hybrid complex activates the excision and mobilization of**

365   **IME*Vch*USA3**

366   Next, we investigated the role of the transcriptional activator genes *acaC* and

367   *sgaC* in the mobilization of IME*Vch*USA3. Deletion of *acaDC* abolished the

368   excision and replication of IME*Vch*USA3^{Kn}, confirming that its excision relies on

369   *rdfN*, whose expression is activated by AcaCD (Figs 4, 5A and 5B). The mutation

370   also confirmed that SgaC provided by IME*Vch*USA3^{Kn} is insufficient by itself to

371   elicit *rdfN* expression. The excision rate remained extremely low in cells that lack

372   the helper plasmid or cells that carry pVCR94^{Sp} Δ*acaDC*. However,

373   IME*Vch*USA3^{Kn} allowed the low-frequency transfer of pVCR94^{Sp} Δ*acaDC* [17]

374   (Fig 5C). Hence SgaC alone can activate to some degree the expression of the

375   transfer genes of the helper plasmid. In contrast, deletion of *acaC* had no

376   significant impact on the excision, replication, and mobilization of

17

377   IME*Vch*USA3^{Kn}, or on the transfer of the helper plasmid (Figs 5A, 5B and 5C).

378   The primary sequences of AcaC and SgaC from IME*Vch*USA3 share 85%

379   identity over 94% coverage, whereas AcaC and SgaC from SGI1 share only 75%

380   identity over 92% coverage. Hence AcaD produced by the plasmid and SgaC

381   produced by the IME likely generate a functional chimeric transcriptional complex

382   that acts as a potent activator of *rdfN* and the transfer genes.

383   The transfer of IME*Vch*USA3^{Kn} Δ*sgaC* decreased nearly 3 logs compared to the

384   wild-type IE, despite the presence of *acaDC* on the helper plasmid (Fig 5C).

385   Moreover, deletion of both *acaC* and *sgaC* nearly abolished all transfer. Ectopic

386   expression of *sgaC* alone from pBAD-*sgaC* complemented these deletions to

387   wild-type levels (Fig 5C). These observations confirm that *sgaC*, not *acaC*,

388   combined with *acaD* produces a hybrid activator complex that is essential for the

389   excision and mobilization of the element.

390   **IME*Vch*USA3 provides a new promoter and N-terminus for *dusA***

391   **expression**

392   Since *dusA*-specific IEs insert within the 5' end of *dusA*, we wondered whether

393   the gene remains expressed after the integration event. Sequence analysis of the

394   *attR* junction of *E. coli* K12 transconjugants revealed that IME*Vch*USA3 provides

395   a new 5' coding sequence that diverges significantly from the native *E. coli dusA*

396   gene (Fig 6A). This alteration of the 5' end of *dusA* results in a novel N-terminus

397   of identical length sharing 61% identity over the 35 initial amino acid residues

398   with native DusA. To test the expression of *dusA*, we constructed a translational

399   *lacZ* fusion to its fortieth codon downstream of the *attR* junction in *E. coli*

18

400   CAG18439 and BW25113 (Fig 6B). β-galactosidase assays revealed that *dusA*

401   remains expressed after integration in both strains, confirming that IME*Vch*USA3

402   provides a new promoter (Fig 6C). However, we observed a statistically

403   significant reduction of *dusA* expression resulting from the integration of the IE in

404   both strains, suggesting that the transcription or translation signals brought by

405   the IE are weaker than the original ones upstream of *E. coli dusA*.

406   **Figure 6. IME*Vch*USA3 drives the expression of *dusA*.** (A) Comparison of the

407   coding sequences of the 5' end of *dusA* in *E. coli* K12 MG1655 before (*attB* site)

408   and after (*attL* junction) the integration of IME*Vch*USA3. The core sequence of

409   the *attB* and *attL* recombination sites is indicated with red shading. The ATG start

410   codon of *dusA* is shown in bold. The sequence shown in blue is internal to

411   IME*Vch*USA3. Amino acid residues shown in red differ from the native N-

412   terminus of DusA. This sequence was obtained by sequencing the *attL* junction

413   of an *E. coli* CAG18439 *dusA*::IME*Vch*USA3^{Kn} transconjugant colony. (B)

414   Schematic representation of the *dusA*'-*lacZ* translational fusion for the detection

415   of *dusA* expression. (C) β-galactosidase activity of the *dusA*'-'*lacZ* fusion before

416   (-) and after (+) insertion of IME*Vch*USA3^{Kn} in *E. coli* CAG18439 (FD034) and

417   BW25113 (FD036). The bars represent the mean and standard error of the mean

418   of three independent experiments. Statistical analyses were performed using an

419   unpaired *t* test to compare the expression before and after integration of

420   IME*Vch*USA3^{Kn} for each strain. Statistical significance is indicated as follows: **,

421   *P* < 0.01; *, *P* < 0.05.

422 **Discussion**

423 SGI1-like elements integrated at the 3' end of *trmE* are widespread in a broad

424 range of *Enterobacteriaceae* and sporadically found in a few *Vibrio* species [7].

425 The integrase of SGI1 and its variants occasionally targets the intergenic region

426 between *sodB* and *purR* genes, a secondary attachment site [36]. Here, we

427 report the identification of distant SGI1-like elements that specifically target the 5'

428 end of *dusA* in multiple *Vibrio* species and the 3' end of *yicC* in

429 *Enterobacteriaceae* and *Balneatrichaceae*. Farrugia *et al.* [30] already described

430 IEs integrated at the 5' end of *dusA*, mostly prophages or phage remnants found

431 exclusively in *Alpha-*, *Beta-* and *Gammaproteobacteria*. These authors identified

432 IE*Vch*Ban1a and IE*Vch*Bra2 in *V. cholerae*, and several other IEs predicted to

433 encode conjugative functions in *Bradyrhizobium*, *Caulobacter*, *Mesorhizobium*,

434 *Paracoccus*, *Pseudomonas*, and *Rhodomicrobium* [30]. Our group recently

435 reported a *dusA*-specific IE in *Aeromonas caviae* 8LM potentially mobilizable by

436 IncC plasmids [32]. GI*Aca*8LM lacks *tra* genes but encodes a mobilization protein

437 (MobI) under the control of an AcaCD-responsive promoter. Together, these

438 reports confirm that *dusA* is an insertion hotspot for distinct families of mobile

439 elements across at least three *Proteobacteria* phyla.

440 Thus far, only the *dusA*-specific IEs in *A. baumannii* D1279779 and *P. protegens*

441 Pf-5 were shown to excise from the chromosome, albeit at a low level [30].

442 Neither IE has been tested for intercellular mobility. Here, we characterized

443 IME*Vch*USA3, a representative member of a subgroup of *dusA*-specific IEs

444 circulating in *Vibrio* species and distantly related to SGI1. We demonstrated that

20

445   IME*Vch*USA3 is mobilizable by IncC conjugative plasmids to *E. coli*. In the

446   presence of an IncC plasmid, this IME excises in practically all cells of the

447   population and becomes highly unstable (Figs 5B and 6A). We showed that its

448   excision was under the control of AcaCD provided by the IncC plasmid and

449   required *rdfN*, a gene whose expression is driven by an AcaCD-responsive

450   promoter (Fig 4). *rdfN* encodes a novel RDF distantly related to the pyocin

451   activator protein PrtN of *Pseudomonas*. *rdfN* seems to be ubiquitous, yet highly

452   divergent, in *dusA*-specific IEs reported by Farrugia *et al.* [30]. For instance,

453   RdfN (PrtN) encoded by the IE of *P. protegens* Pf-5 shares only 29% identity with

454   RdfN of IME*Vch*USA3, and their promoters are unrelated. Hence, the expression

455   of *rdfN* homologs encoded by different families of *dusA*-specific IEs is likely

456   controlled by different factors. Only the IEs that have evolved AcaCD-responsive

457   promoters for their RDF gene are expected to be mobilizable by IncC or related

458   plasmids.

459   Excision and mobilization of IME*Vch*USA3 occurred in the presence of a Δ*acaC*

460   but not a Δ*acaDC* mutant of the helper plasmid (Fig 5), confirming that *sgaC* of

461   the IME produces a functional activator subunit that can interact with AcaD

462   provided by the plasmid. Furthermore, we showed here that, unlike *acaC*, *sgaC*

463   plays a central role in the biology of IME*Vch*USA3 as the absence of *acaC* had

464   no effect on the excision or transfer of the IME, while the absence of *sgaC* in

465   spite of the presence of *acaC*, compromised its mobilization (Figs 5A, 5B and

466   5C). We recently showed that AcaD most likely stabilizes the binding of AcaC to

467   the DNA [22]. Therefore, AcaD and SgaC from IME*Vch*USA3 likely interact to

468   form a chimerical activator complex. This interaction could compensate for the

469   loss of *sgaD* in *yicC*- and type 1 and 2 *dusA*-specific IEs (Fig 1). The primary

470   sequences of AcaC and SgaC of IME*Vch*USA3 (type 1) share 85% identity. In

471   contrast, AcaC only shares 75% identity with SgaC of SGI1 and 64% identity with

472   SgaC of GI*Vch*USA5 (type 3), suggesting that retention of *sgaD* allowed faster

473   divergence of SgaC from AcaC. Retention of *sgaC* in the IEs could result from its

474   essential role as the elicitor of excision and replication reported for SGI1. Indeed,

475   although AcaCD binds to the promoters P_{xis} and P_{rep} of SGI1, it fails to initiate

476   transcription at these two promoters, unlike SgaCD [22]. Nonetheless, P_{xis} and

477   P_{rep} are not conserved in the IEs described here. *S004-rep* is missing, whereas

478   *rdfN* or *rdfM* replaced *xis* with novel AcaCD-responsive promoters (Fig 4 and S3

479   Fig). This observation raises intriguing questions regarding the recruitment of

480   functional gene cassettes and their assimilation in a regulatory pathway. How did

481   *xis*, *rdfN*, and *rdfM* acquire their AcaCD-responsive promoters? Is it by

482   convergent evolution? What are the signals driving *rdfN* expression and IE

483   excision in *dusA*-specific IEs resembling prophages?

484   Approximately 3 copies per cell of IME*Vch*USA3 were detected in the presence

485   of the helper IncC plasmid (Fig 5B), lower than the copy number reported for

486   SGI1 (~8 copies/cell) [20,22]. IME*Vch*USA3 lacks SGI1's replication module

487   (*S004-rep-oriV*); however, one of the multiple genes of unknown function could

488   encode an unidentified replication initiator protein. Notably, GI*Vch*O27-1 encodes

489   a putative replication protein with an N-terminal replicase domain (PF03090) and

490   a C-terminal primase domain (PriCT-1, PF08708) [20]. Multiple IEs also carry

22

491   putative replicons based on *repA* and *trfA* (Fig 1), suggesting that independent

492   replication is crucial in their lifecycle, perhaps to improve their stability in the

493   presence of a helper plasmid [20–24].

494   Farrugia *et al.* [30] hypothesized that *dusA*-specific IEs could restore the

495   functioning of DusA. We demonstrated here that IME*Vch*USA3 provides a new

496   promoter allowing expression of *dusA*, though at a lower level than in IME-free

497   cells, and restores the open reading frame with an altered N-terminus (Fig 6).

498   Similarly, the ICE SXT that targets the 5' end of the peptide chain release factor

499   3 (RF3) gene *prfC* provides a new promoter and N-terminus in both *V. cholerae*,

500   its original host, and *E. coli* [37]. In both cases, the consequences of the

501   alteration of the N-terminus on the activity of the protein remain unknown.

502   The relative positions of *int* and *rdfN*/*rdfM* across the *attP* site suggest that, to

503   remain functional, the recombination modules must be acquired or exchanged

504   when the IEs are in their excised circular form. The promiscuity of different

505   families of IEs targeting *yicC*, *dusA*, and *trmE* and mobilizable by IncC plasmids

506   could act as the catalyst for these recombination events. During entry into a new

507   host cell by conjugation, IncC plasmids elicit the excision of such IEs and

508   promote homologous recombination between short repeated sequences in

509   response to double-stranded break induced by host defense systems (CRISPR-

510   Cas3) [34]. Hence the diversification of IncC plasmid-mobilizable IEs could be a

511   side effect of the DNA repair mechanism used by these plasmids.

512   Unlike SGI1 and its siblings, all *dusA*-specific SGI1-like IEs reported here lack

513   antibiotic resistance genes. Furthermore, SGI1 variants are prevalent in several

514   pathogenic species and relatively well-conserved, whereas their *dusA*-specific

515   relatives are scarce and highly divergent. These observations suggest that

516   despite the considerable functional resemblances between *trmE*- and *dusA*-

517   specific SGI1-like IEs, the epidemiological success of the SGI1 lineage has

518   directly stemmed from the acquisition of class I integrons conferring multidrug

519   resistance by forerunner elements such as SGI0 [38]. Based on the phylogenetic

520   relationships between the core proteins MpsA, TraG, SgaC and TraN, *oriT* loci,

521   and integrase proteins (Figs 2 and 3, and S2A and S3 Figs), we propose a

522   hypothetical evolutionary pathway leading to the emergence of the different types

523   of IEs described here (Fig 7). The diversity of *dusA*-specific IEs and relative

524   homogeneity of the SGI1 group suggest that the latter originated from the

525   progenitor of IncA and IncC plasmids via a *dusA*-specific IE intermediate.

526   **Figure 7. Proposed hypothetical evolutionary pathway of SGI1-like IEs.** The

527   sequence of events was inferred from the phylogenetic trees presented in this

528   study, site of integration and conservation of *traH* and *sgaD* in the IEs. The

529   proposed pathway ignores the gene cargo and presumes that the IE lineages

530   evolved from the progenitor of IncA and IncC plasmids. The *dusA*-specific

531   recombination module was chosen as the progenitor to minimize gain/loss and

532   recombination events. Green and red arrows indicate gene gains and losses,

533   respectively. The orange dashed line indicates a probable recombination event

534   from which stemmed GI*Vch*O27-1.

24

Materials and Methods

Bacterial strains and media

Bacterial strains and plasmids used in this study are described in Table 2. Strains were routinely grown in lysogeny broth at 37°C in an orbital shaker/incubator and were preserved at -75°C in LB broth containing 20% (vol/vol) glycerol. Antibiotics were used at the following concentrations: ampicillin (Ap), 100 µg/ml; chloramphenicol (Cm), 20 µg/ml; erythromycin (Em), 200 µg/ml; kanamycin (Kn), 10 µg/ml for single-copy integrants of pOP*lacZ*-derived constructs, 50 µg/ml otherwise; nalidixic acid (Nx), 40 µg/ml; rifampicin (Rf), 50 µg/ml; spectinomycin (Sp), 50 µg/ml; tetracycline (Tc), 12 µg/ml. Diaminopimelate (DAP) was supplemented to a final concentration of 0.3 mM when necessary.

Mating assays

Conjugation assays were performed as previously described [25]. However, mixtures of donor and recipient cells were incubated on LB agar plates at 37°C for 4 hours. Donors and recipients were selected according to their sole chromosomal markers. When required, mating experiments were performed using LB agar plates supplemented with 0.02% arabinose to induce expression of pBAD30-derived complementation vectors. Frequencies of transconjugant formation were calculated as ratios of transconjugant per donor CFUs from three independent mating experiments.

Molecular biology

Plasmid DNA was prepared using the QIAprep Spin Miniprep Kit (Qiagen), according to manufacturer's instructions. Restriction enzymes used in this study

25

558 were purchased from New England Biolabs. Q5 DNA polymerase (New England

559 Biolabs) and EasyTaq DNA Polymerase (Civic Bioscience) were used for

560 amplifying cloning inserts and verification, respectively. PCR products were

561 purified using the QIAquick PCR Purification Kit (Qiagen), according to

562 manufacturer's instructions. *E. coli* was transformed by electroporation as

563 described by Dower *et al.* [39] in a Bio-Rad GenePulser Xcell apparatus set at 25

564 µF, 200 Ω and 1.8 kV using 1-mm gap electroporation cuvettes. Sanger

565 sequencing reactions were performed by the Plateforme de Séquençage et de

566 Génotypage du Centre de Recherche du CHUL (Québec, QC, Canada).

567 **Plasmids and strains constructions**

568 Plasmids and oligonucleotides used in this study are listed in Table 2 and S2

569 Table, respectively. IME*Vch*USA3^{Cm} was constructed by inserting the *pir*-

570 dependent replication RP4-mobilizable plasmid pSW23T [40] at locus

571 CGT85_RS05425 of *V. cholerae* OYP6G08 (Genbank NZ_NMSY01000009) by

572 homologous recombination. Briefly, CGT85_RS05425 was amplified using primer

573 pair dusAigEcoRIF/dusAigEcoRIR. The amplicon was digested with EcoRI and

574 cloned into EcoRI-digested pSW23T using T4 DNA ligase. The resulting plasmid

575 was confirmed by restriction profiling and DNA sequencing, then introduced into

576 the DAP-auxotrophic *E. coli* β2163 [40] by transformation. This strain was used

577 as a donor in a mating assay to transfer the plasmid into *V. cholerae* OYP6G08,

578 generating IME*Vch*USA3^{Cm}. Single-copy integration of the pSW23T derivative

579 was confirmed by PCR and antibiotic resistance profiling.

580 IME*Vch*USA3^{Kn} was constructed from IME*Vch*USA3^{Cm}. Briefly, pVCR94^{Kn} Δ*acr2*

581 was transferred from the DAP-auxotrophic *E. coli* KH40 into OYP6G08 bearing

582 IME*Vch*USA3^{Cm}. After selection on LB agar medium supplemented with

583 chloramphenicol and kanamycin, Cm^R Kn^R *V. cholerae* OYP6G08

584 transconjugants were confirmed by growth on thiosulfate-citrate-bile salts-

585 sucrose (TCBS) agar medium (Difco). In *V. cholerae*, the integration and excision

586 of the IME were confirmed by amplification of the *attL*, *attR*, *attB*, and *attP* sites

587 with primer pairs oRD4/ORD6, oRD1/oRD3, oRD1/oRD6, and oRD4/oRD3,

588 respectively. IME*Vch*USA3^{Cm} was then mobilized from OYP6G08 to *E. coli*

589 CAG18439. In *E. coli*, the integration and excision of the IME were confirmed by

590 amplification of the *attL*, *attR*, *attB* and *attP* sites with primer pairs oRD4/ORD5,

591 oRD2/oRD3, oRD2/oRD5 and oRD4/oRD3, respectively. IMEVchUSA3^{Kn} was

592 constructed by replacing pSW23T with a single kanamycin resistance marker

593 using the one-step chromosomal gene inactivation technique with primer pair

594 dusAscarNoFRTf/dusAscarNoFRTr and pKD13 as the template. The deletions

595 Δ*sgaC* and Δ*prtN* in IME*Vch*USA3^{Kn} were obtained using the primer pairs

596 oFD26r/oFD26f and DelprtNr/DelprtNf, and pKD3 and pVI36 as the templates,

597 respectively. The Δ*dapA* deletion mutant of *E. coli* MG1655 was constructed

598 using primer pair FwDeltaDapA-MG1655/ RvDeltaDapA-MG1655 and pKD3 as

599 the template. The Δ*lacZ* mutation was introduced in *E. coli* CAG18439 using

600 primer pair lacZW-B/lacZW-F and plasmid pKD4 as the template. The *dusA'*-

601 *'lacZ* fusion was introduced in *E. coli* BW25113 and CAG18439 using primer pair

602 oDF15/oDF16 and pVI42B as the template. The fortieth codon of *dusA* was fused

603 to the eighth codon of *lacZ* downstream of the *attB* site. The λRed recombination

604 system was expressed using either pSIM6, pSIM9 or pKD46 [41,42]. When

605 appropriate, resistance cassettes were excised from the resulting constructions

606 using the Flp-encoding plasmid pCP20 [43]. All deletions were validated by

607 antibiotic profiling and PCR.

608 Fragments encompassing promoter regions upstream of *int*, *traN*, *traG*, *s018* and

609 *rdfN* were amplified using primer pairs oFD6.f/oFD6.r, oFD1.f/oFD1.r,

610 oFD3.f/oFD3.r, oFD5.f/oFD5.r and oFD4.f/oFD4.r, respectively, and genomic

611 DNA from *E. coli* CAG18439 *dusA*::IME*Vch*USA3^{Kn} as the template. The

612 amplicons were digested with PstI/XhoI and cloned into PstI/XhoI-digested

613 pOP*lacZ* [17]. The resulting constructs were single-copy integrated into the *attB$_\lambda$*

614 chromosomal site of *E. coli* BW25113 using pINT-ts [44]. To construct the

615 expression vectors pBAD-*rdfN* and pBAD-*sgaC*, PCR fragments containing *rdfN*

616 or *sgaC* were amplified from genomic DNA of *E. coli* CAG18439 bearing

617 IME*Vch*USA3 as the template and primer pairs prtNEcoRIf/prtNHindIIIrev and

618 oFD38r/oFD44f, respectively. The PCR fragments were digested by either EcoRI

619 or SacI, and HindIII and cloned into pBAD30 cut with the same enzymes.

620 mini-IE was constructed as follows. The 1,591-bp fragment of excised circular

621 IME*Vch*USA3^{Kn} that contains *attP-int* was amplified using primer pair

622 oVB12/oVB10 and genomic DNA from *E. coli* CAG18439 *dusA*::IME*Vch*USA3^{Kn}

623 as the template. The 1,421-bp fragment of pVI36 that contains *aadA7* was

624 amplified using primer pair oVB11/oVB13. Both fragments were joined using the

625 PCR-based overlap extension method [45]. After the final PCR amplification

28

626   using oVB12/oVB13, the amplicon was purified, digested with SacI, and ligated.

627   The ligation mixture was then transformed into *E. coli* EC100. Transformant

628   colonies were selected on LB agar supplemented with spectinomycin. The

629   constitutive expression of *int* and the absence of replicon prompted the

630   spontaneous integration of mini-IE at the 5' end of *dusA* in EC100.

631   All final constructs were verified by PCR and DNA sequencing by the Plateforme

632   de Séquençage et de Génotypage du Centre de Recherche du CHUL (Québec,

633   QC, Canada).

qPCR assays

635   qPCR assays for quantification of excision and copy number of IME*Vch*USA3^{Kn}

636   were carried out as described previously [22] with the following modification.

637   *attB$_{dusA}$* (241 bp) and *higA* (229 bp) of IME*Vch*USA3^{Kn} were quantified using

638   primer pairs attBdusAqPCRfwd/ attBdusAqPCRrev and higAqPCRfwd/

639   higAqPCRrev, respectively (S2 Table). The excision rate and copy number of

640   IME*Vch*USA3^{Kn} were calculated as the ratio of free *attB$_{dusA}$* site per chromosome

641   and as the ratio of *higA* per chromosome, respectively. The data were analyzed

642   and normalized using all three chromosomal genes *dnaB*, *hicB* and *trmE* as

643   references and the qBase framework as described previously [22,46].

β-galactosidase assays

645   The assays were carried out on LB agar plates supplemented with 5-bromo-4-

646   chloro-3-indolyl-β-D-galactopyranoside (X-gal) or in LB broth using *o*-nitrophenyl-

647   β-D-galactopyranoside (ONPG) as the substrate as described previously [32].

29

648   *acaDC* expression from pBAD-*acaDC* was induced by adding 0.2% arabinose to

649   a refreshed culture grown to an OD_{600} of 0.2, followed by a 2-h incubation at

650   37°C with shaking prior to cell sampling.

651   **Comparative analyses**

652   Sequences were obtained using blastp against the Genbank Refseq database

653   with the primary sequences of key proteins MpsA, TraG$_S$, SgaC, TraN$_S$ of SGI1

654   (Genbank AAK02039.1, AAK02037.1, AAK02036.1, AAK02035.1, respectively),

655   and Int$_{dusA}$ of IE*Vch*Bra2 (Genbank EEO15317.1) and Int$_{yicC}$ of IE*Eco*MOD1

656   (Genbank WP_069140142.1). Hits were exported, then sorted by accession

657   number to identify gene clusters that likely belong to complete IEs. Sequences of

658   IEs were manually extracted and the extremities were identified by searching for

659   the direct repeats contained in *attL* and *attR* sites. When an IE sequence

660   spanned across two contigs (e.g., IE*Vch*Hai10 and IE*Ppl*Ind1), the sequence was

661   manually assembled. IE sequences were clustered using cd-hit-est with a 0.95

662   nucleotide sequence identity cut-off [47]. Some of the annotated sequences were

663   manually curated to correct missing small open reading frames such as *mpsB*,

664   and inconsistent start codons. Pairwise comparisons of Int, MpsA, TraG, SgaC

665   and TraN proteins were generated with blastp using sets of representative

666   proteins selected after clustering using cd-hit with a 0.95 sequence identity cut-

667   off (Int, MpsA, TraG, SgaC) or a 0.90 sequence identity cut-off (TraN) [47].

668   Heatmaps showing the blastp identity scores were drawn using the Python library

669   seaborn v0.11.1 [48]. Circular blast representations (blast atlases) were

670   generated with the Blast Ring Image Generator (BRIG) 0.95 [49], with blastn or

30

671   blastp, against SGI1ΔIn104 and IE*Vch*USA2, with an upper identity threshold of

672   80% and a lower identity threshold of 60%. Antibiotic resistance gene prediction

673   was conducted using the Resistance Gene Identifier (RGI) software and CARD

674   3.1.3 database [50]. AcaCD binding motifs were identified using FIMO and MAST

675   [51] with the AcaCD motif matrix (S1 Matrix) described previously [17]. Logos for

676   *attL* and *attR* repeats were generated with MAST [51] using alignments of

677   sequences flanking the IE$_{dusA}$ elements identified in this work.

678   **Phylogenetic analyses**

679   Evolutionary analyses were conducted in MEGA X [52] and inferred by using the

680   maximum likelihood method based on the JTT (MpsA or SgaC proteins), LG

681   (Int$_{dusA}$, Int$_{yicC}$, TraG or RepA$_{IncFII}$ proteins) or WAG (TraN) matrix-based models

682   [53–55]. Protein sequences were aligned with Muscle [56]. Aligned sequences

683   were trimmed using trimal v1.2 using the automated heuristic approach [57].

684   Initial tree(s) for the heuristic search were obtained automatically by applying

685   Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated

686   using a JTT model, and then selecting the topology with the superior log

687   likelihood value. A discrete Gamma distribution was used to model evolutionary

688   rate differences among sites (5 categories) for Int$_{dusA}$ (parameter = 3.5633), Int$_{yicC}$

689   (parameter = 2.6652), SgaC (parameter = 1.4064), TraG (parameter = 1.9005)

690   and TraN (parameter = 1.6476) proteins. For Int$_{dusA}$, MpsA and TraG, the rate

691   variation model allowed for some sites to be evolutionarily invariable ([+I], 7.81%

692   sites for Int$_{dusA}$, 44.62% sites for MpsA and 5.22% sites for TraG$_S$). The trees are

693   all drawn to scale, with branch lengths measured in the number of substitutions

31

694   per site. In all trees, bootstrap supports are shown as percentages at the

695   branching points only when > 80%.

696   *oriT* sequences were obtained manually using the previously identified *oriT* of

697   SGI1 as the reference [19], then clustered using cd-hit-est with a 1.0 nucleotide

698   sequence identity cut-off. Sequences were then aligned using Muscle and a

699   NeighborNet phylogenetic network was built using SplitsTree4 [58] with default

700   parameters (Uncorrected_P method for distances and EqualAngle drawing

701   method). The secondary structures of the aligned *oriT* sequences were predicted

702   using RNAalifold v2.4.17 from the ViennaRNA package [59]. Default options

703   were used (including no RIBOSUM scoring), except for the following: no

704   substituting "T" for "U" (--noconv), no lonely pairs (--noLP), no GU pairs (--noGU)

705   and DNA parameters (-P DNA). The predicted Vienna output and the annotated

706   alignment were merged into a predicted secondary structure of SGI1 *oriT* color-

707   coded to display the inter-island diversity.

708   **Statistical analyses and figures preparation**

709   Numerical data presented in graphs are available in S3 Dataset. Prism 8

710   (GraphPad Software) was used to plot graphics and to carry out statistical

711   analyses. All figures were prepared using Inkscape 1.0 (https://inkscape.org/).

712   **Acknowledgments**

713   We are grateful to Yann Boucher for the kind gift of *Vibrio cholerae* OYP6G08

714   and Kévin T. Huguet for technical assistance. We thank Nicolas Rivard and

715   David Roy for their insightful comments on the manuscript.

32

References

1. Bellanger X, Payot S, Leblond-Bourget N, Guédon G. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol Rev. 2014;38: 720–760. doi:10.1111/1574-6976.12058

2. Guédon G, Libante V, Coluzzi C, Payot S, Leblond-Bourget N. The Obscure World of Integrative and Mobilizable Elements, Highly Widespread Elements that Pirate Bacterial Conjugative Systems. Genes. 2017;8: 337. doi:10.3390/genes8110337

3. Daccord A, Ceccarelli D, Burrus V. Integrating conjugative elements of the SXT/R391 family trigger the excision and drive the mobilization of a new class of *Vibrio* genomic islands. Mol Microbiol. 2010;78: 576–588. doi:10.1111/j.1365-2958.2010.07364.x

4. Daccord A, Ceccarelli D, Rodrigue S, Burrus V. Comparative analysis of mobilizable genomic islands. J Bacteriol. 2013;195: 606–614. doi:10.1128/JB.01985-12

5. Waldor MK. Mobilizable genomic islands: going mobile with *oriT* mimicry. Mol Microbiol. 2010;78: 537–540. doi:10.1111/j.1365-2958.2010.07365.x

6. Carraro N, Rivard N, Ceccarelli D, Colwell RR, Burrus V. IncA/C Conjugative Plasmids Mobilize a New Family of Multidrug Resistance Islands in Clinical *Vibrio cholerae* Non-O1/Non-O139 Isolates from Haiti. mBio. 2016;7: e00509-16. doi:10.1128/mBio.00509-16

7. de Curraize C, Siebor E, Neuwirth C. Genomic islands related to *Salmonella* genomic island 1; integrative mobilisable elements in *trmE* mobilised in trans by A/C plasmids. Plasmid. 2021;114: 102565. doi:10.1016/j.plasmid.2021.102565

8. Mulvey MR, Boyd DA, Olson AB, Doublet B, Cloeckaert A. The genetics of *Salmonella* genomic island 1. Microbes and Infection. 2006;8: 1915–1922. doi:10.1016/j.micinf.2005.12.028

9. Boyd D, Peters GA, Cloeckaert A, Boumedine KS, Chaslus-Dancla E, Imberechts H, et al. Complete nucleotide sequence of a 43-kilobase genomic island associated with the multidrug resistance region of *Salmonella enterica* serovar Typhimurium DT104 and its identification in phage type DT120 and serovar Agona. J Bacteriol. 2001;183: 5725–5732. doi:10.1128/JB.183.19.5725-5732.2001

10. Grim CJ, Hasan NA, Taviani E, Haley B, Chun J, Brettin TS, et al. Genome sequence of hybrid *Vibrio cholerae* O1 MJ-1236, B-33, and CIRS101 and

752   comparative genomics with *V. cholerae*. J Bacteriol. 2010;192: 3524–3533.
753   doi:10.1128/JB.00040-10

754 11. Cummins ML, Hamidian M, Djordjevic SP. *Salmonella* Genomic Island 1 is
755   Broadly Disseminated within Gammaproteobacteriaceae. Microorganisms.
756   2020;8: 161. doi:10.3390/microorganisms8020161

757 12. Hall RM. *Salmonella* genomic islands and antibiotic resistance in *Salmonella*
758   *enterica*. Future Microbiol. 2010;5: 1525–1538. doi:10.2217/fmb.10.122

759 13. Doublet B, Boyd D, Mulvey MR, Cloeckaert A. The *Salmonella* genomic
760   island 1 is an integrative mobilizable element. Mol Microbiol. 2005;55: 1911–
761   1924. doi:10.1111/j.1365-2958.2005.04520.x

762 14. Douard G, Praud K, Cloeckaert A, Doublet B. The *Salmonella* genomic
763   island 1 is specifically mobilized in trans by the IncA/C multidrug resistance
764   plasmid family. PLoS ONE. 2010;5: e15302.
765   doi:10.1371/journal.pone.0015302

766 15. Harmer CJ, Hall RM. The A to Z of A/C plasmids. Plasmid. 2015;80: 63–82.
767   doi:10.1016/j.plasmid.2015.04.003

768 16. Wu W, Feng Y, Tang G, Qiao F, McNally A, Zong Z. NDM Metallo-β-
769   Lactamases and Their Bacterial Producers in Health Care Settings. Clin
770   Microbiol Rev. 2019;32: e00115-18. doi:10.1128/CMR.00115-18

771 17. Carraro N, Matteau D, Luo P, Rodrigue S, Burrus V. The master activator of
772   IncA/C conjugative plasmids stimulates genomic islands and multidrug
773   resistance dissemination. PLoS Genet. 2014;10: e1004714.
774   doi:10.1371/journal.pgen.1004714

775 18. Kiss J, Papp PP, Szabó M, Farkas T, Murányi G, Szakállas E, et al. The
776   master regulator of IncA/C plasmids is recognized by the *Salmonella*
777   Genomic island SGI1 as a signal for excision and conjugal transfer. Nucleic
778   Acids Res. 2015;43: 8735–8745. doi:10.1093/nar/gkv758

779 19. Kiss J, Szabó M, Hegyi A, Douard G, Praud K, Nagy I, et al. Identification
780   and Characterization of *oriT* and Two Mobilization Genes Required for
781   Conjugative Transfer of *Salmonella* Genomic Island 1. Front Microbiol.
782   2019;10: 457. doi:10.3389/fmicb.2019.00457

783 20. Huguet KT, Rivard N, Garneau D, Palanee J, Burrus V. Replication of the
784   *Salmonella* Genomic Island 1 (SGI1) triggered by helper IncC conjugative
785   plasmids promotes incompatibility and plasmid loss. PLoS Genet. 2020;16:
786   e1008965. doi:10.1371/journal.pgen.1008965

34

787 21. Szabó M, Murányi G, Kiss J. IncC helper dependent plasmid-like replication
788   of Salmonella Genomic Island 1. Nucleic Acids Research. 2021;49: 832–
789   846. doi:10.1093/nar/gkaa1257

790 22. Durand R, Huguet KT, Rivard N, Carraro N, Rodrigue S, Burrus V. Crucial
791   role of *Salmonella* genomic island 1 master activator in the parasitism of
792   IncC plasmids. Nucleic Acids Res. 2021; in press. doi:10.1093/nar/gkab204

793 23. Harmer CJ, Hamidian M, Ambrose SJ, Hall RM. Destabilization of IncA and
794   IncC plasmids by SGI1 and SGI2 type *Salmonella* genomic islands.
795   Plasmid. 2016;87–88: 51–57. doi:10.1016/j.plasmid.2016.09.003

796 24. Huguet KT, Gonnet M, Doublet B, Cloeckaert A. A toxin antitoxin system
797   promotes the maintenance of the IncA/C-mobilizable *Salmonella* Genomic
798   Island 1. Sci Rep. 2016;6: 32285. doi:10.1038/srep32285

799 25. Carraro N, Durand R, Rivard N, Anquetil C, Barrette C, Humbert M, et al.
800   *Salmonella* genomic island 1 (SGI1) reshapes the mating apparatus of IncC
801   conjugative plasmids to promote self-propagation. PLOS Genetics. 2017;13:
802   e1006705. doi:10.1371/journal.pgen.1006705

803 26. Humbert M, Huguet KT, Coulombe F, Burrus V. Entry Exclusion of
804   Conjugative Plasmids of the IncA, IncC, and Related Untyped Incompatibility
805   Groups. J Bacteriol. 2019;201: e00731-18. doi:10.1128/JB.00731-18

806 27. Roberts AP, Johanesen PA, Lyras D, Mullany P, Rood JI. Comparison of
807   Tn*5397* from Clostridium difficile, Tn*916* from *Enterococcus faecalis* and the
808   CW459*tet*(M) element from *Clostridium perfringens* shows that they have
809   similar conjugation regions but different insertion and excision modules.
810   Microbiology. 2001;147: 1243–1251. doi:10.1099/00221287-147-5-1243

811 28. Burrus V, Pavlovic G, Decaris B, Guédon G. The ICE*St1* element of
812   *Streptococcus thermophilus* belongs to a large family of integrative and
813   conjugative elements that exchange modules and change their specificity of
814   integration. Plasmid. 2002;48: 77–97. doi:10.1016/s0147-619x(02)00102-6

815 29. Bioteau A, Durand R, Burrus V. Redefinition and unification of the SXT/R391
816   Family of integrative and conjugative elements. Appl Environ Microbiol.
817   2018;84: pii: e00485-18. doi:10.1128/AEM.00485-18

818 30. Farrugia DN, Elbourne LDH, Mabbutt BC, Paulsen IT. A novel family of
819   integrases associated with prophages and genomic islands integrated within
820   the tRNA-dihydrouridine synthase A (*dusA*) gene. Nucleic Acids Res.
821   2015;43: 4547–4557. doi:10.1093/nar/gkv337

822 31. Daccord A, Mursell M, Poulin-Laprade D, Burrus V. Dynamics of the SetCD-
823   regulated integration and excision of genomic islands mobilized by

35

824   integrating conjugative elements of the SXT/R391 family. J Bacteriol.
825   2012;194: 5794–5802. doi:10.1128/JB.01093-12

826 32. Rivard N, Colwell RR, Burrus V. Antibiotic Resistance in *Vibrio cholerae*:
827   Mechanistic Insights from IncC Plasmid-Mediated Dissemination of a Novel
828   Family of Genomic Islands Inserted at *trmE*. mSphere. 2020;5: e00748-20.
829   doi:10.1128/mSphere.00748-20

830 33. Thorsted PB, Shah DS, Macartney D, Kostelidou K, Thomas CM.
831   Conservation of the genetic switch between replication and transfer genes of
832   IncP plasmids but divergence of the replication functions which are major
833   host-range determinants. Plasmid. 1996;36: 95–111.
834   doi:10.1006/plas.1996.0037

835 34. Roy D, Huguet KT, Grenier F, Burrus V. IncC conjugative plasmids and
836   SXT/R391 elements repair double-strand breaks caused by CRISPR-Cas
837   during conjugation. Nucleic Acids Res. 2020;48: 8815–8827.
838   doi:10.1093/nar/gkaa518

839 35. Carraro N, Sauvé M, Matteau D, Lauzon G, Rodrigue S, Burrus V.
840   Development of pVCR94ΔX from *Vibrio cholerae*, a prototype for studying
841   multidrug resistant IncA/C conjugative plasmids. Front Microbiol. 2014;5: 44.
842   doi:10.3389/fmicb.2014.00044

843 36. Doublet B, Golding GR, Mulvey MR, Cloeckaert A. Secondary Chromosomal
844   Attachment Site and Tandem Integration of the Mobilizable *Salmonella*
845   Genomic Island 1. Fairhead C, editor. PLoS ONE. 2008;3: e2060.
846   doi:10.1371/journal.pone.0002060

847 37. Hochhut B, Waldor MK. Site-specific integration of the conjugal *Vibrio
848   cholerae* SXT element into *prfC*. Mol Microbiol. 1999;32: 99–110.

849 38. de Curraize C, Siebor E, Neuwirth C, Hall RM. SGI0, a relative of
850   *Salmonella* genomic islands SGI1 and SGI2, lacking a class 1 integron,
851   found in *Proteus mirabilis*. Plasmid. 2019; 102453.
852   doi:10.1016/j.plasmid.2019.102453

853 39. Dower WJ, Miller JF, Ragsdale CW. High efficiency transformation of *E. coli*
854   by high voltage electroporation. Nucleic Acids Res. 1988;16: 6127–6145.

855 40. Demarre G, Guérout A-M, Matsumoto-Mashimo C, Rowe-Magnus DA,
856   Marlière P, Mazel D. A new family of mobilizable suicide plasmids based on
857   broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalpha)
858   conjugative machineries and their cognate *Escherichia coli* host strains. Res
859   Microbiol. 2005;156: 245–255. doi:10.1016/j.resmic.2004.09.007

860 41. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in
861   *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A.
862   2000;97: 6640–6645.

863 42. Datta S, Costantino N, Court DL. A set of recombineering plasmids for
864   gram-negative bacteria. Gene. 2006;379: 109–115.
865   doi:10.1016/j.gene.2006.04.018

866 43. Cherepanov PP, Wackernagel W. Gene disruption in *Escherichia coli*: TcR
867   and KmR cassettes with the option of Flp-catalyzed excision of the
868   antibiotic-resistance determinant. Gene. 1995;158: 9–14. doi:10.1016/0378-
869   1119(95)00193-a

870 44. Haldimann A, Wanner BL. Conditional-replication, integration, excision, and
871   retrieval plasmid-host systems for gene structure-function studies of
872   bacteria. J Bacteriol. 2001;183: 6384–6393. doi:10.1128/JB.183.21.6384-
873   6393.2001

874 45. Senanayake SD, Brian DA. Precise large deletions by the PCR-based
875   overlap extension method. Mol Biotechnol. 1995;4: 13.
876   doi:10.1007/BF02907467

877 46. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase
878   relative quantification framework and software for management and
879   automated analysis of real-time quantitative PCR data. Genome Biology.
880   2007;8: R19. doi:10.1186/gb-2007-8-2-r19

881 47. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large
882   sets of protein or nucleotide sequences. Bioinformatics. 2006;22: 1658–
883   1659. doi:10.1093/bioinformatics/btl158

884 48. Waskom M. seaborn: statistical data visualization. JOSS. 2021;6: 3021.
885   doi:10.21105/joss.03021

886 49. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image
887   Generator (BRIG): simple prokaryote genome comparisons. BMC
888   Genomics. 2011;12: 402. doi:10.1186/1471-2164-12-402

889 50. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A,
890   et al. CARD 2020: antibiotic resistome surveillance with the comprehensive
891   antibiotic resistance database. Nucleic Acids Res. 2020;48: D517–D525.
892   doi:10.1093/nar/gkz935

893 51. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME
894   SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:
895   W202-208. doi:10.1093/nar/gkp335

896   52.   Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular
897   Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol.
898   2018;35: 1547–1549. doi:10.1093/molbev/msy096

899   53.   Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data
900   matrices from protein sequences. Bioinformatics. 1992;8: 275–282.
901   doi:10.1093/bioinformatics/8.3.275

902   54.   Whelan S, Goldman N. A General Empirical Model of Protein Evolution
903   Derived from Multiple Protein Families Using a Maximum-Likelihood
904   Approach. Mol Biol Evol. 2001;18: 691–699.
905   doi:10.1093/oxfordjournals.molbev.a003851

906   55.   Le SQ, Gascuel O. An Improved General Amino Acid Replacement Matrix.
907   Mol Biol Evol. 2008;25: 1307–1320. doi:10.1093/molbev/msn067

908   56.   Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and
909   high throughput. Nucleic Acids Res. 2004;32: 1792–1797.
910   doi:10.1093/nar/gkh340

911   57.   Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for
912   automated alignment trimming in large-scale phylogenetic analyses.
913   Bioinformatics. 2009;25: 1972–1973. doi:10.1093/bioinformatics/btp348

914   58.   Huson DH, Bryant D. Application of Phylogenetic Networks in Evolutionary
915   Studies. Molecular Biology and Evolution. 2006;23: 254–267.
916   doi:10.1093/molbev/msj030

917   59.   Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C,
918   Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6: 26.
919   doi:10.1186/1748-7188-6-26

920   60.   Kirchberger PC, Orata FD, Nasreen T, Kauffman KM, Tarr CL, Case RJ, et
921   al. Culture-independent tracking of *Vibrio cholerae* lineages reveals complex
922   spatiotemporal dynamics in a natural population. Environ Microbiol. 2020;22:
923   4244–4256. doi:10.1111/1462-2920.14921

924   61.   Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ,
925   et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio
926   cholerae.* Nature. 2000;406: 477–483. doi:10.1038/35020000

927   62.   Singer M, Baker TA, Schnitzler G, Deischel SM, Goel M, Dove W, et al. A
928   collection of strains containing genetically linked alternating antibiotic
929   resistance elements for genetic mapping of *Escherichia coli.* Microbiol Rev.
930   1989;53: 1–24.

931   63.   Ceccarelli D, Daccord A, René M, Burrus V. Identification of the Origin of
932   Transfer (*oriT*) and a New Gene Required for Mobilization of the SXT/R391

933   Family of Integrating Conjugative Elements. J Bacteriol. 2008;190: 5328–
934   5338. doi:10.1128/JB.00150-08

935 64. Grenier F, Matteau D, Baby V, Rodrigue S. Complete genome sequence of
936   *Escherichia coli* BW25113. Genome Announc. 2014;2: pii: e01038-14.
937   doi:10.1128/genomeA.01038-14

938 65. Garriss G, Waldor MK, Burrus V. Mobile antibiotic resistance encoding
939   elements promote their own diversity. PLoS Genet. 2009;5: e1000775.
940   doi:10.1371/journal.pgen.1000775

941 66. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation,
942   and high-level expression by vectors containing the arabinose P_{BAD}
943   promoter. J Bacteriol. 1995;177: 4121–4130.

944

945 **Table 1. Main features of the IEs described in this study.**

IE Name	Organism[1]	Size (bp)	Target site	Genbank accession number
GI*Vch*Rus1	*V. cholerae* 1	30,204	*dusA*	NZ_SMZE01000022
IE*Vch*Aus1	*V. cholerae* A12JL36W30	27,410	*dusA*	NZ_VIOZ01000074
IE*Vch*USA5	*V. cholerae* OYP2C05	28,706	*dusA*	NZ_NMTM01000021
IME*Vch*USA3	*V. cholerae* OYP6G08	30,910	*dusA*	NZ_NMSY01000009
IE*Vch*A215[2]	*V. cholerae* A215 sv Inaba	29,933	*dusA*	CWPR01000020.1
IE*Vch*USA2	*V. cholerae* 692-79	29,931	*dusA*	MIPA01000024
IE*Vch*N2751[2]	*V. cholerae* N2751	30,018	*dusA*	NZ_VSGL01000012
IE*Vch*N2744[2]	*V. cholerae* N2744	30,134	*dusA*	NZ_VSGF01000021.1
IE*Vch*N2708	*V. cholerae* N2708	27,248	*dusA*	NZ_VSFQ01000013
IE*Vch*N2786	*V. cholerae* N2786	24,658	*dusA*	NZ_VSHP01000008
IE*Vch*N2817	*V. cholerae* N2817	28,717	*dusA*	NZ_VSIM01000004
IE*Vch*Chn1	*V. cholerae* N2787	27,195	*dusA*	NZ_VSHQ01000015
IE*Vch*Ban1	*V. cholerae* EM-1676-A	36,519	*dusA*	NZ_KB662834
IE*Vch*Hai10	*V. cholerae* 2012Env-2	37,162	*dusA*	NZ_JSTD01000059/60
IE*Vch*Swe1	*V. cholerae* 11116	28,100	*dusA*	NZ_MDYK01000006
IE*Vch*Bra2	*V. cholerae* TMA-21	28,230	*dusA*	ACHY01000008
IE*Vme*USA1	*V. metoecus* 07-2435	23,518	*dusA*	NZ_LCUE01000016
IE*Vpa*Chn1	*V. parahaemolyticus* GIMxtf41-2013.07	28,589	*dusA*	NZ_MRWJ01000014
IE*Vpa*Chn2	*V. parahaemolyticus* C2_8	25,944	*dusA*	NZ_NNLT01000047
IE*Vpa*Ban1a	*V. parahaemolyticus* NIHCB0757	29,418	*dusA*	AVPX01000004
IE*Vvu*USA1	*V. vulnificus* VA-WGS-18041	30,228	*dusA*	NZ_RBZL01000019
IE*Eco*MOD1	*E. coli* MOD1-EC5437	25,611	*yicC*	NZ_NLPO01000006
IE*Sen*USA1	*S. enterica* Kentucky ARS-CC8289	25,970	*yicC*	NZ_MCPS01000044
IE*Ppl*Ind1	*P. plantistimulans* L1E11	22,777	*yicC*	NZ_LAPT01000132/095
GI*Vch*O27-1	*V. cholerae* 10432-62	26,646	*trmE*	CP010812

946 [1] *V.*, *Vibrio*; *E.*, *Escherichia*; *S.*, *Salmonella*; *P.*, *Pokkaliibacter*

947 [2] Not represented in Fig 1 as nucleotide sequences contain gaps

948 **Table 2. Strains and elements used in this study.**

Strains or elements	Relevant genotype or phenotype[a]	Source or reference
V. cholerae		
OY6PG08	Environmental, Oyster Pond, MA, USA, August 2009	[60]
N16961	O1 El Tor	[61]
E. coli		
β2163	(F−) RP4-2-Tc::Mu ΔdapA::(erm-pir) (Kn Em)	[40]
CAG18439	MG1655 lacZU118 lacI42::Tn10 (Tc)	[62]
BW25113	F⁻ Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻, rph-1, Δ(rhaD-rhaB)568, hsdR514	[41]
EC100	F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ⁻ rpsL (Sm^R) nupG	Epicentre, Madison Wis.
KH40	MG1655 ΔdapA::cat (Cm)	This study
VB112	Rf-derivative of MG1655	[63]
GG56	Nx-derivative of BW25113	[35,64]
FD034	CAG18439 ΔlacZ dusA'-'lacZ-aad7 (Tc Sp)	This study
FD036	GG56 dusA'-'lacZ-aad7 (Nx Sp)	This study
Plasmids		
pKD3	Cm^R PCR template for one-step chromosomal gene inactivation (Cm)	[41]
pKD4	Kn^R PCR template for one-step chromosomal gene inactivation (Kn)	[41]
pKD13	Kn^R PCR template for one-step chromosomal gene inactivation (Kn)	[41]
pVI36	Sp^R PCR template for one-step chromosomal gene inactivation (Sp)	[63]
pVI42B	pVI36 BamHI::P_{lac}-lacZ (Sp)	[65]
pSW23T	pSW23::$oriT_{RP4}$; $oriV_{R6Kγ}$ (Cm)	[40]
pOPlacZ	pAH56 lacZ (Kn)	[17]
pBAD30	ori_{p15A} bla araC P_{BAD} (Ap)	[66]
pBAD-acaDC	pBAD30::acaDC (Ap)	[17]
pBAD-rdfN	pBAD30::rdfN (Ap)	This study
pBAD-sgaC	pBAD30::sgaC (Ap)	This study
pVCR94^{Kn} Δacr2	Δacr2 mutant of pVCR94^{Kn} (Su Kn)	[34]
pVCR94^{Sp}	Sp^R derivative of pVCR94 (Su Sp)	[17]
pVCR94^{Sp} ΔacaC	ΔacaC mutant of pVCR94^{Sp} (Su Sp)	[17]
pVCR94^{Sp} ΔacaDC	ΔacaDC mutant of pVCR94^{Sp} (Su Sp)	[17]
Integrative elements		
IMEVchUSA3		This study
IMEVchUSA3^{Cm}	IMEVchUSA3 CGT85_RS05425ΩpSW23T (Cm)	This study
IMEVchUSA3^{Kn}	Kn^R derivative of IMEVchUSA3 (Kn)	This study
IMEVchUSA3^{Kn} ΔsgaC	ΔsgaC mutant of IMEVchUSA3 (Kn)	This study
IMEVchUSA3^{Kn} ΔrdfN	ΔrdfN mutant of IMEVchUSA3 (Kn)	This study
mini-IE	attP-int-aad7 derived from IMEVchUSA3 (Sp)	This study

949 [a]Ap, ampicillin; Cm, chloramphenicol; Em, erythromycin, Kn, kanamycin; Nx, Nalidixic

950 acid; Rf, rifampin; Sm, streptomycin; Sp, spectinomycin; Su, sulfamethoxazole; Tc,

951 tetracycline; Tm, trimethoprim; ts, thermosensitive.

Supporting information

S1 Fig. Comparative sequence analysis of SGI1-like *dusA*-specific IEs. Blastn and blastp atlases using either SGI1ΔIn104 (A) or IE*Vch*USA2 (B) as the reference. Coding sequences appear on the outermost circle in blue for the positive strand and red for the negative strand, with the *oriT* depicted as a grey arc. All other sequences are represented only according to their homology with the reference, with full opacity corresponding to 100% identity and gaps indicating identity below 60%. The order of the IEs in the atlases is indicated according to the color keys shown in the inset of panel B.

S2 Fig. NeighborNet phylogenetic network and predicted secondary structure of 39 *oriT* loci of SGI1-like IEs. Each IE's integration site and type are annotated. The sequence of canonical SGI1 (Genbank AF261825.2) was used as a reference to show the predicted secondary structure of all *oriT* sequences. Pairs can be perfectly conserved, imperfectly conserved (1/39 not conserved), not conserved (> 1/39), or an A-T or G-C pair only. In the latter case, the sequence is not conserved, but the predicted local secondary structure is.

S3 Figure. Maximum likelihood phylogenetic analysis of key proteins of SGI1-related IEs. The trees for MpsA (A), TraG (B), SgaC (C) and TraN (D) proteins are drawn to scale, with branch lengths measured in the number of substitutions per site over 321, 1,145, 188, and 968 amino acid positions, respectively. For clarity, the lengths of the branches linking the two groups in panels A and C were artificially divided by 8 and 4, respectively. Taxa corresponding to IEs targeting *trmE* and *yicC* are shown by a light blue circle and a red circle, respectively. All other taxa correspond to *dusA*-specific IEs. Proteins accession numbers are provided in S1 Table and S2 Dataset.

42

975   **S4 Figure. Alignment of AcaCD-responsive promoters predicted in IEs targeting**

976   ***dusA*, *yicC* and *trmE*.** Promoter sequences are grouped based on the function of the

977   expressed genes as follows: (A) RDFs; (B) mating pair stabilization; (C) mating pair

978   formation and stabilization; (D) unknown. AcaCD binding sites are shown in green. Logo

979   sequences and *p*-values were generated by MAST [51]. Known transcription start sites

980   are shown in blue [17,22]. Predicted Shine-Dalgarno sequences are shown in pink. The

981   initiation start codon is shown in bold letters.

982   **S5 Figure. Excision of IME*Vch*USA3 is enhanced in IncC$^+$ cells.** (A) Model of

983   excision of IME*Vch*USA3. (B and C) Detection of *attB*, *attP*, *attL* and *attR* sites by PCR

984   in colonies of *V. cholerae* OYO6G08 bearing (lanes 9 to 16) or lacking (lanes 1 to 8)

985   pVCR94^{Kn} Δ*acr2.* Control lanes: L, 1Kb Plus DNA ladder (Transgen Biotech); +,

986   *V. cholerae* N16961 genomic DNA. (D) Detection of *attB*, *attP*, *attL* and *attR* sites by

987   PCR in transconjugant colonies of *E. coli* CAG18439 (lanes 1 to 4). L, 100bp Plus II

988   DNA Ladder (Transgen Biotech)

989   **S1 Table. Features of the identified IEs and associated strains.**

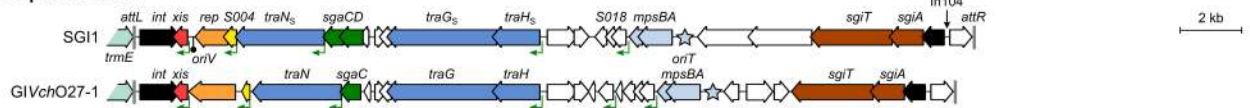990   **S2 Table. Oligonucleotides used in this study.**

991   **S1 Dataset. Features of ORFs in the identified IEs.**

992   **S2 Dataset. Clusters generated by cd-hit for Int, MpsA, TraG, SgaC and TraN.**
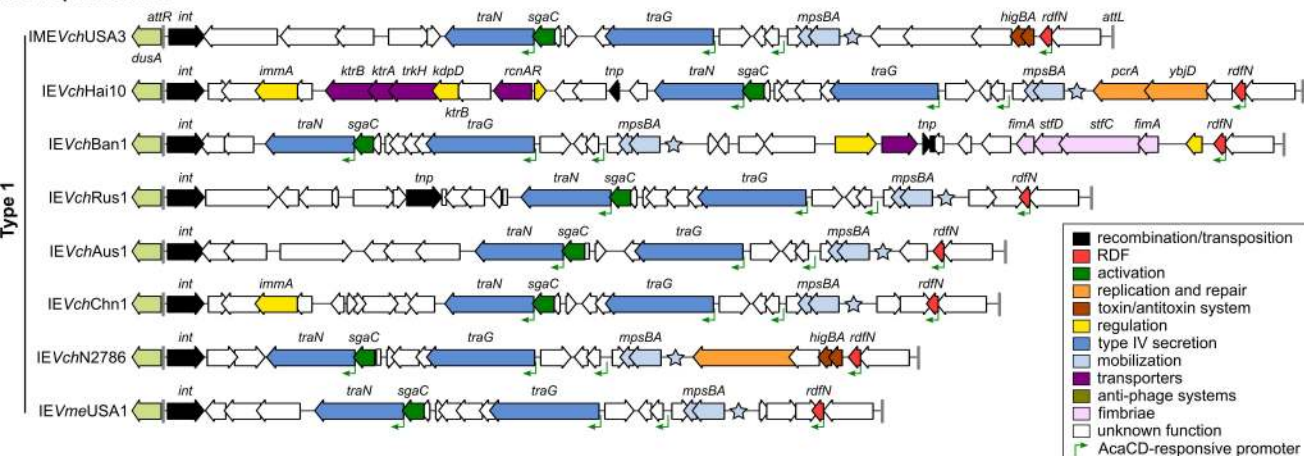
993   **S3 Dataset. Numerical data presented in Figs 4, 5 and 6.**

994   **S1 Matrix. AcaCD motif matrix to identify AcaCD binding sites.**
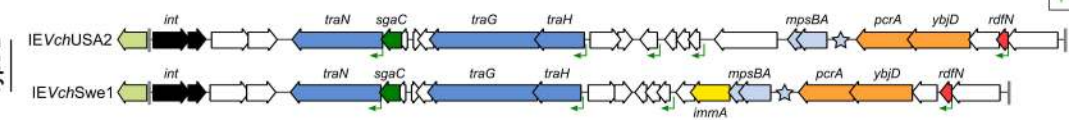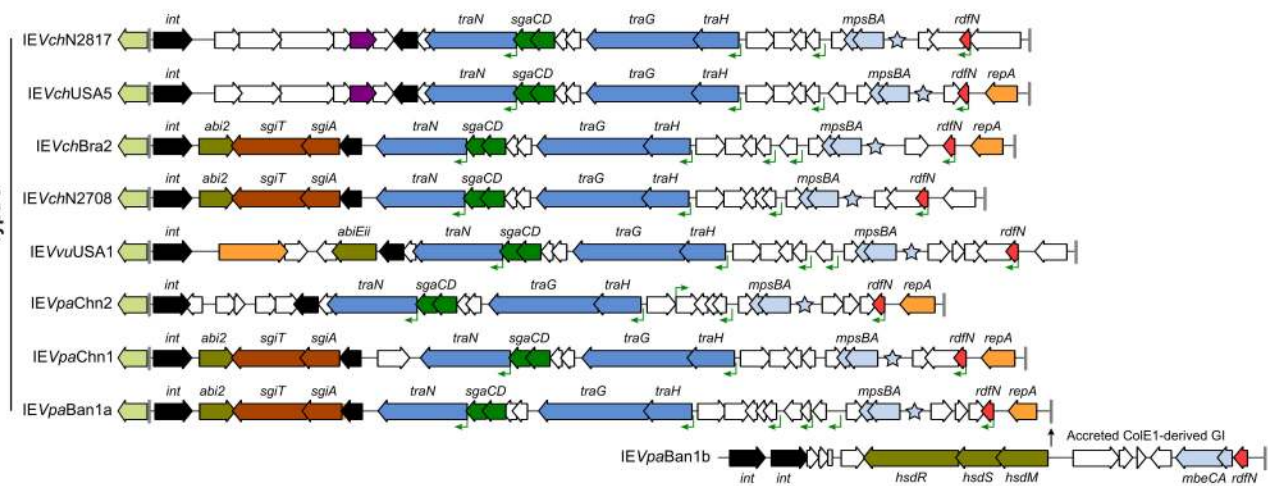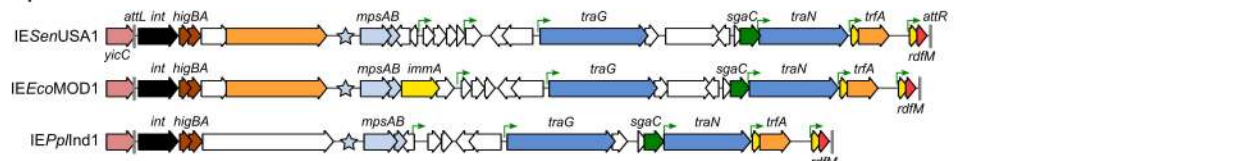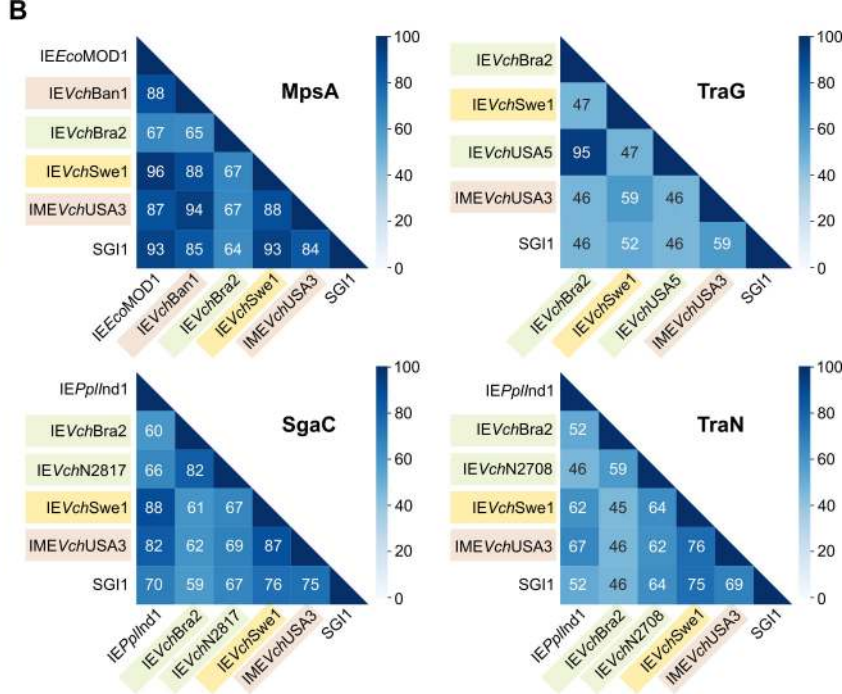
43

trmE-specific IMEs

dusA-specific IMEs

yicC-specific IMEs

A Int$_{yicC}$

B Int$_{dusA}$

C

A MpsA-TraG-SgaC-TraN

- $trmE$ (teal circle)
- $yicC$ (red circle)

Type 1, Type 2, Type 3 phylogenetic tree with the following labels:

Type 1:
IEVchChn1, IMEVchUSA3, IEVchN2786, IEVchBan1, IEVchAus1, IEVmeUSA1, IEVchRus1, IEVchHai10

Type 2:
IEVchUSA2, IEVchSwe1

GIVchO27-1, IESenUSA1, IEPpIInd1, IEEcoMOD1

GI-15, SGI1

Type 3:
IEVchN2708, IEVpaChn2, IEVpaBan1a, IEVchBra2, IEVpaChn1, IEVvuUSA1, IEVchN2817, IEVchUSA5

Scale: 0.10

B Heatmap matrices:

MpsA
	IEEcoMOD1	IEVchBan1	IEVchBra2	IEVchSwe1	IMEVchUSA3	SGI1
IEEcoMOD1						
IEVchBan1	88					
IEVchBra2	67	65				
IEVchSwe1	96	88	67			
IMEVchUSA3	87	94	67	88		
SGI1	93	85	64	93	84	

TraG
	IEVchBra2	IEVchSwe1	IEVchUSA5	IMEVchUSA3	SGI1
IEVchBra2					
IEVchSwe1	47				
IEVchUSA5	95	47			
IMEVchUSA3	46	59	46		
SGI1	46	52	46	59	

SgaC
	IEPpIInd1	IEVchBra2	IEVchN2817	IEVchSwe1	IMEVchUSA3	SGI1
IEPpIInd1						
IEVchBra2	60					
IEVchN2817	66	82				
IEVchSwe1	88	61	67			
IMEVchUSA3	82	62	69	87		
SGI1	70	59	67	76	75	

TraN
	IEPpIInd1	IEVchBra2	IEVchN2708	IEVchSwe1	IMEVchUSA3	SGI1
IEPpIInd1						
IEVchBra2	52					
IEVchN2708	46	59				
IEVchSwe1	62	45	64			
IMEVchUSA3	67	46	62	76		
SGI1	52	46	64	75	69	

A

| | P_{int} | P_{traN} | P_{trsG} | P_{s018} | P_{rdfN} |

Arabinose

-

+

B

A IME*Vch*USA3 excision rate

B IME*Vch*USA3 copy number

C

D

mini-IE

E

	pVCR94^{Sp}	IME*Vch*USA3^{Kn}	pBAD30

Panel A:

pVCR94^{Sp}	WT	-	-	WT	WT	WT	ΔC	ΔDC
IME*Vch*USA3^{Kn}	WT	WT	ΔrN	WT	ΔrN	ΔrN	WT	WT
pBAD30	-	-	-	-	-	rN	-	-

Panel B:

pVCR94^{Sp}	-	-	WT	WT	WT	ΔC	ΔDC
IME*Vch*USA3^{Kn}	WT	ΔrN	WT	ΔrN	ΔrN	WT	WT
pBAD30	-	-	-	-	rN	-	-

Panel C: Log$_{10}$(transconjugants/donor)

pVCR94^{Sp}, IME*Vch*USA3^{Kn}, cotransfer

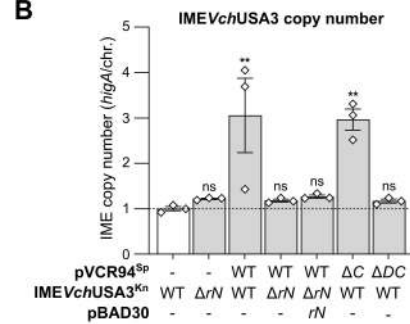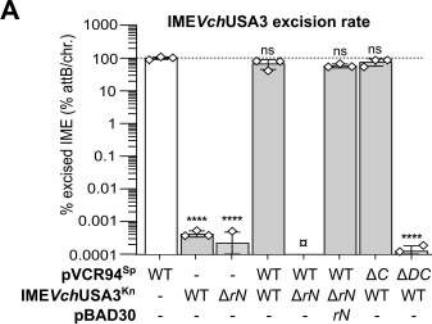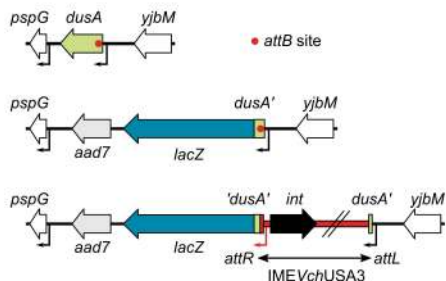| pVCR94^{Sp}	WT	$\Delta acaC$	$\Delta acaDC$	WT	$\Delta acaC$	WT	$\Delta acaC$	WT	WT
IME*Vch*USA3^{Kn}	WT	WT	WT	$\Delta sgaC$	$\Delta sgaC$	$\Delta sgaC$	$\Delta sgaC$	$\Delta rdfN$	$\Delta rdfN$
pBAD30	-	-	-	-	-	$sgaC$	$sgaC$	-	$rdfN$

A

E. coli K12 *dusA attB* site

```
TAATCTCTACATTTGAAA ATG CAC GGT AAT TCT GAA ATG CAA AAA ATC AAC CAA ACC AGC GCA ATG CCT GAA
                    M   H   G   N   S   E   M   Q   K   I   N   Q   T   S   A   M   P   E

AAA ACT GAC GTT CAC TGG AGT GGT CGG TTT AGC GTT GCA CCA ATG CTC GAC TGG A CG GAC AGA CAT
K   T   D   V   H   W   S   G   R   F   S   V   A   P   M   L   D   W   T   D   R   H
```

E. coli K12 *dusA*::IMEVchUSA3^Kn *attR* junction

```
AGTTGAGATTAGTTAAGA ATG AGC CAG AAA CAG AAT ACA GAT AAA TCT TTG CAA GAC AGT GAG TTA TCT AGT
                   M   S   Q   K   Q   N   T   D   K   S   L   Q   D   S   E   L   S   S

AAT ATT GAT GTG TCA TTA AGT CGT AGA TTC AGC TG GCC CCC ATG CTT GAC TGG A CG GAC AGA CAT
N   I   D   V   S   L   S   R   R   F   S   V   A   P   M   L   D   W   T   D   R   H
```

B

C