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

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

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

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

* Corresponding author

E-mail: vincent.burrus@usherbrooke.ca

Short title: *dusA*-specific genomic islands mobilizable by IncC conjugative 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 genomic islands (GIs) distantly related to SGI1 in several species of *Vibrio*, *Aeromonas*, *Salmonella*, *Pokkaliibacter*, and *Escherichia*. Unlike SGI1, these GIs target two alternative chromosomal loci, the 5' end of *dusA* and the 3' end of *yicC*. Although these elements share many features with SGI1, they lack antibiotic resistance genes and carry alternative integration/excision modules. Functional characterization of MGI*Vch*USA3, a *dusA*-specific GI, revealed promoters that respond to AcaCD, the master activator of IncC plasmid transfer genes. Quantitative PCR and mating assays confirmed that MGI*Vch*USA3 excises from the chromosome and is mobilized by an IncC helper plasmid from *Vibrio cholerae* to *Escherichia coli*. MGI*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 GIs from the chromosome. Like *xis* in SGI1, *rdfN* is under the control of an AcaCD-responsive promoter. Although the integration of MGI*Vch*USA3 disrupts *dusA*, the GI provides a new promoter

sequence and restores the reading frame of *dusA* for proper expression of the tRNA-dihydrouridine synthase A. Phylogenetic analysis of the conserved proteins encoded by SGI1-like elements targeting *dusA*, *yicC*, and *trmE* gives a fresh perspective on the possible origin of SGI1 and its variants.

Author summary

We identified genomic islands distantly related to the *Salmonella* Genomic Island 1 (SGI1), a key vector of antibiotic resistance genes in *Gammaproteobacteria*. SGI1 and its variants reside at the 3' end of *trmE*, share a large, highly conserved core of genes, and carry a complex integron that confers multidrug resistance phenotypes to their hosts. Unlike members of the SGI1 group, these novel genomic islands target the 5' end *dusA* or the 3' end of *yicC*, lack multidrug resistance genes, and seem much more diverse. We showed here that, like SGI1, these genomic islands are mobilized by conjugative plasmids of the IncC group. Based on comparative genomics and functional analyses, we propose a hypothetical model of the evolution of SGI1 and its siblings from the progenitor of IncA and IncC conjugative plasmids via an intermediate *dusA*-specific genomic island through gene losses and gain of alternative integration/excision modules.

Introduction

Integrative and mobilizable elements (IMEs) are discrete, mobile chromosomal regions, also known as genomic islands (GIs), 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 also 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 [1,6,7]. *Salmonella* Genomic Island 1 (SGI1) is arguably one of the most atypical and 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 [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

variants have been reported in a wide array of *Gammaproteobacteria*, including *Proteus mirabilis* (PGI1), *Acinetobacter baumannii* (AGI1), *Morganella*, *Providencia*, *Enterobacter*, and *Escherichia coli*, but also in *Vibrio cholerae* and *Klebsiella pneumoniae* [10–12]. Most variants carry a class I integron structurally similar to the In104 integron of SGI1. In104 confers resistance to ampicillin, chloramphenicol, streptomycin/spectinomycin, sulfamethoxazole, and tetracycline [13,14]. SGI1 and its variants are an epidemiological threat exacerbated by their specific mobilization by conjugative plasmids of the incompatibility groups A (IncA) and C (IncC) [15,16]. IncC plasmids contribute to the global circulation of multidrug resistance genes, including NDM metallo-β-lactamase and carbapenemase genes, among a broad range of *Gammaproteobacteria* [17,18]. The transcriptional activator AcaCD encoded by IncC plasmids triggers the excision and mobilization of SGI1 [19,20].

SGI1 and its variants share a conserved set of genes within which insertion sequences and the class 1 integron are inserted at diverse positions [21–24]. The SGI1 backbone is 27.4 kb and contains 28 open reading frames [9]. Members of the SGI1-HKL group seem to have a smaller backbone [23]. Thus far, the function of a few conserved genes has been characterized. Together with the *cis*-acting recombination site *attP,* the genes *int* and *xis* form the recombination module of SGI1 [15]. *int* encodes a site-specific tyrosine recombinase (integrase) that targets the 3' end of *trmE.* The recombination directionality factor (RDF or excisionase) encoded by *xis* enhances the excision reaction catalyzed by Int. The mobilization module includes the mobilization

104 genes *mpsAB* and the *oriT* located upstream of *mpsA* [25]. *mpsA* encodes a
105 tyrosine recombinase that acts as an atypical relaxase. Unlike most
106 characterized IMEs, SGI1 carries a replicon composed of an iteron-based origin
107 of replication (*oriV*) and replication initiator gene *rep* [26,27]. The transcriptional
108 activator complex SgaCD, expressed by SGI1 in response to a coresident IncC
109 plasmid, controls *rep* expression [27,28]. SGI1 also enhances its stability with the
110 help of the *sgiAT* addiction module [29]. At the same time, its replication
111 destabilizes the helper plasmid [26,28,30]. Finally, SGI1 encodes three mating
112 pore subunits, $TraN_S$, $TraH_S$, and $TraG_S$, that actively replace their counterparts
113 in the T4SS encoded by the IncC plasmid [31]. The substitution of TraG allows
114 SGI1 to bypass the IncC-encoded entry exclusion mechanism and transfer
115 between cells carrying conjugative plasmids belonging to the same entry
116 exclusion group [32].

117 Given the atypical functional features conserved in SGI1 variants integrated at
118 *trmE*, we undertook a search for distantly related elements by favoring the
119 retention of conserved functions rather than maximize sequence similarity with
120 SGI1. Using MpsA, $TraG_S$, SgaC, and $TraN_S$ as baits, we searched databases
121 for distant SGI1-like IMEs in bacterial genomes. We report the existence of
122 distantly related IMEs integrated at the 5' end of *dusA* in several species of
123 *Vibrionaceae* and the 3' end of *yicC* in several species of *Gammaproteobacteria*.
124 We have examined the interactions between an IncC plasmid and MGI*Vch*USA3,
125 a *dusA*-specific representative IME *dusA* from an environmental *V. cholerae*
126 strain. The genetic determinants required for the excision of MGI*Vch*USA3 and

its mobilization by IncC plasmids were characterized. Finally, we took a fresh look at the emergence and evolution of SGI1 and its siblings by conducting phylogenetic analyses and proposed a hypothetical evolutionary pathway of IMEs resembling SGI1.

Results

Novel genomic islands resembling SGI1 are inserted in *dusA* and *yicC* in various *Gammaproteobacteria*

To find novel SGI1-like elements, we searched the Refseq database using blastp and the primary sequences of MpsA, $TraG_S$, SgaC, and $TraN_S$. Considering the substitution of integration modules can change the integration site [33–36], the integrase Int_{trmE} was excluded from the analysis. We identified 24 distinct GIs encoding homologs of the four bait proteins in 36 different bacterial strains (Fig 1 and S1 Table). 21 of these GIs are integrated into the 5' end of *dusA* in diverse *Vibrio* species from various origins. The remaining three are located at the 3' end of *yicC* in *E. coli*, *Aeromonas veronii*, *P. mirabilis*, *S. enterica* serovar Kentucky, and *Pokkaliibacter plantistimulans*. The size of the GIs varies from 22.8 kb to 37.1 kb. The conserved genes *mpsA* (together with *mpsB*), *traG*, *traN*, and *sgaC* remain in a syntenic order, though sporadically separated by variable DNA (Fig 1).

Figure 1. Schematic representations of SGI1-related genomic islands. The position and orientation of open reading frames (ORFs) are indicated by arrowed boxes. Colors depict the function deduced from functional analyses and BLAST comparisons. AcaCD binding sites are represented by green angled arrows. Each island is flanked by the *attL* and *attR* (vertical grey lines) attachment sites when integrated into the 3' end of *trmE* (light blue), the 5' end of *dusA* (light green), or the 3' end of *yicC* (pink). Regions that are conserved between two adjacent GIs are depicted in light grey.

Consistent with the change of integration site, the respective *int* genes of SGI1 and the *dusA*- and *yicC*-specific GIs are unrelated. Furthermore, these novel GIs lack *xis* downstream of *int*. Instead, *yicC*-specific GIs carry two small open reading frames (ORF) upstream of the *attR* site. The putative translation product of the second one shares 35% identity over 65% coverage with the excisionase RdfM of MGI*Vfl*Ind1 [37]. Although *dusA*-specific GIs lack *xis* and *rdfM*, all carry an ORF predicted to encode a 76-aminoacyl residue protein containing the pyocin activator protein PrtN domain (Pfam PF11112). Based on its size, position, predicted DNA-binding function, conservation, and evidence presented below, we named this ORF *rdfN*.

None of the reported GIs carries the same replication module (*S004-rep-oriV*) as canonical SGI1. Instead, five *dusA*-specific GIs encode a putative replication initiator protein with the IncFII_repA domain (Pfam PF02387) (Fig 1). The three *yicC*-specific GIs encode a homolog of TrfA (Pfam PF07042), the replication initiator protein of broad-host-range IncP plasmids [38]. No replicative functions could be ascribed with confidence to any gene carried by the other *dusA*-specific GIs.

Although all these GIs seem to lack antibiotic resistance genes, several encode putative functions altering host processes and virulence, including the transport of ions and small molecules, chemotaxis, c-di-GMP signalling, and fimbriae. Nine GIs also encode toxin-antitoxin systems, such as *sgiAT* and *higAB*, which likely enhance their stability (Fig 1).

Finally, two accreted and distinct GIs compose GI*Vch*Ban1: an SGI1-like GI and a second unrelated GI coding for two predicted integrases sharing 44% and 27% identity with Int_{dusA} that targets *dusA*. This second GI encodes a putative type I restriction-modification system, a MobA-like relaxase (MOB_{P1}), the mobilization auxiliary factor MobC, and an RdfN homolog (Fig 1).

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

Blastn and blastp analyses using SGI1ΔIn104 and GI*Vch*USA2 as references confirmed that the identified *dusA*-specific GIs share extensive similarities (Fig 2). Besides the conserved genes encoding MpsA, TraG, SgaC, and TraN, all carry the auxiliary mobilization factor gene *mpsB* and the *oriT* sequence (Fig 2A). Secondary structure prediction of the aligned *oriT* sequences located upstream of *mpsA* using RNAalifold revealed that despite the sequence divergence, the structure of *oriT* with three stem-loops was strictly conserved (S1B Fig). In contrast, *sgaD* is not strictly conserved and highly divergent from *sgaD* of SGI1 when present. Comparison using GI*Vch*USA2 as the reference revealed that *dusA*-specific GIs are distant from SGI1 and cluster into three distinct types as confirmed by the phylogenetic analysis of concatenated MpsA-TraG-SgaC-TraN and Int_{dusA} proteins (Fig 2B, 3A and 3C, and S2 Fig). Type 1 *dusA*-specific GIs such as MGI*Vch*USA3 are mainly found in *V. cholerae* and lack both *traH* and *sgaD*. Type 2 GIs such as GI*Vch*USA2 lack *sgaD* but carry *traH* and are closely related to *yicC*-specific GIs. Finally, type 3 GIs such as GI*Vch*USA5 are the most distant from the two other types and SGI1 (Fig 3A). Type 3 GIs carry both *traH* and *sgaD* and reside in diverse *Vibrio* species.

Figure 2. Comparative sequence analysis of SGI1$_{dusA}$ islands. Blastn and Blastp atlases were constructed using the Blast Ring Image Generator (BRIG) 0.95 [39], using either SGI1ΔIn104 (A) or GI*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 origin of transfer 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 GIs in the atlases in indicated according to the color keys shown in the inset of panel B.

Figure 3. Maximum likelihood phylogenetic analysis of MpsA-TraG-SgaC-TraN (A), Int$_{yicC}$ (B), and Int$_{dusA}$ (C) encoded by SGI1-related GIs. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site over 2,637, 400, and 359 amino acid positions for concatenated MpsA-TraG-SgaC-TraN, Int$_{yicC}$, and Int$_{dusA}$, respectively. Bootstrap supports are indicated as percentages at the branching points only when > 80%. Taxa corresponding to GIs targeting *trmE* and *yicC* are accompanied by a light blue circle and a red circle, respectively. All other taxa correspond to *dusA*-specific GIs. The helper elements and mechanism of mobilization are indicated for each lineage according to the keys shown in the legend box of panel A. The inset of panel C shows logo sequences of *attL* and *attR* attachment sites generated with WebLogo [40], using alignments of sequences flanking the SGI$_{dusA}$ elements presented in Fig 1. The arrows indicate the island termini experimentally determined for the GI D1279779_RGP05 of *A. baumannii* D1279779 by Farrugia

et al. [41]. Phylogenetic relationships of MpsA, TraG, SgaC and TraN proteins are shown separately in S2 Fig.

Phylogenetic analysis of Int_{dusA} proteins confirmed that these GIs are exclusive to the *Vibrionaceae* and distinct from other *dusA*-specific GIs found in other taxa, including GI*Aca*Bra1 that encodes a MobI-like protein and is likely mobilizable by IncC plasmids [42] (Fig 3C). Sequence logos built using alignments of the *attL* and *attR* chromosomal junctions revealed a 21-bp imperfect repeat at the extremities of each GI (Fig 3C). This repeat is similar to the one reported for *dusA*-specific GIs found in a broader range of species [41].

Finally, Int_{yicC} proteins of *yicC*-specific GIs form a cluster distinct from the integrases of GIs mobilizable by IncC plasmids through a MobI protein, and GIs that mimic the *oriT* of SXT/R391 ICEs [3,19,42] (Fig 3B).

Non-canonical SGI1-like GIs carry AcaCD-responsive genes

Considering the divergence of the 24 new GIs from prototypical SGI1, we wondered whether an IncC plasmid could mobilize them like SGI1. The hallmark of IncC-dependent mobilization is the presence of AcaCD-responsive promoters in IncC-mobilizable GIs. Hence, we searched for putative AcaCD-binding sites in the sequences of *trmE*-specific GIs (prototypical SGI1 was used as the positive control) and the *yicC*- and *dusA*-specific GIs. In these GIs, an AcaCD-binding motif was predicted upstream of *traN*, *traHG* (or *traG*), *S018*, and *xis* (or *rdfM* or *rdfN*) (Fig 1 and S3 Fig). Moreover, an AcaCD-binding motif was also predicted upstream of *trfA* in the *yicC*-specific GIs.

We cloned the promoter sequences of *int*, *traN*, *traG*, *S018*, and *rdfN* of MGI*Vch*USA3 upstream of a promoterless *lacZ* reporter gene and monitored the β-galactosidase activity with or without AcaCD. The promoter P_{int} was active regardless of the presence of AcaCD (Fig 4A). In contrast, the four other promoters exhibited weak activity in the absence of AcaCD. Upon induction of *acaDC* expression, P_{traN} and P_{S018} remained unresponsive, while the activities of P_{traG} and P_{rdfN} increased 40 and 400 times, respectively (Fig 4B). The inertia of P_{traN} and P_{S018} toward AcaCD could result from single nucleotide substitutions in the AcaCD binding site previously shown to be essential for recruiting the activator [28]: CCS**A**AAWW instead of CCS**C**AAWW in P_{traN} and CCC**C**AAAA instead of CCC**A**AAAA in P_{S018} (S3 Fig).

Figure 4. β-galactosidase activities of the promoters P_{int}, P_{traN}, P_{traG}, P_{S018} and P_{rdfN} of MGI*Vch*USA3 transcriptionally fused to *lacZ*. (A) Colonies were grown on LB agar with or without arabinose to induce *acaDC* expression from pBAD-*acaDC*. (B) Induction levels of the same promoters in response to AcaCD. β-galactosidase assays were carried out using the strains of panel A. Ratios between the enzymatic activities in Miller units for the arabinose-induced versus non-induced strains containing pBAD-*acaDC* are shown. The bars represent the mean and standard error of the mean of three independent experiments.

Hence, despite their divergence and different integration sites, these GIs share with SGI1 a common activation mechanism elicited by the presence of an IncC plasmid.

IncC plasmids induce the excision and mobilization of MGI*Vch*USA3

Next, we tested whether a coresident IncC plasmid could trigger the excision of MGI*Vch*USA3 from *dusA* in *V. cholerae* OY6PG08. The derepressed IncC plasmid pVCR94^{Kn} Δ*acr2* [43] was transferred by conjugation from *E. coli* KH40 into OY6PG08. OY6PG08 Kn^{R} transconjugants were isolated and tested by PCR to amplify the *attL* and *attR* chromosomal junctions, as well as the *attB* and *attP* sites resulting from the excision of MGI*Vch*USA3 (Fig 5A). Only three out of the eight tested transconjugants retained MGI*Vch*USA3, suggesting it was highly unstable and rapidly lost in IncC^{+} OY6PG08 (Fig 5B). In contrast, the eight control IncC-free OY6PG08 clones yielded strong *attL* and *attR* signals but weaker *attB* and *attP* signals, indicating that MGI*Vch*USA3 spontaneously excised (Fig 5C).

Figure 5. Excision of MGI*Vch*USA3 is enhanced in IncC^{+} cells. (A) Model of excision of MGI*Vch*USA3. (B and C) Detection of *attB*, *attP*, *attL* and *attR* sites by PCR in colonies of *V. cholerae* OYO6G08 bearing (lanes 9 to 16) or lacking (lanes 1 to 8) pVCR94^{Kn} Δ*acr2.* Control lanes: L, 1Kb Plus DNA ladder (Transgen Biotech); +, *V. cholerae* N16961 genomic DNA. (D) Detection of *attB*, *attP*, *attL* and *attR* sites by PCR in transconjugant colonies of *E. coli* CAG18439 (lanes 1 to 4). L, 100bp Plus II DNA Ladder (Transgen Biotech)

To test the mobilization of MGI*Vch*USA3 from *V. cholerae* OY6PG08, we inserted a chloramphenicol resistance marker upstream of *traG*, generating MGI*Vch*USA3^{Cm}, and used pVCR94^{Kn} Δ*acr2* as the helper plasmid. MGI*Vch*USA3^{Cm} readily transferred to *E. coli* CAG18439 used as the recipient

289 strain (7.01 × 10^{-5} transconjugant/donor). Amplification of the *attL* and *attR*
290 junctions using *E. coli*-specific primers confirmed that MGI*Vch*USA3 integrates at
291 *dusA* in *E. coli* (Fig 5D). Again, weak *attB* signal suggests spontaneous excision
292 of MGI*Vch*USA3, whereas the stronger *attP* signals suggests the transient
293 formation of tandem arrays or replication of the element in a subset of cells.

294 **Excision of *dusA*-specific GIs depends on *rdfN***

295 To further characterize the biology of MGI*Vch*USA3, we measured its excision
296 rate and copy number by qPCR, with and without coresident pVCR94^{Sp}. We also
297 monitored its transfer in the same context. Spontaneous excision of the island
298 rarely occurred (<0.001% of the cells) (Fig 6A). In contrast, in the presence of the
299 helper plasmid, the free *attB* site was detected in more than 67% of the cells
300 confirming that the IncC plasmid elicits the excision of MGI*Vch*USA3^{Kn}. Likewise,
301 the presence of the plasmid resulted in a ~3-fold increase of the copy number of
302 MGI*Vch*USA3^{Kn} (Fig 6B), suggesting that the excised form of the island
303 undergoes replication. The frequency of transfer of MGI*Vch*USA3^{Kn} was
304 comparable to that of the helper plasmid (~3.5×10^{-2} transconjugants/donor),
305 while the frequency of cotransfer was more than two logs lower (Fig 6C).

306 **Figure 6. Effect of *acaDC* and *rdfN* on the IncC-dependent excision and**
307 **mobilization of MGI*Vch*USA3.** (A) MGI*Vch*USA3^{Kn} excision rate corresponds to
308 the *attB*/chromosome ratio. (B) MGI*Vch*USA3^{Kn} copy number corresponds to the
309 *higA*/chromosome ratio. For panels A and B, all targets were amplified by qPCR
310 alongside the three reference genes *trmE*, *hicB* and *dnaB*. All ratios were
311 normalized using the control set to 1 and displayed in white. (C) Impact of *acaC*,

acaDC, *sgaC* and *rdfN* deletions on the mobilization of MGI*Vch*USA3. Conjugation assays were performed with CAG18439 (Tc) containing the specified elements as donor strains and VB112 (Rf) as the recipient strain. The bars represent the mean and standard error of the mean obtained from a biological triplicate. ¤ indicates that the excision rate or transfer frequency was below the detection limit. Statistical analyses were performed (on the logarithm of the values for panels A and C) using a one-way ANOVA with Dunnett's multiple comparison test. For panels A and C, statistical significance indicates comparisons to the normalization control. Statistical significance is indicated as follows: ****, $P < 0.0001$; ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$; ns, not significant. (D) Schematic representation of mini-GI inserted at the 5' end of *dusA*. (E) RdfN acts as a recombination directionality factor. Detection of *attB*, *attP*, *attL* and *attR* sites by PCR in colonies of *E. coli* EC100 *dusA*::mini-GI in the presence or absence of *rdfN*. L, 1Kb Plus DNA ladder (Transgen Biotech).

Thus far, the factors required to catalyze the excision of *dusA*-specific GIs have not been examined [41]. Whereas all *dusA*-specific GIs lack *xis* downstream of *int*, they carry a small ORF, here named *rdfN*, coding for a putative PrtN homolog (Fig 1) [41]. The deletion of *rdfN* abolished the excision and replication of MGI*Vch*USA3^{Kn}. Complementation by ectopic expression of *rdfN* from the arabinose-inducible promoter P_{BAD} restored the wild-type excision level but not the replication (Fig 6A and 6B). Likewise, deletion of *rdfN* abolished the mobilization of MGI*Vch*USA3^{Kn} but had no impact on the transfer of the helper plasmid (Fig 6C), confirming the specific role of *rdfN* in the GI's mobility.

To confirm that *rdfN* encodes the sole and only RDF of MGI*Vch*USA3, we constructed mini-GI, a minimal version of MGI*Vch*USA3 that only contains *int* and a spectinomycin-resistance marker. mini-GI is flanked by *attL* and *attR* and is integrated at *dusA* in *E. coli* EC100 (Fig 6D). Using mini-GI, *attB* and *attP* were detected only upon ectopic expression of *rdfN* from pBAD-*rdfN*, confirming that no other MGI*Vch*USA3-encoded protein besides Int and RdfN is required for the excision of the element (Fig 6E). *rdfN* is the essential RDF gene that favors the excision of MGI*Vch*USA3 and, most likely, all *dusA*-specific GIs.

A SgaC/AcaD hybrid complex activates the excision and mobilization of MGI*Vch*USA3

Next, we investigated the role of the transcriptional activator genes *acaC* and *sgaC* in the mobilization of MGI*Vch*USA3. Deletion of *acaDC* abolished the excision and replication of MGI*Vch*USA3^{Kn}, confirming that its excision relies on *rdfN*, whose expression is activated by AcaCD (Fig 4A, 6A and 6B). The mutation also confirmed that SgaC provided by MGI*Vch*USA3^{Kn} is insufficient by itself to elicit *rdfN* expression. The excision rate remained extremely low in cells that lack the helper plasmid or cells that carry pVCR94^{Sp} Δ*acaDC*. However, MGI*Vch*USA3^{Kn} allowed the low-frequency transfer of pVCR94^{Sp} Δ*acaDC* [19,28] (Fig 6C). Hence SgaC alone can activate to some degree the expression of the transfer genes of the helper plasmid. In contrast, deletion of *acaC* had no significant impact on the excision, replication, and mobilization of MGI*Vch*USA3^{Kn}, or on the transfer of the helper plasmid (Fig 6A, 6B and 6C). The primary sequences of AcaC and SgaC from MGI*Vch*USA3 share 85%

identity over 94% coverage, whereas AcaC and SgaC from SGI1 share only 75% identity over 92% coverage. Hence AcaD produced by the plasmid and SgaC produced by the MGI likely generate a functional chimeric transcriptional complex that acts as a potent activator of *rdfN* and the transfer genes.

The transfer of MGI*Vch*USA3^{Kn} Δ*sgaC* decreased nearly 3 logs compared to the wild-type GI, despite the presence of *acaDC* on the helper plasmid (Fig 6C). Moreover, deletion of both *acaC* and *sgaC* nearly abolished all transfer. These observations confirm that *sgaC*, not *acaC*, combined with *acaD* produces a hybrid activator complex that is an essential for the excision and mobilization of the element.

MGI*Vch*USA3 provides a new promoter and N-terminus for *dusA* expression

Since *dusA*-specific GIs insert within the 5' end of *dusA*, we wondered whether the gene remains expressed after the integration event. Sequence analysis of the *attR* junction of *E. coli* K12 transconjugants revealed that MGI*Vch*USA3 provides a new 5' coding sequence that diverges significantly from the native *E. coli dusA* gene (Fig 7A). This alteration of the 5' end of *dusA* results in a novel N-terminus of identical length sharing 61% identity over the 35 initial amino acid residues with native DusA. To test the expression of *dusA*, we constructed a translational *lacZ* fusion to its fortieth codon downstream of the *attR* junction in *E. coli* CAG18439 and BW25113 (Fig 7B). β-galactosidase assays revealed that *dusA* remains expressed after integration in both strains, confirming that MGI*Vch*USA3 provides a new promoter (Fig 7C). However, we observed a statistically

significant reduction of *dusA* expression resulting from the integration of the GI in both strains, suggesting that the transcription or translation signals brought by the GI are weaker than the original ones upstream of *E. coli dusA*.

Figure 7. MGI*Vch*USA3 drives the expression of *dusA*. (A) Comparison of the coding sequences of the 5' end of *dusA* in *E. coli* K12 MG1655 before (*attB* site) and after (*attL* junction) the integration of MGI*Vch*USA3. The core sequence of the *attB* and *attL* recombination sites is indicated with red shading. The ATG start codon of *dusA* is shown in bold. The sequence shown in blue is internal to MGI*Vch*USA3. Amino acid residues shown in red differ from the native N-terminus of DusA. This sequence was obtained by sequencing the *attL* junction of an *E. coli* CAG18439 *dusA*::MGI*Vch*USA1^{Kn} transconjugant colony. (B) Schematic representation of the *dusA*'-*lacZ* translational fusion for the detection of *dusA* expression. The fortieth codon of *dusA* (CAT shown in panel A) was fused to the eighth codon of *lacZ* downstream of the *attB* site. The gene *aadA7* (spectinomycin-resistance) was used for the insertion of *lacZ*. (C) β-galactosidase activity of the *dusA*'-'*lacZ* fusion before (-) and after (+) insertion of MGI*Vch*USA3^{Kn} in *E. coli* CAG18439 (FD034) and BW25113 (FD036). The bars represent the mean and standard error of the mean of three independent experiments. Statistical analyses were performed using an unpaired *t* test to compare the expression before and after integration of MGI*Vch*USA3^{Kn} for each strain. Statistical significance is indicated as follows: **, $P < 0.01$; *, $P < 0.05$.

Discussion

SGI1-like elements integrated at the 3' end of *trmE* are widespread in a broad range of *Enterobacteriaceae* and sporadically found in a few *Vibrio* species [12]. The integrase of SGI1 and its variants occasionally targets the *sodB* gene, a secondary attachment site [44,45]. Here, we report the identification of distant SGI1-like elements that specifically target the 5' end of *dusA* in multiple *Vibrio* species and the 3' end of *yicC* in *Enterobacteriaceae* and *Balneatrichaceae*. Farrugia *et al.* [41] already described GIs integrated at the 5' end of *dusA*, mostly prophages or phage remnants found exclusively in *Alpha*-, *Beta*- and *Gammaproteobacteria*. These authors identified GI*Vch*Ban1 and GI*Vch*Bra2 in *V. cholerae*, and several other GIs predicted to encode conjugative functions in *Bradyrhizobium*, *Caulobacter*, *Mesorhizobium*, *Paracoccus*, *Pseudomonas*, and *Rhodomicrobium* [41]. Our group recently reported a *dusA*-specific GI in *Aeromonas caviae* 8LM potentially mobilizable by IncC plasmids [42]. GI*Aca*8LM lacks *tra* genes but encodes a mobilization protein (MobI) under the control of an AcaCD-responsive promoter. Together, these reports confirm that *dusA* is an insertion hotspot for distinct families of mobile elements across at least three *Proteobacteria* phyla.

Thus far, only the *dusA*-specific GIs in *Acinetobacter baumannii* D1279779 and *Pseudomonas protegens* Pf-5 were shown to excise from the chromosome, albeit at a low level [41]. Neither GI has been tested for intercellular mobility. Here, we characterized MGI*Vch*USA3, a representative member of a subgroup of *dusA*-specific GIs circulating in *Vibrio* species and distantly related to SGI1. We

425 demonstrated that MGI*Vch*USA3 is mobilizable by IncC conjugative plasmids to
426 *E. coli*. In the presence of an IncC plasmid, this GI excises in practically all cells
427 of the population and becomes highly unstable (Figs 5B and 6A). We showed
428 that its excision was under the control of AcaCD provided by the IncC plasmid
429 and required *rdfN*, a gene whose expression is driven by an AcaCD-responsive
430 promoter (Fig 4). *rdfN* encodes a novel RDF distantly related to the pyocin
431 activator protein PrtN of *Pseudomonas*. *rdfN* seems to be ubiquitous, yet highly
432 divergent, in *dusA*-specific GIs reported by Farrugia *et al*. [41]. For instance,
433 RdfN (PrtN) encoded by the GI of *P. protegens* Pf-5 shares only 29% identity
434 with RdfN of MGI*Vch*USA3, and their promoters are unrelated. Hence, the
435 expression of *rdfN* homologs encoded by different families of *dusA*-specific GIs is
436 likely controlled by different factors. Only the *dusA*-specific GIs mobilizable by
437 IncC plasmids are expected be regulated by an AcaCD-like complex.

438 Excision and mobilization of MGI*Vch*USA3 occurred in the presence of a Δ*acaC*
439 but not a Δ*acaDC* mutant of the helper plasmid (Fig 6), confirming that *sgaC* of
440 the GI produces a functional activator subunit that can interact with AcaD
441 provided by the plasmid. Furthermore, we showed here that, unlike *acaC*, *sgaC*
442 plays a central role in the biology of MGI*Vch*USA3 as the absence of *acaC* had
443 no effect on the excision or transfer of the GI, while the absence of *sgaC* in spite
444 of the presence of *acaC*, compromised the mobilization of the GI (Fig. 6A, 6B
445 and 6C). We recently showed that AcaD most likely stabilizes the binding of
446 AcaC to the DNA [28]. Therefore, AcaD and SgaC from MGI*Vch*USA3 likely
447 interact to form a chimerical activator complex. This interaction could

compensate for the loss of *sgaD* in *yicC*- and type 1 and 2 *dusA*-specific GIs (Fig 1). The primary sequences of AcaC and SgaC of MGI*Vch*USA3 (type 1) share 85% identity. In contrast, AcaC only shares 75% identity with SgaC of SGI1 and 64% identity with SgaC of GI*Vch*USA5 (type 3), suggesting that retention of *sgaD* allowed faster divergence of SgaC from AcaC. Retention of *sgaC* in the GIs could result from its essential role as the elicitor of excision and replication reported for SGI1. Indeed, although AcaCD binds to the promoters P_{xis} and P_{rep} of SGI1, it fails to initiate transcription at these two promoters, unlike SgaCD [28]. Nonetheless, P_{xis} and P_{rep} are not conserved in the GIs described here. *S004-rep* is missing, whereas *rdfN* or *rdfM* replaced *xis* with novel AcaCD-responsive promoters (Fig 4 and S3 Fig). This observation raises intriguing questions regarding the recruitment of functional gene cassettes and their assimilation in a regulatory pathway. How did *xis*, *rdfN*, and *rdfM* acquire their AcaCD-responsive promoters? Is it by convergent evolution? What are the signals driving *rdfN* expression and GI excision in *dusA*-specific GIs resembling prophages?

Approximately 3 copies per cell of MGI*Vch*USA3 were detected in the presence of the helper IncC plasmid (Fig 6B), lower than the copy number reported for SGI1 (~8 copies/cell) [26,28]. MGI*Vch*USA3 lacks SGI1's replication (*S004-rep-oriV*); however, one of the multiple genes of unknown function could encode an unidentified replication initiator protein. Notably, GI*Vch*O27-1 encodes a putative replication protein with an N-terminal replicase domain (PF03090) and a C-terminal primase domain (PriCT-1, PF08708) [26]. Multiple GIs also carry putative replicons based on *repA* and *trfA* (Fig 1), suggesting that independent

471 replication is crucial in the lifecycle of these GIs, perhaps to improve their stability
472 in the presence of their helper plasmid [26–30].

473 Farrugia *et al.* [41] hypothesized that *dusA*-specific GIs could restore the
474 functioning of DusA. We demonstrated here that MGI*Vch*USA3 provides a new
475 promoter allowing expression of *dusA*, though at a lower level than in GI-free
476 cells, and restores the open reading frame with an altered N-terminus (Fig 7).
477 Similarly, the ICE SXT that targets the 5' end of the peptide chain release factor
478 3 (RF3) gene *prfC* provides a new promoter and N-terminus in both *V. cholerae*,
479 its original host, and *E. coli* [46]. In both cases, the consequences of the
480 alteration of the N-terminus on the activity of the protein remain unknown.

481 The relative positions of *int* and *rdfN*/*rdfM* across the *attP* site suggest that to
482 remain functional the recombination modules must be acquired or exchanged
483 when the GIs are in their excised circular form. The promiscuity of different
484 families of GIs targeting *yicC*, *dusA*, and *trmE* and mobilizable by IncC plasmids
485 could act as the catalyst for these recombination events. During entry into a new
486 host cell by conjugation, IncC plasmids elicit the excision of such GIs and
487 promote homologous recombination between short repeated sequences in
488 response to double-stranded break induced by host defense systems (CRISPR-
489 Cas3) [43]. Hence the diversification of IncC plasmid-mobilizable GIs could be a
490 side effect of the DNA repair mechanism used by these plasmids.

491 Unlike SGI1 and its siblings, all *dusA*-specific SGI1-like GIs reported here lack
492 antibiotic resistance genes. Furthermore, SGI1 variants are prevalent in several

493 pathogenic species and relatively well-conserved, whereas their *dusA*-specific
494 relatives are scarce and highly divergent. These observations suggest that
495 despite the considerable functional resemblances between *trmE*- and *dusA*-
496 specific SGI1-like GIs, the epidemiological success of the SGI1 lineage has
497 directly stemmed from the acquisition of class I integrons conferring multidrug
498 resistance by forerunner elements such as SGI0 [47]. Based on the analysis of
499 phylogenetic relationships between the core proteins MpsA, TraG, SgaC and
500 TraN, *oriT* loci, and integrase proteins (Fig 3 and S1A Fig), we propose a
501 hypothetical evolutionary pathway leading to the emergence of the different types
502 of GIs described here (Fig 8). The diversity of *dusA*-specific GIs and relative
503 homogeneity of the SGI1 group suggest that the latter originated from the
504 progenitor of IncA and IncC plasmids via a *dusA*-specific GI intermediate.

505 **Figure 8. Proposed hypothetical evolutionary pathway of SGI1-like GIs.** The
506 sequence of events was inferred from the phylogenetic trees presented in this
507 study, site of integration and conservation of *traH* and *sgaD* in the GIs. The
508 proposed pathway ignores the gene cargo and presumes that the GI lineages
509 evolved from the progenitor of IncA and IncC plasmids. The *dusA*-specific
510 recombination module was chosen as the progenitor to minimize gain/loss and
511 recombination events. Green and red arrows indicate gene gains and losses,
512 respectively. The orange dashed line indicates a probable recombination event
513 from which stemmed GI*Vch*O27-1.

Materials and Methods

Bacterial strains and media

Bacterial strains and plasmids used in this study are described in Table 1. 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 [31]. 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

were purchased from New England Biolabs. Q5 DNA polymerase (New England Biolabs) and EasyTaq DNA Polymerase (Civic Bioscience) were used for amplifying cloning inserts and verification, respectively. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen), according to manufacturer's instructions. *E. coli* was transformed by electroporation as described by Dower *et al*. [48] in a Bio-Rad GenePulser Xcell apparatus set at 25 µF, 200 Ω and 1.8 kV using 1-mm gap electroporation cuvettes. Sanger sequencing reactions were performed by the Plateforme de Séquençage et de Génotypage du Centre de Recherche du CHUL (Québec, QC, Canada).

Plasmids and strains constructions

Oligonucleotides used in this study are listed in Table 2. MGI*Vch*USA3^{Cm} was constructed by inserting the *pir*-dependent replication RP4-mobilizable plasmid pSW23T [49] at locus CGT85_RS05425 of *V. cholerae* OYP6G08 (Genbank NZ_NMSY01000009) by homologous recombination. Briefly, CGT85_RS05425 was amplified using primer pair dusAigEcoRIF/dusAigEcoRIR. The amplicon was digested with EcoRI and cloned into EcoRI-digested pSW23T using T4 DNA ligase. The resulting plasmid was confirmed by restriction profiling and DNA sequencing, then introduced into the DAP-auxotrophic *E. coli* β2163 by transformation [49]. This strain was used as a donor in a mating assay to transfer the plasmid into *V. cholerae* OYP6G08, generating MGI*Vch*USA3^{Cm}. Single-copy integration of the pSW23T derivative was confirmed by PCR and antibiotic resistance profiling.

559 MGI*Vch*USA3^{Kn} was constructed from MGI*Vch*USA3^{Cm}. Briefly, pVCR94^{Kn} Δ*acr2*
560 was transferred from the DAP-auxotrophic *E. coli* KH40 into OYP6G08 bearing
561 MGI*Vch*USA3^{Cm}. After selection on LB agar medium supplemented with
562 chloramphenicol and kanamycin, Cm^{R} Kn^{R} *V. cholerae* OYP6G08
563 transconjugants were confirmed by growth on thiosulfate-citrate-bile salts-
564 sucrose (TCBS) agar medium (Difco). MGI*Vch*USA3^{Cm} was then mobilized from
565 OYP6G08 to *E. coli* CAG18439. Integration at the 5' end of *dusA* in *E. coli* was
566 confirmed by amplification of the *attL*, *attR*, *attB* and *attP* sites with primer pairs
567 EcodusAattLf/ dusAattLr, dusAattRf/EcodusAattRr, EcodusAattLf/EcodusAattRr
568 and dusAattRf/ dusAattLr, respectively. MGI*Vch*USA3^{Kn} was constructed by
569 replacing pSW23T with a single kanamycin resistance marker using the one-step
570 chromosomal gene inactivation technique with primer pair
571 dusAscarNoFRTf/dusAscarNoFRTr and pKD13 as the template. The deletions
572 Δ*sgaC* and Δ*prtN* in MGI*Vch*USA3^{Kn} were obtained using the primer pairs
573 oFD26r/oFD26f and DelprtNr/DelprtNf, and pKD3 and pVI36 as the templates,
574 respectively. The Δ*dapA* deletion mutant of *E. coli* MG1655 was constructed
575 using primer pair FwDeltaDapA-MG1655/ RvDeltaDapA-MG1655 and pKD3 as
576 the template. The Δ*lacZ* mutation was introduced in *E. coli* CAG18439 using
577 primer pair lacZW-B/lacZW-F and plasmid pKD4 as the template. The *dusA*'-
578 *'lacZ* fusion was introduced in *E. coli* BW25113 and CAG18439 using primer pair
579 oDF15/oDF16 and pVI42B as the template. The λRed recombination system was
580 expressed using either pSIM6, pSIM9 or pKD46 [50,51]. When appropriate,
581 resistance cassettes were excised from the resulting constructions using the Flp-

582 encoding plasmid pCP20 [52]. All deletions were validated by antibiotic profiling
583 and PCR.

584 Fragments encompassing promoter regions upstream of *int*, *traN*, *traG*, *s018* and
585 *rdfN* were amplified using primer pairs oFD6.f/oFD6.r, oFD1.f/oFD1.r,
586 oFD3.f/oFD3.r, oFD5.f/oFD5.r and oFD4.f/oFD4.r, respectively, and genomic
587 DNA from *E. coli* CAG18439 *dusA*::MGI*Vch*USA3^{Kn} as the template. The
588 amplicons were digested with PstI/XhoI and cloned into PstI/XhoI-digested
589 pOP*lacZ* [19]. The resulting constructs were single-copy integrated into the *attB*$_{\lambda}$
590 chromosomal site of *E. coli* BW25113 using pINT-ts [53].

591 mini-GI was constructed as follows. The 1,591-bp fragment of excised circular
592 MGI*Vch*USA3^{Kn} that contains *attP-int* was amplified using primer pair
593 oVB12/oVB10 and genomic DNA from *E. coli* CAG18439 *dusA*::MGI*Vch*USA3^{Kn}
594 as the template. The 1,421-bp fragment of pVI36 that contains *aadA7* was
595 amplified using primer pair oVB11/oVB13. Both fragments were joined using the
596 PCR-based overlap extension method [54]. After the final PCR amplification
597 using oVB12/oVB13, the amplicon was purified, digested with SacI, and ligated.
598 The ligation mixture was then transformed into *E. coli* EC100. Transformant
599 colonies were selected on LB agar supplemented with spectinomycin. The
600 constitutive expression of *int* and the absence of replicon prompted the
601 spontaneous integration of mini-GI at the 5' end of *dusA* in EC100.

All final constructs were verified by PCR and DNA sequencing by the Plateforme de Séquençage et de Génotypage du Centre de Recherche du CHUL (Québec, QC, Canada).

qPCR assays

qPCR assays for quantification of excision and copy number of MGI*Vch*USA3^{Kn} were carried out as described previously [28] with the following modification. *attB*$_{dusA}$ (241 bp) and *higA* (229 bp) of MGI*Vch*USA3^{Kn} were quantified using primer pairs attBdusAqPCRfwd/ attBdusAqPCRrev and higAqPCRfwd/ higAqPCRrev, respectively (Table 1). The excision rate and copy number of MGI*Vch*USA3^{Kn} were calculated as the ratio of free *attB*$_{dusA}$ site per chromosome and as the ratio of *higA* per chromosome, respectively. The data were analyzed and normalized using all three chromosomal genes *dnaB*, *hicB* and *trmE* as references and the qBase framework as described previously [28,55].

β-galactosidase assays

The assays were carried out on LB agar plates supplemented with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) or in LB broth using *o*-nitrophenyl-β-D-galactopyranoside (ONPG) as the substrate as described previously [42]. *acaDC* expression from pBAD-*acaDC* was induced by adding 0.2% arabinose to a refreshed culture grown to an OD_{600} of 0.2, followed by a 2-h incubation at 37°C with shaking prior to cell sampling.

Comparative analyses

Sequences were obtained using Blastp against the Genbank Refseq database with the primary sequences of key proteins MpsA, $TraG_S$, SgaC, $TraN_S$ of SGI1 (Genbank AAK02039.1, AAK02037.1, AAK02036.1, AAK02035.1, respectively), and ultimately Int_{dusA} of GI*Vch*Bra2 (Genbank EEO15317.1) and Int_{yicC} of GI*Eco*MOD1 (Genbank WP_069140142.1). Hits were exported, then sorted by accession number to identify gene clusters that likely belong to complete GIs. Sequences of GIs were manually extracted and the extremities were identified by searching for the direct repeats contained in *attL* and *attR* sites. When a GI sequence spanned across two contigs (e.g., GI*Vch*Hai10 and GI*Ppl*Ind1), the sequence was manually assembled. GI sequences were clustered using cd-hit-est with a 0.95 nucleotide sequence identity cut-off [56]. Some of the annotated sequences were manually curated to correct missing small open reading frames such as *mpsB*, and inconsistent start codons.

Circular blast representations were generated with the Blast Ring Image Generator (BRIG) 0.95 [39], with blastn or blastp, against SGI1ΔIn104 and GI*Vch*USA2, with an upper identity threshold of 80% and a lower identity threshold of 60%.

Evolutionary analyses were conducted in MEGA X [57] and inferred by using the maximum likelihood method based on the JTT (MpsA or SgaC proteins), LG (Int_{dusA}, Int_{yicC}, TraG or $RepA_{IncFII}$ proteins) or WAG (TraN) matrix-based models [58–60]. Protein sequences were aligned with Muscle [61]. Aligned sequences were trimmed using trimal version 1.2 using the automated heuristic approach

[62]. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with the superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories) for Int_{dusA} (parameter = 3.5633), Int_{yicC} (parameter = 2.6652), SgaC (parameter = 1.4064), TraG (parameter = 1.9005) and TraN (parameter = 1.6476) proteins. For Int_{dusA}, MpsA and TraG, the rate variation model allowed for some sites to be evolutionarily invariable ([+I], 7.81% sites for Int_{dusA}, 44.62% sites for MpsA and 5.22% sites for $TraG_S$). The trees are all drawn to scale, with branch lengths measured in the number of substitutions per site.

oriT sequences were obtained manually using the previously identified *oriT* of SGI1 as the reference [25], then clustered using cd-hit-est with a 1.0 nucleotide sequence identity cut-off. Sequences were then aligned using Muscle and a NeighborNet phylogenetic network was built using SplitsTree4 [63] with default parameters (Uncorrected_P method for distances and EqualAngle drawing method). The secondary structures of the aligned *oriT* sequences were predicted using RNAalifold 2.4.17 from the ViennaRNA package [64]. Default options were used (including no RIBOSUM scoring), except for the following: no substituting "T" for "U" (--noconv), no lonely pairs (--noLP), no GU pairs (--noGU) and DNA parameters (-P DNA). The predicted Vienna output and the annotated alignment were merged into a predicted secondary structure of SGI1 *oriT* color-coded to display the inter-island diversity.

Statistical analyses and figures preparation

Prism 8 (GraphPad Software) was used to plot graphics and to carry out statistical analyses. All figures were prepared using Inkscape 1.0 (https://inkscape.org/).

Acknowledgments

We are grateful to Yann Boucher for the kind gift of *Vibrio cholerae* OYP6G08 and Kévin T. Huguet for technical assistance. We thank Nicolas Rivard and David Roy for their insightful comments on the manuscript.

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. Brouwer MSM, Warburton PJ, Roberts AP, Mullany P, Allan E. Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of Clostridium difficile. PLoS ONE. 2011;6: e23014. doi:10.1371/journal.pone.0023014

7. 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

8. Boyd DA, Peters GA, Ng L, Mulvey MR. Partial characterization of a genomic island associated with the multidrug resistance region of *Salmonella enterica* Typhymurium DT104. FEMS Microbiol Lett. 2000;189: 285–291. doi:10.1111/j.1574-6968.2000.tb09245.x

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

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

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

12. 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

13. 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

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

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

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

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

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

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

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

21. Doublet B, Praud K, Weill F-X, Cloeckaert A. Association of IS*26*-composite transposons and complex In4-type integrons generates novel multidrug

resistance loci in *Salmonella* genomic island 1. J Antimicrob Chemother. 2009;63: 282–289. doi:10.1093/jac/dkn500

22. de Curraize C, Neuwirth C, Bador J, Chapuis A, Amoureux L, Siebor E. Two new *Salmonella* genomic islands 1 from *Proteus mirabilis* and description of blaCTX-M-15 on a variant (SGI1-K7). J Antimicrob Chemother. 2018;73: 1804–1807. doi:10.1093/jac/dky108

23. de Curraize C, Siebor E, Varin V, Neuwirth C, Hall RM. Two New SGI1-LK Variants Found in *Proteus mirabilis* and Evolution of the SGI1-HKL Group of *Salmonella* Genomic Islands. mSphere. 2020;5: e00875-19. doi:10.1128/mSphere.00875-19

24. Lei C-W, Chen Y-P, Kong L-H, Zeng J-X, Wang Y-X, Zhang A-Y, et al. PGI2 Is a Novel SGI1-Relative Multidrug-Resistant Genomic Island Characterized in *Proteus mirabilis*. Antimicrob Agents Chemother. 2018;62: e00019-18, /aac/62/5/e00019-18.atom. doi:10.1128/AAC.00019-18

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

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

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

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

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

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

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

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

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

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

35. Taviani E, Spagnoletti M, Ceccarelli D, Haley BJ, Hasan NA, Chen A, et al. Genomic analysis of ICE*Vch*Ban8: An atypical genetic element in *Vibrio cholerae*. FEBS Lett. 2012;586: 1617–1621. doi:10.1016/j.febslet.2012.03.064

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

37. Daccord A, Mursell M, Poulin-Laprade D, Burrus V. Dynamics of the SetCD-regulated integration and excision of genomic islands mobilized by integrating conjugative elements of the SXT/R391 family. J Bacteriol. 2012;194: 5794–5802. doi:10.1128/JB.01093-12

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

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

40. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14: 1188–1190. doi:10.1101/gr.849004

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

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

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

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

45. Siebor E, Neuwirth C. Proteus genomic island 1 (PGI1), a new resistance genomic island from two Proteus mirabilis French clinical isolates. J Antimicrob Chemother. 2014;69: 3216–3220. doi:10.1093/jac/dku314

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

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

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

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

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

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

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

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

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

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

56. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics. 2010;26: 680–682. doi:10.1093/bioinformatics/btq003

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

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

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

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

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

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

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

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

65. Kirchberger PC, Orata FD, Nasreen T, Kauffman KM, Tarr CL, Case RJ, et al. Culture-independent tracking of *Vibrio cholerae* lineages reveals complex spatiotemporal dynamics in a natural population. Environmental Microbiology. 2020 [cited 21 Sep 2020]. doi:10.1111/1462-2920.14921

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

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

68. Ceccarelli D, Daccord A, René M, Burrus V. Identification of the Origin of Transfer (*oriT*) and a New Gene Required for Mobilization of the SXT/R391 Family of Integrating Conjugative Elements. J Bacteriol. 2008;190: 5328–5338. doi:10.1128/JB.00150-08

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

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

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

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

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

933 **Table 1. 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	[65]
N16961	O1 El Tor	[66]
E. coli		
β2163	(F−) RP4-2-Tc::Mu Δ*dapA*::(*erm-pir*) (Kn Em)	[49]
CAG18439	MG1655 *lacZU118 lacI42*::Tn*10* (Tc)	[67]
BW25113	F$^-$ Δ(*araD-araB*)*567*, Δ*lacZ4787*(::*rrnB-3*), λ$^-$, *rph-1*, Δ(*rhaD-rhaB*)*568*, *hsdR514*	[50]
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 Δ*dapA*::*cat* (Cm)	This study
VB112	Rf-derivative of MG1655	[68]
GG56	Nx-derivative of BW25113	[69,70]
FD034	CAG18439 Δ*lacZ 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)	[50]
pKD4	Kn^R PCR template for one-step chromosomal gene inactivation (Kn)	[50]
pKD13	Kn^R PCR template for one-step chromosomal gene inactivation (Kn)	[50]
pVI36	Sp^R PCR template for one-step chromosomal gene inactivation (Sp)	[68]
pVI42B	pVI36 BamHI::P_{lac}-*lacZ* (Sp)	[71]
pSW23T	pSW23::*oriT*$_{RP4}$; *oriV*$_{R6Kγ}$ (Cm)	[49]
pOP*lacZ*	pAH56 *lacZ* (Kn)	[19]
pBAD30	*ori*$_{p15A}$ *bla araC P*$_{BAD}$ (Ap)	[72]
pBAD-*acaDC*	pBAD30::*acaDC* (Ap)	[19]
pBAD-*rdfN*	pBAD30::*rdfN* (Ap)	This study
pVCR94^{Kn} Δ*acr2*	Δ*acr2* mutant of pVCR94^{Kn} (Su Kn)	[43]
pVCR94^{Sp}	Sp^R derivative of pVCR94 (Su Sp)	[19]
pVCR94^{Sp} Δ*acaC*	Δ*acaC* mutant of pVCR94^{Sp} (Su Sp)	[19]
pVCR94^{Sp} Δ*acaDC*	Δ*acaDC* mutant of pVCR94^{Sp} (Su Sp)	[19]
Genomic Islands		
MGI*Vch*USA3		This study
MGI*Vch*USA3^{Cm}	MGI*Vch*USA3 CGT85_RS05425ΩpSW23T (Cm)	This study
MGI*Vch*USA3^{Kn}	Kn^R derivative of MGI*Vch*USA3 (Kn)	This study
MGI*Vch*USA3^{Kn} Δ*sgaC*	Δ*sgaC* mutant of MGI*Vch*USA3 (Kn)	This study
MGI*Vch*USA3^{Kn} Δ*rdfN*	Δ*rdfN* mutant of MGI*Vch*USA3 (Kn)	This study
mini-GI	*attP-int-aad7* derived from MGI*Vch*USA3 (Sp)	This study

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

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

936 tetracycline; Tm, trimethoprim; ts, thermosensitive.

938 **Table 2. Oligonucleotides used in this study.**

Primer name	Nucleotide sequence (5' to 3')	Reference
dusAigEcoRIF	NNNNGAATTCACAAGTTATCGCTCTATCACTG	This study
dusAigEcoRIR	NNNNGAATTCCCTTTGATGTGGGGCATG	This study
VchdusAattLf	CAGCCGACGTTGAGGTTAA	This study
EcodusAattLf	CAGCCGACATTCAGGTTG	This study
dusAattLr	AAGCGAATGATGCCTTTACTG	This study
dusAattRf	GTGTGTGTAGCTTCAGGTG	This study
EcodusAattRr	GTCGAACCTGGATTGTTTATCATTG	This study
VchdusAattRr	TCACCATCACTGTTGGACTT	This study
dusAscarNoFRTf	CAGCAGTCCTATCATGCCCCACATCAAAGGGAATTCAGAGCGCTTTTGAAG CTCA	This study
dusAscarNoFRTr	GATCATAATCAATATCGATGGAGAAAAGCAATGACATCGGAATAGGAACTT CAAGA	This study
DelprtNf	ATCGTTGGAAATTGTTGAGAATGATTGAGGATAGCTGTGTAGGCTGGAGCT GCTTCG	This study
DelprtNr	GGGATGGGATAATATTTGGCATTCAGACCCAGGTAGTTAATTCCGGGGATC CGTCGACC	This study
FwDeltaDapA-MG1655	ATGTTCACGGGAAGTATTGTCGCGATTGTTACTCCGGTGTAGGCTGGAGCT GCTTCG	This study
RvDeltaDapA-MG1655	TTACAGCAAACCGGCATGCTTAAGCGCCGCTCTGACCATATGAATATCCTC CTTA	This study
lacZW-B	GCGAAATACGGGCAGACATGGCCTGCCCGGTTATTACATATGAATATCCTC CTTA	This study
lacZW-F	TTGTGAGCGGATAACAATTTCACACAGGAAACAGCTGTGTAGGCTGGAGCT GCTTCG	This study
oDF15	AGCGTTGCACCAATGCTCGACTGGACGGACAGACATCTGGCCGTCGTTTTA CAACGTCG	This study
oDF16	AGGGCGTGGTGAATTTGACTACTTTTTGGTGAAAAGGCAGCATTACACGTC TTGAG	This study
oFD26f	GCAGAACGGGCATTCGACACAAGTTCGCTGATTAACGTGTAGGCTGGAGCT GCTTC	This study
oFD26r	CCAGGTCTTTGGCCGCAAAAATGAGGATGAGTAGTCCATATGAATATCCTC CTTA	This study
oFD1r	NNNNCTCGAGCACATGATTTCCGGAAATAAAAGC	This study
oFD1f	NNCTGCAGTTAATCAGCGAACTTGTGTCGAAT	This study
oFD3r	NNNNCTCGAGAGACAAATACTCCCGACTTGATCC	This study
oFD4r	NNNNCTCGAGAGCTATCCTCAATCATTCTCAACA	This study
oFD4f	NNCTGCAGTAAAAACATTTGAGAGGTCATTCGG	This study
oFD5r	NNNNCTCGAGAGACACCTCCAAAAAGTTGAAGG	This study
oFD5f	NNCTGCAGGCAGCTTATAGCATGAATCTGTAC	This study
oFD6r	NNNNCTCGAGGAAGCGAATGATGCCTTTACTGG	This study
oFD6f	NNCTGCAGCGCTGAATCTACGACTTAATGACA	This study
prtNEcoRIf	NNNGAATTCAAGGAGGAATAATAAATGAATACCGCATTTCTTCTG	This study
prtNHindIIIrev	NNNAAGCTTCAGGTAGTTACAGTTCTCTC	This study
oVB10	CAGGTGGCACTTTTCGGGGTCAGTCTTTCCCAACACTCATCCCCTTCTG	This study
oVB11	GAGTGTTGGGAAAGACTGACCCCGAAAAGTGCCACCTGCATCGATG	This study
oVB12	NNNGAGCTCCGGGCGAGAAGTAGCGTTGA	This study
oVB13	NNNGAGCTCGATAGCTAGACTGGGCGGT	This study
attBdusAqPCRfwd	CCTGAATAGTGATGCTGAATAAC	This study
attBdusAqPCRrev	CCATTTCGGTATACAGCAAC	This study
higAqPCRfwd	CTTCCTGCTCAAAGACTCTATG	This study
higAqPCRrev	CTGGTGACCGAGTTTCTG	This study
qFwpVCR	AAGAGAACCAAAGACAAAGACC	[26]
qRvpVCR	CACCTTCACCGTGAAATGC	[26]

qdnaBFw	ACGATTTTTACACCCGCCCAC	[26]
qdnaBRv	ATCATCTCACGGACAACGGCAC	[26]
qhicBFw	GCTTATCCCTTTACCTTCGCC	[26]
qhicBRv	TAACTCTTTGCCAAGCGCC	[26]
qthdFFw	GATAATGACACTATCGTAGCCC	[26]
qthdFRv	GCAGTTCCAGCACATCTTC	[26]

939

Supporting information

S1 Figure. NeigborNet phylogenetic network and predicted secondary structure of 39 *oriT* loci of SGI1-like islands. Each island's integration site and type are annotated. The sequence of canonical SGI1 was used as a reference to describe the predicted secondary structure of all *oriT* sequences. Pairs can be perfectly conserved, imperfectly conserved (1/39 not conserved), not conserved (> 1/39) or can be either an A-T or G-C pair only. In this last case, the sequence is not conserved, but the predicted local secondary structure is. “Not conserved” means that in some instances it will correspond to an A-T/G-C pair, which will either tighten a stem, form a bubble or widen an existing bubble.

S2 Figure. Maximum likelihood phylogenetic analysis of key proteins of SGI1-related GIs. 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. Bootstrap supports are indicated as percentages at the branching points only when > 80%. 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 SGI1$_{trmE}$ islands and SGI1$_{yicC}$ islands are accompanied by a light blue circle and a red circle, respectively. All other taxa correspond to SGI1$_{dusA}$ islands.

S3 Figure. Alignment of AcaCD-responsive promoters predicted in GIs targeting *dusA*, *yicC* and *trmE*. Promoter sequences are grouped according to the function of the expressed genes as follows: (A) recombination directionality factors; (B) mating pair stabilization; (C) mating pair formation and stabilization; (D) unknown. AcaCD binding

963 sites are shown in green. Logo sequences and *p*-values were generated by MAST [73].
964 When known, the transcription start sites are shown in blue based on previously
965 published works [19,28]. Predicted Shine-Dalgarno sequences are shown in pink. The
966 initiation start codon is shown in bold letters.

967 **S1 Table. Features of the identified GIs and associated strains.**

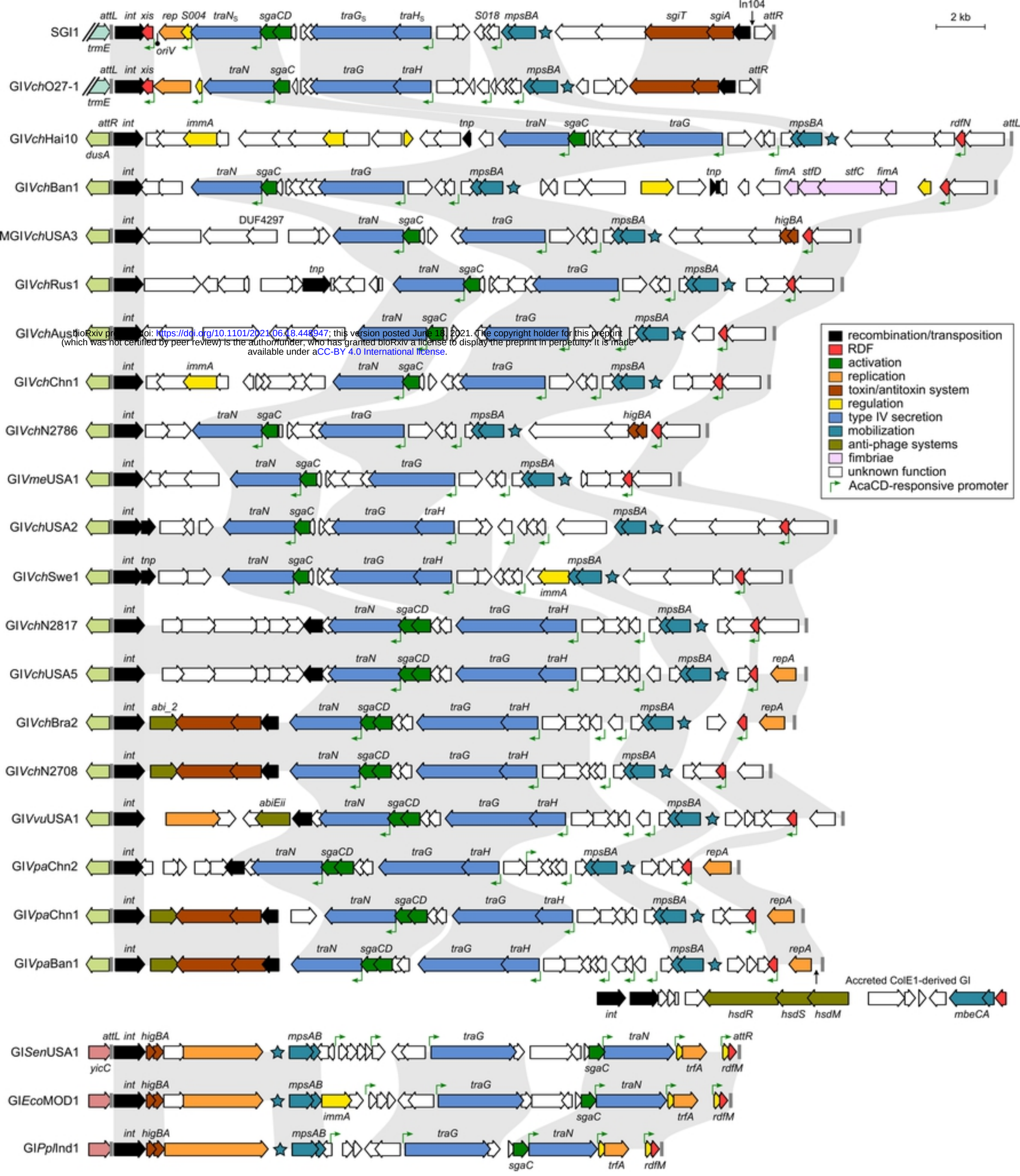
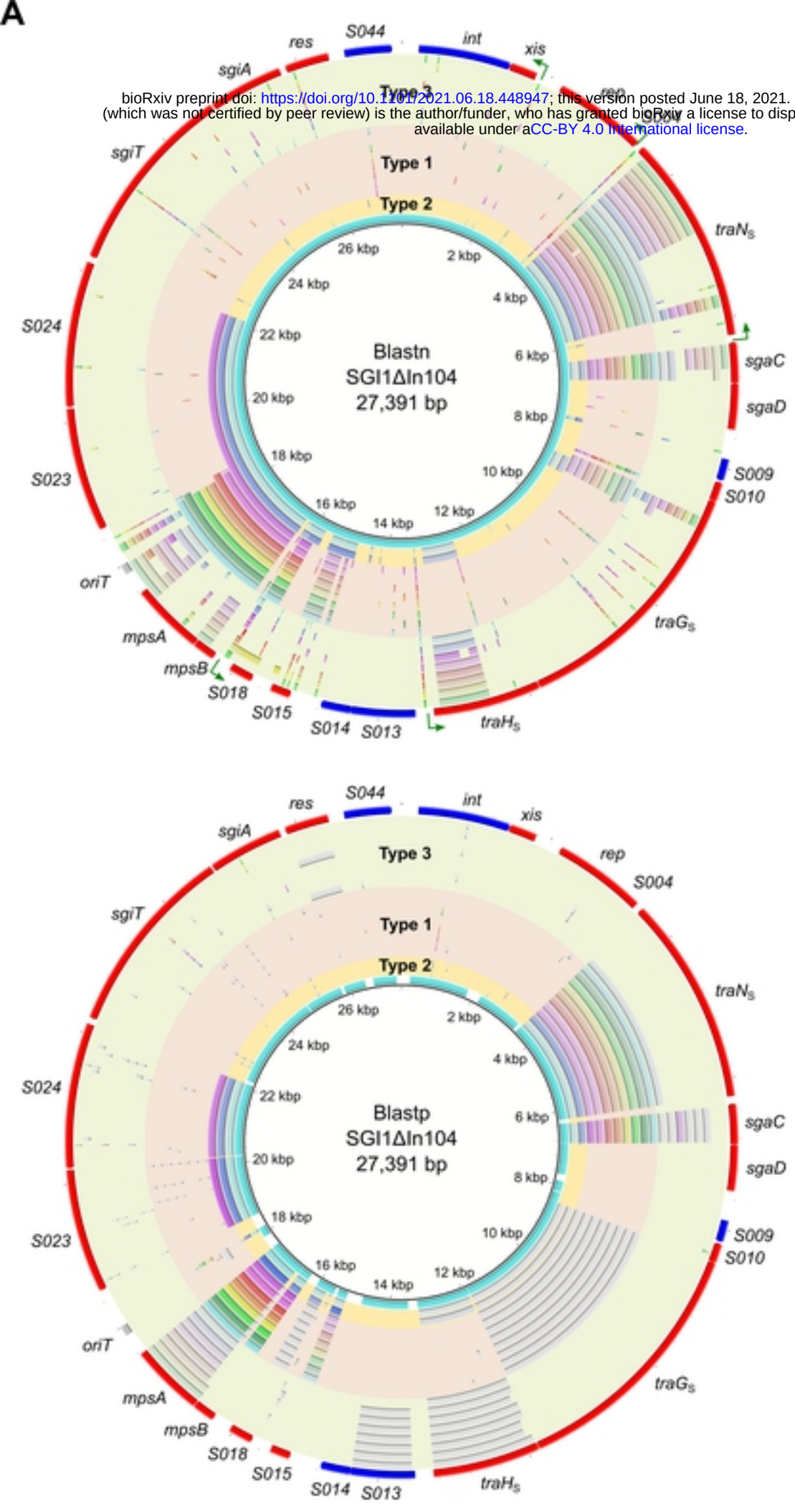


Fig 1



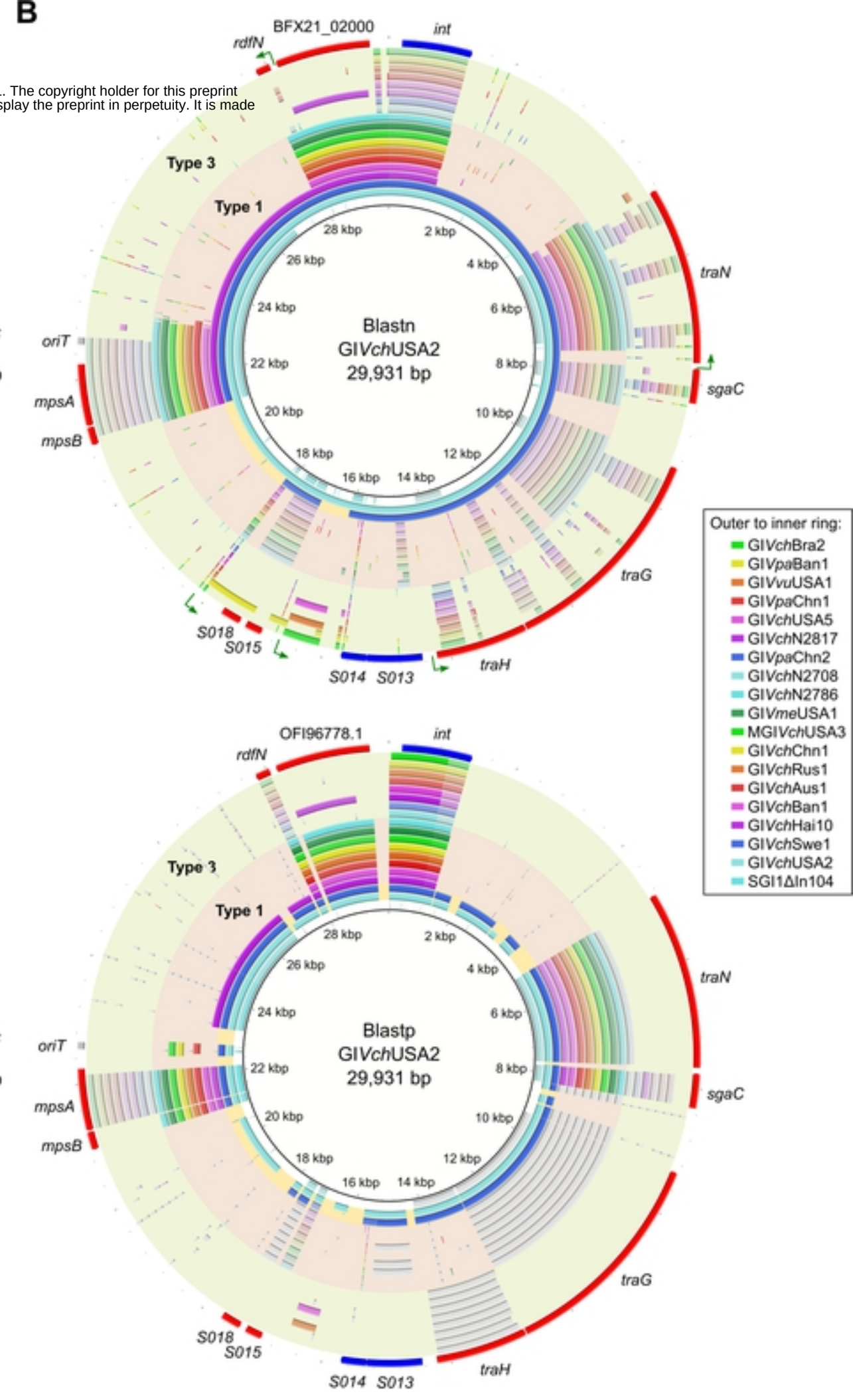


Fig 2

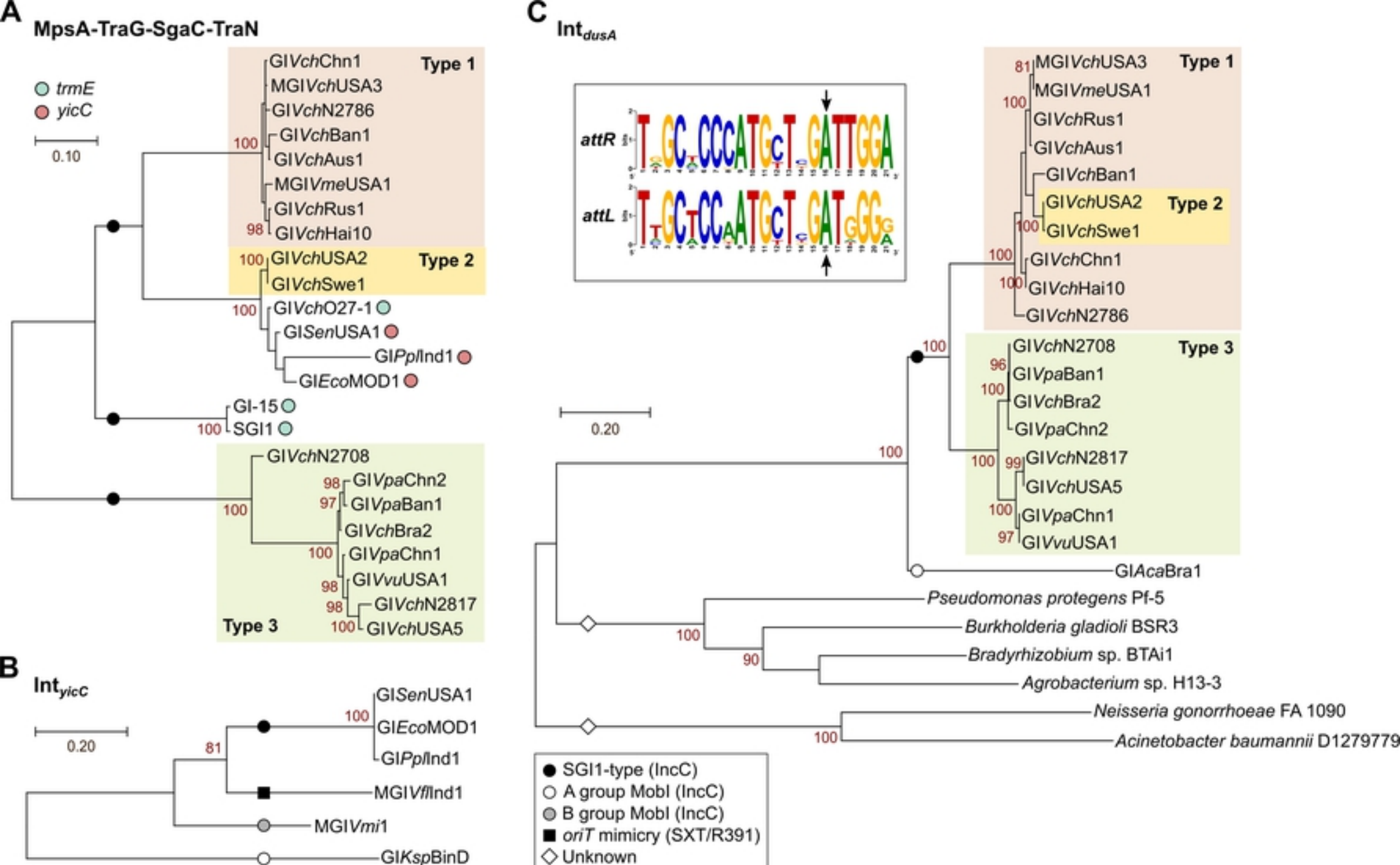

A
MpsA-TraG-SgaC-TraN
trmE
yicC
0.10
GIVchChn1
MGIVchUSA3
GIVchN2786
GIVchBan1
GIVchAus1
MGIVmeUSA1
GIVchRus1
GIVchHai10
Type 1
GIVchUSA2
GIVchSwe1
Type 2
GIVchO27-1
GISenUSA1
GIPplInd1
GIEcoMOD1
GI-15
SGI1
GIVchN2708
GIVpaChn2
GIVpaBan1
GIVchBra2
GIVpaChn1
GIVvuUSA1
GIVchN2817
GIVchUSA5
Type 3
B
Int yicC
0.20
GISenUSA1
GIEcoMOD1
GIPplInd1
MGIVflInd1
MGIVmi1
GIKspBinD
C
Int dusA
attR
attL
0.20
MGIVchUSA3
MGIVmeUSA1
GIVchRus1
GIVchAus1
GIVchBan1
GIVchUSA2
GIVchSwe1
GIVchChn1
GIVchHai10
GIVchN2786
GIVchN2708
GIVpaBan1
GIVchBra2
GIVpaChn2
GIVchN2817
GIVchUSA5
GIVpaChn1
GIVvuUSA1
GIAcaBra1
Pseudomonas protegens Pf-5
Burkholderia gladioli BSR3
Bradyrhizobium sp. BTAi1
Agrobacterium sp. H13-3
Neisseria gonorrhoeae FA 1090
Acinetobacter baumannii D1279779
SGI1-type (IncC)
A group MobI (IncC)
B group MobI (IncC)
oriT mimicry (SXT/R391)
Unknown

Fig 3

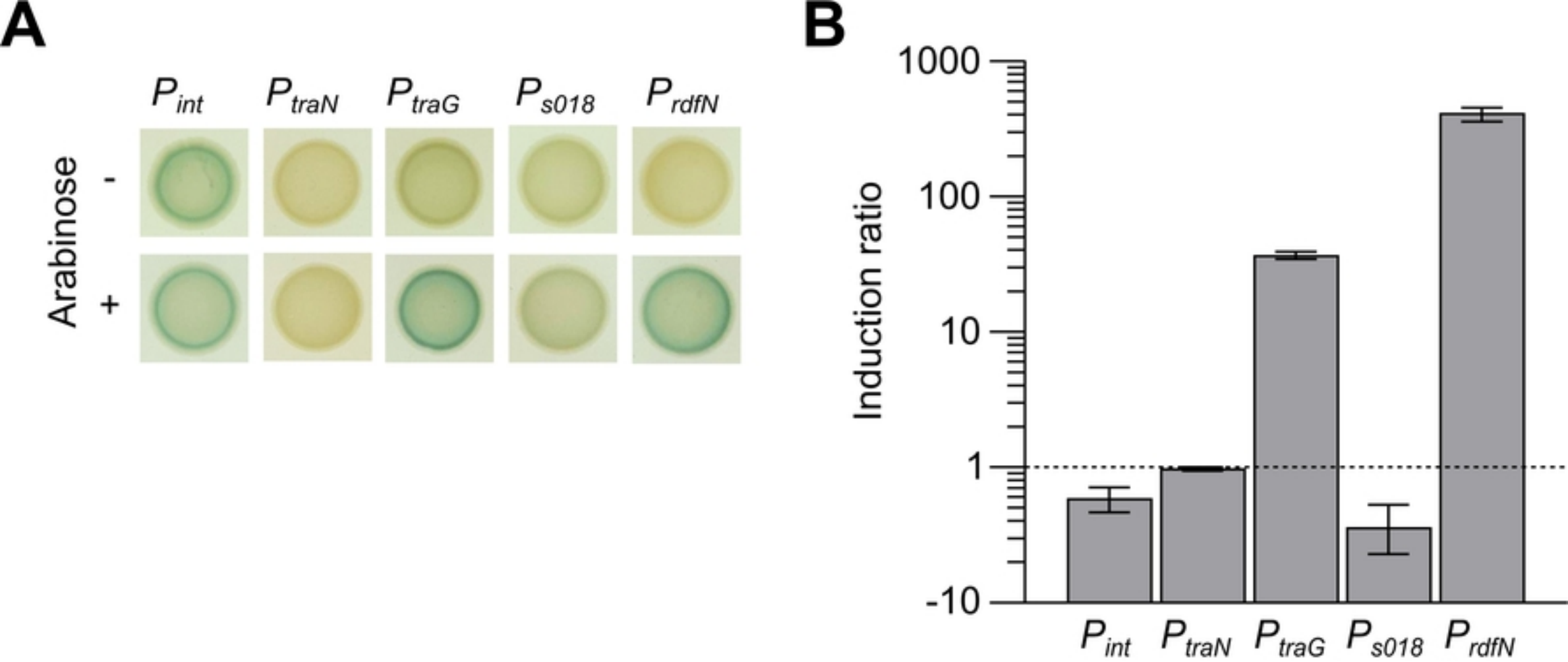

A
P_{int}
P_{traN}
P_{traG}
P_{s018}
P_{rdfN}
Arabinose
-
+
B
Induction ratio
1000
100
10
1
-10
P_{int}
P_{traN}
P_{traG}
P_{s018}
P_{rdfN}

Fig 4

A

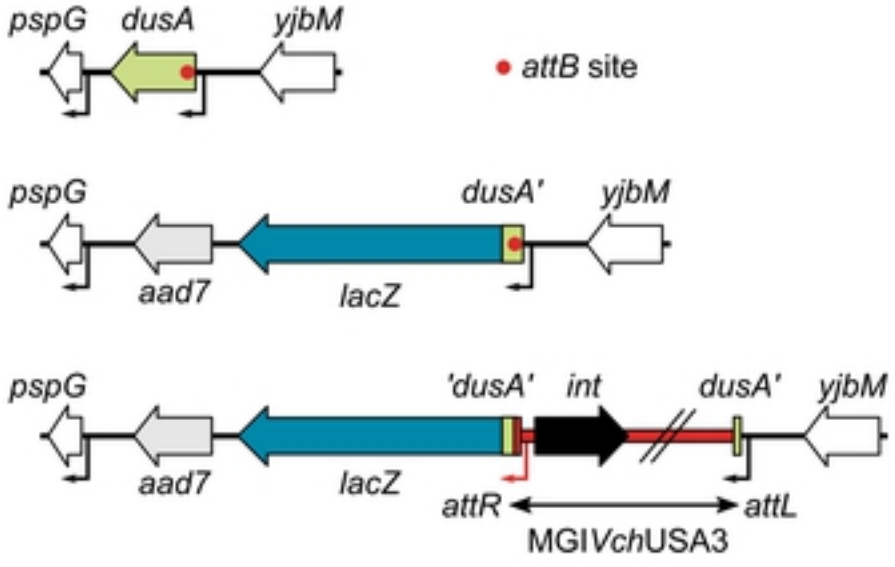
E. coli K12 *dusA attB* site

```
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 ACG 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*::MGI*Vch*USA3^{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 GTG GCC CCC ATG CTT GAC TGG ACG 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

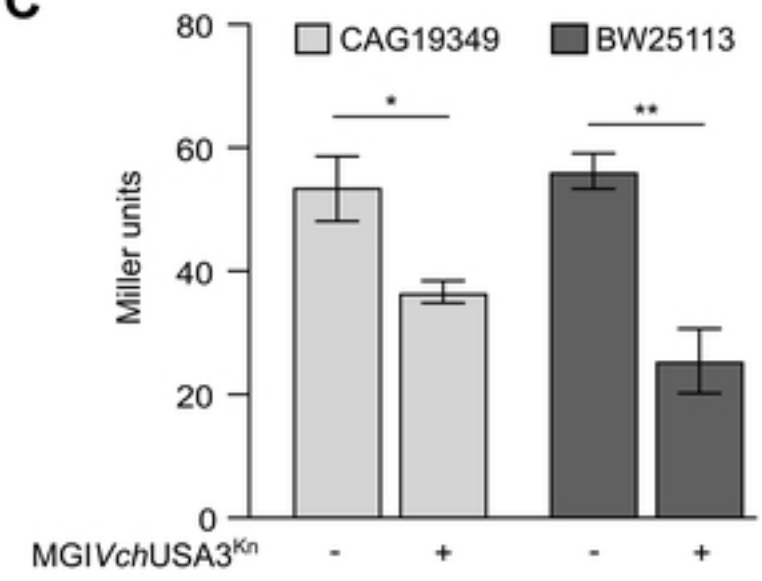


Fig 7

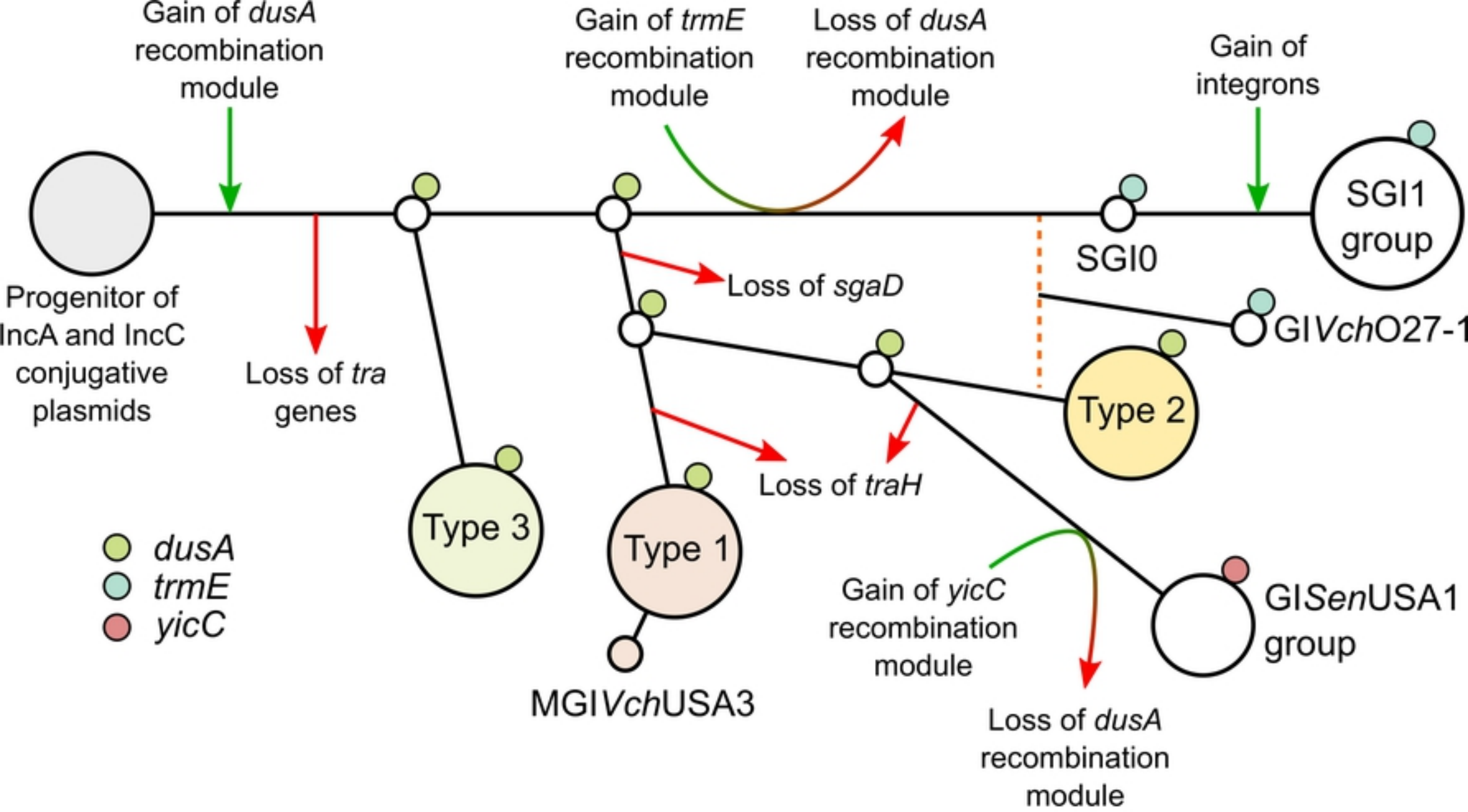
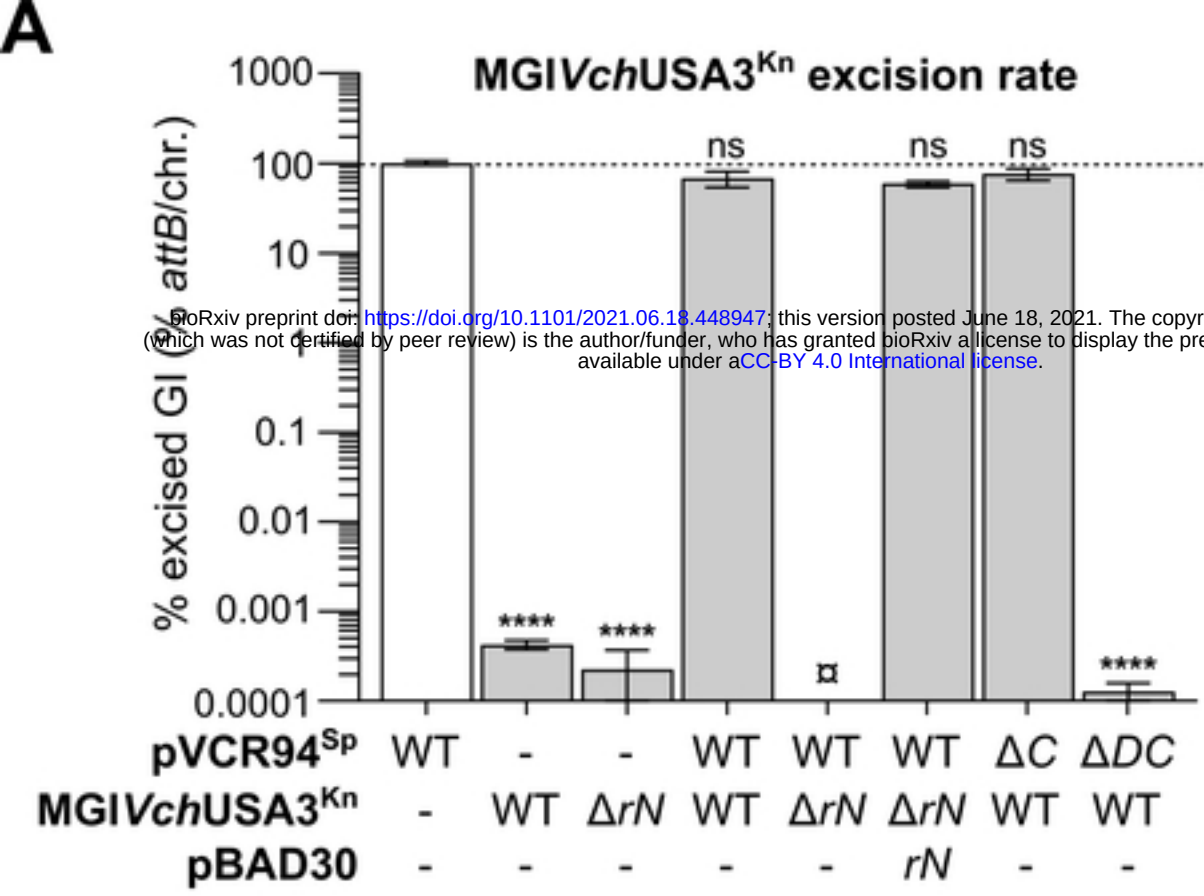
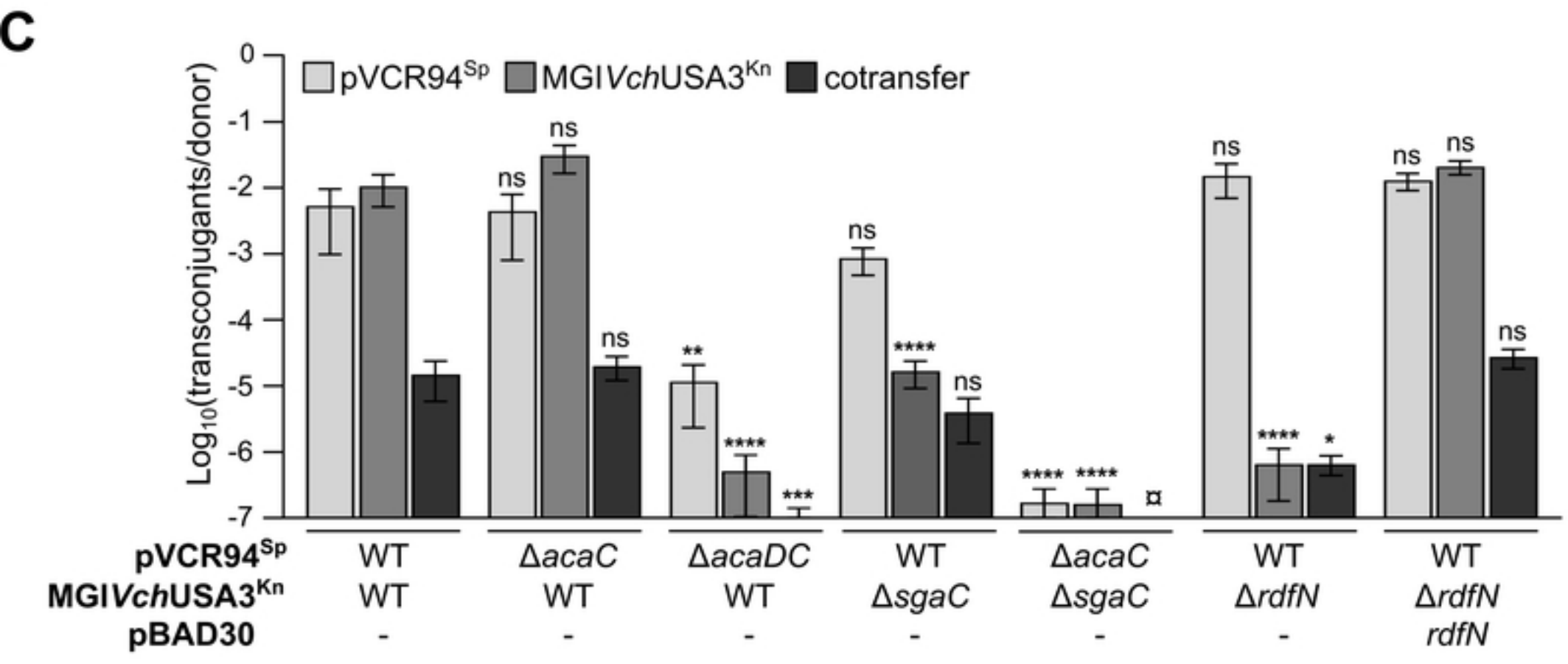
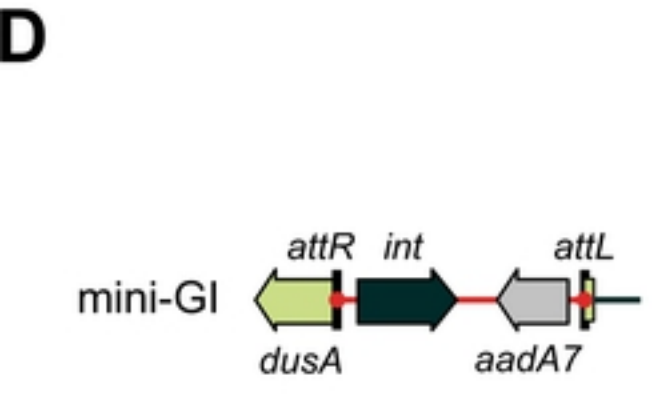

Gain of *dusA* recombination module
Gain of *trmE* recombination module
Loss of *dusA* recombination module
Gain of integrons
Progenitor of IncA and IncC conjugative plasmids
Loss of *tra* genes
Loss of *sgaD*
SGI0
SGI1 group
GI*Vch*O27-1
Type 2
Loss of *traH*
Type 3
Type 1
dusA
trmE
yicC
MGI*Vch*USA3
Gain of *yicC* recombination module
GI*Sen*USA1 group
Loss of *dusA* recombination module

Fig 8







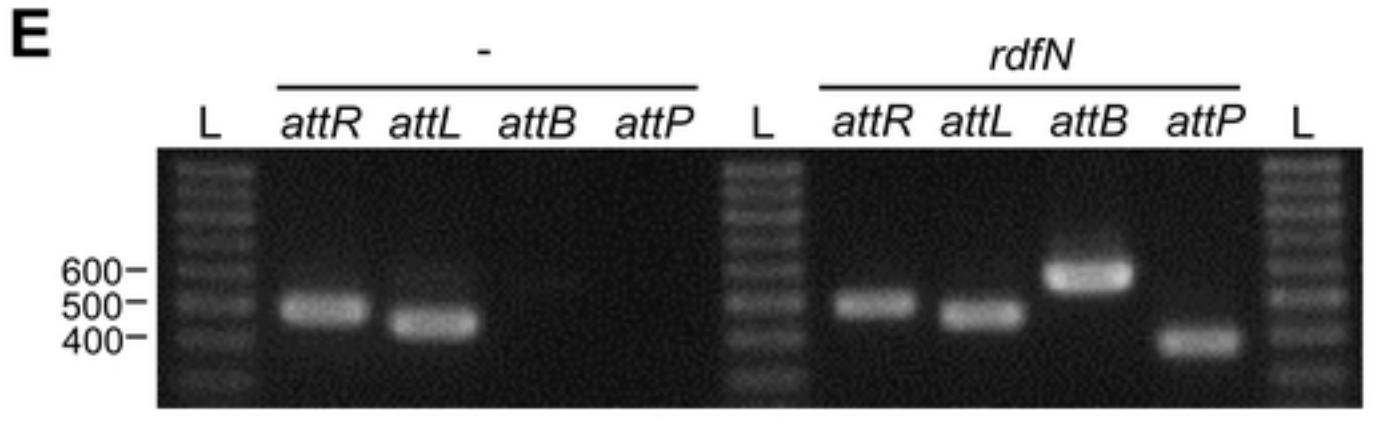


Fig 6

Fig 5