ÉTUDE DE LA RÉGULATION TRANSCRIPTIONNELLE DES ÉLÉMENTS INTÉGRATIFS ET CONJUGATIFS DE LA FAMILLE SXT/R391

par

Dominic Poulin-Laprade

thèse présentée au Département de biologie en vue
de l'obtention du grade de docteur ès sciences (Ph.D.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, Septembre 2015

Le 17 septembre 2015

le jury a accepté la thèse de madame Dominic Poulin-Laprade
dans sa version finale.

Membres du jury

Professeur Vincent Burrus
Directeur de recherche
Département de biologie

Professeur Jan Roelof van der Meer
Évaluateur externe
Université de Lausanne

Professeur Sébastien Rodrigue
Évaluateur interne
Département de biologie

Professeur François Malouin
Évaluateur interne
Département de biologie

Professeur Ryszard Brzezinski
Président-rapporteur
Département de biologie

Evolution is a process of constant branching and expansion.

- Stephen Jay Gould

SOMMAIRE

Les éléments intégratifs et conjugatifs (ICE) de la famille SXT/R391 sont reconnus pour leur rôle prépondérant dans la propagation de la résistance aux antibiotiques parmi des populations de Gammaproteobactéries, en particulier chez *Vibrio cholerae*, l'agent pathogène causant le choléra. Ces éléments génétiques autonomes possèdent tous les gènes nécessaires à leur dissémination au sein d'une population bactérienne et s'intègrent normalement dans un site précis du chromosome bactérien. L'activateur SetCD et la machinerie de conjugaison encodée par les ICE permettent non seulement leur transfert conjugatif, mais également la mobilisation d'îlots génomiques, les MGI (*mobilizable genomic islands*). Lorsque leur transfert est enclenché sans excision au préalable, les MGI et les ICE peuvent mobiliser plusieurs centaines de kb d'ADN chromosomique adjacent à leurs sites d'insertions. Cet ADN mobilisé peut alors recombiner avec le génome de la cellule réceptrice, aboutissant à des remplacements d'allèles. En plus du squelette de gènes conservés de cette famille d'ICE, ces éléments portent une cargaison d'ADN variable qui peut coder pour des fonctions adaptatives potentiellement avantageuses pour l'hôte bactérien. Les ICE SXT/R391 portent également les gènes codant pour un système de recombinaison qui promeut la diversité de la famille en générant des ICE hybrides. Ces éléments mobiles sont extrêmement stables dans les populations bactériennes. Cette stabilité est attribuable à leur intégration au chromosome et à plusieurs composantes qu'ils contiennent, par exemple les systèmes toxine-antitoxine de la cargaison d'ADN variable ou encore le système conservé de partition des éléments excisés.

La majorité des gènes portés par les ICE SXT/R391 est contrôlée par leur système de régulation qui se situe au cœur de ce projet doctoral. Ce système de régulation

comprend SetR, le répresseur responsable du maintien de l'état quiescent dans lequel l'ICE est intégré au chromosome et est propagé verticalement dans la population bactérienne, c'est-à-dire au rythme de la réplication chromosomique et de la division cellulaire. Lorsque l'ADN bactérien est endommagé, il y a activation de la réponse SOS de réparation de l'ADN par RecA, un facteur de l'hôte, qui induit parallèlement l'autoprotéolyse de SetR, levant ainsi la répression exercée sur les gènes *setC* et *setD.* Ces derniers codent pour SetCD, le complexe activateur des ICE SXT/R391 qui active l'expression de la machinerie de conjugaison ainsi que d'autres fonctions codées par ces ICE.

Ce projet doctoral a permis l'identification de nouvelles composantes importantes pour la régulation des ICE SXT/R391. Premièrement, nous avons généré par génie génétique plusieurs mutants qui ont permis de caractériser CroS par des essais de transfert conjugatif, de PCR quantitatif en temps réel (qRT-PCR) et d'expression avec le gène rapporteur *lacZ*. Nous avons déterminé que CroS est un régulateur transcriptionnel qui, avec SetR, constitue un interrupteur génétique permettant l'induction du transfert conjugatif dépendante de RecA. Nous avons également validé par gel à retardement la liaison par SetR et CroS d'un site opérateur additionnel. Des essais β-galactosidase ont montré que ce site contribue à la répression des gènes *croS*, *setC* et *setD*. De plus, les résultats de ce projet doctoral ont clarifié certains points concernant la régulation par SetCD. Des essais d'immunoprécipitation de la chromatine couplée à la digestion avec une exonucléase (ChIP-exo) combinés avec le séquençage de l'ARN (RNA-seq) et la détermination des sites +1 d'initiation de la transcription (5'-RACE et extension d'amorces) ont permis d'établir le régulon de SetCD chez les ICE SXT/R391 et chez les MGI qu'ils mobilisent. La nécessité de SetCD dans la cellule réceptrice pour qu'il y ait intégration de l'ICE de manière site-spécifique dans l'extrémité 5' du gène *prfC* a été mise en évidence à l'aide de la construction de mutants, d'essais de transfert conjugatif, de buvardage de type Southern, d'électrophorèse en champs pulsés et de PCR en temps réel. Nous avons

également observé, grâce à des essais de PCR quantitatif et d'activité β-galactosidase, une boucle de rétroaction positive médiée par l'activation de l'excision et de la réplication de l'ICE par SetCD. En somme, ce projet doctoral a mené à une meilleure compréhension des composantes et des mécanismes en scène pour la gouvernance de cette famille d'ICE qui sont, entres autres, d'importants vecteurs de la dissémination des résistances aux antibiotiques.

Mots-clés : résistance aux antibiotiques, éléments intégratifs et conjugatifs, îlots génomiques, régulation transcriptionnelle, Gammaproteobacteria, Vibrio cholerae, ICE SXT/R391

REMERCIEMENTS

Je commence par remercier Vincent qui a rendu ces travaux possibles. Il m'a insufflé rigueur et intégrité, et m'a permis d'apprendre à mon rythme et de façon autonome. Vive tes eurêka qui ont débloqué le projet à plusieurs reprises et ont contribué à l'élaboration des conclusions de cette étude.

Je remercie le Fonds de recherche du Québec – Nature et technologies pour m'avoir financé pour une partie de ce projet doctoral.

Je remercie également Ryszard pour ton mentorat qui m'a guidée à des moments clés dès ma première année universitaire. Merci aussi à François et Sébastien pour vos précieux conseils.

Merci aux membres présents et passés du laboratoire, spécialement Geneviève, Daniela, Christelle, Eric, Nicolas. Il est tellement agréable de travailler avec une belle équipe. Je remercie également Dominick Matteau pour sa contribution au 2^{e} article.

Je serai éternellement reconnaissante envers ma famille, Marie-Thérèse et Réjean, Amélie et Evelyne, qui m'ont toujours encouragée et supportée tout au long de ce parcours. Je tiens aussi à remercier les Brousseau pour leur chaleur et leur accueil.

Merci à mes fidèles amies, particulièrement Roxanne, Francisca, Marie-Sol, Nancy et Gabrielle qui m'ont aidée à garder l'équilibre.

Le dernier, mais non le moindre, merci à Jean-Philippe, mon support quotidien, mon partenaire.

TABLE DES MATIÈRES

LISTE DES ABRÉVIATIONS

ADN : acide désoxyribonucléique

ADNsb : acide désoxyribonucléique simple-brin

ADNc : acide désoxyribonucléique complémentaire

ARNm : ARN messager

ATP : adénosine triphosphate

CHEF-PFGE : *contour-clamped homogeneous electric field pulsed field gel electrophoresis*

ChIP-exo : *chromatin immunoprecipitation coupled with exonuclease digestion*

croS : *cro-like repressor of SXT*

CTXΦ : bactériophage filamenteux et lysogène retrouvé chez *V. cholerae.* Code pour la toxine cholérique (*ctxAB*).

DNase I : *deoxyribonuclease I*

HGT : *horizontal gene tranfer*

ICE : *integrative and conjugative element*

kDa : kilo Dalton

pb : paire de base

kb : kilo paire de base

Mb : méga paire de base

MGI : *mobilisable genomic island*

OMS : organisation mondiale de la santé

OL : *operator left*. Opérateur gauche du bactériophage λ composé des sites OL1, OL2 et OL3.

OR : *operator right*. Opérateur droit du bactériophage λ composé des sites OR1, OR2 et OR3.

ORF : *open reading frame*

qRT-PCR : *quantitative real time polymerase chain reaction*

RNAP : *RNA polymerase*

RNA-seq : *RNA sequencing*

setR : *SXT excision and transfer repressor*

setCD : *SXT excision and transfer activator subunits C and D*

SXT : Sulfaméthoxazole et triméthoprime, aussi appelé cotrimoxazole lorsqu'utilisé en combinaison. L'ICE SXT a été nommé ainsi, car il possède les gènes de résistance à ces deux antibiotiques.

TCP : *toxin co-regulated pilus*

VPI : *Vibrio pathogenicity island.*

VSP-1 : *Vibrio seventh pandemic island-1.*

VSP-2 : *Vibrio seventh pandemic island-2*

5'-RACE : *5' rapid amplification of cDNA ends*

LISTE DES TABLEAUX

Chapitre 2

1er article

2e article

Annexes

LISTE DES FIGURES

Chapitre 1

Chapitre 2

1er article

2[e] article

Chapitre 3

Annexes

CHAPITRE 1

INTRODUCTION GÉNÉRALE

1.1. Éléments intégratifs et conjugatifs bactériens

Les bactéries sont parmi les premiers organismes vivants à avoir colonisé la Terre. Elles ont traversé toutes les ères et sont encore à ce jour omniprésentes sur terre et en mer. Leur succès est attribuable à leur redoutable capacité d'adaptation aux environnements hostiles, médié en partie par leur force du nombre, mais aussi grandement grâce à un processus d'échange de matériel génétique entre individus appelé le transfert horizontal de gènes (*Horizontal Gene Transfer*, HGT). Le HGT correspond à l'acquisition d'information génétique provenant d'un organisme sans en être le descendant direct, par opposition au transfert vertical qui correspond à l'héritage de l'ensemble de l'information génétique de la cellule mère. Le HGT peut se produire selon trois voies distinctes : la transformation, la transduction et la conjugaison (Thomas and Nielsen, 2005). La transformation consiste en l'internalisation d'ADN libre situé à proximité de la bactérie (Lorenz and Wackernagel, 1994; Meibom, 2005). La transduction implique l'infection d'une bactérie par une capside de phage contenant le génome phagique ou une portion de celui de la bactérie donneuse (Zinder and Lederberg, 1952). La conjugaison, quant à elle, est une voie de HGT reposant sur le transfert de matériel génétique au travers d'un pore de conjugaison établi entre une bactérie donneuse et une ou plusieurs réceptrices en contact direct. Ce mode de transfert est particulièrement efficace étant donné la grande quantité de matériel génétique transloquée par évènement de transfert et le large spectre d'hôtes possibles (Halary *et al*., 2010; Kloesges *et al*., 2011; Norman *et al*., 2009). La conjugaison peut être orchestrée par des plasmides conjugatifs, mais

aussi par des éléments intégratifs et conjugatifs (*Integrative and Conjugative Elements,* ICE) (Burrus *et al.*, 2002). Les ICE sont très abondants chez les procaryotes (Guglielmini *et al*., 2011) et sont présents autant chez les bactéries à Gram positif à haut et à bas G+C que chez les protéobactéries et les Bacteroidetes (Bordeleau *et al*., 2012; Burrus *et al*., 2002; Ghinet *et al*., 2011; Guglielmini *et al*., 2011).

Les ICE ont deux états, soit dormant au sein du chromosome comme le ferait un prophage, soit excisé et engagé dans leur transfert de cellule à cellule comme le ferait un plasmide conjugatif (Burrus *et al*., 2002). Les ICE assurent leur hérédité verticale en tirant avantage de la machinerie de réplication de leur hôte en s'intégrant dans un de ses réplicons (Burrus *et al*., 2002). Sous la forme linéaire intégrée, l'ICE est bordé de séquences *attL* et *attR* (Figure 1). Ces sites sont les cibles de l'intégrase (Int) et l'excisionase (Xis), toutes deux codées par l'ICE, qui permettent la recombinaison entre *attL* et *attR* menant à l'excision de l'ICE sous sa forme circulaire qui porte un site *attP*, et laissant un site *attB* dans le chromosome de l'hôte (Figure 1) (Burrus and Waldor, 2003; Hochhut and Waldor, 1999). Une fois l'élément circularisé, il y a assemblage du relaxosome, un complexe nucléoprotéique contenant la relaxase, des facteurs protéiques auxiliaires et un locus de l'ICE appelé l'*oriT* (*origin of transfer*). Le transfert est amorcé par la relaxase qui reconnaît et scinde un des deux brins de l'*oriT* au site *nic* (Lanka and Wilkins, 1995). Le relaxosome dirige ensuite, avec l'aide d'une protéine de couplage, le transport de l'ICE jusqu'au pore de conjugaison qui relie les cellules donneuse et réceptrice (Alvarez-Martinez and Christie, 2009; Llosa *et al*., 2002). Le pore est généralement un système de sécrétion de type IV (Alvarez-Martinez and Christie, 2009). Une fois dans la cellule réceptrice, l'élément simple-brin parasite encore une fois la machinerie réplicative de l'hôte pour l'obtention d'un élément double-brin (Lanka and Wilkins, 1995). L'intégrase catalyse la recombinaison entre la séquence *attP* présente sur l'ICE circulaire et la séquence *attB* présente dans le génome de l'hôte, régénérant ainsi les séquences *attL* et *attR*

qui bordent l'ICE intégré (Burrus and Waldor, 2003; Hochhut and Waldor, 1999). La réplication en cercle-roulant intercellulaire associée à la conjugaison supporte la conservation d'une copie de l'ICE dans la cellule donneuse (Garcillán-Barcia *et al.*, 2009). Le matériel génétique contenu dans l'élément peut donc être disséminé à un rythme exponentiel, souvent encouragé par une pression sélectionnant une caractéristique donnée, par exemple la résistance à un ou plusieurs antibiotiques. En opposition à cela, une étude a montré que les donneuses de l'ICE*clc,* qui appartient à une autre famille d'ICE, meurent peu après le transfert (Reinhard *et al.*, 2013). En effet, des expériences de microscopie ont montré que le transfert de ICE*clc* est initié par des cellules tc (*transfer competent cells*) qui croissent très lentement, qui ont une morphologie anormalement longue et qui finissent par lyser (Reinhard *et al*. 2013). L'inhibition de la croissance des cellules tc est due à un petit gène appelé *shi,* ainsi qu'à la région 5' d'un gène avec une homologie partielle pour *parA*. Le mode d'action de *parA*-*shi* est encore mal compris, mais Shi possède une faible homologie pour un domaine eucaryote de canal calcique voltage-dépendant ce qui suggère que Shi influence le potentiel membranaire. Le sacrifice des cellules donneuses tc est confiné à une petite proportion de la population (3 % à 5 %), donc la perte de *fitness* est indétectable à l'échelle de la population.

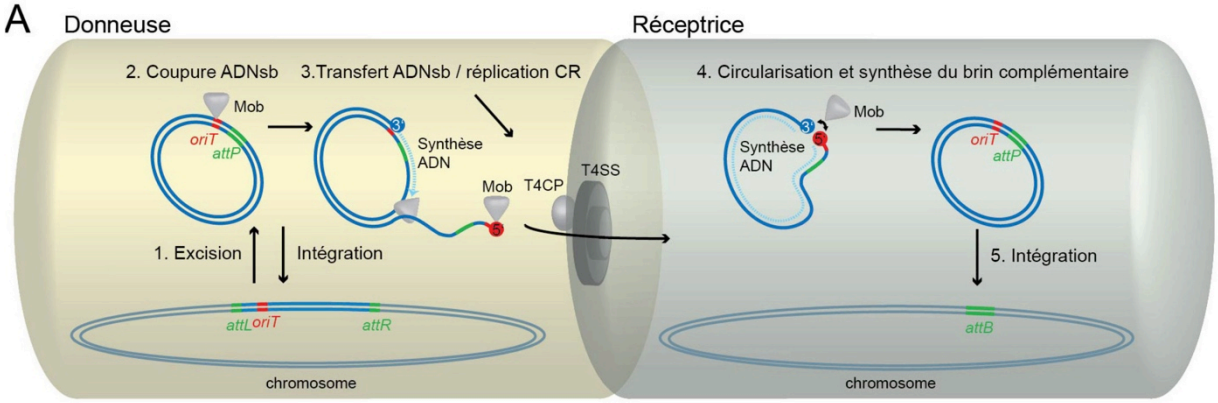

Figure 1. Modèle de transfert par conjugaison des ICE. Dans la cellule donneuse, l'excision de l'ICE résulte de la recombination entre les sites *attL* et *attR* (1). À la suite de la circularisation, le relaxosome (Mob) reconnaît l'origine de transfert (*oriT*) et scinde un des deux brins (2), tout en restant lié de manière covalente à l'extrémité 5' du brin scindé (2). La relaxosome recrute des facteurs réplicatifs de l'hôte, initiant ainsi la réplication en cercle-roulant (3). Le complexe nucléoprotéique interagit ensuite avec une protéine de couplage qui génère l'énergie pour la translocation de l'ICE à travers le pore de conjugaison (T4SS). Une fois dans la cellule réceptrice, Mob relie la molécule d'ADN simple-brin et le brin complémentaire est synthétisé (5). L'intégration dans le chromosome de la cellule réceptrice résulte de la recombinaison entre *attP*, le site d'attachement situé sur la molécule circulaire et *attB*, le site spécifique d'intégration dans le chromosome bactérien. Adapté de Bordeleau *et al.* (2012).

Les ICE sont modulaires et codent, entre autres, pour les protéines nécessaires à leur fonctions de base, c'est-à-dire à leur insertion, leur excision, leur transfert par conjugaison et à la régulation de tous ces processus (Burrus and Waldor, 2004a; Wozniak *et al.*, 2009). Les ICE peuvent également contenir une cargaison de gènes variables pouvant aider les bactéries à faire face à une variété d'environnements potentiellement hostiles. Par exemple, certains de ces éléments codent pour :

1) des régulateurs d'un messager secondaire (c-di-GMP) important pour la virulence, la mobilité et la formation de biofilm chez *V. cholerae* (Bordeleau *et al.*, 2010);

2) des enzymes de voies de dégradation de composés aromatiques (van der Meer and Sentchilo, 2003; Nishi *et al.*, 2000; Ravatn *et al.*, 1998);
3) des enzymes de biosynthèse de thiamine, biotine et nicotinate, et de fixation d'azote (Sullivan and Ronson, 1998);
4) des gènes de résistance à des antibiotiques (Spagnoletti *et al.*, 2014) ou à des métaux lourds (Coetzee *et al.*, 1972).

De plus, les ICE ont fréquemment un large spectre d'hôtes, facilitant la transmission de ces gènes fonctionnels parmi les populations bactériennes. Par exemples, les ICE de type Tn*916* ont été retrouvés chez des *Proteobacteria*, des *Actinobacteria* et des *Firmicutes* (Roberts and Mullany, 2009), tandis que ceux de type ICE*clc* peuvent se transférer vers des *Gammaproteobacteria* et vers des *Betaproteobacteria* (Springael *et al.*, 2002). Les ICE SXT/R391, quant à eux, ont été retrouvés dans plusieurs genres appartenant aux *Gammaproteobacteria* (Wozniak *et al.*, 2009). La famille d'ICE SXT/R391 a joué un rôle clé dans la dissémination de la résistance aux antibiotiques chez l'agent causal du choléra. On comprend donc facilement l'engouement pour l'étude de cette classe d'éléments génétiques mobiles qui contribuent à l'évolution et à l'adaptation des populations bactériennes.

1.2. Hôtes bactériens des ICE SXT/R391

1.2.1. *Vibrio cholerae*

1.2.1.1. Généralités

Vibrio cholerae est un bacille incurvé à Gram négatif, mobile grâce à son flagelle polaire. C'est en 1854 qu'il a été décrit pour la première fois par Filipo Pacini, un anatomiste italien, comme l'agent étiologique du choléra (Pacini, 1854). Contrairement à la majorité des bactéries qui possède un seul chromosome, *V. cholerae* en possède deux (Figure 2). La plupart des gènes métaboliques sont retrouvés sur le chromosome I (*large chromosome*, ~3 Mb). Le chromosome II, quant à lui, est plus petit (*small chromosome*, ~1 Mb) et contient également des gènes importants pour les fonctions métaboliques des cellules (Heidelberg *et al.*, 2000). Selon les travaux de Heidelberg *et al.* (2000), le chromosome II aurait initialement été un méga plasmide acquis avant la différenciation de la famille des *Vibrionaceae*.

Le génome de *V. cholerae* a été grandement façonné par le transfert horizontal de matériel génétique comme le montre la Figure 2 qui indique la position d'îlots génomiques présents dans différentes souches (Chun *et al.*, 2009). Les îlots génomiques sont des segments d'ADN présents dans certaines souches d'une même espèce et auxquels on attribue des propriétés de mobilité passées ou présentes (Juhas *et al.*, 2009). Parmi ces îlots, on retrouve un prophage (CTXΦ), des îlots de pathogénicité (VPI, VSP-1, VSP-2), ainsi que des éléments conjugatifs et intégratifs (SXT). Certains îlots non autonomes pourraient avoir été acquis par mobilisation ou par transformation naturelle, c'est-à-dire par acquisition d'ADN présent dans

l'environnement. En effet, la compétence naturelle de *V. cholerae* est stimulée par un contact avec la chitine, un polymère naturel présent dans les carapaces de crustacés et d'arthropodes, ses hôtes environnementaux (Meibom, 2005).

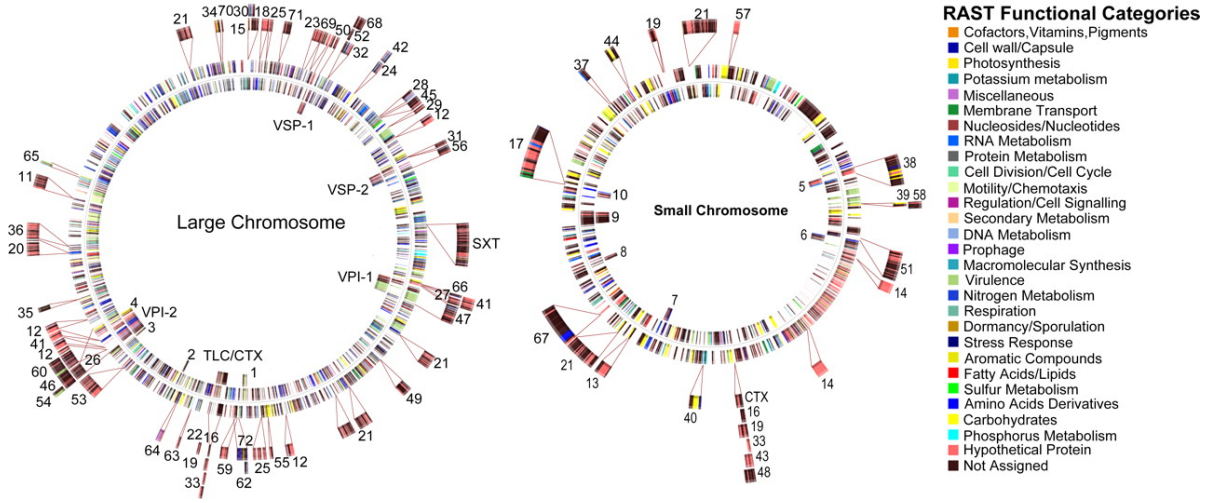

Figure 2. Représentation des chromosomes de *V. cholerae*. Pour chaque chromosome, les cercles centraux montrent les gènes de *V. cholerae* O1 El Tor N16961. Les îlots génomiques retrouvés chez N16961 sont indiqués à l'intérieur des cercles, tandis que les îlots absents chez N16961, mais présents chez d'autres *V. cholerae* sont représentés à l'extérieur des cercles. Tiré de Chun *et al*. (2009).

1.2.1.2. Habitats

V. cholerae habite normalement les mers et océans. La salinité et la chaleur des estuaires et des côtes lui conviennent particulièrement bien. Elle peut également se développer en eau douce, si celle-ci est assez chaude et riche en nutriments organiques (Huq *et al.*, 1984; Singleton *et al.*, 1982a, 1982b). Il a été démontré dans de nombreuses études que le réchauffement climatique favorise le développement des *Vibrio* potentiellement pathogènes dans l'hémisphère Nord (Böer *et al.*, 2013;

Sterk *et al.*, 2015; Vezzulli *et al.*, 2012, 2013). Lorsque son environnement n'est pas propice à son développement, *V. cholerae* entre dans un état de dormance appelé VBNC (*Viable But Not Cultivable*) (Roszak and Colwell, 1987). Son métabolisme ralentit et elle prend alors une forme rétrécie et sphérique qui lui permet de résister à des variations de température, de salinité et de disponibilité en matière organique, tout en demeurant infectieuse (Roszak and Colwell, 1987).

En milieux aquatiques, *V. cholerae* peut être sous forme planctonique, c'est-à-dire en forme de virgule et mobile grâce à son flagelle polaire (Figure 3). Cependant, elle est principalement retrouvée au sein de biofilms en association avec ses réservoirs environnementaux, c'est-à-dire les crustacés, les algues, les coraux ou encore les poissons (Butler and Camilli, 2005). Son attachement à la chitine du plancton lui permet d'être emportée sur de grandes distances grâce aux courants marins (Colwell, 1996). De cette façon, les épidémies de choléra peuvent se propager d'un continent à l'autre. *V. cholerae* peut également se développer transitoirement dans le tractus gastro-intestinal humain (Kaper *et al.*, 1995). L'ingestion de souches pathogènes peut mener au développement du choléra, une diarrhée aigüe dont on peut mourir en quelques heures en absence de traitement (Organisation mondiale de la Santé, 2014). Dans les régions où les conditions sanitaires ne sont pas adéquates, les épidémies de choléra permettent la multiplication et la dissémination efficace de *V. cholerae*. Généralement, cette bactérie n'est que de passage chez l'humain, mais certains porteurs sains en région endémique ont été déclarés (Colwell, 1996). En Amérique du Nord, bien que nos conditions climatiques s'opposent au développement endémique du choléra et/ou que notre système de traitement d'eau soit efficace, nous ne sommes pas à l'abri des souches de *V. cholerae* pathogènes qui peuvent se retrouver en grand nombre sur les fruits de mer que nous consommons. Quelques cas de choléra acquis sur le territoire des États-Unis sont déclarés chaque année. Ces cas sporadiques de choléra sont généralement attribués

à la consommation de fruits de mer crus en provenance de la côte américaine du golfe du Mexique (Loharikar *et al.*, 2015).

1.2.1.3. Choléra

1.2.1.3.1. Généralités

Le choléra est rencontré sur tous les continents et est un indicateur clé de l'insuffisance du développement social (Organisation mondiale de la Santé, 2014). La majorité des épidémies survient dans des zones en voie de développement où l'accès à l'eau potable est problématique et où les conditions sanitaires sont défaillantes. C'est en 1849 que John Snow, un médecin londonien, a établi que les épidémies de choléra se propageaient principalement par la consommation d'eau contaminée par des matières fécales (Snow, 1855). Un patient atteint de choléra peut mourir en quelques heures s'il n'est pas hydraté correctement. Généralement, les pays touchés signalent un taux de mortalité de moins de 5 %. Cependant ce taux peut atteindre 50 % lors d'épidémie dans des zones où la population est la plus vulnérable. L'OMS estime qu'il y aurait entre 3 et 5 millions de cas et entre 100 000 et 120 000 décès par an (Organisation mondiale de la Santé, 2014; Zuckerman *et al.*, 2007).

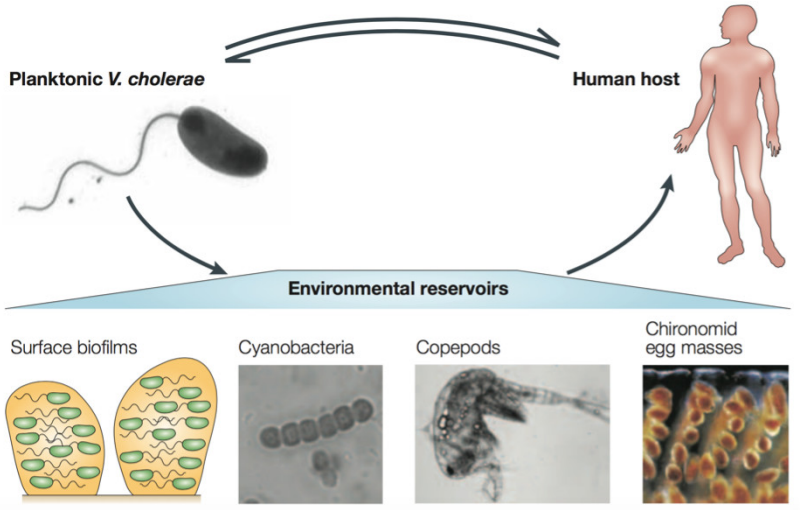

Figure 3. Cycle de vie de *V. cholerae*. La forme planctonique *de V. cholerae*, qui est relâchée dans les selles liquides humaines, est centrale à son cycle de vie. Ces bactéries sont très mobiles et peuvent être ingérées par un nouvel hôte humain, ou s'associer à des surfaces abiotiques, à des copépodes, à des algues ou encore à des œufs de Chironomides présents dans l'environnement. Les *V. cholerae* environnementaux peuvent se dissocier de leur hôte et être ingérées par un humain ou encore coloniser un nouvel hôte environnemental. Ceux-ci peuvent également être ingérés par un humain, causer une infection et mener au relâchement de *V. cholerae* planctoniques, complétant ainsi le cycle de vie. Tiré de Butler et Camilli (2005).

1.2.1.3.2. Pathogenèse et traitements

Le choléra est contracté par l'ingestion de 10^6 à 10^{11} bactéries toxinogènes, selon l'acidité gastrique (Krauss *et al.*, 2003). *V. cholerae* est une bactérie non invasive et sa virulence est attribuable à la toxine cholérique codée par le bactériophage filamenteux CTXΦ (Waldor and Mekalanos, 1996). Celui-ci infecte son hôte bactérien et s'intègre dans son génome, établissant un programme lysogène stable. Une fois ingérées, les bactéries migrent vers l'épithélium intestinal grâce à leur flagelle, puis un pili spécifique (*Toxin Co-Regulated Pilus*, TCP) est synthétisé et permet l'adhésion

de la bactérie aux cellules mucosales (Herrington *et al.*, 1988). Par la suite, la toxine est sécrétée dans la lumière intestinale et lie les entérocytes qui l'internalisent par endocytose, ce qui aboutit à une excrétion massive d'eau et d'électrolytes, d'où la diarrhée sévère. Le meilleur traitement pour survivre au choléra est l'apport en eau et en électrolytes par perfusion. Certains vaccins oraux sont disponibles, mais ils sont onéreux et la protection immunitaire est limitée dans le temps (Hall and Sack, 2015). Ces vaccins sont parfois administrés aux populations à risque, mais ils sont principalement utilisés par les voyageurs ou le personnel hospitalier.

1.2.1.3.3. Épidémiologie du choléra

Les premières descriptions du choléra ont été élaborées dans un traité de chirurgie de l'Inde ancienne appelé *Sushruta Samshita* environ 500 à 400 ans av. J.-C. (Colwell, 1996). Ces évidences montrent que le choléra existait déjà au temps de l'Antiquité, mais c'est seulement en 1817 que les relevés de pandémie ont vu le jour (Colwell, 1996). La deuxième pandémie (1829 à 1851) a sévi en Russie, puis a traversé l'Atlantique jusqu'au fleuve Saint-Laurent. À l'été 1832, une épidémie de choléra a fait rage au Québec, entraînant la mort d'environ 2000 personnes (Sendzik, 1997). Cette épidémie s'est ensuite étendue jusqu'à La Nouvelle-Orléans, ce qui est un indicateur des mauvaises conditions sanitaires de l'époque. Heureusement, nos conditions climatiques s'opposent au développement endémique du choléra.

La 7^{e} pandémie de choléra a débuté en 1961 lorsque les souches de biotype El Tor sont devenues endémiques en Indonésie. Ces souches se sont propagées en Asie et en Afrique, remplaçant les souches de biotype classique de choléra endémique (Faruque *et al.*, 1998). En 1991, *V. cholerae* El Tor a causé des épidémies majeures en Amérique du Sud et en Afrique, tuant plusieurs milliers de personnes (Colwell,

1996). À cette époque, aucune souche de sérogroupe non-O1 n'avait causé une épidémie. En 1992, des épidémiologistes ont observé l'émergence d'une nouvelle souche de *V. cholerae* pathogène portant le sérogroupe O139 (Ramamurthy *et al.*, 1993). Initialement observée dans la ville portuaire de Madras, en Inde, *V. cholerae* O139 s'est propagée dans le nord de l'Inde, puis au Bangladesh et en Thaïlande (Calia *et al.*, 1994). Depuis 1993, *V. cholerae* O139 a également été rapportée au Népal, au Burma, en Malaisie, en Arabie Saoudite, en Chine et au Pakistan (Ramamurthy *et al.*, 1993). *V. cholerae* O139 ressemble grandement à *V. cholerae* El tor O1 (Cholera Working Group, 1993), mais s'en distingue par deux caractéristiques principales : le remplacement du sérogroupe O1 par le sérogroupe O139, impliquant une capsule (Waldor *et al.*, 1994) et un patron distinct de résistance aux antibiotiques (Nair *et al.*, 1994; Waldor *et al.*, 1996). C'est en 1996 que ce nouveau patron de résistances a été associé à la présence d'un élément génétique mobile nommé SXT (Waldor *et al.*, 1996). De nos jours, les éléments de type SXT, regroupés dans la famille SXT/R391, sont ubiquitaires chez les souches cliniques de *V. cholerae*, et des variants de SXT avec le même patron de résistance ont été retrouvés chez des *V. cholerae* de sérotypes O139, O1 et non-O1 (Burrus *et al.*, 2006). ICE*Vch*Ind5 est reconnu comme étant l'ICE le plus rencontré chez les souches de *V. cholerae* El Tor circulant à travers le monde (Ceccarelli *et al.*, 2011a, 2011b, 2011c; Spagnoletti *et al.*, 2014). Ces éléments mobiles ont grandement contribué à la dissémination de multirésistances aux antibiotiques et sont reconnus comme un trait distinctif des souches de la 7^{e} pandémie de choléra (Figure 4) (Mutreja *et al.*, 2011; Spagnoletti *et al.*, 2014).

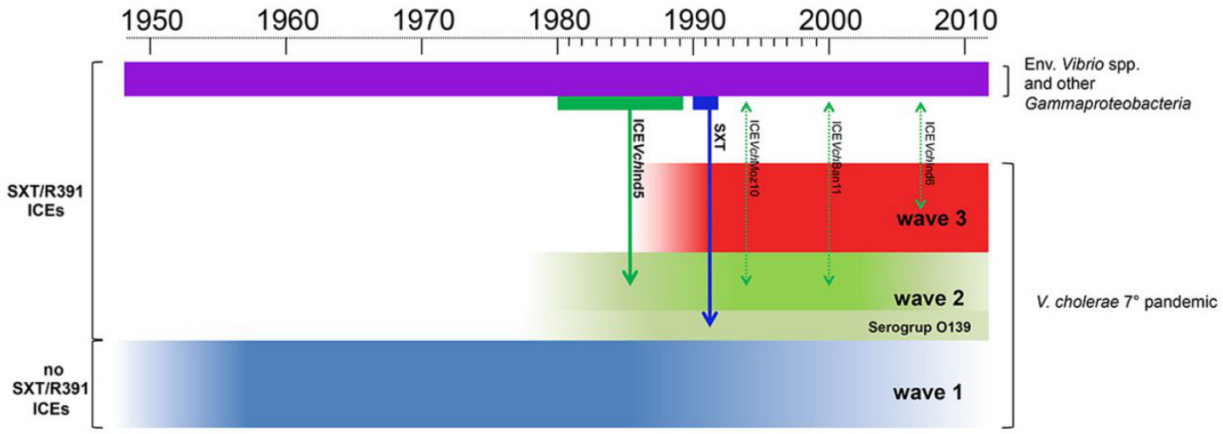

Figure 4. Schématisation de l'acquisition des ICE SXT/R391 dans la lignée de *V. cholerae* causant la 7e pandémie. Les trois vagues pandémiques de la transmission de choléra sont respectivement représentées en bleu, en vert et en rouge, tel qu'établi par Mutreja *et al.* (2011). Le sérogoupe O139 est représenté comme un sous-groupe de la 2e vague en vert pâle. Les lignes verticales indiquent les évènements d'acquisition d'ICE de l'environnement (ICE*Vch*Ind5 et SXT) et des ICE dérivés de réarrangements par recombinaison (ICE*Vch*Moz10, ICE*Vch*Ban11, ICE*Vch*Ind6). Les souches donneuses environnementales hypothétiques sont représentées par la barre mauve. Tiré de Spagnoletti *et al.* (2014).

En janvier 2010, un fort tremblement de terre secoua Haïti et dévasta ses infrastructures sanitaires déjà chancelantes. Neuf mois plus tard, une épidémie de choléra particulièrement meurtrière fut déclarée et s'étendit sur tout le territoire d'Hispaniola. Le choléra n'ayant pas été rapporté sur cette île depuis près d'un siècle, la population était particulièrement vulnérable de par l'absence d'immunité acquise (Ceccarelli *et al.*, 2011c). En date du 19 juin 2015, 744 147 cas et 8964 morts ont été répertoriés (Ministère de la santé publique et de la population d'Haïti, 2015). Le choléra est maintenant endémique sur l'île d'Hispaniola (Alam *et al.*, 2014, 2015; Kahler *et al.*, 2015). Comme ailleurs dans le monde, la majorité des souches cliniques possède un ICE SXT/R391 <99 % identique à ICE*Vch*Ind5 (Annexe 1) (Ceccarelli *et al.*, 2013).

1.2.2. Autres hôtes des ICE SXT/R391

Bien que la majorité ait été retrouvée chez *V. cholerae*, de nombreux ICE SXT/R391 résident dans le chromosome d'hôtes appartenant aux familles des *Enterobacteriaceae* et des *Vibrionaceae* (Annexe 1). Les hôtes des ICE SXT/R391 sont variés et peuvent se développer dans les tractus gastro-intestinal et urinaire de l'homme et/ou encore dans l'environnement en association avec la vie marine, sous forme planctonique ou dans les sédiments. À ce jour, la base de données ICEberg (Bi *et al.*, 2012) recense 89 ICE SXT/R391. Leur caractérisation est souvent sommaire (présence/absence du gène de l'intégrase, antibiogramme), mais plusieurs ont été séquencés (Rodríguez-Blanco *et al.*, 2012; Taviani *et al.*, 2008; Wozniak *et al.*, 2009). Ces éléments partagent un même ensemble de gènes conservés assurant leurs fonctions de base, mais se distinguent entre eux par leur cargaison d'ADN variable.

1.2.2.1. Incidence clinique

Les premiers ICE SXT/R391 découverts ont attiré l'attention à cause de la résistance aux antibitioques qu'ils confèrent et propagent (Coetzee *et al.*, 1972; Waldor *et al.*, 1996). Ces éléments ont été retrouvés dans plusieurs espèces de *Vibrio* pathogènes autre que *V. cholerae*. Par exemple, ICE*Vfl*Ind1 a été isolé de *V. fluvialis* ayant causé une diarrhée infantile (Ahmed *et al.*, 2005). R391, l'autre élément prototype de la famille SXT/R391, a été découvert en 1967 à Pretoria, en Afrique du Sud, dans un isolat clinique de *Providencia rettgeri* (Coetzee *et al.*, 1972). Celui-ci est un pathogène opportuniste habitant le tractus gastro-intestinal humain et pouvant causer des infections urinaires et des diarrhées (Jones and Mobley, 1987; Yoh *et al.*, 2005). R997, ICE*Pmi*Jpn1 et ICE*Pmi*USA1 ont été découverts respectivement en Inde, au Japon et aux États-Unis chez *Proteus mirabilis*, une bactérie pouvant causer des

infections du tractus urinaire (Harada *et al.*, 2010; Matthew *et al.*, 1979; Pearson *et al.*, 2008). Des ICE SXT/R391 ont également été dépistés dans des souches pathogènes infectant les poissons, contribuant ainsi au problème de la résistance aux antibiotiques en pisciculture. En voici quelques exemples : ICE*Pda*Spa1 (*Photobacterium damselae*), ICE*Eni*Spa1 (*Enterovibrio nigricans*), ICE*Val*Por1 (*V. alginolyticus*), ICE*Vsp*Por2 (*V. splendidus*), ICE*Vsc*Spa1 (*V. scophthalmi*), ICE*Sha*Por1 (*Shewanella haliotis*) (Osorio *et al.*, 2008; Rodríguez-Blanco *et al.*, 2012).

Bien que n'ayant pas encore été naturellement retrouvés dans ces souches, les ICE SXT/R391 sont facilement transférables à *Escherichia coli*, *Klebsiella pneumoniae*, *Salmonella enterica* serovar Typhimurium et *Citrobacter koseri* en laboratoire (Harada *et al.*, 2010; Waldor *et al.*, 1996).

1.2.2.2. Incidence environnementale

Certains hôtes des ICE SXT/R391 ont un potentiel pathogène, mais sont également communément rencontrés dans les milieux aquatiques. Tel que décrit précédemment (Figure 3), *V. cholerae* peut coloniser le zooplancton, le phytoplancton et les œufs de chironomidés (Butler and Camilli, 2005). Les éléments ICE*Vfo*And1 et ICE*Msp*And1 ont été retrouvés dans l'océan Indien, respectivement chez *Vibrio fortis* et *Marinomonas ostreistagni* en association avec du mucus corallien (Badhai *et al.*, 2013). pMERPH et ICE*Spu*PO1 ont été identifiés chez *Shewanella putrefaciens* provenant respectivement de sédiments du fleuve Mersey au Royaume-Uni et de l'océan Pacifique (Murray *et al.*, 2001; Peters *et al.*, 1991). Plusieurs isolats d'*Alteromonas macleodii* en provenance de la mer Méditerranée portaient

ICE*Ama*AS1 (López-Pérez *et al.*, 2013; Pinhassi and Berman, 2003; Sass *et al.*, 2001).

1.3. Particularités des ICE SXT/R391

1.3.1. Modules de gènes conservés

Comme la majorité des éléments génétiques mobiles, les ICE SXT/R391 ont une organisation modulaire, c'est-à-dire que les gènes et séquences impliqués dans une fonction donnée sont regroupés dans une même région (Burrus *et al.*, 2002; Toussaint and Merlin, 2002). En effet, cette famille possède un squelette de gènes communs hautement conservés et distribués en modules dont les fonctions sont liées au transfert par conjugaison et à la régulation génétique responsable du bon déroulement du transfert (Figure 5). Les modules d'intégration/excision et de régulation ressemblent à ceux des prophages. En particulier, les fonctions de SetR, Int et Xis des ICE SXT/R391 ressemblent à celles de CI, Int et Xis du bactériophage λ. Ces protéines agissent respectivement comme des répresseurs maintenant l'état quiescent, des intégrases et des facteurs de recombinaison directionnels (excisionases). Les modules permettant la réplication et la conjugaison ressemblent à ceux retrouvés chez les plasmides conjugatifs (Beaber *et al.*, 2002), tandis que le régulateur transcriptionnel qui active l'expression de ces modules, SetCD, partage de l'homologie avec le complexe activateur FlhDC qui gouverne la synthèse du flagelle chez de nombreuses bactéries. Étant au coeur de ma thèse, la région régulatrice et les mécanismes de régulation gouvernant les deux états des ICE SXT/R391 seront extensivement développés plus loin dans l'introduction, dans les chapitres 2 et 3, ainsi que dans la conclusion.

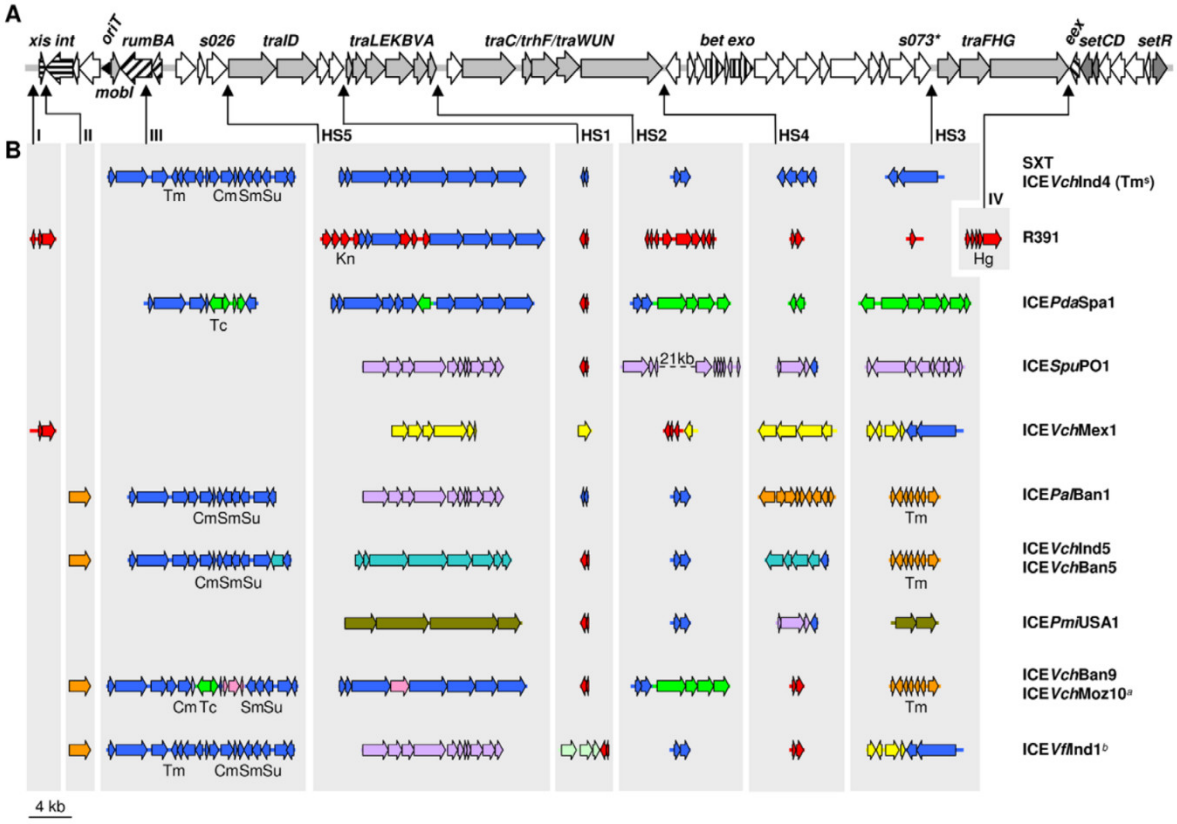

Figure 5. Structure génique des ICE SXT/R391. (A) Le squelette de gènes conservés est représenté par les flèches larges. Les ORF lignés sont impliqués dans l'excision et l'intégration site-spécifique (*xis* et *int*), la réparation mutagène de l'ADN (*rumAB*), la recombinaison homologue (*bet* et *exo*) ou l'exclusion d'entrée (*eex*). Les ORF gris foncé correspondent aux gènes impliqués dans la régulation (*setCDR*). Les ORF gris pâle représentent les gènes codants pour la machinerie de conjugaison. Les ORF blancs sont de fonction inconnue. **(B)** Les régions d'ADN variable de 13 ICE séquencés sont représentées. Les couleurs correspondent à l'ICE dans lequel elles ont initialement été décrites : SXT (bleu), R391 (rouge), ICE*Pda*Spa1 (vert), ICE*Spu*PO1 (violet), ICE*Vch*Mex1 (jaune), ICE*Pal*Ban1 (orange), ICE*Vch*Ind5 (turquoise), ICE*Pmi*USA1 (olive), ICE*Vch*Ban9 (rose), ICE*Vfl*Ind1 (vert pâle). Les flèches minces indiquent les sites des régions variables (I-IV) et points chauds d'insertion (HS1-HS5). Cm, chloramphénicol; Hg, mercure; Kn, kanamycine; Sm, streptomycine; Su, sulfaméthoxazole; Tc, tétracycline; Tm, triméthoprime. * indique que *s073* est absent de ICE*Pda*Spa1. [a] ICE*Vch*Moz10 ne confère pas la résistance au Tm. [b] Les gènes en violet de ICE*Vfl*Ind1 sont déduits du séquençage partiel, d'analyse par PCR et de la comparaison avec ICE*Spu*PO1. Tiré de Wozniak *et al.* (2009).

1.3.1.1. Intégration et excision

Une des caractéristiques principales des membres de la famille SXT/R391 est leur intégration à un site spécifique du chromosome bactérien. Chez *V. cholerae* et *E. coli* K12, SXT s'intègre de manière spécifique et indépendante de *recA* dans la portion 5' du gène *prfC*, un gène codant pour un facteur de relargage de chaînes peptidiques (RF3) impliqué dans la terminaison de la traduction (Hochhut and Waldor, 1999). L'intégrase (Int) est essentielle et suffisante pour l'intégration de SXT dans le chromosome bactérien d'une cellule réceptrice (Hochhut and Waldor, 1999), mais requiert l'excisionase (Xis), un facteur de recombinaison directionnel, pour une excision plus efficace (Burrus and Waldor, 2003). Int est une recombinase à tyrosine de la famille des recombinases site-spécifiques P4 (Marchler-Bauer *et al.*, 2015). Dans la cellule donneuse, Int reconnaît et catalyse la recombinaison entre les séquences *attL* et *attR* qui bordent l'élément intégré, menant à la circularisation de l'élément portant un site *attP*. L'excision a lieu environ dans 1 cellule sur 500 chez R391 et 1 cellule sur 60 chez SXT (Carraro *et al.*, 2015a). L'expression de Int est sous le contrôle de SetCD, mais les facteurs permettant l'expression de Xis sont encore mal compris (Beaber *et al.*, 2002; Burrus and Waldor, 2003).

Fait intéressant, l'insertion de SXT, un élément d'environ 100 kb, dans le gène *prfC* n'altère pas sa fonction. En effet, l'ICE intégré restitue le cadre de lecture de *prfC*, maintenant la protéine correspondante active, en plus de lui procurer un site de fixation au ribosome et un promoteur (Hochhut and Waldor, 1999). Lorsque *prfC* est éliminé du génome de la réceptrice, SXT s'intègre dans des sites alternatifs, préférentiellement *pntB*, un gène codant pour une pyridine nucléotide transhydrogénase (Burrus and Waldor, 2003). La fréquence de transfert de SXT est diminuée de 10 fois quand il est intégré dans *pntB* au lieu de *prfC* (Burrus and Waldor, 2003). Deux possibilités pourraient expliquer ce phénomène. Soit la

régulation traductionnelle par RF3 a un effet subtil sur le transfert de SXT, soit les différences entre les séquences d'insertion dans *prfC* et *pntB* pourraient affecter la fonction de Int et Xis et donc la recombinaison menant à l'excision de l'élément (Burrus and Waldor, 2003).

1.3.1.2. Système de partition

Une fois installés dans une population bactérienne, les ICE SXT/R391 sont très rarement perdus (Carraro *et al.*, 2015a; Wozniak and Waldor, 2009). La stabilité de ces ICE est principalement attribuable à leur intégration dans le chromosome et à leur descendance verticale subséquente (Burrus *et al.*, 2002). Les ICE SXT/R391 possèdent aussi, dans leurs insertions d'ADN variable, des systèmes toxine-antitoxine menant à la mort ou à une diminution du taux de croissance des bactéries ayant perdu l'ICE. Une étude récente (Carraro *et al.*, 2015a) montre deux nouveaux mécanismes contribuant au maintien des ICE SXT/R391 (Figure 6). Le premier relève de la réplication en cercle-roulant dépendante de TraI et de *l'oriT*, et sera décrit plus précisément dans la prochaine section. Le deuxième est un système de partition codé par des gènes appartenant au squelette conservé de la famille SXT/R391.

Par analogie avec le système de partitionnement ParMRC de type II (Salje *et al.*, 2010), il a été suggéré que le gène *srpM* code pour une protéine de type actine qui, en polymérisant, médie la poussée des ICE circulaires, tandis que le produit du gène *srpR* lie l'ADN au site *srpC*, qui ressemble à un centromère, et agit à titre de connecteur entre *srpC* et les filaments de SrpM en cours de polymérisation (Carraro *et al.*, 2015a). Ce système permet donc la ségrégation active des ICE circulaires entre cellules-filles en cours de division. Ce système est sous le contrôle d'un promoteur dépendant de SetCD qui régule aussi l'expression de l'intégrase (Beaber

et al., 2002; Burrus and Waldor, 2003; Carraro *et al.*, 2015a), assurant ainsi une co-expression des systèmes médiant l'excision et le partitionnement des ICE SXT/R391.

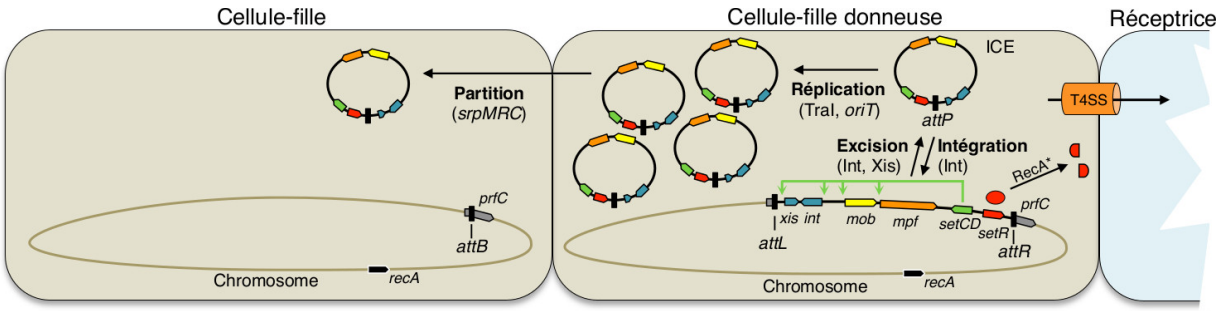

Figure 6. Résumé des mécanismes de maintien des ICE SXT/R391. Représentation d'une paire conjugale dont la cellule donneuse est en cours de division. La cellule réceptrice est partielle pour des considérations d'espace. L'ICE peut être maintenu dans la population par intégration dans le chromosome, par réplication en cercle-roulant, par partitionnement entre cellules-filles ou grâce à des systèmes toxine-antitoxine (ces derniers sont omis pour simplifier la figure). Le locus *srpMRC* n'est pas représenté, mais fait partie de l'opéron contenant *int.* Bien que les processus de conjugaison et de maintien se chevauchent dans le temps (levée de la répression par SetR), la conjugaison en elle-même n'augmente pas la stabilité des ICE SXT/R391 (Carraro *et al.*, 2015a).

1.3.1.3. Machinerie de conjugaison, réplication et exclusion

En opposition aux AICE (*actinomycete integrative and conjugative elements*) pour qui l'ADN est transféré d'une donneuse vers une réceptrice sous forme double-brin, il est admis que les ICE SXT/R391 se transfèrent sous forme d'ADN simple-brin (Beaber *et al.*, 2002; Poele *et al.*, 2008). La machinerie de conjugaison permet la préparation de l'ICE circulaire et l'assemblage d'un pilus et d'un pore de conjugaison assurant le transfert d'un brin de l'ICE vers la cellule réceptrice. La fréquence de transfert de SXT et R391 oscille entre $1x10^{-3}$ et $1x10^{-5}$ (Carraro *et al.*, 2015a; Ceccarelli *et al.*, 2008; Daccord *et al.*, 2010). La machinerie de conjugaison est divisée en deux modules de

gènes. Premièrement, le module de mobilisation (*mob*), qui code pour TraI et MobI, des facteurs cruciaux du relaxosome. Deuxièmement, le module de formation de la paire conjugale (*mpf*, *mating pair formation*), qui contient les autres gènes du système de sécrétion de type IV, c'est-à-dire la protéine de couplage, les gènes codant pour le pilus qui stabilise le premier contact entre cellules partenaires, et le pore de conjugaison (Alvarez-Martinez and Christie, 2009). Les modules *mob* et *mpf* sont entrecoupés par des insertions d'ADN variables, reflétant les multiples évènements de recombinaison dont ces éléments ont été la cible.

La relaxase TraI, avec l'aide de la protéine auxiliaire MobI, reconnaitrait l'origine de transfert (*oriT*) située immédiatement en amont du gène *mobI* (Ceccarelli *et al.*, 2008). Par analogie avec d'autres systèmes mieux connus, tels que F (Lawley *et al.*, 2003), il est présumé que le relaxosome est un complexe nucléoprotéique qui reconnaît *l'oriT* et coupe un des deux brins d'ADN, auquel la relaxase reste liée de manière covalente. TraI agirait ensuite comme un facteur d'initiation de la réplication, recrutant des composantes de la machinerie de réplication de la bactérie hôte (Carraro *et al.*, 2015a; Lee *et al.*, 2010a; Llosa *et al.*, 2002). Cette réplication en cercle-roulant est conditionnelle à l'activation de l'ICE et il a été suggéré qu'elle contribue au maintien des ICE en permettant leur passage d'une cellule donneuse vers une cellule réceptrice, mais également en permettant leur hérédité verticale dans des cellules en division active dans lesquelles l'ICE est excisé (Carraro *et al.*, 2015a; Lee *et al.*, 2010a). La relaxase liée de manière covalente à l'ICE simple-brin interagit ensuite avec la protéine de couplage qui fournit l'énergie nécessaire au pompage de l'ICE vers la cellule réceptrice au travers du pore de conjugaison (Figure 1). Une fois dans la cellule réceptrice, la relaxase médierait la recircularisation du brin et sa réplication pour générer un élément double-brin servant de substrat à l'intégrase.

L'interaction entre cellules donneuse et réceptrice ne mène pas toujours au transfert de l'ICE. Des protéines d'exclusion peuvent interférer dans ce processus. Les cellules qui portent SXT excluent les copies entrantes de SXT, mais pas celles de R391 et vice versa. Ce phénomène d'exclusion dépend de variants de TraG et Eex, des protéines de la membrane interne respectivement des cellules donneuses et réceptrices (Marrero and Waldor, 2005, 2007a). Les régions C-terminales de TraG et Eex sont localisées dans les cytoplasmes des cellules donneuses et réceptrices et contiennent les résidus essentiels au processus d'exclusion et à sa spécificité (Marrero and Waldor, 2007b). Le mécanisme d'interaction entre TraG et Eex est encore inconnu, en particulier le mode d'interaction de protéines cytoplasmiques de deux partenaires conjugaux. Bien qu'imparfait (Marrero and Waldor, 2007a), ce système d'exclusion permet d'amenuiser la redondance dans la propagation des ICE SXT/R391 au sein d'une population bactérienne.

La conjugaison est un mécanisme de stabilisation de nombreux plasmides conjugatifs, en permettant la réinfection de cellules qui auraient perdu le plasmide (Bahl *et al.*, 2009). Une étude récente (Carraro *et al.*, 2015a) a montré que ce n'est pas le cas pour les ICE SXT/R391, étant donné que la délétion de *traG*, une composante importante du pore de conjugaison, n'affecte pas la fréquence de perte de R391.

1.3.1.4. Système de recombinaison

Un système de recombinaison homologue indépendant de RecA, codé par les gènes *bet* et *exo*, est également retrouvé dans le squelette de gènes conservés des ICE SXT/R391 (Garriss *et al.*, 2009). Bet a 55 % d'identité avec la recombinase du système Red du bactériophage λ, tandis qu'Exo montre 27 % d'identité avec

l'exonucléase du système Red (Garriss *et al.*, 2009). Bet et Exo ont la capacité de recombiner différents éléments de la famille SXT/R391 retrouvés en tandem, générant ainsi des ICE hybrides. Ces évènements de recombinaison mènent à la diversification des patrons de fonctions accessoires, dont la résistance aux antibiotiques (Figure 7) (Burrus and Waldor, 2004b; Garriss *et al.*, 2009; Spagnoletti *et al.*, 2014). Les gènes *bet* et *exo* sont activés par SetCD (Annexe 2) (Garriss *et al.*, 2013a).

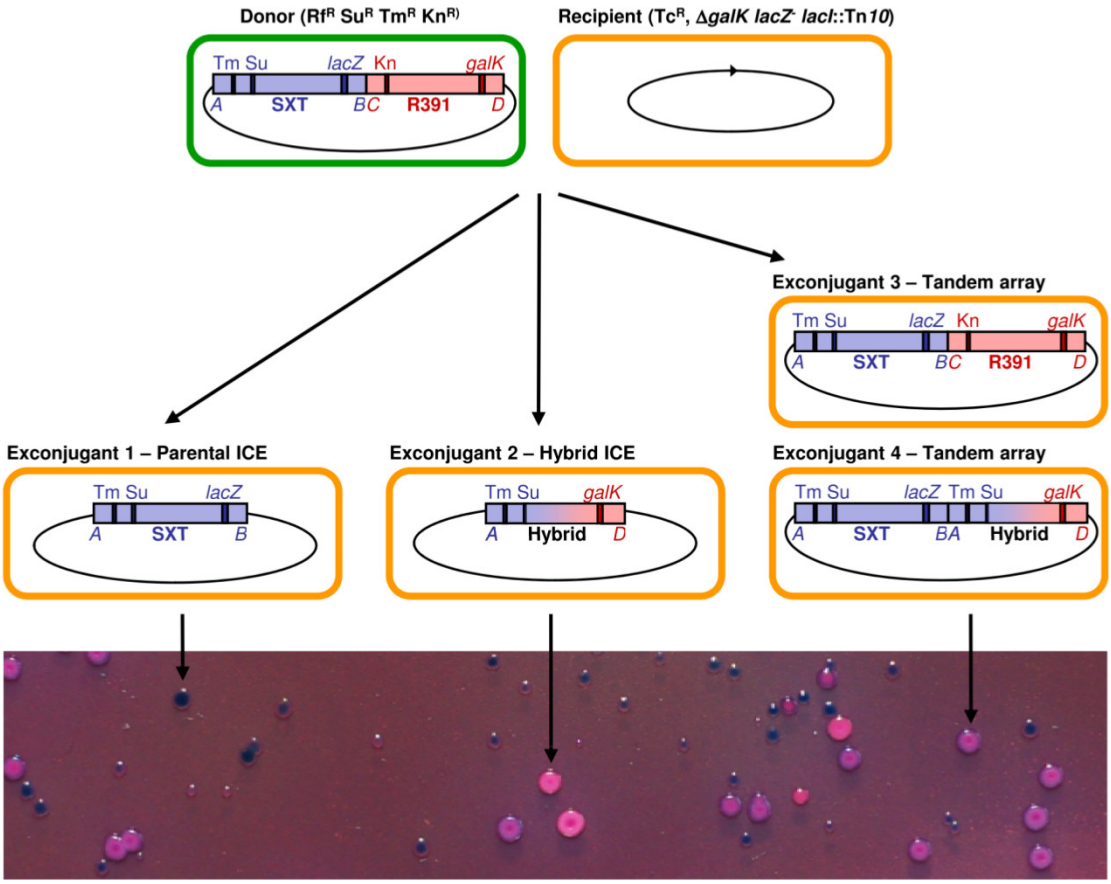

Figure 7. Schématisation d'un essai semi-quantitatif basé sur la couleur des colonies permettant la détection d'ICE hybrides. La position relative des marqueurs antibiotiques (Tm, Su, Kn) et phénotypiques (*lacZ* et *galK*) est indiquée dans SXT et R391. L'utilisation d'une réceptrice Δ*galK lacZ*⁻ *lacI*::Tn*10* permet l'expression constitutive des marqueurs *lacZ* et *galK* insérés dans les ICE. La conjugaison entre une cellule donneuse (verte) qui contient un tandem SXT-R391 et une cellule réceptrice (orange) donne des exconjugants qui peuvent contenir soit un élément simple, un tandem ou un élément hybride. Une gélose MacConkey contenant du X-gal, du D-galactose, de la Tm, du Su et de la tétracycline (panneau du bas) révèle des colonies qui portent soit SXT (colonies bleues), soit un ICE hybride (colonies rouges) ou encore un tandem SXT-R391 (colonies mauves). Les colonies mauves peuvent également contenir un tandem composé d'un SXT et d'un élément hybride (exconjugant 4). Les colonies rouges et mauves sont plus grosses, car elles utilisent le D-galactose comme source de carbone. (A,B), les extrémités gauche et droite de SXT; (C,D), les extrémités gauche et droite R391. Tiré de Garriss *et al.* (2009).

1.3.2. Cargaison d'ADN variable

1.3.2.1. Éléments mobiles et gènes de résistance

Les régions variables qui distinguent les membres de la famille SXT/R391 sont dispersées en des points précis au sein du squelette de gènes conservés (Figure 5) (Wozniak *et al.*, 2009). Par exemple, SXT et plusieurs autres éléments de la famille possèdent une insertion dans le gène *rumB,* laquelle confère la résistance au sulfaméthoxazole, au trimethoprime, au chloramphénicol et à la streptomycine (Wozniak *et al.*, 2009). Chez SXT, cette région d'environ 17,2 kb est un transposon composite portant plus de sept transposases (Hochhut *et al.*, 2001; Toleman and Walsh, 2010). Il est intéressant de mentionner que chez les ICE qui ne portent pas cette insertion, les gènes *rumAB* codent pour une polymérase mutagène activée par le régulateur LexA lors de l'activation de la réponse SOS (Kulaeva *et al.*, 1995). Un autre exemple est ICE*Pmi*Jpn1 qui porte un gène conférant la résistance à un spectre étendu de céphalosporines, $bla_{CMY\text{-}2}$, qui est localisé sur un transposon composite médié par IS10 (Harada *et al.*, 2010). De plus, un intégron conférant la résistance au triméthoprime est retrouvé, entre autres, dans le point chaud d'insertion 3 (Figure 5, HS3) de ICE*Vch*Ind5 et ICE*Vch*Moz10 (Hochhut *et al.*, 2001; Wozniak *et al.*, 2009). Plusieurs autres exemples de gènes de résistance aux antibiotiques et au mercure propagés par les ICE SXT/R391 ont été décrits (Ceccarelli *et al.*, 2011d; Garriss *et al.*, 2013b; Rodríguez-Blanco *et al.*, 2012; Wozniak *et al.*, 2009). Les transposons composites et les intégrons permettent le mouvement, le recrutement et l'accumulation des gènes de résistance (Cambray *et al.*, 2010; Roberts *et al.*, 2008).

1.3.2.2. Systèmes toxine-antitoxine

Les systèmes toxine-antitoxine (TA) dits 'de dépendance' sont retrouvés autant chez les éléments mobiles qu'au sein des chromosomes procaryotes (Van Melderen and De Bast, 2009). Ils sont composés de deux gènes, généralement co-transcrits, qui codent respectivement pour une toxine stable et une antitoxine labile. Dès qu'une bactérie perd les gènes codant pour le système, l'antitoxine est rapidement dégradée, libérant ainsi l'activité destructrice de la toxine. Un arrêt de la croissance ou la mort des cellules ayant perdu l'antitoxine encourage le développement des bactéries ayant conservé le système dont la proportion augmente dans la population.

Lorsque l'ICE est sous sa forme circulaire, il y a un risque de perte de l'élément si la bactérie se divise avant que la réintégration dans le chromosome n'ait lieu (Carraro *et al.*, 2015a; Wozniak and Waldor, 2009). Deux systèmes toxine-antitoxine ont été décrits chez SXT et peuvent expliquer en partie la perte extrêmement faible de SXT. Le premier est codé par les gènes *s044* et *s045* qui sont situés dans un point chaud d'insertion entre *traD* et *tral* (Figure 5 HS1). Ce système toxine-antitoxine de type *tad-ata* est fonctionnel pour la rétention de plasmides dans *E. coli* et *Paracoccus versutus* (Dziewit *et al.*, 2006). *s044* et *s045* sont organisés en opéron, mais peuvent être transcrits individuellement à partir de leur propre promoteur (Dziewit *et al.*, 2006). Aucun régulateur se liant à ses promoteurs n'a été identifié. Le deuxième système TA de SXT est codé par les gènes *mosA* et *mosT* situés dans le point chaud d'insertion entre *traL* et *traF* (Figure 5 HS2) (Wozniak and Waldor, 2009). Lorsque l'élément est sous sa forme intégrée, l'expression de *mosAT* est maintenue à un niveau très faible par l'autorépression exercée par MosA. Il semble y avoir levée de la répression de *mosAT* par Xis et SetCD (Wozniak and Waldor, 2009). L'expression de ce système est donc limitée au moment ou SXT est vulnérable, sous sa forme extrachromosomique. Bien que le mécanisme d'action de la toxine MosT soit toujours

inconnu, il a été démontré que son expression inhibe la croissance de *V. cholerae* et *E. coli*. R391 possède son propre système toxine-antitoxine de type HipAB, différent des deux retrouvés chez SXT. La toxine et l'antitoxine sont respectivement codées par les gènes *hipA* et *hipB*, lesquels sont localisés dans la région variable en amont de *xis* (Figure 5 VRI). La délétion de *hipA* mène à une augmentation de la perte de R391 de 10 fois (Carraro *et al.*, 2015a). La toxine HipA possède un domaine kinase Ser/Thr-like eucaryote qui est responsable de la phosphorylation du résidu Thr^{382} de EF-Tu, un facteur d'élongation, ce qui mène à un état de stase étant donné que EF-$Tu\text{-}Thr^{382}$ ne peut plus lier les aminoacyl-tRNA (Schumacher *et al*., 2009). La stase permet la persistance aux traitements antibiotiques qui ciblent les fonctions des cellules actives. HipB est l'antitoxine qui se lie et inhibe HipA. HipB est sujette à la protéolyse par Lon et réprime l'opéron *hipBA* en liant de manière coopérative quatre operateurs en aval de *hipBA* (Hansen *et al*., 2012).

Il n'est pas exclu que les systèmes TA codés par les ICE SXT/R391 puissent avoir un rôle à jouer dans la persistance bactérienne, un phénomène reconnu comme menant à la tolérance à de nombreux agents antimicrobiens. En effet, HipA a été montré comme étant un facteur de tolérance chez *E. coli* (Lewis, 2008; Schumacher *et al.*, 2009).

1.3.2.3. Systèmes restriction-modification

Certains membres de la famille SXT/R391 possèdent des systèmes de restriction-modifications (RM) qui peuvent conférer la résistance à l'infection de l'hôte bactérien par des phages (Rodríguez-Blanco *et al.*, 2012; Wilson and Murray, 1991). La prédaction de *V. cholerae* par les phages dans un contexte de maladie humaine a été

montrée comme pouvant altérer son potentiel virulent (Seed *et al.*, 2014). Les phages sont extrêmement abondants en milieu aquatique et au niveau du tractus gastro-intestinal humain. De plus, les systèmes RM sont également reconnus pour stabiliser les îlots génomiques et les chromosomes (Wilson and Murray, 1991). Il a même été avancé que les RM contribueraient au maintient des espèces bactériennes en limitant le transfert horizontal (Sneppen *et al*., 2015). Ils peuvent aussi participer à la plasticité génomique étant donné que les endonucléases génèrent des coupures, donc des substrats pour la recombination homologue menant au 'brassage' d'allèles (Vasu and Nagaraja, 2013). Finalement, le rôle régulateur de la méthylation a longtemps été sous-estimé chez les bactéries et est de plus en plus étudié (Sánchez-Romero *et al*., 2015). L'avantage conféré par les systèmes RM portés par ces éléments est sûrement très important.

1.3.3. Mobilisation de matériel génétique par les ICE SXT/R391

1.3.3.1. Mobilisation en *cis*

La capacité des éléments SXT/R391 à transférer du matériel génétique va au-delà de leur propre transfert. En effet, il été démontré que SXT peut mobiliser jusqu'à 500 kb d'ADN chromosomique en *cis* d'une manière similaire à celle observée chez les souches Hfr de *E. coli* (Hochhut *et al.*, 2000). Chez *V. cholerae*, l'îlot de pathogénicité qui code pour le pili TCP, un facteur de virulence important, est situé environ à 200 kb en 3' de *prfC*. Étant donnée l'orientation du transfert de SXT, il pourrait donc être mobilisé et transféré par le système de conjugaison de SXT (Hochhut *et al.*, 2000). Cependant, la région *wfb*, qui code pour le sérotype O139, est située environ à 460 kb en 5' de *prfC*. Il semble donc improbable que SXT ait joué un rôle dans le mobilisation de *wfb* occasionnant le changement de sérotype de *V. cholerae* El Tor O1 à *V. cholerae* O139 en 1992 (Hochhut *et al.*, 2000). Il est important de noter que l'excision et la circularisation de SXT ne sont pas nécessaires à l'expression de ses gènes de transfert, c'est pourquoi il est capable de mobiliser de l'ADN chromosomique.

1.3.3.2. Mobilisation en *trans*

La caractérisation de l'origine de transfert (*oriT*) des ICE SXT/R391 a permis l'identification de séquences chromosomiques ressemblant à *l'oriT* avec plus de 63 % d'identité (Ceccarelli *et al.*, 2008; Daccord *et al.*, 2010). Une étude plus poussée de ces *oriT* chromosomiques a révélé qu'elles appartiennent à une famille d'îlots génomiques mobilisables (MGI) intégrés à l'extrémité 3' de *yicC* dans le chromosome

d'espèces de *Vibrio*, d'*Alteromonas*, de *Pseudoalteromonas* et de *Methylophaga* (Daccord *et al.*, 2010, 2013). Les MGI ont une taille comprise entre 18 et 33 kb et sont dépourvus d'un système de sécrétion de type IV. Ils ont un squelette de gènes conservé d'environ 5,5 kb qui contient les gènes int_{MGI}, *rdfM*, *cds4* et *cds8* (Figure 8). Les deux premiers codent respectivement pour l'intégrase et l'excisionase des MGI, qui médient l'intégration et l'excision du chromosome de leur hôte bactérien, tandis que les deux derniers sont de fonction inconnue. Le squelette de gènes conservé est interrompu par des insertions d'ADN variable en contenu et en taille. La plupart de ces gènes codent pour des fonctions adaptatives telles que des systèmes de restriction-modification de type I et III qui confèrent la résistance contre l'infection par des phages (Daccord *et al.*, 2013).

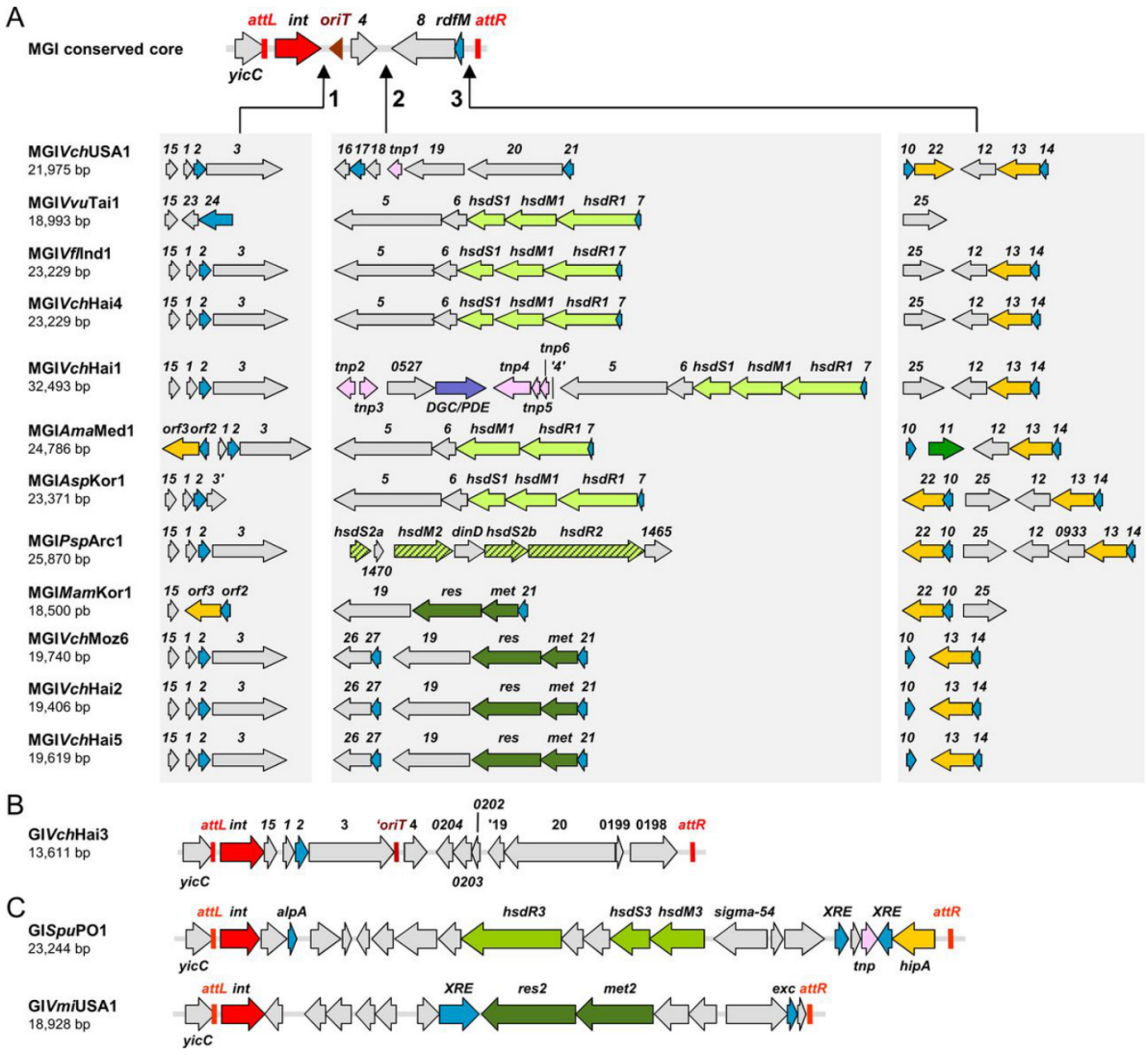

Figure 8. Diversité des MGI mobilisés par les ICE SXT/R391. (A) Squelette conservé et insertions d'ADN variable de 12 MGI séquencés. **(B)** Structure atypique de GI*Vch*Hai3 issue de *V. cholerae* H-09; **(C)** Contenu génétique de GI*Spu*PO1 et GI*Vmi*USA1 résidants respectivement dans *Shewanella* sp. W3-18-1 et *V. mimicus* VM573. Les couleurs des cadres de lectures ouverts indiquent les fonctions putatives : rose et rouge, recombinaison; violet, diguanylate cyclase/ phosphodiestérase; bleu, régulateur transcriptionnel; vert, systèmes de restriction-modification; orange, toxine putative; gris, fonction inconnue. Tiré de Daccord *et al.* (2013).

L'étape initiale de la mobilisation d'un MGI par un ICE SXT/R391 est son excision du chromosome. L'excision requiert l'activation par SetCD des gènes int_{MGI} et *rdfM* suivie de la recombinaison entre les sites d'attachement $attL_{MGI}$ et $attR_{MGI}$ qui bordent le MGI intégré (Figure 9 et Annexe 3) (Daccord *et al.*, 2012). Le MGI circulaire extrachromosomique résultant porte l'$oriT_{MGI}$ qui imite *l'oriT* des ICE SXT/R391 et parasite le relaxosome codé par ceux-ci. Ultimement, le MGI est transloqué vers la cellule réceptrice à travers le pore de conjugaison codé par l'ICE. Une fois dans la cellule réceptrice, le MGI devient complètement indépendant de son ICE mobilisateur et des activateurs SetCD. En effet, le MGI exprime constitutivement son intégrase (int_{MGI}) à un niveau bas, mais suffisant pour l'intégration du MGI à l'extrémité 3' de *yicC* (Daccord *et al.*, 2012). MGI*Vfl*Ind1, découvert chez *Vibrio vulnificus*, a été utilisé comme prototype pour l'étude de cette classe de MGI. MGI*Vfl*Ind1 est mobilisé à haute fréquence entre souches d'*E. coli*, autant par ICE*Vfl*Ind1 que par SXT (10^{-3} exconjugants par cellule donneuse). Cette fréquence est augmentée à 10^{-1} exconjugants par cellule donneuse à la suite de la surexpression de *setCD* ou de l'induction à la mitomycine C (Daccord *et al.*, 2010, 2012). MGI*Vfl*Ind1 est aussi capable de mobiliser en *cis* plus de 1 Mb d'ADN chromosomique situé en 5' de *yicC* (Daccord *et al.*, 2010). L'orientation du transfert étant adéquat, et le gène *yicC* étant à proximité de la région qui code pour l'antigène

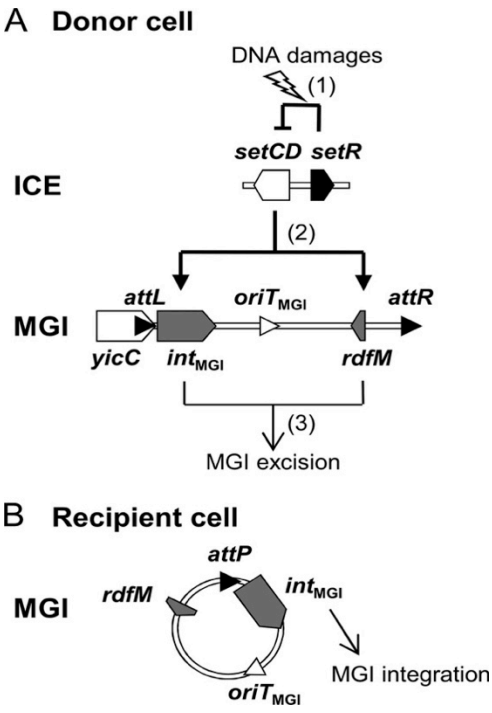

Figure 9. Dynamique d'intégration de d'excision des MGI. (A) Excision du chromosome de la cellule donneuse. (1) La réponse SOS est activée par de l'ADN endommagée, levant la répression exercée par SetR sur *setCD*. (2) SetCD active l'expression de int_{MGI} et *rdfM*. (3) Int_{MGI} et RdfM procèdent à l'excision du MGI. **(B)** L'intégration dans le chromosome de la cellule réceptrice. int_{MGI} est exprimé à un niveau bas permettant l'intégration du MGI de manière indépendante des ICE. Tiré de Daccord *et al.* (2012).

O chez *V. cholerae*, le locus correspondant à l'antigène O139 aurait pu être transféré vers *V. cholerae* El tor O1 par l'entremise d'un MGI, puis un évènement de recombinaison entre les deux loci aurait permis l'acquisition du nouveau sérotype. La mobilisation d'ADN chromosomique par MGIV*fl*Ind1 implique la reconnaissance de $oriT_{MGI}$ par le relaxosome d'un ICE SXT/R391 sans excision préalable du MGI.

1.4. Modèles de régulation génétique

La répression des phénomènes d'excision, de transfert et d'intégration des ICE SXT/R391 est médiée par SetR et est semblable à la régulation précoce du bactériophage λ (Beaber *et al.*, 2004; Ptashne, 2004). λ est un bactériophage tempéré, c'est-à-dire qu'il peut se développer soit dans un état lytique, aboutissant à la lyse cellulaire et au relâchement de centaines de virus, soit devenir latent en s'intégrant sous forme de prophage dans le chromosome de sa bactérie hôte, la rendant lysogène. Par analogie, le transfert conjugatif de SXT correspondrait au cycle lytique, tandis que sa forme intégrée représenterait le cycle lysogène. Lorsque λ infecte une bactérie, typiquement *E. coli*, une voie de régulation commune aux états lytique et lysogène est mise en branle, c'est l'activation des gènes précoces. Ensuite, plusieurs facteurs environnementaux entrent en jeu pour orienter le phage vers la voie lytique ou vers la voie lysogène. De nombreux activateurs et répresseurs transcriptionnels et post-transcriptionnels ont été découverts au sein de ces voies de régulation chez λ et d'autres phages lambdoïdes (Ptashne, 2004). Ces réseaux de régulation sont étudiés depuis plus de 40 ans et sont considérés comme un paradigme de régulation génétique (Herskowitz and Hagen, 1980; Oppenheim *et al.*, 2005).

L'activation de l'excision, du transfert et de l'intégration des ICE SXT/R391 est médiée par le complexe activateur SetCD (Beaber *et al.*, 2002). Celui-ci est homologue à AcaCD, l'activateur des plasmides conjugatifs IncA/C (Annexe 4), ainsi qu'à FlhDC, un complexe régulateur bien caractérisé responsable de l'activation de la synthèse du flagelle bactérien. Nous allons donc nous attarder brièvement à la description de ces modèles génétiques afin de mettre en évidence les rapprochements possibles avec la régulation des ICE SXT/R391. Pour éviter toute ambiguïté entre les composantes des ICE SXT/R391 et celles des bactériophages lambdoïdes, une convention différentielle a été adoptée. Les promoteurs de SXT sont écrits en italique et marqués par un grand P (P_L, P_R, P_{s003}, etc.), contrairement à ceux de λ (pR, pRM, pL, pI, etc.). Notez que le promoteur P_L équivaut à pR, tandis que P_R équivaut à pRM (Beaber and Waldor, 2004). Les sites opérateurs de SXT sont écrits en italique (*O4*, *OL*, *O1*, *O2*, *O3*) contrairement à ceux de λ (OR1, OR2, OR3, OL1, OL2, OL3). Les sites *O1*, *O2* et *O3* sont respectivement les équivalents de OR1, OR2 et OR3.

1.4.1. Régulation de la transcription bactérienne

Une étape clé de la transcription à partir d'un promoteur est la liaison de l'ARN polymérase (*RNA Polymerase*, RNAP) aux boîtes -10 et -35 (Snyder *et al.*, 2013). Celles-ci sont reconnues par le facteur σ, une sous-unité interchangeable de la RNAP qui dicte sa spécificité (Figure 10) (Snyder *et al.*, 2013). Le complexe formé de la RNAP et du facteur σ est appelé l'holoenzyme. La formation d'un complexe ouvert compétent pour la transcription résulte de l'isomérisation menant à la séparation des brins d'ADN près du site d'initiation de la transcription (Figure 10). Après la polymérisation d'environ une dizaine de nucléotides, l'holoenzyme s'évade du promoteur, relâche le facteur σ et l'élongation se poursuit jusqu'à la terminaison du transcrit (Snyder *et al.*, 2013).

Chacune des étapes de la transcription est sujette à des mécanismes de régulation pour éviter le gaspillage de ressources cellulaires et/ou l'expression inopportune de certains gènes (Snyder *et al.*, 2013). La transcription est principalement régulée au niveau du promoteur, et la force de celui-ci est définie par les facteurs suivants :

1) la proximité de la séquence des boîtes -10 et -35 à celles consensuelles d'un facteur sigma donné;

2) la présence d'autres éléments de promoteur qui participent au recrutement de l'holoenzyme (ex. élément UP, boîte -10 étendue);

3) la distance entre les éléments de promoteur;

4) l'action de facteurs de transcription (répresseur ou activateur).

Un répresseur peut réprimer la transcription de plusieurs façons. Il peut bloquer l'accès de l'holoenzyme aux boîtes -10 et -35, sa liaison peut occasionner un recourbement du promoteur défavorisant la liaison de l'holoenzyme ou encore il peut affecter l'isomérisation de l'holoenzyme (Snyder *et al.*, 2013). Les modes d'action des activateurs transcriptionnels, quant à eux, peuvent permettre d'augmenter l'affinité de l'holoenzyme pour le promoteur, de faciliter l'isomérisation ou encore de recourber l'ADN de manière à former le site de reconnaissance (Snyder *et al.*, 2013).

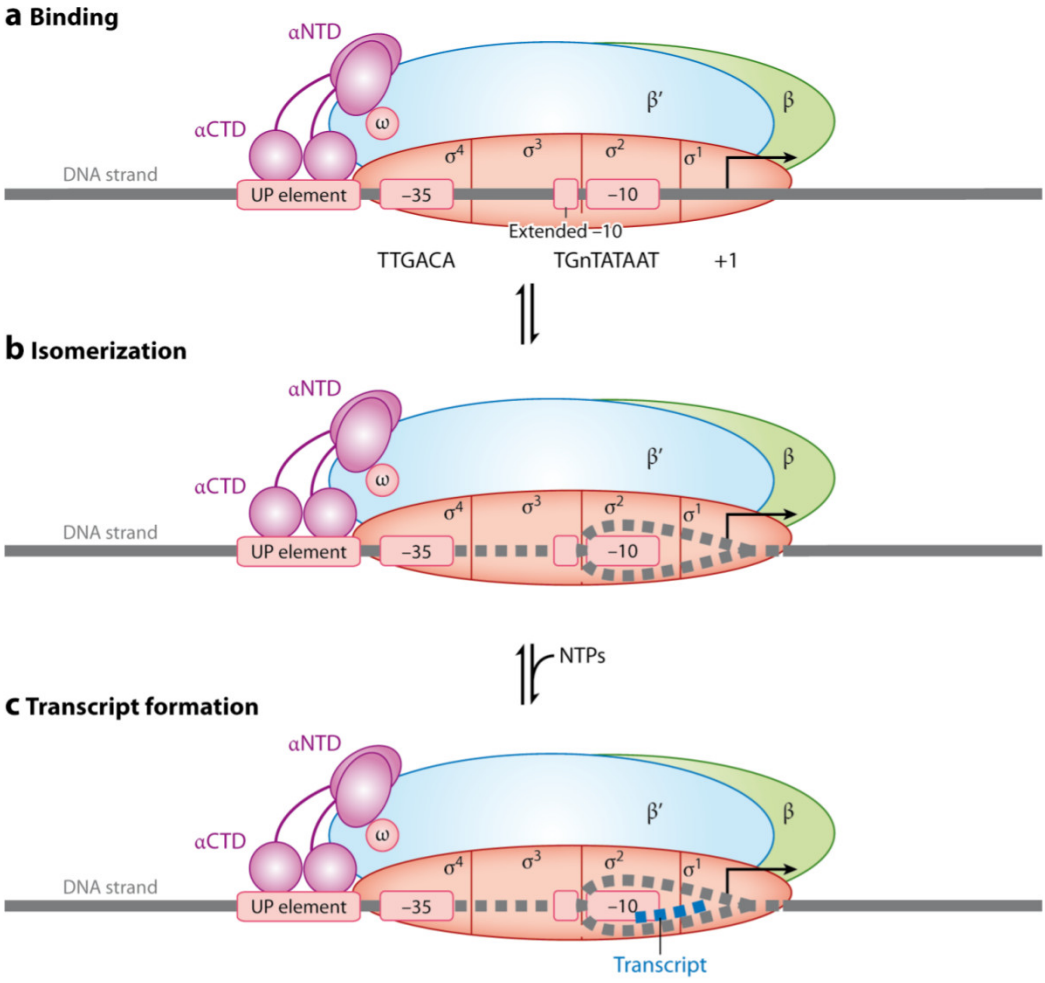

Figure 10. Interaction de l'holoenzyme avec un promoteur et les étapes d'initiation de la transcription. (A) Représentation de l'holoenzyme liée au promoteur. Le site d'initiation de la transcription est indiqué par une flèche coudée. **(B)** L'isomérisation mène au déroulement et à la séparation des deux brins d'ADN autour du site d'initiation de la transcription. **(C)** Formation d'un transcrit (ligne bleue pointillée) en présence de nucléotides triphosphatés. Abréviations: NTD, domaine N-terminal (*N-terminal domain*); CTD, domaine c-terminal (C-*terminal domain*); UP, élément en amont (*upstream element*). Tiré de Lee *et al.* (2012).

1.4.2. Bactériophage λ

1.4.2.1. Voie lytique

La voie lytique est celle par défaut lors d'une infection par λ. En effet, l'intégrase (Int) et CI, le répresseur des promoteurs précoces (pL et pR), ne sont pas initialement exprimés. Rapidement après l'infection, la RNAP bactérienne lie les promoteurs pL et pR et transcrit les gènes situés en amont des terminateurs de transcription *tL1* et *tR1* (Figure 11) (Roberts, 1969; Rosenberg and Court, 1979). Il y a donc expression du gène codant le facteur d'antiterminaison N à partir de pL et expression de *cro* à partir de pR.

Des facteurs N se lient aux sites *nutL* et *nutR* de l'ARN naissant, puis interagissent avec les protéines Nus. Celles-ci sont impliquées à plusieurs niveaux dans les procédés de transcription et de traduction chez les bactéries (Cardinale *et al*., 2008; McGary and Nudler, 2013; Peters *et al*., 2012) et participent à l'antiterminaison chez λ. Le complexe d'antiterminaison ainsi formé permet à la RNAP de transcrire au-delà des terminateurs *tL1* et *tR1* (Das, 1993; Das *et al.*, 1985; Friedman and Court, 1995). Ainsi, des groupes de gènes distaux peuvent alors être exprimés. Ceux-ci sont impliqués autant dans la voie lytique que dans la voie lysogène. Les gènes de l'opéron dont la transcription démarre à partir de pL sont impliqués dans le mécanisme de recombinaison site-spécifique (*int* et *xis*), tandis que les gènes exprimés à partir de pR sont *O*, *P*, *ren* (des gènes de réplication) et un autre facteur d'antiterminaison appelé *Q* (Figure 11) (Court *et al.*, 2006). L'expression de *Q* permet la transcription au-delà de *tR'* à partir de pR', un promoteur constitutif. Les gènes en aval de *tR'* sont impliqués dans le cycle lytique, donc lorsqu'ils sont exprimés grâce à

Q, le bactériophage λ passe un point de non-retour menant à la lyse cellulaire et au relâchement de centaines de bactériophages.

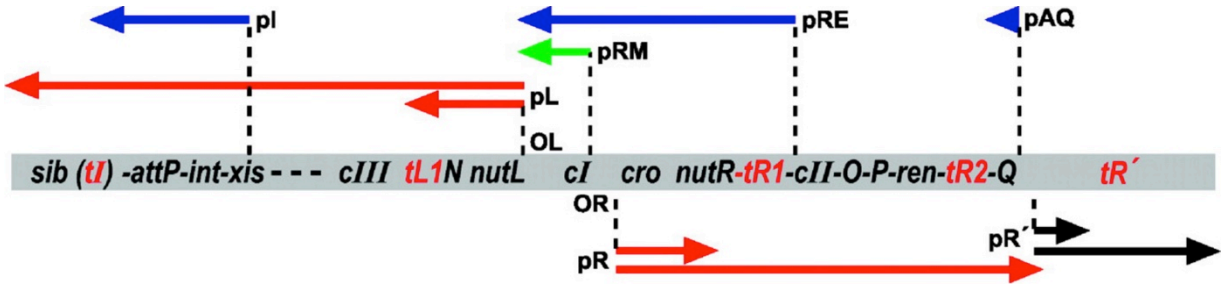

Figure 11. Les gènes et transcrits du bactériophage λ. Les gènes sont représentés dans le rectangle gris. Les transcrits hâtifs à partir de pL et pR sont indiqués par les flèches rouges. Les transcrits tardifs à partir de pR' sont indiqués par les flèches noires. Les transcrits initiés aux promoteurs pI, pRE, pAQ à la suite de l'activation par CII sont indiqués par les flèches bleues. Le transcrit inité par pRM et dépendant de l'activation par CI est indiqué par une flèche verte. Les terminateurs sont indiqués par les lettres rouges à travers les gènes. Les opérateurs OL et OR, qui sont liés par CI et Cro, sont indiqués à côté des promoteurs pL et pR. Tiré de Court *et al.* (2007).

1.4.2.2. Voie lysogène

Le répresseur CI est suffisant pour maintenir le prophage intégré dans le chromosome, préservant l'hôte dans un état lysogène (Jacob and Monod, 1961; Lwoff, 1953). Dans un contexte d'infection, le gène *cI* doit préalablement être activé par CII, dont l'intégrité est assurée par CIII (Kaiser, 1957). *cII* et c*III* sont exprimés à la suite de l'antiterminaison par N des terminateurs *tL1* et *tR1* (Figure 11). HflB (FtsH), une protéase de l'hôte, se lie à la portion C-terminale de CII, ce qui provoque sa dégradation (Datta *et al.*, 2005; Kobiler *et al.*, 2002). CIII inhibe l'activité protéasique de HflB, permettant ainsi d'augmenter la concentration en CII, donc les chances d'un établissement de l'état lysogène (Banuett *et al.*, 1986; Hoyt *et al.*,

1982). Lorsqu'une quantité suffisante de CII est accumulée, cet activateur induit l'expression coordonnée de trois promoteurs : pI qui conduit l'expression de *int*, pRE à partir duquel est exprimé *cI*, et pAQ d'où est généré un ARN antisens dirigé contre le gène *Q*. L'ARN antisens contre *Q* empêche l'expression de gènes lytiques, tandis que l'intégrase et CI sont essentiels à l'établissement de l'état lysogène (Kobiler *et al.*, 2005). Une fois CI traduit, CII et CIII ne sont plus nécessaires, car CI active sa propre expression à partir de pRM, un promoteur faible.

1.4.2.3. Boucle entre OL et OR

L'activation de *cI* par CII, à partir du promoteur pRE, induit la formation d'une boucle dans l'ADN, maintenue par un octamère de CI (Figure 12) (Court *et al.*, 2007). Ce repliement permet le maintien stable de l'état lysogène du bactériophage. La liaison simultanée de CI aux opérateurs OL et OR assure, d'une part, la répression des promoteurs précoces pL et pR et d'autre part, le maintien du niveau de CI optimal pour la répression lysogène à partir de pRM (Figure 11 et 12) (Dodd, 2004; Dodd *et al.*, 2001; Svenningsen *et al.*, 2005). Des études ont également montré que la formation d'une boucle entre OL et OR contribue à l'activation de *cI* à partir de pRM en permettant au domaine C-terminal de la sous-unité α de la RNAP liée à pRM d'entrer en contact avec un site adjacent à l'opérateur OL (Anderson and Yang, 2008; Cui *et al.*, 2013).

1.4.2.4. Induction dépendante de RecA*

Le signal induisant l'activation de l'excision, de la réplication et de la propagation d'un prophage est lié à la réponse SOS, soit la réponse bactérienne aux dommages à

l'ADN, et implique l'inactivation du répresseur CI (Roberts *et al.*, 1978). La conjugaison d'ADN simple-brin a également été montrée comme un activateur de la réponse SOS (Baharoglu *et al.*, 2010). En utilisant l'énergie de l'ATP, des monomères de la recombinase de l'hôte RecA polymérisent sur de l'ADNsb générant des filaments RecA-ADNsb (RecA*). Ceux-ci sont compétents pour la recombinaison homologue et sont aussi des effecteurs qui libèrent l'activité protéolytique latente des répresseurs LexA et CI (Little, 1984; Chen *et al.*, 2008). Tandis que CI est le répresseur maintenant l'état lysogène de λ, LexA est le répresseur de la réponse SOS (Bailone *et al.*, 1979; Little, 1984). Le répresseur CI est donc une sentinelle aux aguets d'un signal de stress cellulaire bactérien. Cette fonction assure l'évasion des prophages λ qui résident dans des cellules endommagées et potentiellement mourantes.

1.4.2.5. Fonction de Cro

Le clivage de CI permet le relâchement de son emprise sur les opérateurs OL et OR ce qui entraîne l'expression du gène *cro* (Figure 12). Ce dernier code pour un co-répresseur qui, avec CI, forment un interrupteur génétique qui gouverne le passage de l'état lysogène à l'état lytique (Eisen *et al.*, 1970; Folkmanis *et al.*, 1977; Meyer and Ptashne, 1980; Meyer *et al.*, 1980; Schubert *et al.*, 2007). Dès la levée de la répression par CI, *cro* est exprimé à partir de pR et la protéine Cro inhibe l'expression de *cI*, probablement grâce à sa grande affinité pour OR3, d'où elle peut réprimer pRM (Figure 13) (Johnson *et al.*, 1978; Meyer *et al.*, 1980; Takeda, 1979). Plusieurs études ont montré que la répression de pRM par Cro est cruciale lors de l'induction du prophage (Johnson *et al.*, 1981; Schubert *et al.*, 2007). Le mode d'action de Cro à pRM semble prévenir la recrudescence du niveau de CI lors d'une induction suboptimale de la réponse SOS, telle qu'une exposition brève à des rayons UV. La

répression de pRM par Cro rend le développement lytique plus efficace en améliorant la transition de l'état lysogène à l'état lytique.

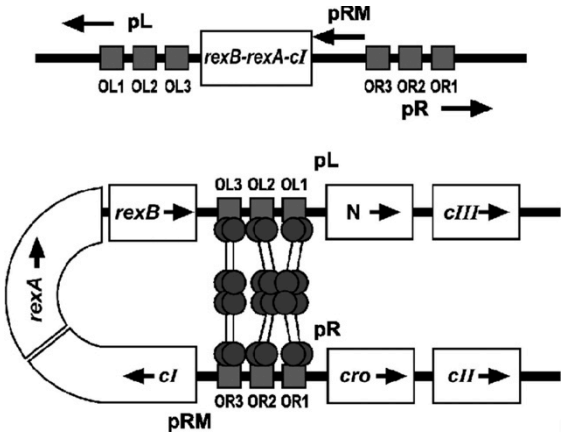

Figure 12. La liaison de CI aux opérateurs *OL* et *OR*. Lorsque le bactériophage λ est lysogène, le promoteur pRM transcrit les gènes *cI*, *rexA et rexB*. Le panneau du bas représente la liaison simultanée de CI aux opérateurs OL et OR, formant un complexe stable. La liaison de CI à OL1-OL2 et OR1-OR2 bloque la transcription à partir de pL et pR, tandis que la liaison de CI à OR3-OL3 réprime pRM et diminue l'expression de *cI*. Lorsque le niveau de CI chute, les sites OL3 et OR3 sont libérés, ce qui permet l'autoactivation par CI. Tiré de Court *et al.* (2007).

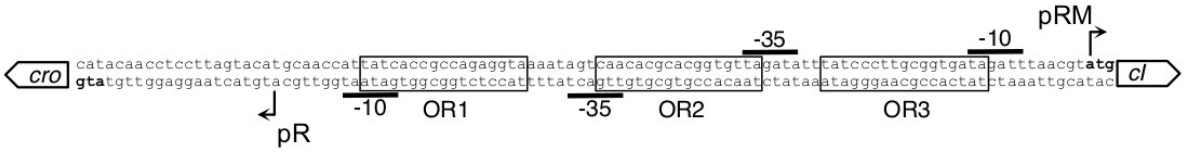

Figure 13. La région intergénique entre *cro* et *cI* chez λ. Les sites d'initiation de la transcription ainsi que les boîtes -10 et -35 des promoteurs pR et pRM conduisant respectivement l'expression de *cro* et *cI* sont indiqués. Les codons d'initiation de *cro* et *cI* sont en gras. Les sites de liaison de CI et Cro dans l'opérateur OR (OR1, OR2 et OR3) sont encadrés.

Cro lie également, mais avec une affinité moins grande, les sites OR1 et OR2, bloquant ainsi sa propre expression et celle de CII (Figures 11, 12 et 13) (Darling *et al.*, 2000; Johnson *et al.*, 1978; Takeda, 1979). En effet, il a été démontré que la

liaison de Cro à pL et pR diminue de 2 à 4 fois l'expression des gènes précoces lytiques, dont les régulateurs *N*, *cIII*, *cII* et *Q* (Eisen *et al.*, 1970; Svenningsen *et al.*, 2005). Le développement lytique des mutants λ*cI*$^{+}$*cro*$^{-}$ est grandement affecté (Eisen *et al.*, 1970; Folkmanis *et al.*, 1977). Cette défectuosité semble causée en partie par la surexpression de CII et CIII, étant donné qu'il y a suppression de la mutation *cro*$^{-}$ par les mutations *cII*$^{-}$ *cIII*$^{-}$ (Folkmanis *et al.*, 1977). La répression de pL et pR par Cro est donc importante pour éviter une résurgence trop hâtive de CI, mais également l'augmentation du niveau cellulaire de CII.

1.4.3. Activation de la synthèse du flagelle

Le complexe FlhDC est reconnu comme étant le facteur de transcription à la tête de la régulation d'une quarantaine de gènes responsables de la synthèse du flagelle chez plusieurs bactéries, telles qu'*E. coli*, *Proteus mirabilis* et *Salmonella typhimurium* (Chilcott and Hughes, 2000; Dufour *et al.*, 1998). Ces gènes sont exprimés de manière hiérarchique à partir de promoteurs divisés en trois classes. La classe I possède un seul promoteur qui conduit l'expression de *flhDC*. La classe II comprend plusieurs promoteurs activés par FlhDC, qui mènent à l'expression du crochet et du corps basal du flagelle, ainsi que d'un facteur sigma alternatif essentiel à l'activation des promoteurs de classe III (Soutourina and Bertin, 2003). L'assemblage du flagelle étant très énergivore et crucial pour la motilité des bactéries, l'expression de *flhDC* est fortement régulée par une pléthore de régulateurs cellulaires globaux répondant à des variations environnementales (Soutourina and Bertin, 2003; Yakhnin *et al.*, 2013) ainsi que par des petits ARN (De Lay and Gottesman, 2012). L'activation des promoteurs flagellaires de classe II est observée lorsqu'ils sont liés par le complexe hétérotétramérique $FlhD_2C_2$ (Liu and Matsumura, 1994). Cependant, la structure du complexe a été résolue par Wang *et al.* (2006), et il semblerait que le complexe soit plutôt sous la forme $FlhD_4C_2$ (Figure 13) (Wang *et al.*, 2006). Le dimère $FlhC_2$ peut

lier l'ADN de manière indépendante, mais $FlhD_2$ augmente la stabilité et la spécificité du complexe (Claret and Hughes, 2002). Le complexe $FlhD_4C_2$ lie l'ADN par des structures hélice-tour-hélice ainsi que par une région liant le zinc dans FlhC (Lee *et al.*, 2011).

Il a également été montré que FlhDC régule l'expression d'autres processus que la synthèse du flagelle, par exemple l'expression de facteurs de virulence (Allison *et al.*, 1994; Stafford *et al.*, 2005). Contrairement à ce qui est observé chez les eucaryotes, les facteurs de transcription sous forme d'hétéromères sont rares chez les procaryotes.

Figure 14. Le complexe activateur $FlhD_4C_2$. Représentation schématique d'un hexamère de FlhDC lié à un fragment d'ADN de forme B de 49 pb. Les résidus de FlhD, dont la mutation en une alanine affecte la formation du complexe FlhDC ou la liaison spécifique à l'ADN, sont indiqués respectivement en orange et en rouge. Les sites de liaison du zinc dans les monomères de FlhC sont indiqués. Le squelette phosphodiester de l'ADN est représenté par un ruban vert. Les portions qui sont suggérées pour l'interaction avec le complexe sont en magenta. Tiré de Wang *et al.* (2006).

1.5. Module de régulation des ICE SXT/R391

La région régulatrice des ICE SXT/R391 est située à l'extrémité 3' de l'élément intégré et est composée de huit gènes. Sept d'entre eux sont orientés dans un même sens et sont exprimés à partir du promoteur P_L (*s086, s084, s083, s082, setC*, *setD* et *eex*), tandis que le huitième (*setR*) est d'orientation contraire et transcrit à partir de P_R (Figure 15) (Beaber *et al.*, 2002). Le gène *s086* code pour un régulateur transcriptionnel putatif et partage de l'homologie avec *cro* du bactériophage λ. *s084* et *s083* sont de fonction inconnue. La protéine codée par *s082* possède un domaine conservé de transglycosylase lytique principalement retrouvé chez les phages, mais sa fonctionnalité n'a pas été vérifiée. SetC et SetD sont les activateurs des gènes de transfert (Beaber and Waldor, 2004; Beaber *et al.*, 2002). Le produit du gène *eex* est un facteur d'exclusion entre ICE de la même famille (Marrero and Waldor, 2005). SetR, quant à lui, est le répresseur permettant le maintien de la quiescence des ICE SXT/R391 (Beaber *et al.*, 2002, 2004). Les gènes de régulation et leurs promoteurs sont extrêmement bien conservés au sein de la famille SXT/R391, ce qui reflète leur importance. En effet, bien que les membres de la famille aient été découverts aux quatre coins du globe et à des décennies d'intervalle, les gènes *setR* et *s086* sont conservés à plus de 97 % au niveau nucléotidique, tandis que *setC* et *setD* sont conservés à plus de 95 % (Wozniak *et al.*, 2009).

Figure 15. Le module de régulation des ICE SXT/R391. Les gènes sont indiqués par les flèches et sont colorés selon leur fonction : vert, activateur transcriptionnel; orange, régulateur putatif; rouge, répresseur; bleu, exclusion; blanc, fonction inconnue.

1.5.1. Répresseur SetR

À l'instar de ce qui est observé chez le bactériophage λ, les gènes impliqués dans le transfert des ICE SXT/R391 ne sont pas exprimés de façon constitutive. La majorité du temps, l'élément est intégré dans le chromosome, et les gènes de recombinaison et de transfert sont silencieux (Beaber *et al.*, 2004). Dans cet état de dormance, les seuls gènes exprimés sont *setR* et d'autres appartenant aux régions d'ADN variables qui sont régulées de façon indépendante, par exemple les gènes de résistance aux antibiotiques. Le gène *setR* code pour une protéine de 215 acides aminés avec un motif de liaison à l'ADN hélice-tour-hélice (HTH_3, PF01381) dans sa portion N-terminale et un motif d'autoprotéolyse dans sa région C-terminale (Peptidase_S24, PF00717) (Finn *et al.*, 2014). SetR est homologue aux répresseurs de la famille CI, principalement retrouvés chez les phages lambdoïdes (Beaber *et al.*, 2002). La nécessité de SetR chez les ICE SXT/R391 est reflétée par l'impossibilité de générer un mutant de *setR* sans 1) simultanément le complémenter en *trans* avec une copie de *setR* ou encore sans 2) effectuer la mutation de *setR* dans une souche où les gènes *setC* et *setD* ont préalablement été inactivés (Beaber *et al.*, 2002, 2004). SetR est nécessaire pour réprimer SetCD et éviter la surexpression de TraV, une composante du pore de conjugaison dont la surexpression entraîne la lyse cellulaire (Armshaw and Pembroke, 2015).

1.5.2. Régulation des promoteurs P_L et P_R exercée par SetR

SetR maintient la quiescence des ICE SXT/R391 en liant quatre sites opérateurs (*OL*, *O1*, *O2*, *O3*) distribués dans la région intergénique entre *s086* et *setR* (Figure 16) (Beaber and Waldor, 2004). Des essais d'empreinte à la DNase I (*deoxyribonuclease I*) ont révélé que l'affinité relative de SetR pour ses opérateurs va comme suit :

O1>*O2*≈*O3*>*OL* (Beaber and Waldor, 2004). Un site additionnel localisé à 800 pb en aval du promoteur P_L a été suggéré, mais jamais vérifié (Beaber and Waldor, 2004). Une étude a proposé que la liaison de SetR à ses quatre sites opérateurs mène à son autorégulation à partir du promoteur P_R (Beaber and Waldor, 2004). Lorsque le niveau de SetR est bas, sa liaison à *O1* mènerait à son autoactivation à partir du promoteur P_R. Lorsque la concentration cellulaire en SetR dépasse une certaine limite, SetR réprimerait sa propre expression par sa liaison au site *O3* de moindre affinité, bloquant ainsi la boîte -10 du promoteur P_R (Beaber and Waldor, 2004). Beaber et Waldor (2004) ont observé l'effet répresseur de SetR sur P_R en introduisant un plasmide portant une fusion transcriptionnelle P_R-*lacZ* dans des souches possédant ou pas SXT, ou possédant les mutants SXT Δ*setCD* ou SXT Δ*setCD* Δ*setR*. La quantification de l'activité β-galactosidase de ces souches a montré que la présence de SXT diminue de 30 % l'activité du promoteur P_R (SXT^- versus SXT^+) (Beaber and Waldor, 2004). La délétion de *setCD* n'affecte pas significativement l'expression de P_R (SXT^+ versus SXT Δ*setCD*), tandis que l'expression de P_R n'est pas altérée par la présence d'un mutant SXT Δ*setCD* Δ*setR* en comparaison avec une souche ne possédant pas l'ICE (SXT^- versus SXT Δ*setCD* Δ*setR*) (Beaber and Waldor, 2004). Ces résultats confirment que SetR réprime P_R. La liaison de SetR aux sites opérateurs *O1* et *OL* encombre le promoteur P_L empêchant ainsi la liaison de la RNAP et l'expression subséquente de *setCD*.

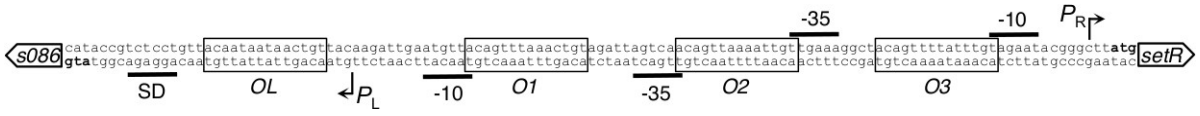

Figure 16. La région intergénique entre *s086* et *setR*. Les sites de liaison de SetR *OL*, *O1*, *O2* et *O3* sont indiqués par des rectangles. Les promoteurs P_L et P_R, ainsi que les boîtes -10 et -35 correspondantes sont indiqués en aval des gènes *setR* et *s086*. Les codons d'initiation sont en gras.

1.5.3. Levée de la répression par SetR

En laboratoire, le transfert de SXT de *E. coli* vers *E. coli* se produit à une fréquence d'environ $1x10^{-4}$ exconjugants par cellule donneuse. Lorsque la réponse SOS est induite, on constate une augmentation d'environ 400 fois de la fréquence de formation d'exconjugants (Beaber *et al.*, 2004). Le signal qui déclenche l'activation et la propagation des ICE SXT/R391 est connecté à la réponse SOS par RecA* selon un processus très similaire à l'activation des phages lambdoïdes (Beaber *et al.*, 2004; Waldor *et al.*, 1996). Une étude inspirée des découvertes faites chez λ a permis de tisser le lien entre RecA* et SetR grâce au mutant *setR*^{G49E} dans lequel le site de clivage Ala-Gly normalement activé par RecA* a été détruit (Beaber *et al.*, 2004; Gimble and Sauer, 1985). Tel qu'anticipé, le mutant *setR*^{G49E} de SXT ne répond plus à la mitomycine C, un agent de réticulation de l'ADN qui est reconnu pour être un inducteur de la réponse SOS chez les bactéries (Szybalski and Iyer, 1964). RecA* est donc le facteur central coordonnant la perception d'ADN dont l'intégrité est compromise (ce qui peut résulter de l'exposition à des antibiotiques), à la réponse bactérienne de stress cellulaire SOS (réparation mutagène de l'ADN) et à l'induction du transfert par conjugaison des ICE SXT/R391 (d'importants vecteurs de dissémination de la résistance aux antibiotiques) (Baharoglu *et al.*, 2010; Beaber *et al.*, 2004).

En résumé, à la suite de dommages à l'ADN, RecA* stimule l'activité latente d'autoprotéolyse de SetR, levant ainsi la répression de SetR sur le promoteur P_L, ce qui entraîne l'expression de *setCD* qui code pour le complexe activateur de l'excision et du transfert conjugatif. SetR perçoit une réponse de stress cellulaire et déclenche l'évasion des ICE SXT/R391 vers des cellules réceptrices potentiellement saines. Pour une réactivité optimale ainsi que pour éviter le gaspillage de ressources cellulaires, l'expression de SetR est fortement régulée et maintenue à un niveau bas

(Beaber and Waldor, 2004). Le transcrit contenant *setR* n'a pas de séquence Shine-Dalgarno qui permet la fixation de l'ARNm à la petite sous-unité du ribosome, un mécanisme post-transcriptionnel qui contribue à maintenir un niveau intracellulaire bas du répresseur SetR (Beaber and Waldor, 2004; Van Etten and Janssen, 1998). L'induction spontanée de la réponse SOS dans une sous-population de cellules expliquerait le niveau basal d'induction du transfert des ICE SXT/R391 (Beaber and Waldor, 2004; McCool *et al.*, 2004).

1.5.4. Complexe activateur SetCD

Les activateurs SetC et SetD sont apparentés respectivement à FlhC (25 % d'identité et 37 % de similarité) et FlhD (23 % d'identité et 41 % de similarité), les sous-unités de l'activateur principal de la synthèse du flagelle. Il a été établi que les délétions simples de *setC* ou de *setD* abolissent l'excision de l'élément et son transfert (Beaber *et al.*, 2002). Ce phénotype a pu être complémenté par la présence de plasmides contenant les gènes individuels sous le contrôle d'un promoteur inductible à l'arabinose (Beaber *et al.*, 2002). Le rôle activateur de SetCD sur l'excision et le transfert a été confirmé avec des fusions *lacZ* transcriptionnelles impliquant les promoteurs des gènes *int*, codant pour l'intégrase, et *traL* et *traG*, les premiers gènes de deux opérons putatifs d'assemblage du pilus et du pore de conjugaison (Beaber *et al.*, 2002). Waldor et son équipe ont également montré que la surexpression de *setCD* chez une cellule possédant SXT est létale et ils ont suggéré que ce phénomène résulte de la surexpression des gènes *tra* entraînée par la surexpression de *setCD* (Beaber *et al.*, 2002). Armshaw et Pembroke ont montré que la surexpression de *traV* cause des dommages à la membrane cellulaire, ce qui supporte cette hypothèse (Armshaw and Pembroke, 2013, 2015).

Des évidences conflictuelles suggèrent une autorégulation potentielle par SetCD. D'un côté, la surexpression de *setCD* mène à une augmentation de quarante fois de l'expression d'une fusion *setD*::*lacZ* dans SXT (Beaber *et al.*, 2002). De l'autre côté, La délétion de *setCD* n'a aucun effet sur l'expression observée du promoteur P_L, ce qui suggère que setCD ne réprime ni active P_L (Beaber *et al.*, 2004).

D'autres évidences conflictuelles ont été rapportées concernant l'activation de l'expression du gène codant l'intégrase dans la cellule réceptrice. En effet, les travaux de Beaber *et al.* (2002) suggèrent que *setCD* n'est pas nécessaire dans les cellules réceptrices étant donné que des mutants Δ*setC* et Δ*setD* de SXT sont capables de se transférer et de s'établir dans les cellules réceptrices, et ce, même si les délétions sont complémentées seulement dans les cellules donneuses. Ces évidences suggèrent soit que l'intégrase est exprimée de manière constitutive, permettant ainsi l'intégration de l'ICE dans une cellule réceptrice qui ne contient pas SetCD, soit que SetCD ou Int sont transloqués dans la cellule réceptrice au travers du pore conjugatif. Paradoxalement, Burrus et Waldor (2003) ont observé qu'un vecteur suicide, possédant seulement l'intégrase exprimée à partir de son propre promoteur et le site d'attachement *attP*, était incapable de s'intégrer dans le chromosome d'une réceptrice qui ne contient pas SetCD. Ce résultat suggère que SetCD est essentiel à l'expression *de novo* de l'intégrase dans une cellule réceptrice.

SetCD active également des gènes dont la fonction n'est pas directement liée à la conjugaison, par exemple le système de recombinaison homologue *bet*/*exo* codé par les ICE SXT/R391 qui génère des hybrides d'ICE intégrés en tandem (Annexe 2) (Garriss *et al.*, 2013a). Les travaux de Wozniak et Waldor (2009) ont montré une levée de la répression de l'opéron *mosAT* par Xis, SetC et SetD (Wozniak and Waldor, 2009). Malgré toutes les évidences portant à croire que le complexe SetCD

est un activateur de la transcription, la liaison directe de SetCD à l'ADN n'a pas encore été vérifiée.

1.6. Hypothèses du projet de doctorat

Les hypothèses concernant la régulation des ICE SXT/R391 qui ont menées aux objectifs de ce projet de doctorat sont les suivantes :

1) Le gène *s086*, l'homologue de *cro* qui code pour un répresseur du bactériophage λ, code pour un répresseur dont l'activité, conjointement avec celle de SetR, constitue un interrupteur génétique régissant la transition entre les états quiescent et conjugatif des ICE SXT/R391. Pour ce faire, S086 lie directement et réprime les promoteurs P_L et P_R;

2) Le cinquième opérateur putatif de SetR participe à la répression du transfert conjugatif;

3) Le complexe activateur SetCD lie directement et active les promoteurs des gènes de transfert ainsi que ceux responsables de la propagation des MGI;

4) SetCD n'active pas directement sa propre expression;

5) L'expression *de novo* de SetCD est essentielle dans la cellule réceptrice pour l'intégration typique de SXT dans *prfC*.

1.7. Objectifs du projet de doctorat

L'objectif global du projet de doctorat était d'affiner notre compréhension de la régulation des ICE SXT/R391 pour mieux comprendre les déterminants de la

dissémination efficace de ces vecteurs de matériel génétique. En utilisant SXT comme prototype de la famille, les objectifs généraux et spécifiques ci-dessous ont été élaborés afin de vérifier les hypothèses du projet.

1) Déterminer la fonction de *s086*
    a. Établir si le produit de *s086* affecte l'activité du promoteur P_L qui conduit, entre autres, l'expression de *s086* et de *setCD* (Figure 15);
    b. Établir si le produit de *s086* affecte l'activité du promoteur P_R, qui conduit l'expression de *setR* (Figure 15);
    c. Vérifier si le rôle de *s086* est dépendant de l'induction de la réponse SOS;
    d. Vérifier la liaison directe de S086 aux promoteurs P_L et P_R.

2) Montrer la fonctionnalité du cinquième site opérateur
    a. Générer une souche reportrice de l'activité du promoteur P_L en présence et en absence du cinquième opérateur.
    b. Vérifier si le rôle du cinquième opérateur est dépendant de l'induction de la réponse SOS;

3) Établir le régulon de SetCD
    a. Identifier les promoteurs dépendants de SetCD dans les ICE SXT/R391 et les MGI qu'ils mobilisent;
    b. Décrire l'organisation des promoteurs dépendants de SetCD;
    c. Déterminer le motif de reconnaissance de SetCD;
    d. Déterminer si SetCD active directement sa propre expression;
    e. Déterminer si *setCD* est essentiel dans les cellules réceptrices pour l'établissement typique de SXT dans *prfC*.

CHAPITRE 2

RÉSULTATS

La majorité des résultats obtenus au cours du projet de doctorat sont contenus dans deux articles scientifiques, l'un publié et l'autre soumis dans un journal avec comité de lecture. Le chapitre 2 présente ces deux articles, qui répondent aux objectifs du projet de doctorat présentés dans la section précédente. Notez que les Annexes 2 et 3 contiennent des articles scientifiques publiés dans lesquels certains résultats de ce projet concernant l'activation par SetCD ont été inclus.

L'annexe 5, quant à elle, contient un chapitre de livre auquel j'ai contribué grâce à mon expertise en gel à retardement.

2.1. CroS est essentiel à l'induction du transfert des ICE SXT/R391 dépendante de la réponse SOS.

2.1.1. Présentation de l'article

Au moment où le projet a débuté, le modèle de régulation des ICE SXT/R391 faisait état seulement de deux régulateurs, SetR et SetCD. Une partie du projet de doctorat consistait à définir le rôle de deux composantes additionnelles qui contribuent à la régulation des ICE SXT/R391. Premièrement, la fonction de *s086*, un régulateur transcriptionnel renommé *croS* dans cet article, a été investiguée (Objectif 1).

Deuxièmement, l'implication dans le contrôle du transfert d'un cinquième site opérateur, précédemment suggéré pour la liaison par SetR, a été évaluée (Objectif 2).

La fonction de CroS a pu être mise en lumière en vérifiant l'effet de sa surexpression et de sa délétion sur la fréquence de transfert de l'élément prototype SXT ainsi que sur l'activité des promoteurs P_L et P_R qui régulent respectivement l'expression des gènes *croS-setCD* et *setR*. Les résultats ont montré que CroS est essentiel à l'induction optimale du transfert à la suite de l'activation de la réponse SOS. Cette fonction est médiée par la liaison directe de CroS aux sites opérateurs également liés par SetR. La liaison de CroS à ses sites opérateurs résulte en la répression des promoteurs P_L et P_R, évitant ainsi la recrudescence prématurée de SetR et la surexpression de SetCD.

Des essais de gel à retardement ont permis d'observer la liaison de SetR et de CroS au cinquième opérateur dont la fonctionnalité était incertaine. Dans la première version soumise de l'article, des essais avec le gène rapporteur *lacZ* ont également montré que cet opérateur participe à la répression du transcrit débutant au promoteur P_L qui contient *setCD*. Ce résultat suggère que le cinquième site opérateur contribue à la répression du transfert conjugatif par SetR. Cependant, les évaluateurs de *Journal of Bacteriology* n'étaient pas convaincus de ces résultats, ils ont donc été enlevés de l'article pour la version publiée.

Dans cet article, nous nous sommes également intéressés aux évidences conflictuelles concernant un processus potentiel d'autoactivation par SetCD. Le niveau d'expression d'une fusion *setD*'-'*lacZ* ainsi que les niveaux d'ARNm portant *setR* et *croS* sont augmentés dans des cellules soumises à une surexpression de *setCD*. Ce phénotype est complémenté par la délétion de *traI*. La surexpression de

setCD mène à une surexpression de TraI qui est, entre autres, un facteur d'initiation de la réplication des ICE SXT/R391. Cette augmentation des niveaux d'ARNm de *setR*, *croS* et *setCD* serait donc causée par un nombre de copies de SXT plus élevé, ce qui suggère fortement que SetCD ne régule pas directement sa propre expression. Ces résultats sont présentés dans l'article qui suit :

Poulin-Laprade, D. et Burrus, V. (2015) A λ Cro-like repressor is essential for the induction of conjugative transfer of SXT/R391 elements in response to DNA damage. *Journal of Bacteriology* (Numéro de soumission : JB00638-15R1).

2.1.2. Contribution à l'article

L'approche expérimentale a été définie conjointement avec Vincent Burrus. J'ai réalisé la totalité des expériences qui ont mené aux conclusions de cet article. J'ai également rédigé le manuscrit et généré les figures qui ont par la suite été révisées par Vincent Burrus, Nicolas Carraro et Alain Lavigueur.

2.1.3. Page titre

A λ Cro-like repressor is essential for the induction of conjugative transfer of SXT/R391 elements in response to DNA damage

Dominic Poulin-Laprade and Vincent Burrus

Running Head: CroS is a key factor for induction of SXT transfer

2.1.4. Abstract

Integrative and Conjugative Elements (ICEs) of the SXT/R391 family are the main contributors to acquired multidrug resistance in the seventh pandemic lineage of *Vibrio cholerae*, the etiological agent of the diarrhoeal disease cholera. Conjugative transfer of SXT/R391 ICEs is triggered by antibiotics and agents promoting DNA damage through RecA-dependent autoproteolysis of SetR, an ICE-encoded λ CI-like repressor. Here, we describe the role of CroS, a distant λ Cro homolog, as a key component contributing to the regulation of expression of the activator SetCD that orchestrates the expression of the conjugative transfer genes. We show that deletion of *croS* abolishes the SOS-response dependent induction of SXT despite the presence of a functional *setR* gene. Using qRT-PCR and *lacZ* reporter assays, we also show that CroS represses *setR* and *setCD* expression by binding to operator sites shared with SetR. Furthermore, we provide evidence of an additional operator site bound by SetR and CroS. Finally, we show that SetCD expression generates a positive feedback loop due to SXT excision and replication in a fraction of the cell population. Together, these results refine our understanding of the genetic regulation governing the propagation of major vectors of multidrug resistance.

2.1.5. Importance

Healthcare systems worldwide are challenged by an alarming drug resistance crisis caused by the massive and rapid propagation of antibiotic resistance genes, and associated emergence of multidrug resistant pathogenic bacteria. SXT/R391 ICEs contribute to this phenomenon not only in clinical and environmental Vibrios but also in several species of *Enterobacteriaceae*. We have identified and characterized here

the regulator CroS as a key factor for the stimulation of conjugative transfer of these ICEs in response to DNA-damaging agents. We have also untangled conflicting evidence regarding autoactivation of transfer by the master activator of SXT/R391 ICEs, SetCD. Discovery of CroS provides a clearer and more complete understanding of the regulatory network that governs the dissemination of SXT/R391 ICEs in bacterial populations.

2.1.6. Introduction

Cholera is a severe infectious disease caused by the ingestion of food or water contaminated by *Vibrio cholerae*. This disease remains widespread in regions with limited access to clean water and where poor sanitation allows for easy dissemination of the bacterium in drinking water sources. Cholera is characterized by a profuse watery diarrhoea that rapidly induces massive fluid loss, causing a severe dehydration of the patient that may lead to death within 24 hours of symptom onset. Although epidemic cholera is usually caused by *Vibrio cholerae* O1, the unusual serogroup O139 emerged in the early 90s as the cause of a cholera outbreak in India (1). O139 clinical isolates were found to be resistant to sulfamethoxazole and trimethoprim, two antibiotics commonly used for the treatment of severe cases of cholera (2). This resistance was found to be transmissible and linked to an Integrative and Conjugative Element (ICE) named SXT (3). ICEs are self-transmissible bacterial mobile elements that play a major role in gene exchange in bacterial populations as they are horizontally transferred via conjugation by a process similar to the one used by many conjugative plasmids (4, 5). Unlike plasmids, ICEs do not stably maintain in an episomal form and are rather found integrated into the chromosome. SXT integrates into the chromosome of *V. cholerae* in a site-specific manner into the 5' end of *prfC*, a gene coding for the peptide release factor RF3 (6). Since the discovery

of SXT, SXT or related ICEs have been found to be prevalent in the seventh pandemic isolates of *V. cholerae*, sporadically present in other *Vibrio* species (7-9) and other *Gammaproteobacteria* of clinical origin or isolated from the aquatic environment, such as *Photobacterium* (10), *Proteus* (11), *Alteromonas* (12), *Marinomonas* (13), and *Shewanella* (9, 14). SXT is closely related to R391, an ICE conferring resistance to kanamycin and mercury, originally detected in a 1967 South African isolate of *Providencia rettgeri* (15, 16). All ICEs related to SXT and R391 are grouped into a single family, namely the SXT/R391 family, because they all share the same chromosomal integration site and a set of conserved genes essential for site-specific integration, conjugative transfer and regulation (6, 8). SXT/R391 ICEs also contain variable DNA insertions conferring adaptive traits including resistance to antibiotics, heavy metals and bacteriophage infection (8, 17, 18), synthesis of the second messenger c-di-GMP (19) and homologous recombination and mutagenic repair systems (20, 21). Beside their own transfer, SXT/R391 ICEs have been shown to mobilize plasmids, phylogenetically unrelated genomic islands (MGIs) and up to 1.5 Mb of chromosomal DNA by Hfr transfer (22, 23).

The most conserved genes (97% identity at the nucleotide level) shared by SXT/R391 ICEs are *s086* and *setR*, which belong to the regulatory module located at the rightmost end of the integrated ICE, near the *attR* attachment site (Figure 1A) (8). The regulatory module contains eight ORFs, seven of which are in the same orientation (*s086*, *s084*, *s083*, *s082*, *setD*, *setC*, *eex*), while the last gene, *setR*, is divergently transcribed (Figure 1A and B) (24). *eex* together with the convergent gene *traG* code for the entry exclusion system of SXT/R391 ICEs (25). The overlapping genes *setC* and *setD* encode the SetCD activator complex. SetCD was shown to bind upstream of the -35 sequence of 11 promoters in SXT/R391 ICEs, activating the expression of over 40 genes essential for site-specific and homologous recombination, ICE replication and partition, and conjugative transfer (26-28). The function of *s082, s083*,

s084 and *s086* remains unknown. *In silico* analysis has revealed that *s086* codes for a putative small basic protein of 9.4 kDa with a predicted helix-turn-helix (HTH) DNA binding domain. *setR* codes for the main repressor of SXT/R391 ICEs (24, 29). SetR contains a HTH-XRE (PF01381) motif in its N-terminal moiety and a C-terminal LexA-like autoproteolysis domain (PF00717). SetR is related to 434 CI and other lambdoid phage repressors and has been shown to bind to four operator sites located between *s086* and *setR* (Figure 1B) (24, 30). These operators, *O1*, *O2*, *O3,* and *OL*, are part of P_L and P_R, two divergent overlapping promoters located in the intergenic region between *s086* and *setR*. The relative affinities of SetR for these four operator sites and their positions suggest an autoregulation of *setR* expression from P_R (30) and a repression of the operon containing *s086* and *setCD* driven from P_L (29). Repression at P_L is alleviated when the cellular pool of SetR drops, most likely as a result of autoproteolysis stimulated by DNA damage-induced activation of RecA (29).

The relative positions of *setR* and *s086* are reminiscent of *cI* and *cro* carried by the bacteriophage λ. CI and Cro form a pair that governs the transition between the lysogenic and lytic pathways of λ lifecycle (31, 32). To date, SXT/R391 ICEs have been known to be regulated by only two transcriptional regulators: the repressor SetR and the activator complex SetCD (24). In this study, we identify S086 as a third regulator, which will be referred to as CroS for Cro-like repressor of SXT. We demonstrate that CroS is a key factor for alleviation of SetR repression and induction of conjugative transfer of SXT/R391 ICEs. By binding to five operator sites shared with SetR, CroS acts as a repressor of P_L and P_R, thereby repressing both *setCD* and *setR* in a mechanism that is evocative of λ Cro repression on CI and CII (33). We also show that SetCD increases its own expression not by activating P_L but by indirectly generating a positive feedback loop triggered by activation of SXT replication, which increases SXT copy number and ultimately, the level of *setCD* mRNA transcript. Finally, we propose a new model of SXT transfer regulation that includes CroS as a

repressor of P_L and P_R, thereby promoting the SOS-dependent induction of SXT/R391 ICEs transfer.

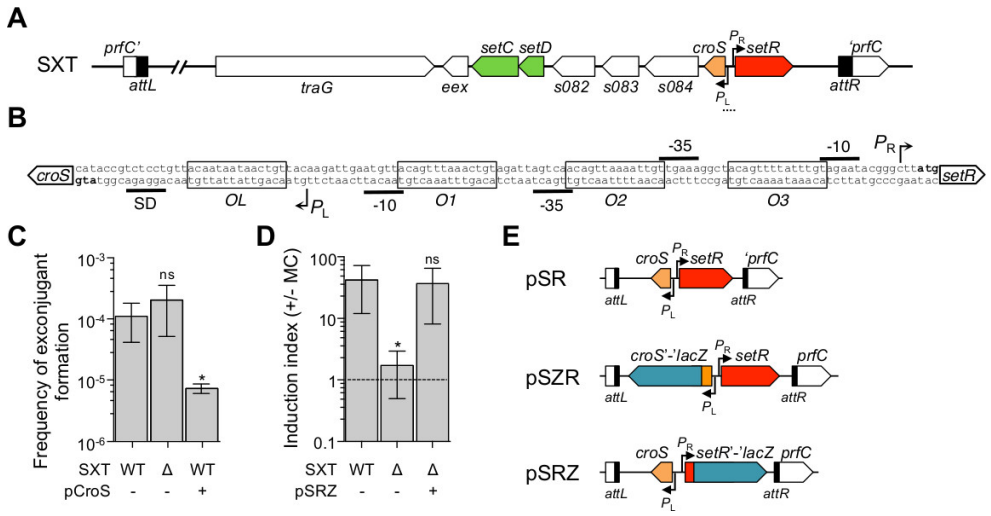

Figure 1. Effect of *croS* on SXT conjugative transfer. (A) Schematic representation of the regulatory module of SXT. The dotted line indicates the region enlarged in panel B. (B) The intergenic region between *croS*, previously referred to as *s086*, and *setR* in SXT as represented by Beaber and Waldor (30). The P_L and P_R promoters, the *OL*, *O1*, *O2* and *O3* operators, the -10 and -35 elements, the Shine-Dalgarno (SD) sequence of P_L, and start codons of *croS* and *setR* are represented. (C) Effect of the deletion or overexpression of *croS* on SXT transfer. Conjugation assays were carried out using as donors *E. coli* BW25113 SXT (VB17, WT) or its Δ*croS* mutant (DPL353, Δ). Overexpression assays were carried out by expressing *croS* from the arabinose-inducible P_{BAD} promoter in pCroS (DPL525). *E. coli* CAG18439 (Tc^r) was used as the recipient. Frequencies of exconjugant formation were expressed as a number of exconjugant CFUs (Su^r Tm^r Tc^r) per recipient CFUs (Tc^r). (D) Effect of the deletion of *croS* on the induction of SXT transfer by mitomycin C (MC). Donor strains were VB17, DPL353 and DPL263 in which *croS* was expressed from its native P_L promoter in pSRZ integrated in single copy in tandem with SXT Δ*croS.* The recipient strain was CAG18439. The induction index corresponds to the ratio between the frequencies of exconjugant formation obtained with and without MC treatment prior to transfer. The dashed line indicates an induction index of 1, i.e. no difference between the induced and non-induced conditions. The bars represent the mean and standard deviation values obtained from at least three independent experiments. Statistical analyses were performed on the logarithm value of the means using two-tailed unpaired t-tests. Asterisks indicate that P-values are <0.05 when compared to the WT. (E) Schematic representation of the insert of the pSR, pSZR and pSRZ plasmids site-specifically integrated into the 5' end of *prfC* alone or in tandem with SXT. The plasmid backbone is not represented. In pSZR and pSRZ, *lacZ* was translationally fused to the 6th codon of *croS* and *setR*, respectively.

2.1.7. Materials and methods

2.1.7.1. Bacterial strains and media

The bacterial strains used in this study are described in Table 1. The strains were routinely grown in lysogeny broth (LB-Miller, EMD) at 37°C in an orbital shaker/incubator and were maintained at -80°C in LB broth containing 15% (vol/vol) glycerol. Antibiotics were were used at the following concentrations: ampicillin (Ap), 100 μg ml^{-1}; chloramphenicol (Cm), 20 μg ml^{-1}; erythromycin (Em), 10 μg ml^{-1}; kanamycin (Kn), 50 μg ml^{-1}or 10 μg ml^{-1} for single copy integrants of pOPlacZ; spectinomycin (Sp), 50 μg ml^{-1}; sulfamethoxazole (Su), 160 μg ml^{-1}; tetracycline (Tc), 12 μg ml^{-1}; trimethoprim (Tm), 32 μg ml^{-1}. When required, bacterial cultures were supplemented with 0.1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG), 0.02% L-arabinose or 100 ng ml^{-1} mitomycin C (MC).

2.1.7.2. Bacterial conjugations

Conjugation assays were performed as described in Burrus and Waldor (34) with the following modifications. When induction with mitomycin C (MC) was performed, overnight cultures of donor and recipient cells were diluted 1:4, the donor cultures were supplemented or not with MC and incubated for 1 hour at 37°C with shaking. The cells were harvested by centrifugation for 3 min at 1200g, washed in 1 volume of LB broth and resuspended in 1/20 volume of LB broth. Mating mixtures were then deposited on LB agar plates and incubated at 37˚C for 2 hours.

Table 1. Strains and plasmids used in this study

Strain or plasmid	Relevant genotype and phenotype[a]	References
E. coli strains		
CAG18439	MG1655 *lacZU118 lacI42*::Tn*10* (Tc^r)	(76)
HW220	CAG18439 *prfC*::SXT (Su^r Tm^r)	(6)
DC147	CAG18439 *prfC*::SXT Δ*traI*::*aph* (Su^r Tm^r Kn^r)	This study
BW25113	F^-, Δ(*araD-araB*)*567*, Δ*lacZ4787*(::*rrnB-3*), λ^-, *rph-1*, Δ(*rhaD-rhaB*)*568*, *hsdR514*	(37, 77)
β2163	(F^-) RP4-2-Tc::Mu Δ*dapA*::(*erm-pir*) (Kn^r Em^r)	(36)
BL21(DE3)	F^-, *ompT* Δ*hsdS gal dcm* λDE3	Novagen
VB17	BW25113 *prfC*::SXT (Su^r Tm^r)	(28)
DPL353	BW25113 *prfC*::SXT Δ*croS* (Su^r Tm^r)	This study
DPL263	BW25113 *prfC*::[SXT Δ*croS*]-pSRZ (Su^r Tm^r Cm^r)	This study
DPL525	BW25113 *prfC*::SXT pCroS (Su^r Tm^r Ap^r)	This study
DPL297	BW25113 *prfC*::SXT *setD'-'lacZ-aad7* Δ*setC* (Su^r Tm^r Sp^r)	This study
DPL548	BW25113 *prfC*::[SXT Δ*traI*::*aph setD'-'lacZ-aad7* Δ*setC*] (Su^r Tm^r Sp^r Kn^r)	This study
DPL231	BW25113 *prfC*::pSR (*croS-setR*) (Su^r Tm^r)	This study
DPL241	BW25113 *prfC*::pSZR (*croS*'-'*lacZ-setR*) (Su^r Tm^r)	This study
DPL246	BW25113 *prfC*::pSRZ (*croS-setR*'-'*lacZ*) (Su^r Tm^r)	This study
Plasmids		
pVI36	PCR template for one-step chromosomal gene inactivation (Sp^r)	(78)
pKD13	PCR template for one-step chromosomal gene inactivation (Kn^r)	(37)
pVI42B	pVI36 BamHI::P_{lac}-*lacZ* (Sp^r)	(20)
pKD46	λRed recombination arabinose-inducible encoding plasmid (Ap^r)	(37)
pSW23T	$oriT_{RP4}$ $oriV_{R6K\gamma}$ (Cm^r)	(36)
pSR	pSW23T *croS-setR* $attP_{SXT}$ (Cm^r)	This study
pSZR	pSW23T *croS'-'lacZ-aad7-setR* $attP_{SXT}$ (Sp^r Cm^r)	This study
pSRZ	pSW23T *croS-setR'-'lacZ-aad7* $attP_{SXT}$ (Sp^r Cm^r)	This study
pVI68	pAH57 Δ(xis_{λ}-int_{λ})::int_{SXT} (Ts)	(20)
pBAD-TOPO	ori_{pBR322} *bla araC* P_{BAD} (Ap^r)	Invitrogen
pBAD30	ori_{p15A} *bla araC* P_{BAD} (Ap^r)	(79)
pGG2B	pBAD30::*setDC* (Ap^r)	(28)
pCroS	pBAD-TOPO::*croS* (5'-3') (Ap^r)	This study
pCroS-inv	pBAD-TOPO::*croS* (3'-5') (Ap^r)	This study
pSetR	pBAD-TOPO::*setR* (Ap^r)	This study
pTraI	pBAD-TOPO::*traI* (Ap^r)	(27)
pCR2.1-TOPO	ori_{pUC} (Ap^r Kn^r)	Invitrogen
pDPL128	pCR2.1-TOPO::P_L-P_R (Ap^r Kn^r)	This study
pGEX-croS	pGEX-6P-1::*croS* (Ap^r)	This study
pGEX-setR	pGEX-6P-1::*setR* (Ap^r)	This study

[a] Ap^r, ampicillin resistant; Cm^r, chloramphenicol resistant; Em^r, erythromycin resistant; Kn^r, kanamycin resistant; Sp^r, spectinomycin resistant; Su^r, sulfamethoxazole resistant; Tc^r, tetracycline resistant; Tm^r, trimethoprim resistant; Ts, thermosensitive.

2.1.7.3. Molecular biology methods

Plasmid DNA was prepared using the EZ-10 Spin Column Plasmid DNA Minipreps Kit (Biobasic) according to manufacturer's instructions. All the enzymes used in this study were purchased from New England BioLabs. PCR assays were performed using either the Taq, the Q5 or the Phusion polymerase according to manufacturer's instructions. When necessary, PCR products were purified using an EZ-10 Spin Column PCR Products Purification Kit (Biobasic) according to manufacturer's instructions. *E. coli* was transformed by electroporation as described by Dower *et al*. (35) in a BioRad GenePulser Xcell apparatus set at 25 µF, 200 Ohms and 1.8 kV using 1-mm gap electroporation cuvettes. Sequencing reactions were performed by the Plateforme de Séquençage et de Génotypage du Centre de Recherche du CHUL (Québec, QC, Canada).

2.1.7.4. Plasmid and strain constructions

Plasmids and oligonucleotides used in this study are listed in Tables 1 and S1. Plasmids pCroS and pSetR were constructed by amplifying *croS* and *setR* using primer pairs s086pBADf/s086pBADr.s or setRpBADf/setRpBADr.s and genomic DNA of *E. coli* HW220 as the template, and cloning into the TA cloning expression vector pBAD-TOPO (Invitrogen) according to the manufacturer's instructions. pSR was constructed by PCR amplification of the *croS-setR-attP* region resulting from the excision of SXT in *E. coli* HW220 using the primer pair PEs086/VISLR3. The 2 022-bp PCR product was first cloned into pCR2.1-TOPO (Invitrogen) according to the manufacturer's instructions. The cloned fragment was then recovered by NotI/BamHI digestion and then ligated to NotI/BamHI-digested pSW23T, yielding pSR. pSW23T

contains the RP4 mobilization locus, the R6K *pir*-dependent conditional origin of replication and a chloramphenicol resistance gene (36). Plasmid pDPL128 was constructed by cloning into pCR2.1-TOPO (Invitrogen) the intergenic region between *setR* and *croS*, which was amplified by PCR using primer pair RRs086-4/RRs086setRr and genomic DNA of *E. coli* HW220 as the template.

Δ*croS* and Δ*traI* derivatives of SXT were constructed in *E. coli* HW220 by using the one-step chromosomal gene inactivation technique (37) with primer pairs delS086F/dels086R and TraIWF2/TraIWR3, respectively, and the pKD13 template. SXT Δ*traI*::*aph* complemented with pTraI was transferred by conjugation into *E. coli* BW25113. To construct *E. coli* DPL548, the Δ*traI*::*aph* mutation was then transduced into *E. coli* DPL297 (BW25113 *prfC*::SXT Δ*setCD*::*lacZ-aad7*) using P1vir generalized transduction. Strains containing pSR integrated into the 5' end of *prfC* were constructed by mating the RP4^{+} *pir*$^{+}$ strain *E. coli* β2163 pSR with *E. coli* BW25113 containing pVI68 as the recipient strain. pVI68 expresses the integrase of SXT, thereby allowing site-specific integration of pSR, which was verified by amplifying the *attL* and *attR* sites by PCR using primer pairs EattBR/VISRF and EattBF/VISLR3, respectively. The absence of multiple integrations of pSR in tandem arrays was verified with primers VISLR3 and VISRF. The resulting strain, *E. coli* DPL231, contains a unique copy of the *croS-setR* region integrated site-specifically into *prfC*. Translational *lacZ* fusions *setR*'-'*lacZ*, *croS*'-'*lacZ* and *setCD*'-'*lacZ* were constructed using the one-step chromosomal gene inactivation technique (37) using primer pairs DATsetR-lacZf/DATsetR-lacZr, DATsetQ-lacZf/DATsetQ-lacZr and DATsetCD-lacZf/DATsetCD-lacZr, respectively and pVI42B as the template. In each construct, the 6th codon of the gene of interest was fused to the 9th codon of *lacZ*. The *setR*'-'*lacZ* and *croS*'-'*lacZ* mutations targeted BW25113 derivative strains containing pSR integrated into *prfC* (DPL231), whereas the *setCD*'-'*lacZ* mutation targeted SXT in *E.*

coli VB17. SXT Δ*croS* was transferred into DPL246 by conjugation to generate DPL263.

2.1.7.5. RNA isolation and qRT-PCR

Total RNA was extracted and cDNA were synthesized as described previously (19, 38). RNA integrity was verified using a 2100 Bioanalyzer (Agilent). Quantitative amplification of *setR*, *croS*, *setD* and *setC* was carried out using primer pairs RTsetR-F/RTsetR-R, RTcroS-F/RTcroS-R, RTsetD-F/RTsetD-R and RTsetC-F/RTsetC-R, respectively. For normalization, *rpoZ* was amplified using primers RTrpoZcoli-F/ RTrpoZcoli-R and the ΔΔCt calculation method. Experiments were repeated three times in triplicate and combined. Melting curves were carried out on the final reaction products (156-165 bp) to confirm that amplification was specific to targets. All primer pairs exhibited efficiencies between 95% and 97%.

2.1.7.6. Real-time quantitative PCR assays for relative quantification of *attB* and *attP*

The frequency of excision and copy number of the excised circular form of SXT were assessed by real-time quantitative PCR (qPCR) as described elsewhere (39). Genomic DNA was obtained from cell cultures of *E. coli* VB17, DPL297 pGG2B and DPL548 pGG2B grown in conditions described for the qRT-PCR assays. Bacterial cultures were grown in LB medium at 37°C to an OD_{600} of 0.2, then for an additional 2 hours at 37°C. *prfC*, *attB* and *attP* were quantified using primer pairs prfC.qec.F1/prfC.qec.R1, attB.qec.F2/attB.qec.R2 and attP.qec.F2/attP.qec.R2,

respectively (27) (Table S1). Three biological replicates of qPCR experiments were performed on the RNomics platform of the Laboratoire de Génomique Fonctionnelle de l'Université de Sherbrooke (http://lgfus.ca) (Sherbrooke, QC, Canada). The normalization was performed as described previously (39).

2.1.7.7. β-galactosidase assays

The substrate used to determine the LacZ levels was the o-2-nitrophenyl-β-D-galactopyranoside (ONPG) and the assays were carried out as described previously (40) in LB medium and 100 µg ml^{-1} ampicillin for maintenance of pCroS and pGG2B. The induction of *croS* and *setCD* expression and of the SOS response were done by supplementing respectively with 0.2% arabinose and/or 100 ng ml^{-1} MC a refreshed culture grown to an OD_{600nm} of 0.2 followed by 2-h incubation at 37°C with shaking prior cell sampling.

2.1.7.8. Expression and purification of the SetR and CroS proteins

Overnight-grown cultures of *E. coli* BL21(DE3) bearing pGEX-croS or pGEX-setR were diluted 1/1000 in fresh 2× YTA broth and incubated at 37°C with agitation. Protein expression was induced at mid-exponential phase (OD_{600} of 0.6) with 0.1 mM IPTG. The cultures were grown for an additional 3 h at 37°C with agitation. Cells were collected by centrifugation, re-suspended in PBS containing 1% Triton X-100, 1 mM PMSF and a cocktail of protease inhibitors (Protease Inhibitor Cocktail, Sigma). Cells were lysed by sonication, centrifuged to pellet the debris and CroS and SetR were recovered by affinity chromatography using the GST purification module (GE

Healthcare) with the PreScission protease (GE Healthcare) according to the manufacturer's instructions. Protein concentration was estimated using the Bradford protein assay (Bio-Rad) and purity was determined by SDS-PAGE analysis.

2.1.7.9. Dimerization assay

Prior to dimerization assay, purity of CroS was estimated by SDS-PAGE analysis to 95%. No contaminant proteins with a size similar to the one of CroS was observed. 4.7 µg of purified CroS were incubated with 0.625% glutaraldehyde for 30 min at room temperature. The reaction mix was then denatured at 100°C for 3 min, loaded on a 15% SDS-PAGE gel, and run for 75 min at 150 V (41). The gel was then stained with Coomassie Brilliant Blue R-250. The Precision Plus Protein Kaleidoscope Standards ladder (Bio-Rad) was used as a molecular weight marker.

2.1.7.10. Electrophoretic mobility shift assays (EMSA)

The probes used in EMSA assays were either annealed oligonucleotides or PCR products. Purified PCR products and a single oligonucleotide for each annealed pairs were labeled with [γ-32P]-dATP using the T4 polynucleotide kinase. An equimolar quantity of the complementary oligonucleotide was added, the volume brought to 100 µl in 20 mM Tris-HCl pH 8.0, 20 mM KCl, denatured at 95˚C for 5 min, then cooled to room temperature over 1 hour. The [γ-^{32}P]-labeled double stranded DNA fragments were separated from unincorporated [γ-^{32}P]-dATP using Illustra MicroSpin G-25 Columns (GE Healthcare) according to the manufacturer's instructions. Purified CroS and SetR proteins were pre-incubated in 24 µl of buffer I (20 mM Tris-HCl pH 8.0, 20

mM KCl, 1 mM $MgCl_2$, 5% glycerol) for 20 min at 4˚C. For PCR probes 1 to 6 (Figure 4D), non-specific competitor DNA was added by supplementing buffer I with 50 ng ml^{-1} of sonicated salmon sperm DNA. Radiolabeled probes (2 000 CPM) were added and samples were incubated for 30 minutes at 37˚C, then 10 minutes at 4˚C. Samples were immediately loaded onto a pre-run 4% polyacrylamide gel containing 0.5× Tris-Borate-EDTA buffer and migrated by electrophoresis at a constant voltage of 120 V at 4°C for 1 h 15. Gels were dried, exposed to a Phosphor Screen (Kodak) and results were visualized using a Typhoon FLA 9500 imager (GE Healthcare). The densitometry values (intensity units/mm^2) of each shifted operator were quantified using the Quantity One software (BioRad). The background was subtracted, then percentages of maximal intensity were calculated by dividing the density of each band by the one of strongest density, both obtained using 1 µM of either SetR or CroS.

2.1.7.11. DNase I protection assays

To analyze the top strand, the probe was prepared first by digesting pDPL128 with HindIII. The linearized plasmid was then dephosphorylated using the Antartic phosphatase, and subsequently digested with NsiI. To analyze the bottom strand, the probe was prepared by digesting pDPL128 with XhoI, Antartic phosphatase dephosphorylation, and digestion with KpnI. The probes were then end-labeled using the T4 polynucleotide kinase and [γ-^{32}P]-ATP 6 000 Ci $mmol^{-1}$. 50 µl binding reactions between CroS and the radiolabeled probe (20 000 CPM) were performed as described for EMSA assays. 50 µl of cofactor solution (10 mM $MgCl_2$, 5 mM $CaCl_2$) and 0.1 U of DNaseI was added and incubated for 2 min at room temperature. The reaction was terminated by addition of 100 µl of stop solution (1% SDS, 200 mM NaCl, 20 mM EDTA pH 8.0, 40 µg ml^{-1} of tRNA) and extracted with 200 µl of phenol/chloroform/isoamyl alcohol (25:24:1). DNA was precipitated with three

volumes of ethanol, washed with 70% ethanol, dried and resuspended in sequencing gel loading buffer and denatured for 5 min at 95˚C. Sequencing reactions that serve as ladder were done using the Sequenase 2.0 DNA sequencing kit (Affymetrix-usb) and primers FP-frag2-HindIII or FP-frag2-XhoI for the top and bottom strands, respectively. After electrophoresis through a denaturing 10% polyacrylamide sequencing gel in 0.8× glycerol tolerant gel buffer, the gel was dried, and detection was carried out using a Phosphor Screen (Kodak) and Typhoon FLA 9500 imager (GE Healthcare).

2.1.8. Results

2.1.8.1. CroS is important for the DNA damage-induced activation of SXT transfer

To investigate the role of CroS in the biology of SXT/R391 ICEs, we first monitored the effect of its deletion or overexpression on the transfer of SXT in non-induced conditions and in the presence of the DNA-damaging agent mitomycin C (MC). First, we observed that the basal level of SXT transfer remained unaffected by *croS* deletion in non-induced conditions (Figure 1C). However, while transfer of wild-type SXT was induced more than 30 folds by MC, the Δ*croS* mutant remained virtually unresponsive to MC (Figure 1D). This phenotype could result from a polar effect of *croS* deletion on the expression of the downstream genes *setCD*, which are part of the same operon. To rule out this possibility, we carried out a complementation assay by expressing *croS* from its native promoter P_L from pSRZ inserted as a single copy into the chromosomal gene *prfC* in a tandem fashion with SXT Δ*croS* (Figure 1E). pSRZ restored the 30-fold MC-dependent induction to SXT Δ*croS*, thereby confirming that the lack of response to MC exhibited by the Δ*croS* mutant was not due to a polar

effect on the expression of *setCD* (Figure 1D). Moreover, overexpression of *croS* from a P_{BAD} promoter provided by pCroS led to a 10-fold reduction of SXT transfer (Figure 1C). Altogether, these results suggest that CroS is a crucial regulator for stimulation of SXT transfer in response to DNA damage, and suggest the existence of a λ-like SetR-CroS switch in the regulatory network of SXT/R391 ICEs.

2.1.8.2. Predicted structure and DNA binding domain of CroS

Extensive studies of many members of the Cro-like family have provided valuable insights for predicting distinctive features of newly discovered Cro-like repressors, which allowed this family to be used as a model in evolutionary studies of protein structure and DNA recognition (42, 43). CroS is predicted to contain a DNA-binding helix-turn-helix domain (HTH_XRE, COG4197) (Figure 2A). HHpred analyses revealed that CroS shares 25% identity with λ Cro and 35% identity with Xfaso 1, a putative regulator encoded by a prophage of *Xylella fastidiosa*. Although these two homologous regulators have similar regulatory functions and are encoded by genes that are found in a similar context, λ Cro and Xfaso 1 have been shown to be structurally different. λ Cro is a dimeric protein with a mixed β-sheet/α-helix fold, while Xfaso 1 is a monomeric all-α helical fold protein in solution (42). A high-accuracy model of the predicted structure of CroS based on the crystal structure of Xfaso 1 was obtained using Phyre2 (44) (Figure 2B and C). This model suggests that CroS exhibits an all-α structure and exists in solution either as monomer or as dimer. Despite extended similarities, CroS does not contain the cysteine residues C42 and C55 forming the intra-subunit cystine bond observed in Xfaso 1 (42).

We carried out an oligomerization assay using glutaraldehyde cross-linking and observed two bands with sizes compatible with either monomeric or dimeric forms of CroS in solution (Figure 2D) indicating that CroS can form higher order complexes even in the absence of target DNA.

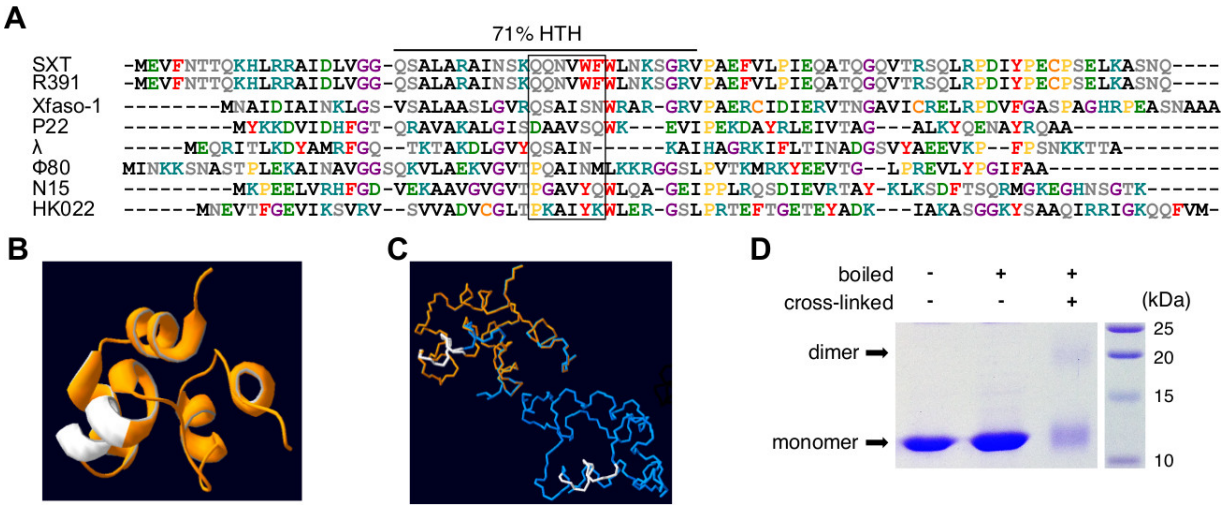

Figure 2. Predicted structure and DNA binding domain of CroS. (A) Amino acid sequence alignment of Cro-like repressors computed by Clustal Omega (72). The source ICE or phage is indicated on the left side. Sequences are sorted according to phylogenetic proximity. Amino acids are color-coded according to their chemical properties: pink, positively charged; green, negatively charged; gray, no charge, polar and hydrophilic; red, aromatic; black, hydrophobic; orange, cysteine; yellow, proline; purple, glycine. DNA recognition helices as described by Hall *et al.* (43) are indicated by a black box. The Dodd and Egan method predicted a 71% chance of a helix-turn-helix (HTH) motif (73). (B) 3D ribbon structure of CroS as modeled by Phyre2 with template c3bd1B (Xfaso 1 in *Xylella fastidiosa* strain ann-1) with 99.9% confidence, 76% coverage and 38% identity (44). The putative DNA recognition helix is shown in white. (C) Superposition of dot surfaces of PBDs of a monomer of CroS (orange) and a dimer of Xfaso 1 (blue) as crystallized by Roessler *et al.* (42) with the predicted DNA recognition helices shown in white. (D) SDS-PAGE analysis of purified CroS protein cross-linked or not with 0.625% glutaraldehyde.

2.1.8.3. CroS acts as a repressor of *setR*

CroS similarity with other Cro-like repressors and the relative positions of *croS* and *setR* led us to hypothesize that the inability of a Δ*croS* mutant to respond to MC induction was at least in part due to the lack of repression of the P_R promoter that drives *setR* expression. To test this hypothesis, we used RT-qPCR to measure and compare the relative mRNA levels of *setR* in wild-type SXT and its Δ*croS* mutant, with or without overexpression of *croS* and in the presence or absence of MC (Figure 3A and B *setR*). We observed that while *setR* expression only increased threefold upon MC induction in wild-type SXT, it increased tenfold in the Δ*croS* mutant (Figure 3A *setR*). Complementation with pSRZ expressing *croS* from its native promoter restored to near wild-type level the induction of *setR* expression. In contrast, overexpression of *croS* abolished the MC-mediated induction of *setR* expression (Figure 3B *setR*). This phenotype results from *croS* overexpression as the control plasmid pCroS-inv, in which the open reading frame of *croS* is in the reverse orientation, did not lead to abolition of MC-dependent induction (Figure S1). To isolate the CroS-SetR locus from the rest of SXT-encoded functions, we also quantified *setR* from pSR to mimic wild-type SXT and from pSZR to mimic SXT Δ*croS*. Although the observed differences were not statistically significant in these simplified genetic contexts, MC-mediated induction of *setR* expression followed similar trends as observed with the complete SXT suggesting that, besides CroS and SetR, no other ICE-encoded factor regulates *setR* expression (Figure 3A and B *setR*). To confirm the repression of *setR* expression by CroS, we also measured the β-galactosidase activity of a *setR'*-'*lacZ* fusion (pSRZ) in the presence or absence of arabinose-induced pCroS (Figure 3C *setR*). Expression of the *setR'*-'*lacZ* fusion was decreased twofold when *croS* was overexpressed. Similar results were obtained regardless of the absence or presence of MC, thereby indicating that *setR* expression is not altered by MC in the absence of

the SetR protein. Together, these results confirm that CroS is a *setR* repressor, likely by acting on the P_R promoter.

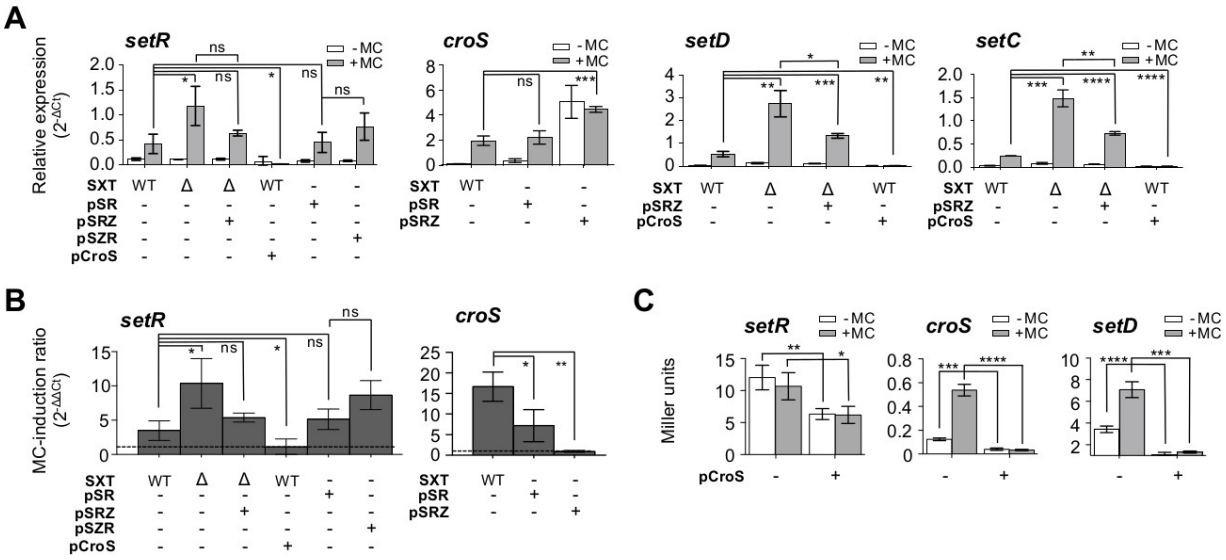

Figure 3. CroS represses expression from P_L and P_R. (A) Effect of the deletion and overexpression of *croS* on the mRNA level of *setR*, *croS*, *setD* and *setC*. *setR*, *croS*, *setD* and *setC* mRNA transcript levels in the presence or absence of MC were measured by quantitative RT-PCR (qRT-PCR). The strains carried SXT (VB17, WT) or its Δ*croS* mutant (DPL353, Δ) and/or the simplified pSR (DPL231) or its pSZR (DPL241) and pSRZ (DPL263, DPL246) derivatives. (B) Effect of the deletion and overexpression of *croS* on the SOS-dependent induction of *setR* and *croS* expression. The strains used are the same as those in panel A for *setR* and *croS*, but the results are shown as induction ratios obtained using ΔΔCt calculations. The level of expression without MC was subtracted from the level with MC. The dashed line indicates an MC-induction ratio of 1, i.e. no difference between the induced and non-induced conditions. (C) Effect of the overexpression of *croS* on the translation of *setR*, *croS* and *setD*. β-galactosidase assays with strains harboring translational *lacZ* fusions to *setR* (DPL246) and *croS* (DPL241) derived from pSR, or to *setD* derived from SXT (DPL297). The bars represent the mean and standard deviation values obtained from at least three independent experiments. Statistical analyses were performed on the means using two-tailed unpaired t-tests. P-values are indicated as follows: ns, P>0.05; *, P≤0.05; **, P≤0.01; ***, P≤0.001; ****, P≤0.0001.

2.1.8.4. CroS represses the promoter P_L

In λ, Cro not only represses the expression of the associated *cI* repressor gene, but it also represses the expression of lytic genes (33). In SXT, integration/excision and conjugative transfer genes are the counterpart of λ lytic genes. Their expression is activated by the class II transcriptional activator complex SetCD (24, 26). Previously, expression of SetCD was shown to be driven from P_L, in the same mRNA transcript that harbors *croS* (30). To assess whether CroS can repress P_L, we measured the relative mRNA level of transcripts containing *setC* and *setD* in cells bearing wild-type SXT or its Δ*croS* derivative mutant, in the presence or absence of pSRZ, pCroS and MC (Figure 3A *setD* and *setC*). In a MC-induced Δ*croS* context, we observed that the level of mRNA containing *setD* and *setC* rose by fivefold when compared to wild-type SXT. This enhanced expression of *setCD* was partially complemented by *croS* provided from its native promoter in pSRZ. Moreover, overexpression of *croS* from pCroS virtually abolished *setD* and *setC* expression (Figure 3A *setD* and *setC*). The repressive effect of overexpression of CroS was also observed at the translational level since the *croS'*-'*lacZ* and *setD'*-'*lacZ* fusions yielded very low β-galactosidase activities even when the cells were exposed to MC (Figure 3C *croS* and *setD*). Expression of *croS* was similar in SXT and pSR contexts, while removal of the *setR* gene in pSRZ boosted expression of *croS*, thereby confirming the repressive function of SetR on P_L (Figure 3A *croS*). Abolition of the MC responsiveness in pSRZ confirms the pivotal role of SetR for integration of the SOS signal (RecA*) to the regulatory network. *croS* expressed from its native promoter is not sufficient to completely repress P_L upon MC induction. These expression assays confirm the importance of SetR for the MC-dependent induction of SXT/R391 ICEs transfer, and that CroS acts as a mild repressor of the P_L promoter, which drives the expression of the master activator complex SetCD.

2.1.8.5. DNA binding motif recognized by SetR

To confirm the motif bound by SetR (30), the ~4.8 kb nucleotide sequence overlapping *s084* and *attR* was submitted to the motif-based sequence analysis tool (MEME) (45). This analysis led to the identification of the same operators previously described for SetR located between *croS* and *setR* (*OL*, *O1*, *O2*, and *O3*), as well as the remote fifth operator *O4* (Figure 4A) (30). A Motif Alignment and Search Tool (MAST (46)) analysis revealed that this motif is not found anywhere else in the SXT sequence.

SetR binding to each independent operator, including *O4*, was confirmed by electrophoretic mobility shift assays (EMSA) experiments with purified SetR and a DNA probe corresponding to the single operator (Figure 4B and C). We observed a differential affinity of SetR for its operator sites in the following order: *O4≈O1>O3>OL>O2.* To gain a better understanding of SetR DNA binding pattern, we mixed purified SetR to permutated radiolabeled probes*,* i.e. probes of identical size partially overlapping each other, containing *OL*, *O1*, *O2*, and *O3* (Figure 4D). SetR binding led to the formation of one predominant complex for all the probes tested, suggesting that it binds cooperatively to the operators. The variations in the migration distance of the protein-DNA complexes suggests that SetR binding to its operator leads to DNA bending as shown for other CI-like repressors (47, 48). Indeed, the position of the bending angle along the length of the DNA probe modifies the overall structure of the migrating protein:DNA complex, and hence its migration distance (49).

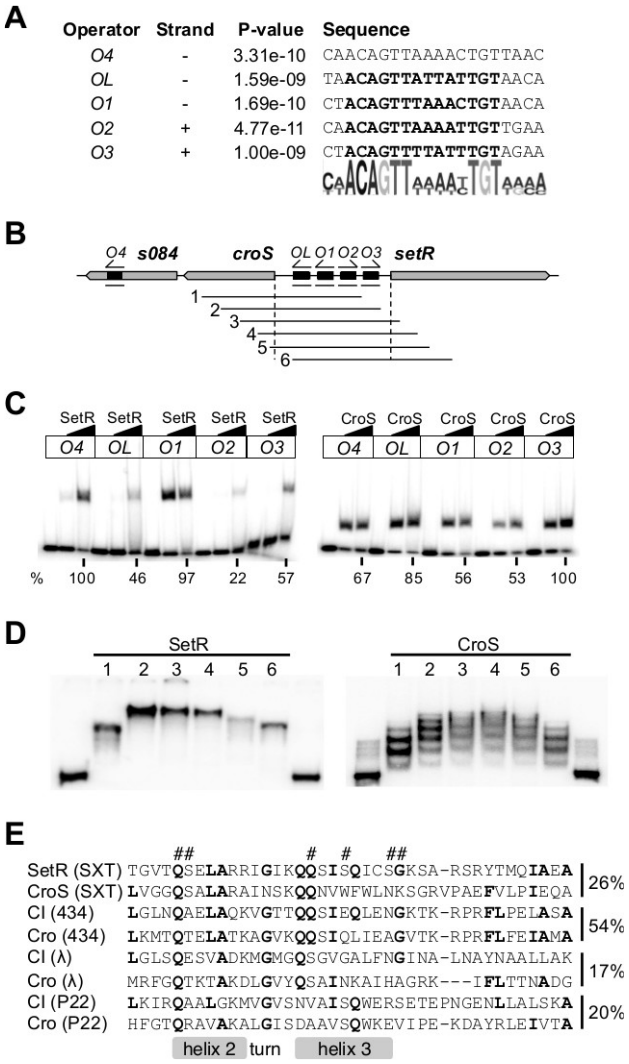

Figure 4. CroS and SetR binding to the switch locus of SXT. (A) Sequences of the five SetR and CroS operators and the associated logo computed by the Multiple Em for Motif Elicitation (MEME). The sites previously described by Beaber and Waldor (30) are indicated in bold letters. (B) Representation of the switch locus and the ^{32}P-labeled probes used for the gel shifts assays in C and D. Orientation of the motifs is indicated on the top of the operators. (C) Binding of 0.6 and 1 µM of purified SetR or CroS to annealed oligonucleotides containing a single operator (length of 31 to 36 bp). The left-most lane for each operator contains the probe alone. The percentage of intensity of each shifted band in relation to the band of maximal intensity (*O4* for SetR and *O3* for CroS using 1 µM of SetR or CroS) are indicated below the autoradiographs. (D) Bending assays with 1 µM of SetR or CroS and the permuted PCR probes 1 to 6. The border lanes contain probe 1 with no added protein. (E) Alignment of the HTH_XRE domains (74) of co-repressor pairs computed by Clustal Omega (72). Amino acids conserved in at least four of the aligned repressors are shown in bold. Hash-marks (#) on top of the alignment display indicate the specific residues of SetR involved in DNA binding that have been annotated on an NCBI-curated HTH_XRE domain (74). The helix-turn-helix components are indicated as described elsewhere (51, 75). Percentages of identity between the represented amino acids of the co-repressors are indicated on the right of the alignment.

2.1.8.6. DNA binding motif recognized by CroS

It has been reported for numerous lambdoid bacteriophages that CI-like and cognate Cro-like repressors share operator sites due to similarity in their HTH DNA binding domains (50-52). The HTH domains of CroS and SetR share 26% identity, which is higher than other well-studied λ-like repressor pairs such as λ and P22 (Figure 4E), suggesting that both protein share the same operator sites. We tested this hypothesis by carrying out EMSA experiments with purified CroS and DNA probes corresponding to individual SetR operator sites and confirmed that CroS independently binds to the exact same five operators (Figure 4C). We observed a differential affinity of CroS for its operator sites in the following order: *O3>OL>O4>O1≈O2.* Like SetR, purified CroS specifically binds P_L and P_R as neither protein was able to shift probes containing the promoter P_{s089} or P_{s003} in the presence of a specific competitor DNA (annealed oligonucleotides containing either *O1* or *O2*, Figure S2). Incubation of purified CroS with permutated radiolabeled probes containing *OL*, *O1*, *O2*, and *O3* (Figure 4D) led to the formation of multiple shifted CroS-DNA complexes, suggesting that, unlike SetR, CroS binds multiple sites in a non-cooperative manner. The large footprints observed following CroS binding to *OL*, *O1* and *O2* operators, confirms the concomitant binding of CroS to these sites (Figure S3). Altogether, our results of SetR and CroS binding assays suggest that both repressors bind the same operators located in the SXT locus that contains the promoters P_L and P_R.

2.1.8.7. SetCD indirectly generates a positive feedback loop by activating SXT replication

Previously published work suggests that SetCD regulates its own expression (24). To test this hypothesis, we quantified the β-galactosidase activity of a *setD*'-'*lacZ* fusion in SXT (DPL297) with or without overexpression of *setCD* from a P_{BAD} promoter (pGG2B) (Figure 5A *setD*). Upon overexpression of *setCD*, the β-galactosidase activity of the *setD'-'lacZ* fusion rose by 75-fold. Although *croS* and *setD* are part of the same mRNA transcript driven by P_L (Figure 5A *croS*), this drastic increase was not observed for the *croS'*-'*lacZ* fusion measured from pSZR (DPL241), which is integrated and locked into the chromosome. We also measured the *setR* and *croS* mRNA transcript levels in cells bearing SXT or its Δ*setCD* mutant, in the absence of MC, with or without pGG2B (Figure 5B). Deletion of *setCD* did not significantly alter the expression levels of *setR* and *croS*. In contrast, overexpression of *setCD* increased both *setR* and *croS* mRNA levels, which is seemingly inconsistent with the results obtained by measuring β-galactosidase activities (Figure 5A and B). These results suggested that *setCD* expression could be driven from an unidentified alternative SetCD-dependent promoter located between P_L and *setD*. However, despite several attempts, we failed to detect any additional promoter between these two loci using 5' RACE and primer extension assays (data not shown). Furthermore, genome-wide 5' RACE and ChIP-exo experiments recently confirmed that no promoter is detectable downstream of P_L and that P_L itself is not a SetCD-dependent promoter (26). Another SXT-encoded factor absent from pSZR could explain this discrepancy.

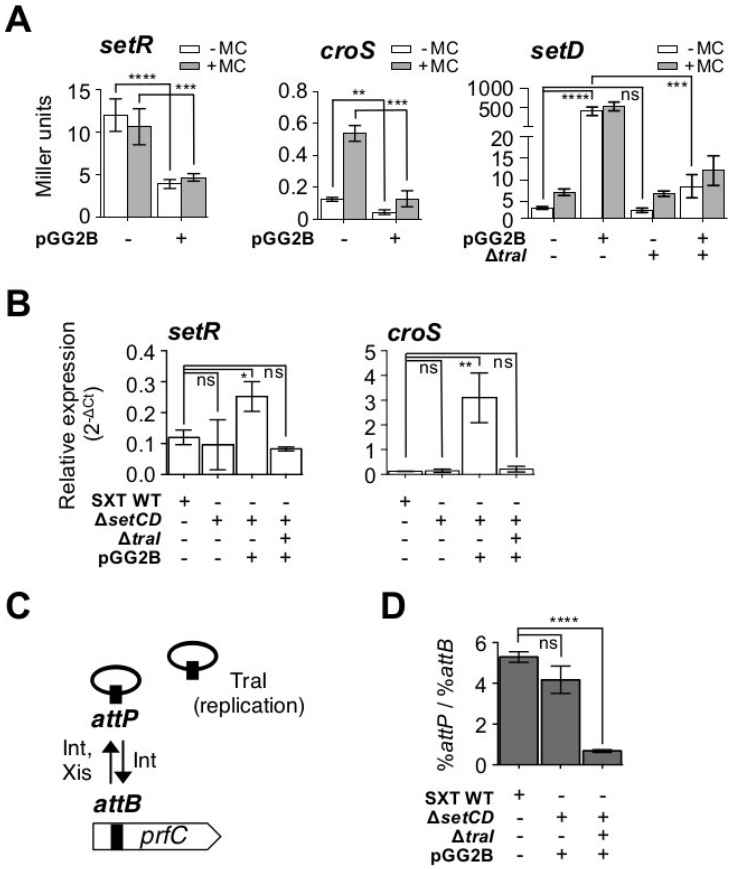

Figure 5. Overexpression of *setCD* leads to an indirect positive feedback loop boosting the expression from P_L and P_R. (A) Effect of SetCD on the expression of *setR*, *croS* and *setD* translationally fused to *lacZ* measured by β-galactosidase assays in the presence or absence of MC. The strains contain pSRZ (DPL246), pSZR (DPL241), SXT *setD*'-'*lacZ* Δ*setC* (DPL297) or SXT Δ*traI setD*'-'*lacZ* Δ*setC* (DPL548). Overexpression assays were carried out by expressing *setCD* from the arabinose-inducible P_{BAD} promoter in pGG2B. (B) Effect of the deletion and overexpression of *setCD* on *setR* and *croS* mRNA levels. qRT-PCR assays carried out on *E. coli* BW25113 carrying SXT (VB17, WT) or its Δ*setCD* or Δ*traI* Δ*setCD* (DPL297 and DPL548) derivatives without MC induction. Overexpression assays were carried out by expressing *setCD* from pGG2B. (C) Representation of integration/excision and replication of SXT. (D) Measurement of SXT copy number by qRT-PCR (%*attP* / %*attB* with *prfC* as a reference) in SXT derivatives (identical to those in A and B) and with or without overexpression of *setCD*. Bars represent the mean and standard deviation values from three independent experiments. Statistical analyses were performed on the means using two-tailed unpaired t-tests, with the exception of the results in A *setD* for which the logarithm value of the means were used. P-values are indicated as follows: ns, P>0.05; *, P≤0.05; **, P≤0.01; ***, P≤0.001; ****, P≤0.0001.

Recent work has established that in inducing conditions, SXT/R391 ICEs replicates by a rolling-circle mechanism initiated by the relaxase TraI. This replication step is important for conjugative transfer as well as the stability of the ICE in the cell and in the bacterial population (27). To test whether activation of *traI* expression by SetCD (26) is responsible for the increased *setCD* expression in a SXT background, we measured the β-galactosidase activity of the *setD*'-'*lacZ* fusion in a strain carrying SXT Δ*traI* (DPL548) and pGG2B. We observed that deletion of *traI* abolished the pGG2B-induced increase of expression of *setR*, *croS* and *setD* (Figure 5A and B). To confirm that *traI* deletion has an effect on the copy number of SXT, we measured the mean copy number of extrachromosomal circular forms of SXT per cell by establishing the ratio between the amount of *attP* recombination sites resulting from SXT excision and the amount of unoccupied chromosomal *attB* sites by real-time qPCR. In theory, each event of SXT excision yields one unoccupied *attB* site on the chromosome and one *attP* site on the circular excised SXT if it does not replicate (*attP*/*attB* = 1). Consistent with previous reports (27, 39), the ratio *attP*/*attB* was 5 for wild type SXT. While this ratio was similar at 4 for SXT Δ*setD*'-'*lacZ* pGG2B, it decreased to 0.7 for SXT Δ*traI setD'-'lacZ* pGG2B, confirming the importance of TraI in SXT replication (Figure 5D). Altogether, these results demonstrate that the apparent involvement of SetCD in the expression of genes of the regulatory region is not due to direct activation of P_L and P_R by SetCD. Instead, SetCD triggers *int*, *xis*, and *traI* expression, thereby driving SXT excision and replication, which increases the copy number of SXT molecules in a larger fraction of the cell population, leading to increased *setR* and *croS-setCD* mRNAs.

2.1.9. Discussion

SXT/R391 ICEs are major vectors of multidrug resistance dissemination across a plethora of bacterial genera encountered in clinical and environmental settings (7-9). Their highly conserved regulatory region governs the transition between the ICE quiescent and conjugative states (24, 29, 30). Here, we report several new levels of genetic regulation controlling SXT/R391 ICE behavior. Using SXT as a prototypical member of the family, we have identified and characterized the function of CroS, an ICE-encoded Cro-like repressor that serves as a key component in this genetic switch together with SetR, the main repressor of SXT/R391 ICEs (24). We also show that SetR and CroS are able to bind the *O4* operator, likely consolidating the repressed state of these ICEs. Finally, expression of SetCD results in a positive feedback mechanism that is dependent on *traI*, which encodes the presumed relaxase of SXT/R391 ICEs. Given the high conservation of the core genes, especially those of the regulatory region of SXT/R391 ICEs (8), our findings are most likely applicable to all members of the family.

2.1.9.1. The SetR-CroS switch

In λ genetic regulation, the nature of the binding patterns of CI and Cro to the O_R1, O_R2 and O_R3 operators leads either to the activation or repression of *cI* expression from P_{RM}, or to the repression or derepression of *cro* from P_R (52). The relative affinities of both repressors for each operator and their respective cell abundance, combined with intrinsic cellular stochasticity, determine the fate of the switch. The closest structural relative of SetR is the CI repressor encoded by lambdoid bacteriophage 434. The N-terminal dimer interface of 434 CI was shown to drive the

central base preferences, which defines the specific affinity of the repressor for each operator independently (53). Even though extensive thermodynamic studies are mandatory to infer quantitative relative affinities, the conditions used in binding reactions allowed us to observe variations in complex formation between the independent operators. Using it as an indicator of affinity, the inferred relative affinities of SetR for the five operator sites are *O4≈O1>O3>OL>O2*, and are consistent with what was observed by DNase I footprinting (30). The relative affinities of CroS for the same operators were *O3>OL>O4>O1≈O2*. This binding pattern is consistent with λ Cro which binds more tightly to O_R3 that O_R1 and O_R2.

In SXT, the center-to-center spacing between *O1-O2* and *O2-O3* is 24 and 23 bp (Figure 1B), the ideal distance for cooperative binding given that two adjacent bound repressor dimers are located on the same side of the DNA helix (54). The *OL-O1* spacing of 30 bp suggests that binding of the repressor to *O1* cannot yield cooperative binding to *OL,* the orientation of the adjacent repressors being inadequate. Given the observed affinity of SetR, and in agreement with λ regulation, the likely preferred binding pattern of SetR is to *O1-O2*. Organization of operator sites and promoters of the regulatory locus of SXT (Figure 1B) suggests that SetR activates its own expression from P_R when cooperatively bound to *O1-O2*, likely by contacting the σ^{70} subunit of the RNA polymerase (RNAP) (30, 55-57). Binding of SetR to *O1-O2* not only triggers its own expression from P_R, but also blocks access of the RNAP to P_L, which drives expression of *setCD* and the subsequent activation of conjugative genes of SXT (26, 29, 30, 58). By analogy with λ *cI* regulation, SetR would repress its own expression by further binding to *O3* in addition to *O1-O2* when its intracellular level increases (30, 55, 56).

In our assays, translation of SetR appeared to be higher than translation of CroS, suggesting that SetR protein levels are higher that CroS levels (Figure 3C). In fact, in non-induced conditions, only 1 cell out of 11 500 on average acts as a donor cell (frequency of exconjugant formation = 8.65×10^{-5} ± 9.88×10^{-6} exconjugant CFUs / donor CFUs, n=3), suggesting that the quantity of *croS* mRNA transcripts per cell is lower than the quantity of SetR mRNA transcript in most cells of the population. The primer efficiency bias in qRT-PCR precludes us from directly comparing the amount of mRNA containing *setR* and *croS* (Figure 3A). CroS is presumably dominant in only 1 out of 11 500 cells transitioning from the quiescent to transfer-proficient state of SXT. We have established that CroS is mandatory for SOS-induction of SXT transfer, which is consistent with λ regulation (33). CroS represses P_L and P_R when SetR repression is alleviated. This double repression likely delays recovery of the SetR pool and displaces the balance towards P_L expression long enough for activation of conjugative transfer to occur properly. In our assays, P_L repression by CroS was only observed upon overexpression of *croS*, not when *croS* was expressed from its native promoter (Figure 3A *croS*). This observation suggests that CroS is a weak repressor of P_L. In wild-type induced SXT, conditional replication mediated by TraI (27) increases *croS* mRNA level (Figure 5B), which could allow partial repression of P_L. CroS would then act at P_L to mitigate expression of *setCD*, which has been shown to be toxic when overexpressed (24, 59).

The closest structural relative of CroS is Xfaso 1, a Cro-like repressor for which the structure has been determined (42). To our knowledge, experimental data supporting the functionality of Xfaso 1 remains limited (60). Secondary structure predictions suggest that Xfaso 1 and CroS have an all-α-helix structure as do the well-studied Cro protein encoded by the lambdoid bacteriophage P22 isolated from *Salmonella* (Figure 2B) (61). The genetic regulation network of P22 is more complex than the one of λ as P22 codes for additional repressors to maintain lysogeny (62). Despite their

structural differences (all-α versus α-β folds), Cro proteins likely play highly similar role towards the P_R and P_{RM} promoters in P22 and λ (63, 64). Indeed, opposite affinities of Cro-like ($O_R3 > O_R1 > O_R2$) and CI-like ($O_R1 > O_R2 > O_R3$) repressors for each operator were observed for both phages. Consistent with our results, the affinity of the Cro-like repressors for each operators exhibits less variation when compared to the affinity of CI-like repressors (63). Study of many natural and synthetic versions of the λ genetic network led to the conclusion that while most components are not essential, altogether they contribute to the overall efficiency and stability of the system (65). We showed that the role of CroS is particularly important in SXT regulation as SXT lacks two components that are important for noise reduction and stabilization of the λ system, the activator CII and two additional operators in the *OL* region.

2.1.9.2. Pleiotropic effect of SetCD overexpression

In λ and in SXT the *cro* genes are co-transcribed with genes coding for a transcriptional activator, *cII* in λ and *setCD* in SXT. These activators are not related. Instead, SetCD is related to FlhCD and AcaCD, the master activators of flagella synthesis in Gram-negative bacteria (66) and of conjugative transfer functions of IncA/C plasmids (67), respectively. One role of CII is the activation of *cI* by acting at the promoter P_{RE} (52). The intracellular level of CII ultimately sets the decision between lysis and lysogeny. Similarly, our results show that mRNA levels of *setR* rise upon *setCD* overexpression (Figure 5B *setR*). However, our results revealed that it is not due to direct activation of P_R by SetCD. Instead expression of *setCD* enhances markedly the mRNA level of *int*, *xis* and *traI*, presumably leading to a higher portion of the cell population in which the ICE undergoes excision and conjugation-associated replication (26, 27). The elevated mRNA levels of *setR* and *croS* observed in SetCD-

induced bacterial cultures (Figure 5B) also support this scenario. Paradoxically, overexpression of *setCD* lowered the β-galactosidase activity of the *setR-lacZ* and *croS-lacZ* fusions constructed in the minimal pSR (without the SXT backbone) (Figure 5A). This collateral repressive effect of SetCD overexpression could be caused by illegitimate SetCD DNA binding. In addition to being overexpressed, none of the SetCD target promoters were present in such cells to titrate SetCD excess. Alternatively, overexpression of SetCD might reduce *lacZ* expression levels through non-specific effects on cellular growth rates.

The level of translation was higher for *setD* than for *croS* (Figure 3C). This difference could be attributable to spacing between their respective Shine-Dalgarno sequences and start codons, which is closer to the optimum for *setD* than for *croS* (68).

2.1.9.3. A revised model of SXT early regulation

Since the preferred substrate for RecA*-activated self-cleavage of 434 CI is the DNA-bound dimer (69), one could argue that upon DNA damage the cooperatively bound SetR-DNA complex becomes substrate of RecA*, unleashing SetR's latent autoproteolysis activity (29). Release of the operators alleviates the repression exerted by SetR on P_L allowing *croS* and *setCD* expression (Figure 6B). CroS, the first translated protein, would bind to *O3* to prevent premature resurgence of the SetR pool (Figure 6C). Then, CroS would bind in a non-cooperative fashion to *OL*, *O1*, *O2* and *O4*, thereby preventing adverse overexpression of *setCD* (Figure 6D). This repression by CroS would allow exclusion of SetR's strong binding to the intergenic region between *croS* and *setR* long enough to observe the SOS-mediated induction of transfer. At some point, CroS would vacate *OL*, *O1*, *O2* and *O3*, allowing SetR to

take over the intergenic region and reset the quiescent state (Figure 6E). Our results show that both CroS and SetR are able to bind to *O4*. Their binding could generate a roadblock that compromises transcription of the whole *croS-eex* operon driven from P_L beyond *O4*. Alternatively, because SetR is able to bind cooperatively, its binding could lead to the formation of a long-range loop between a SetR-bound *O4* and the intergenic region between the *croS* and *setR* genes. Such a role is unlikely for CroS, which lacks the C-terminal domain mediating self-assembly, cooperative DNA binding and concomitant loop formation ability (70, 71). This long-range loop or roadblock at *O4* mediated by SetR could stabilize the SetR-dominant state. Additional experiments are required to decipher the regulatory mechanism(s) involving the *O4* operator site.

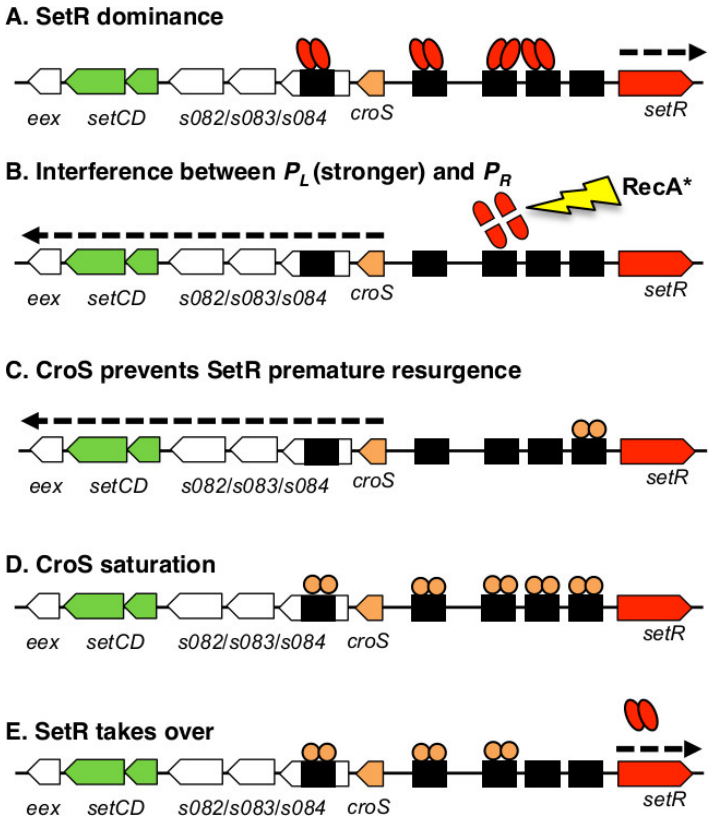

Figure 6. Representation of proposed key intermediate states (A-E) of *O4*, *OL*, *O1*, *O2* and *O3* binding by SetR or CroS. A is a presumed intermediate of the dominant quiescent behavior, while B to E are most likely transient. The dotted arrow indicates active transcription.

The genetic network of SXT/R391 ICEs elegantly sets at play regulatory components reminiscent of lambdoid phages (SetR-CroS epigenetic switch) and bacterial motility (SetCD activation) to rule the propagation of DNA material. The diversity of bacterial hosts in which SXT/R391 ICEs dwell suggests that the regulation system they encode was selected for minimal interactions with host genetic networks apart from their induction stimulated by RecA*. SXT/R391 ICEs take advantage of a host regulatory stress response that can be caused by DNA targeting antibiotics and UV light, to trigger the dissemination of multidrug resistance and other adaptive functions through *Gammaproteobacteria* populations.

2.1.10. Acknowledgements

This work was supported by a Discovery Grant and Discovery Acceleration Supplement from the Natural Sciences and Engineering Council of Canada [326810 and 412288 to V.B.]. V.B. holds a Canada Research Chair in molecular bacterial genetics. D.P.L was supported by scholarships from the Fonds de recherche du Québec. We are grateful to C. Jouogo Noumsi, V. Côté-Breton, M. Paquette-D'Avignon, D. Ceccarelli and the Centre de calcul scientifique of the Université de Sherbrooke for their technical assistance and to N. Carraro, A. Lavigueur and J.P. Brousseau for their insightful comments on the manuscript.

2.1.11. References

1. **Ramamurthy T, Garg S, Sharma R, Bhattacharya SK, Nair GB, Shimada T, Takeda T, Karasawa T, Kurazano H, Pal A, *et al*.** 1993. Emergence of novel strain of *Vibrio cholerae* with epidemic potential in southern and eastern India. Lancet **341:**703-704.

2. **Kaper JB, Morris JG, Jr., Levine MM.** 1995. Cholera. Clin. Microbiol. Rev. **8:**48-86.

3. **Waldor MK, Tschape H, Mekalanos JJ.** 1996. A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. J. Bacteriol. **178:**4157-4165.

4. **Wozniak RA, Waldor MK.** 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat. Rev. Microbiol. **8:**552-563.

5. **Burrus V, Pavlovic G, Decaris B, Guedon G.** 2002. Conjugative transposons: the tip of the iceberg. Mol. Microbiol. **46:**601-610.

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

7. **Spagnoletti M, Ceccarelli D, Rieux A, Fondi M, Taviani E, Fani R, Colombo MM, Colwell RR, Balloux F.** 2014. Acquisition and evolution of SXT-R391 integrative conjugative elements in the seventh-pandemic *Vibrio cholerae* lineage. MBio **5**.

8. **Wozniak RA, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, Garriss G, Dery C, Burrus V, Waldor MK.** 2009. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. PLoS Genet. **5:**e1000786.

9. **Rodriguez-Blanco A, Lemos ML, Osorio CR.** 2012. Integrating conjugative elements as vectors of antibiotic, mercury, and quaternary ammonium compound resistance in marine aquaculture environments. Antimicrob. Agents Chemother. **56:**2619-2626.

10. **Osorio CR, Marrero J, Wozniak RA, Lemos ML, Burrus V, Waldor MK.** 2008. Genomic and functional analysis of ICE*Pda*Spa1, a fish-pathogen-derived SXT-related integrating conjugative element that can mobilize a virulence plasmid. J. Bacteriol. **190:**3353-3361.

11. **Harada S, Ishii Y, Saga T, Tateda K, Yamaguchi K.** 2010. Chromosomally encoded *blaCMY*-2 located on a novel SXT/R391-related integrating conjugative element in a *Proteus mirabilis* clinical isolate. Antimicrob. Agents Chemother. **54:**3545-3550.

12. **Lopez-Perez M, Gonzaga A, Rodriguez-Valera F.** 2013. Genomic diversity of "deep ecotype" *Alteromonas macleodii* isolates: evidence for Pan-Mediterranean clonal frames. Genome Biol Evol **5:**1220-1232.

13. **Badhai J, Kumari P, Krishnan P, Ramamurthy T, Das SK.** 2013. Presence of SXT integrating conjugative element in marine bacteria isolated from the mucus of the coral *Fungia echinata* from Andaman Sea. FEMS Microbiol. Lett. **338:**118-123.

14. **Pembroke JT, Piterina AV.** 2006. A novel ICE in the genome of *Shewanella putrefaciens* W3-18-1: comparison with the SXT/R391 ICE-like elements. FEMS Microbiol. Lett. **264:**80-88.

15. **Coetzee JN, Datta N, Hedges RW.** 1972. R factors from *Proteus rettgeri*. J Gen Microbiol **72:**543-552.

16. **Beaber JW, Burrus V, Hochhut B, Waldor MK.** 2002. Comparison of SXT and R391, two conjugative integrating elements: definition of a genetic backbone for the mobilization of resistance determinants. Cell. Mol. Life Sci. **59:**2065-2070.

17. **Peters SE, Hobman JL, Strike P, Ritchie DA.** 1991. Novel mercury resistance determinants carried by IncJ plasmids pMERPH and R391. Mol. Gen. Genet. **228:**294-299.

18. **Balado M, Lemos ML, Osorio CR.** 2013. Integrating conjugative elements of the SXT/R391 family from fish-isolated Vibrios encode restriction-modification systems that confer resistance to bacteriophages. FEMS Microbiol. Ecol. **83:**457-467.

19. **Bordeleau E, Brouillette E, Robichaud N, Burrus V.** 2010. Beyond antibiotic resistance: integrating conjugative elements of the SXT/R391 family that encode

novel diguanylate cyclases participate to c-di-GMP signalling in *Vibrio cholerae*. Environ. Microbiol. **12:**510-523.

20. **Garriss G, Waldor MK, Burrus V.** 2009. Mobile antibiotic resistance encoding elements promote their own diversity. PLoS Genet. **5:**e1000775.

21. **Kulaeva OI, Wootton JC, Levine AS, Woodgate R.** 1995. Characterization of the *umu*-complementing operon from R391. J. Bacteriol. **177:**2737-2743.

22. **Hochhut B, Marrero J, Waldor MK.** 2000. Mobilization of plasmids and chromosomal DNA mediated by the SXT element, a constin found in *Vibrio cholerae* O139. J. Bacteriol. **182:**2043-2047.

23. **Daccord A, Ceccarelli D, Burrus V.** 2010. 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. **78:**576-588.

24. **Beaber JW, Hochhut B, Waldor MK.** 2002. Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. J. Bacteriol. **184:**4259-4269.

25. **Marrero J, Waldor MK.** 2005. Interactions between inner membrane proteins in donor and recipient cells limit conjugal DNA transfer. Dev. Cell **8:**963-970.

26. **Poulin-Laprade D, Matteau D, Jacques PE, Rodrigue S, Burrus V.** 2015. Transfer activation of SXT/R391 integrative and conjugative elements: unraveling the SetCD regulon. Nucleic Acids Res. **43:**2045-2056.

27. **Carraro N, Poulin D, Burrus V.** 2015. Replication and active partition of integrative and conjugative elements (ICEs) of the SXT/R391 family: The line between ICEs and conjugative plasmids is getting thinner. PLoS Genet. **11:**e1005298.

28. **Garriss G, Poulin-Laprade D, Burrus V.** 2013. DNA-damaging agents induce the RecA-independent homologous recombination functions of integrating conjugative elements of the SXT/R391 family. J. Bacteriol. **195:**1991-2003.

29. **Beaber JW, Hochhut B, Waldor MK.** 2004. SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature **427:**72-74.

30. **Beaber JW, Waldor MK.** 2004. Identification of operators and promoters that control SXT conjugative transfer. J. Bacteriol. **186:**5945-5949.

31. **Svenningsen SL, Costantino N, Court DL, Adhya S.** 2005. On the role of Cro in λ prophage induction. Proc Natl Acad Sci U S A **102:**4465-4469.

32. **Johnson A, Meyer BJ, Ptashne M.** 1978. Mechanism of action of the Cro protein of bacteriophage λ. Proc Natl Acad Sci U S A **75:**1783-1787.

33. **Schubert RA, Dodd IB, Egan JB, Shearwin KE.** 2007. Cro's role in the CI Cro bistable switch is critical for λ's transition from lysogeny to lytic development. Genes Dev. **21:**2461-2472.

34. **Burrus V, Waldor MK.** 2004. Formation of SXT tandem arrays and SXT-R391 hybrids. J. Bacteriol. **186:**2636-2645.

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

36. **Demarre G, Guerout AM, Matsumoto-Mashimo C, Rowe-Magnus DA, Marliere P, Mazel D.** 2005. 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. **156:**245-255.

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

38. **Daccord A, Mursell M, Poulin-Laprade D, Burrus V.** 2012. Dynamics of the SetCD-regulated integration and excision of genomic islands mobilized by integrating conjugative elements of the SXT/R391 family. J. Bacteriol. **194:**5794-5802.

39. **Burrus V, Waldor MK.** 2003. Control of SXT integration and excision. J. Bacteriol. **185:**5045-5054.

40. **Miller JF.** 1992. A short course in bacterial genetics. Cold Spring Harbor Laboratory Press, Plainview, NY.

41. **Pradervand N, Delavat F, Sulser S, Miyazaki R, van der Meer JR.** 2014. The TetR-type MfsR protein of the integrative and conjugative element (ICE) ICE*clc* controls both a putative efflux system and initiation of ICE transfer. J. Bacteriol. **196:**3971-3979.

42. **Roessler CG, Hall BM, Anderson WJ, Ingram WM, Roberts SA, Montfort WR, Cordes MH.** 2008. Transitive homology-guided structural studies lead to discovery of Cro proteins with 40% sequence identity but different folds. Proc Natl Acad Sci U S A **105:**2343-2348.

43. **Hall BM, Lefevre KR, Cordes MH.** 2005. Sequence correlations between Cro recognition helices and cognate O(R) consensus half-sites suggest conserved rules of protein-DNA recognition. J. Mol. Biol. **350:**667-681.

44. **Kelley LA, Sternberg MJ.** 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc **4:**363-371.

45. **Bailey TL, Elkan C.** 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol **2:**28-36.

46. **Bailey TL, Gribskov M.** 1998. Combining evidence using p-values: application to sequence homology searches. Bioinformatics **14:**48-54.

47. **Griffith J, Hochschild A, Ptashne M.** 1986. DNA loops induced by cooperative binding of λ repressor. Nature **322:**750-752.

48. **Koudelka GB.** 1991. Bending of synthetic bacteriophage 434 operators by bacteriophage 434 proteins. Nucleic Acids Res. **19:**4115-4119.

49. **Kim J, Zwieb C, Wu C, Adhya S.** 1989. Bending of DNA by gene-regulatory proteins: Construction and use of a DNA bending vector. Gene **85:**15-23.

50. **Koudelka GB, Lam CY.** 1993. Differential recognition of OR1 and OR3 by bacteriophage 434 repressor and Cro. J. Biol. Chem. **268:**23812-23817.

51. **Sauer RT, Yocum RR, Doolittle RF, Lewis M, Pabo CO.** 1982. Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. Nature **298:**447-451.

52. **Ptashne M.** 2004. A genetic switch: phage λ revisited, 3rd edition ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

53. **Koudelka GB.** 1998. Recognition of DNA structure by 434 repressor. Nucleic Acids Res. **26:**669-675.

54. **Mao C, Carlson NG, Little JW.** 1994. Cooperative DNA-protein interactions. Effects of changing the spacing between adjacent binding sites. J. Mol. Biol. **235:**532-544.

55. **Li M, Moyle H, Susskind MM.** 1994. Target of the transcriptional activation function of phage λ CI protein. Science **263:**75-77.

56. **Xu J, Koudelka GB.** 2000. DNA sequence requirements for the activation of 434 P_{RM} transcription by 434 repressor. DNA Cell Biol. **19:**621-630.

57. **Kuldell N, Hochschild A.** 1994. Amino acid substitutions in the -35 recognition motif of σ^{70} that result in defects in phage λ repressor-stimulated transcription. J. Bacteriol. **176:**2991-2998.

58. **Meyer BJ, Maurer R, Ptashne M.** 1980. Gene regulation at the right operator (OR) of bacteriophage λ. II. OR1, OR2, and OR3: their roles in mediating the effects of repressor and Cro. J. Mol. Biol. **139:**163-194.

59. **Armshaw P, Pembroke JT.** 2013. Generation and analysis of an ICE R391 deletion library identifies genes involved in the element encoded UV-inducible cell-sensitising function. FEMS Microbiol. Lett. **342:**45-53.

60. **de Mello Varani A, Souza RC, Nakaya HI, de Lima WC, Paula de Almeida LG, Kitajima EW, Chen J, Civerolo E, Vasconcelos AT, Van Sluys MA.** 2008. Origins of the *Xylella fastidiosa* prophage-like regions and their impact in genome differentiation. PLoS One **3:**e4059.

61. **Newlove T, Konieczka JH, Cordes MH.** 2004. Secondary structure switching in Cro protein evolution. Structure **12:**569-581.

62. **Susskind MM, Botstein D.** 1978. Molecular genetics of bacteriophage P22. Microbiol Rev **42:**385-413.

63. **Poteete AR, Hehir K, Sauer RT.** 1986. Bacteriophage P22 Cro protein: sequence, purification, and properties. Biochemistry **25:**251-256.

64. **Poteete AR, Ptashne M.** 1982. Control of transcription by the bacteriophage P22 repressor. J. Mol. Biol. **157:**21-48.

65. **Little JW.** 2010. Evolution of complex gene regulatory circuits by addition of refinements. Curr. Biol. **20:**R724-734.

66. **Soutourina OA, Bertin PN.** 2003. Regulation cascade of flagellar expression in Gram-negative bacteria. FEMS Microbiol. Rev. **27:**505-523.

67. **Carraro N, Matteau D, Luo P, Rodrigue S, Burrus V.** 2014. The master activator of IncA/C conjugative plasmids stimulates genomic islands and multidrug resistance dissemination. PLoS Genet. **10:**e1004714.

68. **Chen H, Bjerknes M, Kumar R, Jay E.** 1994. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. Nucleic Acids Res. **22:**4953-4957.

69. **Pawlowski DR, Koudelka GB.** 2004. The preferred substrate for RecA-mediated cleavage of bacteriophage 434 repressor is the DNA-bound dimer. J. Bacteriol. **186:**1-7.

70. **Bell CE, Frescura P, Hochschild A, Lewis M.** 2000. Crystal structure of the λ repressor C-terminal domain provides a model for cooperative operator binding. Cell **101:**801-811.

71. **Bell CE, Lewis M.** 2001. Crystal structure of the λ repressor C-terminal domain octamer. J. Mol. Biol. **314:**1127-1136.

72. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG.** 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. **7:**539.

73. **Dodd IB, Egan JB.** 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18:**5019-5026.

74. **Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH.** 2013. CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res. **41:**D348-352.

75. **Harrison SC, Aggarwal AK.** 1990. DNA recognition by proteins with the helix-turn-helix motif. Annu. Rev. Biochem. **59:**933-969.

76. **Singer M, Baker TA, Schnitzler G, Deischel SM, Goel M, Dove W, Jaacks KJ, Grossman AD, Erickson JW, Gross CA.** 1989. A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*. Microbiol Rev **53:**1-24.

77. **Grenier F, Matteau D, Baby V, Rodrigue S.** 2014. Complete Genome Sequence of *Escherichia coli* BW25113. Genome Announc **2**.

78. **Ceccarelli D, Daccord A, Rene M, Burrus V.** 2008. 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. **190:**5328-5338.

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

2.1.12. Supplementary data

Title: A λ Cro-like repressor is essential for the induction of conjugative transfer of SXT/R391 elements in response to DNA damage

Authors: Dominic Poulin-Laprade and Vincent Burrus

Table S1. Primers used in this study

Primer	Nucleotide sequence (5' to 3')	Purpose in this study
delS086F	ACTGTAACATTCAATCTTGTAACAGTTATTATTGTA ACAGGAGACGGTTGTGTAGGCTGGAGCTGCTTC	Deletion of *croS*
delS086R	TGTGTTTACGTAAAATGCTGACCATTATTTGCCCTT AATCCTAGATCCGTCGACCTGCAGTTC	Deletion of *croS*
TraIWF2	TTGAACCTTCACACTATCGTGTCGGAGGTTCACTAT GTGTGTAGGCTGGAGCTGCTTCG	Deletion of *traI*
TraIWR3	GCGTGAGCGGTAAACAGAACCGTTTCCCCCTCATCA AACTATTCCGGGGATCCGTCGACC	Deletion of *traI*
RTrpoZcoli-F	GCTCGTCAGATGCAGGTAGG	Amplification of *rpoZ*
RTrpoZcoli-R	GCTTGTAATTCAGCGGCTTC	Amplification of *rpoZ*
RTsetR-F	GACAGTATGGAAAATCCAAG	Amplification of *setR*
RTsetR-R	AGGTACTTCTGTTCACCATC	Amplification of *setR*
RTsetQ-F	GGGCAATCAGCATTAGCAC	Amplification of *croS*
RTsetQ-R	TTGGGCATTCGGGATAGATG	Amplification of *croS*
RTs084-F	CAAAGAAGAATGGCTTCGCA	Amplification of *s084*
RTs084-R	ACGTACTCGTCTTCCCATTG	Amplification of *s084*
RTs082-F	CACAAAGTGTTGGTATTGAT	Amplification of *s082*
RTs082-R	GCTCTGCTTCTGTCTTAGAT	Amplification of *s082*
RTsetD-F	GTCTTGCGCTTTGGACTTAC	Amplification of *setD*
RTsetD-R	GCGATTCTTGGGATACCTGT	Amplification of *setD*
RTsetC-F	ATGCTTTGATGAAGGCTTACG	Amplification of *setC*
RTsetC-R	ATACAGACTACCGCATCGAC	Amplification of *setC*
RTlacZ-F	GTGACGTCTCGTTGCTGCAT	Amplification of *lacZ*
RTlacZ-R	CACCCTGCCATAAAGAAACTG	Amplification of *lacZ*
VISLR3	GCATTCTCCTGAAAATCAATG	Assessment of SR tandem arrangements
EattBR	AGCAGCACCTTCTCGGTGAT	Assessment of SR tandem arrangements
EattBF	GCCGCACTTTTGCCATTATT	Assessment of SR tandem arrangements
VISRF	CTCTCATTAACTGGGTTCAGG	Assessment of SR tandem arrangements
DATsetR-lacZf	TTTATTTGTAGAATACGGGCTTATGAAAACTTTATC CGAAGCCGTCGTTTTACAACGTCGT	Construction of *setR::lacZ* fusion
DATsetR-lacZr	CAGCTCATGAACAACTGTGGCTAAACCACAAATTCC TTCGGTGTAGGCTGGAGCTGCTTCG	Construction of *setR::lacZ* fusion
DATsetQ-lacZf	TATTATTGTAACAGGAGACGGTATGGAAGTGTTCAA CACAGCCGTCGTTTTACAACGTCGT	Construction of *croS::lacZ* fusion
DATsetQ-lacZr	TGTGTTTACGTAAAATGCTGACCATTATTTGCCCTT AATCGTGTAGGCTGGAGCTGCTTCG	Construction of *croS::lacZ* fusion
DATsetD-lacZf	TACTACATGAACTGGAGGTCCGTCAATGACAACCAA TCAACAAGCCGTCGTTTTACAACGTCGT	Construction of *setCD::lacZ* fusion
DATsetC-lacZr	AGTAGTGCACGGGCGGTGCACAATCAAATCATGTAT CAGCATGGTGTAGGCTGGAGCTGCTTCG	Construction of *setCD::lacZ* fusion
RRso86setRf	TGTTGTGTTGAACACTTCC	Amplification of EMSA probe for *croS-setR*
RRso86setRr	AGTAAGCTGCAAGGCAT	Amplification of EMSA probe for *croS-setR*
RRs086-4	GTTACAAGATTGAATGTTACAG	For footprinting and bending probe 6
5ftOp-F	TCATGGTCTTCATTGAGGCCCAT	Amplification of EMSA probe for 5^{ft} operator
5ftOp-R	TCAAGCTCGGCTTCCAGAAGAAT	Amplification of EMSA probe for 5^{ft} operator
RRs063s089F	ATCCGAAGATATATTCCATTGGT	Amplification of EMSA probe for *s063-s089*
RRs063s089R	TGAATGATTGTTCAAAGATGGGT	Amplification of EMSA probe for *s063-s089*
RRintF	ACATACTTATCTCCTGTGCACTCA	Amplification of EMSA probe for *s003-mobI*
RRintR	ACGACGTTTGGCGTCTCGAT	Amplification of EMSA probe for *s003-mobI*
RRsetQ-O1f	GATTGAATGTTACAGTTTAAACTGTAGATTAGTCA	Annealed oligonucleotides for EMSA
RRsetQ-O1r	TGACTAATCTACAGTTTAAACTGTAACATTCAATC	Annealed oligonucleotides for EMSA
RRsetQ-O2f	ATTAGTCAACAGTTAAAATTGTTGAAAGGCT	Annealed oligonucleotides for EMSA
RRsetQ-O2r	AGCCTTTCAACAATTTTAACTGTTGACTAAT	Annealed oligonucleotides for EMSA

Table S1. Primers used in this study (continued)

Primer	Nucleotide sequence (5' to 3')	Purpose in this study
RRsetQ-O3f	TGTTGAAAGGCTACAGTTTTATTTGTAGAATACG	Annealed oligonucleotides for EMSA
RRsetQ-O3r	CGTATTCTACAAATAAAACTGTAGCCTTTCAACA	Annealed oligonucleotides for EMSA
RRsetQ-OLf	ACCGTCTCCTGTTACAATAATAACTGTTACAAGAT	Annealed oligonucleotides for EMSA
RRsetQ-OLr	ATCTTGTAACAGTTATTATTGTAACAGGAGACGGT	Annealed oligonucleotides for EMSA
RRsetQ-O4f	TGTATCTGGTTAACAGTTTTAACTGTTGGCTTATT	Annealed oligonucleotides for EMSA
RRsetQ-O4r	AATAAGCCAACAGTTAAAACTGTTAACCAGATACA	Annealed oligonucleotides for EMSA
RRsetQ-6	AGCATTCACGCGAAGCG	For bending probe 6
RRsetQ-7	CATGGTGTAACGAGACCTA	For bending probe 5
RRsetQ-8	TCCATACCGTCTCCTGTTA	For bending probe 5
RRsetQ-9	AGAGCAAATCTGGCTGATC	For bending probe 4
RRsetQ-10	CGGAGATGTTTTTGTGTTGT	For bending probe 4
RRsetQ-15	TAGTCGTTCGGATAAAGTT	For bending probe 3
RRsetQ-16	GATTTATTCAACCAAAACCA	For bending probe 3
RRsetQ-17	GTATTCTACAAATAAAACTGTAGC	For bending probe 2
RRsetQ-18	GGTAAAACGAATTCAGCGG	For bending probe 2
RRsetQ-19	ACAATTTTAACTGTTGACTAATCTACAG	For bending probe 1
RRsetQ-20	GTTACCTGTCCTTGCGTAGCT	For bending probe 1
prfC.qec.F1	AGTCAACGTTGCCACTGCCC	Quantification of *prfC*
prfC.qec.R1	AAGCGCCAGTTGGCTTTCGT	Quantification of *prfC*
attB.qec.F2	GCGCGATGCCGCTTACTCAA	Quantification of *attB*
attB.qec.R2	GCGGTCTGAATGGCCTGTCC	Quantification of *attB*
attP.qec.F2	AGATCAGCGAAAATAGCGGCCA	Quantification of *attP*
attP.qec.R2	GCGTTGAAAGGCTGGGCGG	Quantification of *attP*
FP-frag2-HindIII	AGCTTGGTACCGAGCTCG	For footprint's sequencing ladder
FP-frag2-XhoI	TCGAGCGGCCGCCAGTGT	For footprint's sequencing ladder
S086pBADf	TGAGAGGAATAATAAATGGAAGTGTTCAAC	Cloning of *croS* in pBADTOPO
S086pBADr.s	CTACTGGTTACTGGCTTTCAGCTCACTT	Cloning of *croS* in pBADTOPO
SetRpBADf	TGAGAGGAATAATAAATGAAAACTTTATC	Cloning of *setR* in pBADTOPO
SetRpBADr.s	CTACCAGAAATCGATGATAGCTTGTC	Cloning of *setR* in pBADTOPO
SetRpGEXf	NNNNNGGATCCAAAACTTTATCCGAACGACT	Production of the protein SetR
SetRpGEXr	NNNNNCTCGAGCTACCAGAAATCGATGATAGCT	Production of the protein SetR
S086pGEXf	NNNNNGGATCCGAAGTGTTCAACACA	Production of the protein CroS
S086pGEXr	NNNNNCTCGAGCTACTGGTTACTGGCT	Production of the protein CroS

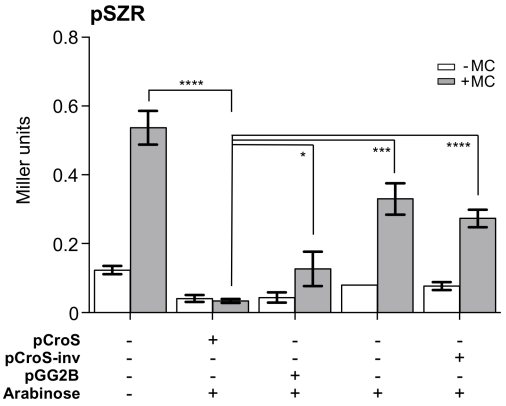

Figure S1. The effect of arabinose and a P_{BAD} vector on the translation of *croS*. β-galactosidase assays with strains harboring the translational *croS*'-'*lacZ* fusion (pSZR, DPL241) with or without pCros overexpressing *croS*, pCroS-inv in which *croS* is in the reverse orientation and pGG2B overexpressing *setCD.* The assays were carried out in presence or absence of mitomycin C (MC) and/or arabinose. The bars represent the mean and standard deviation values obtained from at least three independent experiments. Statistical analyses were performed on the means using two-tailed unpaired t-tests. P-values are indicated as follows: P>0.05; *, P≤0.01; ***, P≤0.001; ****, P≤0.0001.

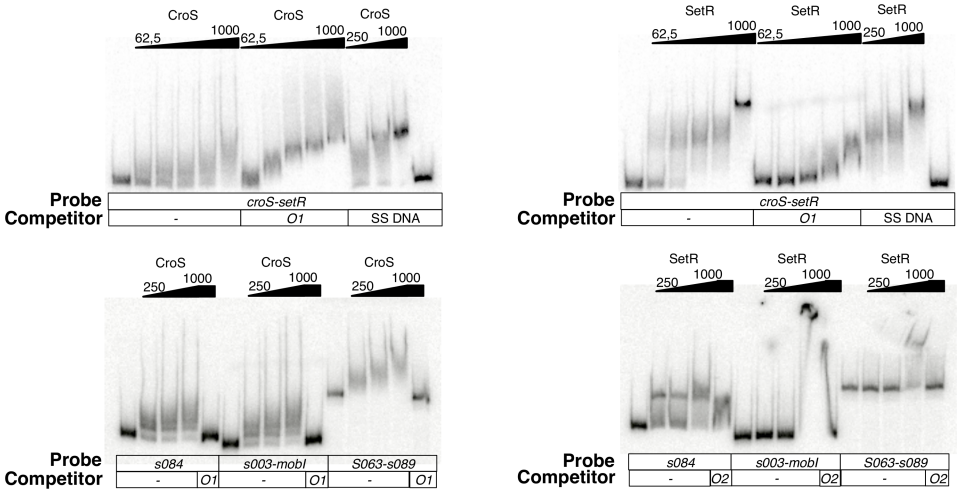

Figure S2. Control samples for the specificity of binding of purified SetR and CroS to four intergenic regions of SXT. In each gel, the left-most and right-most lanes are the probe alone. The added protein concentration is indicated in nM on top each autoradiograph. As indicated, competitor DNA (50 ng ml^{-1} of sonicated salmon sperm DNA (SS DNA) or 5 pmol of 30-bp annealed primers harboring the indicated operator (*O1* or *O2*)) was incubated 20 min with the protein prior the addition of the labeled probe. The *s003-mobI* and *s063-s089* regions are known to be bound and activated by SetCD (1).

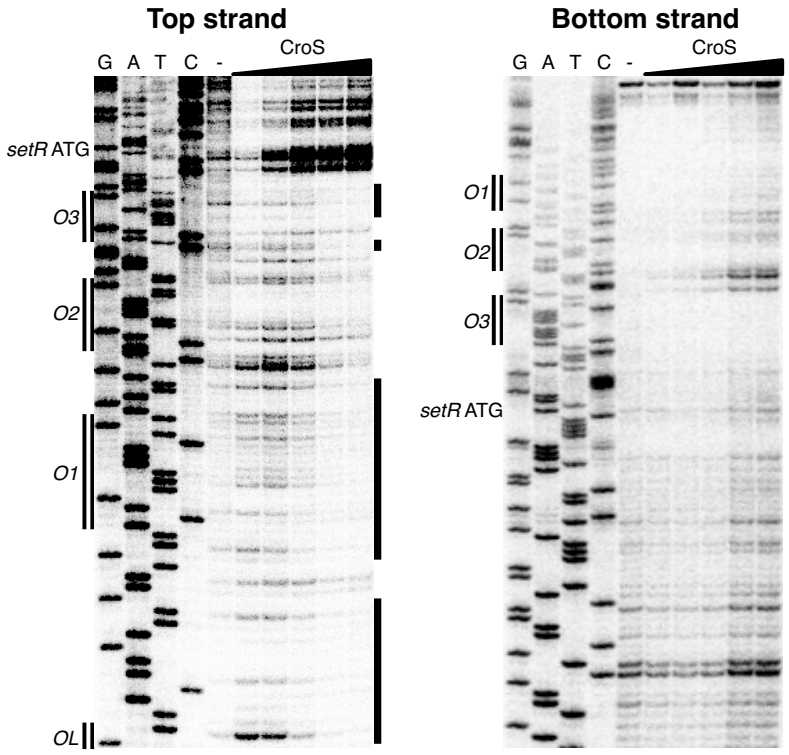

Figure S3. DNase I footprint on both strands of the switch locus of SXT. Regions protected by 0.5 µM to 8 µM of CroS (twofold increments) are indicated by vertical lines on the right side of the autoradiographs.

2.1.12.1. References

1. **Poulin-Laprade D, Matteau D, Jacques PE, Rodrigue S, Burrus V.** 2015. Transfer activation of SXT/R391 integrative and conjugative elements: unraveling the SetCD regulon. Nucleic Acids Res. **43:**2045-2056.

2.2. Activation du transfert des ICE SXT/R391 : le régulon SetCD

2.2.1. Présentation de l'article

Comme nous l'avons décrit en introduction, des études ont montré que SetCD est important chez les ICE SXT/R391 pour l'expression de l'intégrase (Beaber *et al.*, 2002; Burrus and Waldor, 2003), de deux groupes de gènes impliqués dans la synthèse du pilus et du pore de conjugaison (Beaber *et al.*, 2002) ainsi que pour l'expression d'un système de recombinaison homologue générant des ICE hybrides (Annexe 2) (Garriss *et al.*, 2013a). Les travaux de Daccord *et al.* (2012) (Annexe 3) ont également montré que l'intégrase et l'excisionase des MGI mobilisés par les ICE SXT/R391 sont activées par SetCD. Malgré ces évidences, les promoteurs ciblés par SetCD, son motif de reconnaissance de l'ADN ainsi que son mécanisme d'action étaient toujours inconnus.

L'article qui suit répond à ces lacunes en dressant une liste exhaustive des promoteurs ciblés par SetCD chez trois représentants de la famille d'ICE SXT/R391 ainsi que chez un MGI mobilisé par cette famille (Objectif 3). Nos expérimentations ont confirmé que les gènes impliqués dans la recombinaison et le transfert sont exprimés à partir de promoteurs dépendants de SetCD. De plus, nous avons établi que SetCD agit comme un activateur de classe II, c'est-à-dire que son motif de reconnaissance chevauche la boîte -35 de ses promoteurs cibles, de manière à recruter plus efficacement la RNAP bactérienne. En plus d'identifier les promoteurs cibles, cette étude a permis de définir expérimentalement l'arrangement opéronique des ICE SXT/R391. Les résultats exposés dans cet article supportent également la nécessité de SetCD dans la cellule réceptrice pour l'activation adéquate de la

transcription du gène de l'intégrase de l'ICE dans la cellule réceptrice et l'intégration subséquente dans *prfC*, réglant les évidences conflictuelles de travaux antérieurs. L'ensemble de ces résultats est présenté dans l'article qui suit, accepté pour publication en 2015 dans la revue *Nucleic Acids Research* publiée par *Oxford University Press*, dont la référence complète est la suivante :

Poulin-Laprade, D., Matteau, D., Jacques, P.E., Rodrigue, S. and Burrus, V. (2015). Transfer Activation of SXT/R391 Integrative and Conjugative Elements: Unraveling the SetCD Regulon. Nucleic Acids Research. Feb 27;43(4):2045-56.

2.2.2. Contribution à l'article

L'approche expérimentale a été développée conjointement avec Vincent Burrus, Dominick Matteau et Sébastien Rodrigue. J'ai effectué les expériences menant aux résultats présentés aux Figures 3 et 5 et préparé les Figures 3, 5 et 6B. J'ai également fourni les constructions génétiques et établi les conditions d'induction utilisées pour l'immunoprécipitation de la chromatine couplée à la digestion par une exonucléase (ChIP-exo) et le séquençage de l'ARN (RNA-seq). Ces expériences ont été effectuées et analysées par Dominick Matteau avec l'aide de Sébastien Rodrigue et Pierre-Étienne Jacques, et les résultats sont exposés dans les Figures 1, 2 et 4. Dominick Matteau a préparé les figures 1, 2, 4 et 6A. J'ai rédigé le manuscrit conjointement avec Vincent Burrus, Sébastien Rodrigue et Dominick Matteau.

2.2.3. Page titre

Transfer Activation of SXT/R391 Integrative and Conjugative Elements: Unraveling the SetCD Regulon

Dominic Poulin-Laprade[1,#], Dominick Matteau[2,#], Pierre-Étienne Jacques[3], Sébastien Rodrigue[2], and Vincent Burrus[1]

[1] Laboratory of Bacterial Molecular Genetics

[2] Laboratory of Microbial Systems and Synthetic Biology

[3] Laboratory of Bioinformatics and Genomics

[#] The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors.

2.2.4. Abstract

Integrative and conjugative elements (ICEs) of the SXT/R391 family have been recognized as key drivers of antibiotic resistance dissemination in the seventh-pandemic lineage of *Vibrio cholerae*. SXT/R391 ICEs propagate by conjugation and integrate site-specifically into the chromosome of a wide range of environmental and clinical *Gammaproteobacteria*. SXT/R391 ICEs bear *setC* and *setD*, two conserved genes coding for a transcriptional activator complex that is essential for activation of conjugative transfer. We used chromatin immunoprecipitation coupled with exonuclease digestion (ChIP-exo) and RNA sequencing (RNA-seq) to characterize the SetCD regulon of three representative members of the SXT/R391 family. We also identified the DNA sequences bound by SetCD in MGI*Vfl*Ind1, a mobilizable genomic island phylogenetically unrelated to SXT/R391 ICEs that hijacks the conjugative machinery of these ICEs to drive its own transfer. SetCD was found to bind a 19-bp sequence that is consistently located near the promoter -35 element of SetCD-activated genes, a position typical of class II transcriptional activators. Furthermore, we refined our understanding of the regulation of excision from and integration into the chromosome for SXT/R391 ICEs and demonstrated that *de novo* expression of SetCD is crucial to allow integration of the incoming ICE DNA into a naive host following conjugative transfer.

2.2.5. Introduction

Integrating conjugative elements (ICEs) have recently been shown to be the most abundant conjugative elements in practically all prokaryotic clades (1,2). As such, ICEs are a major driving force of bacterial genome evolution allowing rapid acquisition

of a variety of new traits and adaptive functions such as virulence, metabolic pathways, and resistance to antimicrobial compounds, heavy metals or bacteriophage infection (3-5). For instance, ICEs of the SXT/R391 family largely contribute to the spread of antibiotic resistance genes in the seventh-pandemic lineage of *Vibrio cholerae*, the etiologic agent of cholera, which remains a major cause of mortality and morbidity on a global scale (6-9). Most SXT/R391 ICEs found in *V. cholerae* clinical isolates confer resistance to sulfamethoxazole and trimethoprim, two antibiotics commonly used for the treatment of cholera (10,11). Since the early 90's, SXT/R391 ICEs have become widespread in environmental and clinical *V. cholerae* isolates from Asia and Africa (7,8). SXT/R391 ICEs are also present in all isolates recovered from cholera patients in Haiti (12-15), are naturally occurring in many other *Gammaproteobacteria* (6,16-18), and are easily transferred to *Escherichia coli* in the laboratory (19). The SXT/R391 ICEs are grouped together because they share a common set of 52 highly conserved genes, among which ~25 are important for their maintenance, dissemination by conjugation and regulation (6,20). Highly conserved genes in SXT/R391 ICEs are distributed in seven distinct clusters separated by variable cargo DNA (HS1 to 5 and VRI to IV) (Figure 1A). These conserved clusters consist of *int* (integration/excision), *mob1-2* (DNA processing), *mpf1-3* (mating pair formation modules 1, 2 and 3), and *reg* (regulation) (Figure 1A) (6).

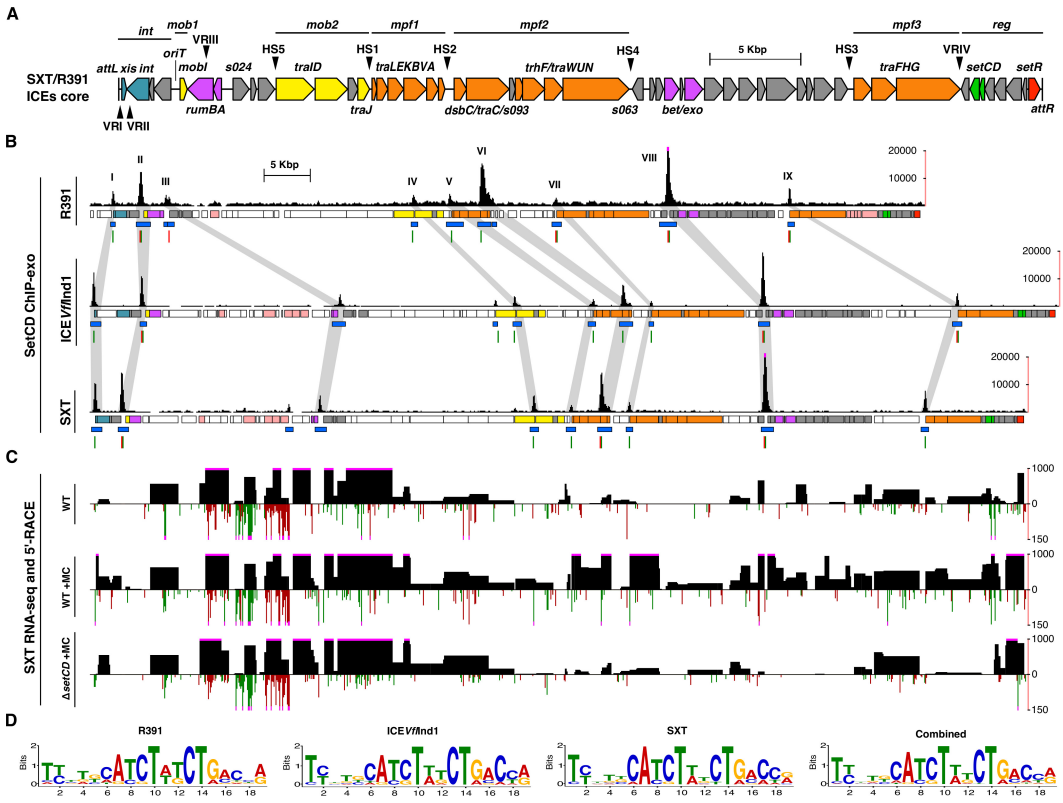

Figure 1. ***In vivo*** **identification of SetCD targets. (A)** Schematic representation of the conserved genes of SXT/R391 ICEs adapted from Wozniak *et al* (6). Genes are represented by arrows and are color coded as follow: blue, integration and excision; yellow, DNA processing; orange, mating pair formation; purple, RecA-independent homologous recombination and Umu-like mutagenic repair; green, transcriptional activator; red, transcriptional repressor; grey, other or hypothetical functions. Variable cargo DNA inserted in the conserved core of SXT/R391 ICEs is marked by arrowheads (HS1 to 5 and VRI to IV). The left and right chromosomal attachment sites *attL* and *attR* are also shown. **(B)** SetCD ChIP-exo analysis for R391, ICE*Vfl*Ind1 and SXT. For each ICE, four tracks are shown. First track: number of ChIP-exo reads mapped on ICE DNA sequence (pink dots at the top of black bars indicate a signal beyond the represented y-axis maximal value). Roman numbers indicate ChIP-exo peaks conserved between the three ICEs. Second track: genes from conserved core (same color code as in panel A), variable DNA regions (white), and antibiotic/heavy-metal resistance genes (pink). Third and fourth tracks: position of ChIP-exo enrichment peaks and the position of SetCD-binding motifs, respectively. SetCD motifs were identified for each ICE using the corresponding logo shown in panel D, and represented by green and red tick marks on positive and negative strands, respectively. **(C)** RNA-seq experiments on wild-type SXT in the presence or absence of mitomycin C, and SXT Δ*setCD* in the presence mitomycin C. For each condition, the upper track shows the reads per kilobase of transcript per million mapped reads value (RPKM) for each gene (black boxes) and the lower track shows the genome-wide 5'-RACE signals (positive strand in green, negative strand in red). Pink dots are as in B. **(D)** Logo sequences recognized by SetCD in R391, ICE*Vfl*Ind1, SXT, as well as the consensus logo of all three ICEs.

Mobilizable genomic islands (MGIs) are small (<33 kb) genomic islands found in several species of marine *Gammaproteobacteria* (21). SXT/R391 ICEs can mediate the transfer in *trans* of a particular class of MGIs at high frequency by an unusual mechanism. ICE-encoded relaxosome proteins recognize a *cis*-acting locus in MGIs that mimics the origin of transfer (*oriT*) of SXT/R391 ICEs (22). Unlike conjugative plasmids, ICEs and MGIs do not stably maintain by extra-chromosomal replication and must integrate into the host cell's chromosome to be vertically inherited (3,22,23). The gene pairs *int/xis* and int_{MGI}/*rdfM* are key components for the maintenance of SXT/R391 ICEs and MGIs, respectively. *int* and int_{MGI} code for two distinct and unrelated integrases, which mediate the integration of SXT/R391 ICEs into the 5' end of *prfC* (peptide chain release factor RF3), and the integration of MGIs into the 3' end of *yicC* (protein of unknown function), respectively (22,24). *xis* and *rdfM* encode recombination directionality factors (RDFs), which facilitate the integrase-mediated excision from the chromosome of SXT/R391 ICEs and MGIs as circular molecules that serve as substrates for the conjugative transfer machinery (23,25).

The conjugative transfer of SXT/R391 ICEs is regulated by three conserved genes located near the *attR* attachment site (Figure 1A) (20). *setR* encodes a λ cI-related transcriptional repressor, which prevents the expression of *setC* and *setD* (26,27). Agents that damage DNA and induce the host SOS response (UV light, mitomycin C, ciprofloxacin) are thought to induce the RecA*-stimulated autoproteolysis and inactivation of SetR, thereby alleviating the repression of *setCD* and allowing excision and transfer of SXT/R391 ICEs (26) (Figure S1A). The proteins SetC and SetD are thought to assemble as a heteromeric complex that activates the expression of SXT/R391 genes important for conjugative transfer (20,25). SetCD has also been reported to activate the expression of the *mosAT* toxin-antitoxin system carried by SXT (28) as well as the expression of int_{MGI} and *rdfM* of MGIs, thereby triggering the excision of MGIs from the chromosome (22,23). The number of promoters activated

by SetCD in SXT/R391 ICEs and the nature of the SetCD operator sites are currently unknown.

In this study we characterized the SetCD regulon and DNA motifs bound by SetCD in three SXT/R391 ICEs and one MGI originating from three different pathogens using chromatin immunoprecipitation coupled to exonuclease digestion (ChIP-exo) and RNA sequencing (RNA-seq). From this analysis, we identified and validated sequences of the SetCD operators. Finally, we investigated the dynamics of integration and excision of SXT to address the regulation of expression of *xis* and *int* in both donor and recipient cells. We demonstrated that SetCD must be expressed *de novo* in the recipient cells to allow the establishment of SXT in a new host.

2.2.6. Material and methods

2.2.6.1. Bacterial strains and bacterial conjugation assays

The *E. coli* strains used in this study are described in Table 1. The strains were routinely grown in Luria-Bertani (LB) broth at 37°C in an orbital shaker/incubator and were maintained at -80°C in LB broth containing 15% (vol/vol) glycerol. Antibiotics were used as described in Text S1. Conjugation assays were performed as described elsewhere (29). To induce expression of Int from pInt33 and SetCD from pGG2B in complementation assays, mating experiments were carried on LB agar plates supplemented with 0.02% arabinose.

2.2.6.2. Molecular biology methods

Genomic and plasmid DNA preparation, PCR products amplification and purification, electro-transformation of *E. coli*, gene expression analysis by qRT-PCR and β-galactosidase assays, Southern blotting, CHEF-PFGE and sequencing were performed using standard molecular biology techniques. Details are provided in Text S1.

2.2.6.3. Plasmid and strain constructions

Plasmids and primers used in this study are described in Table 1 and S2, respectively. Mutants of SXT, R391 and ICE*Vfl*Ind1 were constructed using the one-step chromosomal gene inactivation technique (30). Constructions of reporter and expression vectors were done using conventional molecular methods. Detailed methodology is described in Text S1.

2.2.6.4. ChIP-exo experiments and RNA sequencing

A description of the ChIP-exo, RNA seq and 5' RACE were conducted as described in Carraro *et al.* (31). Additional details are provided in Text S1. Sequenced libraries are described in Table S3.

2.2.6.5. Data availability

Fastq files for each experiment were deposited at the NCBI Sequence Read Archive (SRA) under accession numbers SRX708080 and SRR1583172 for SXT ChIP-exo, SRX708425 and SRR1583516 for R391 ChIP-exo, SRX708426 and SRR1583532 for ICE*Vfl*Ind1/MGI*Vfl*Ind1 ChIP-exo, SRX708086 and SRR1583199 for SXT RNA-seq as well as SRX708115 and SRR1583202 for SXT 5'-RACE. Complete data from aligned reads for ChIP-exo, RNA-seq and 5'-RACE can also be visualized using the UCSC genome browser at http://bioinfo.ccs.usherbrooke.ca/setCD.html

Table 1. Strains and plasmids used in this study.

Strain or plasmid	Relevant genotype and phenotype[a]	Reference
E. coli strains		
CAG18439	MG1655 *lacZU118 lacI42*::Tn*10* (Tc^r)	(55)
BW25113	F-, Δ(*araD-araB*)567, Δ*lacZ4787*(::*rrnB-3*), λ-, *rph-1*, Δ(*rhaD-rhaB*)*568*, *hsdR514*	(30,56)
HW220	CAG18439 *prfC*::SXT (Tc^r Su^r Tm^r)	(24)
VB17	BW25113 *prfC*::SXT (Su^r Tm^r)	This study
AD57	CAG18439 *prfC*::ICE*Vfl*Ind1 *yicC::*DUP(MGI*Vfl*Ind1) (Tc^r Su^r Tm^r)	(22)
AD72	CAG18439 *prfC*::SXT *yicC*::MGI*Vfl*Ind1*::aph* (Tc^r Su^r Tm^r Kn^r)	(22)
AD133	CAG18439 *prfC*::[SXT Δ*setCD*] *yicC*::MGI*Vfl*Ind1::*aph* (Tc^r Su^r Tm^r Kn^r)	(23)
AD132	CAG18439 *yicC*::MGI*Vfl*Ind1::*aph* pGG2B (Tc^r Kn^r Ap^r)	(23)
DPL2	CAG18439 *prfC*::SXT pGG2B (Tc^r Su^r Tm^r Ap^r)	This study
DPL3	CAG18439 *prfC*::[SXT Δ*setCD*] (Tc^r Su^r Tm^r)	This study
DPL5	CAG18439 *prfC*::[SXT Δ*setCD*] pGG2B (Tc^r Su^r Tm^r Ap^r)	This study
VB111	MG1655 (Nx^r)	(57)
DPL9	VB111 pInt33 (Nx^r Ap^r)	This study
DPL246	BW25113 *prfC*::[pDPL227 Δ*setR*::*lacZ-aad7*] (Cm^r Sp^r)	This study
DPL453	BW25113 $attB_\lambda$::pDPL440 pGG2B (Kn^r Ap^r)	This study
DPL400	BW25113 $attB_\lambda$::pDPL382 pGG2B (Kn^r Ap^r)	This study
DPL490	BW25113 $attB_\lambda$::pDPL465 pGG2B (Kn^r Ap^r)	This study
DPL393	BW25113 $attB_\lambda$::pDPL384 pGG2B (Kn^r Ap^r)	This study
DPL394	BW25113 $attB_\lambda$::pDPL385 pGG2B (Kn^r Ap^r)	This study
DPL489	BW25113 $attB_\lambda$::pDPL467 pGG2B (Kn^r Ap^r)	This study
DPL494	MG1655 $lacZo^{CD1}$	This study
DPL501	MG1655 $lacZo^{CD2}$	This study
DPL513	BW25113 *prfC*::R391 (Kn^r)	This study
DPL491	BW25113 *prfC*::[R391 $orf90^{3xFLAG}$] (Kn^r)	This study
DPL492	BW25113 *prfC*::[SXT $setC^{3xFLAG}$] (Su^r Tm^r)	This study
DPL493	CAG18439 *prfC*::[ICE*Vfl*Ind1 $setC^{3xFLAG}$] *yicC*::DUP(MGI*Vfl*Ind1) (Tc^r Su^r Tm^r)	This study
Plasmids		
pInt33	pBAD33 int_{SXT} (Cm^r)	(25)
pBAD30	*ori* p15A *bla araC* P_{BAD}-MCS (Ap^r)	(58)
pGG2B	pBAD30 *setCD* (Ap^r)	(35)
pSW23T	$oriT_{RP4}$ $ori_{R6K\gamma}$ (Cm^r)	(59)
pDPL189	pCR2.1-TOPO $attP_{SXT}$ (Ap^r Kn^r)	This study
pKD3	PCR template for one-step chromosomal gene inactivation (Cm^r)	(30)
pKD13	PCR template for one-step chromosomal gene inactivation (Kn^r)	(30)
pVI36	PCR template for one-step chromosomal gene inactivation (Sp^r)	(57)
pDPL458	pKD3 containing -73 to -31 of P_{s003} (Cm^r)	This study
pNC12	pVI36-SPATagCt (Sp^r)	This study
pOPlacZ	$ori_{R6K\gamma}$ $attP_\lambda$ *aph lacZ* (Kn^r)	(31)
pDPL440	pOPlacZ containing -100 to +81 of SXT P_{s003} (Kn^r)	This study
pDPL382	pOPlacZ containing -52 to +81 of SXT P_{s003} (Kn^r)	This study
pDPL465	pOPlacZ containing -35 to +81 of SXT P_{s003} (Kn^r)	This study
pDPL384	pOPlacZ containing -180 to +11 of SXT P_{xis} (Kn^r)	This study
pDPL385	pOPlacZ containing -52 to +11 of SXT P_{xis} (Kn^r)	This study
pDPL467	pOPlacZ containing -35 to +11 of SXT P_{xis} (Kn^r)	This study

[a] Ap^r, ampicillin resistant; Cm^r, chloramphenicol resistant; Em^r, erythromycin resistant; Kn^r, kanamycin resistant; Nx^r, nalidixic acid resistant; Sp^r, spectinomycin resistant; Su^r, sulfamethoxazole resistant; Tc^r, tetracycline resistant; Tm^r, trimethoprim resistant; Ts, thermosensitive.

2.2.7. Results

2.2.7.1. Characterization of the SetCD regulon in three SXT/R391 ICEs

The exact target genes and sequence motif recognized by the SetCD complex, which plays an essential role in ICE conjugative transfer activation, has yet to be determined. Using RNA-seq and ChIP-exo (31,32), we undertook an exhaustive characterization of the SetCD regulon in three ICEs found in clinical isolates of three different pathogens: the prototypical ICEs SXT from *V. cholerae* O139 (19) and R391 from *Providencia rettgeri* (33) as well as ICE*Vfl*Ind1 from *Vibrio fluvialis* (34). *E. coli* strains DPL492, DPL491 and DPL493 (Table 1) bearing derivatives of SXT, R391 and ICE*Vfl*Ind1, each expressing a native SetD subunit along with a SetC subunit C-terminally fused to the 3xFLAG tag ($SetC^{3xFLAG}$) were used in these experiments (Figure S1A). The 3xFLAG tag did not affect the function of the SetC orthologs based on the similar transfer frequency of each ICE relative to its wild-type counterpart (Figure S1B).

The ChIP-exo and RNA-seq assays were carried out after induction of the cell cultures using mitomycin C to trigger expression of SetCD from its native promoter. ChIP-exo data analyses revealed nine major SetCD enrichment peaks located upstream of the same genes and operons – most of which playing a key role in conjugative transfer – in the conserved core sequence shared by the three ICEs (Figure 1A, 1B and Table S1). Four of these peaks are located upstream of genes that are predicted to be involved in the formation of the mating pore: *traL* (conjugal transfer pilus assembly protein, peak V), *traV* (outer membrane lipoprotein, peak VI), *dsbC* (conjugative disulfide bond isomerase, peak VII) and *traF* (conjugal pilus

assembly protein, peak IX). One peak was also present inside the 3' end of the predicted relaxase gene *traI*, upstream of the gene *traD* (type IV coupling protein, peak IV). Two additional peaks were observed, one upstream of *xis* (peak I) and one (peak II) in the intergenic region between *mobI* (auxiliary component of the relaxosome) and *s003*, which is part of the operon containing *int*. The strongest peak (peak VIII) was detected in the intergenic region between the two divergent genes *s063* and *s089*. *s089* is the first gene of a large operon coding for a RecA-independent homologous recombination system (35,36). The last statistically significant peak (peak III) is located in the intergenic region between *rumA* (UmuD-like protein) and *s024*. Transcriptional activity measured by RNA-seq in *E. coli* HW220 (wild-type SXT) and DPL3 (SXT Δ*setCD*) correlates with the presence of a SetCD-binding site as the expression of 29 out of 52 core genes in SXT, including genes located downstream of SetCD-binding sites, is significantly increased upon mitomycin C induction compared to a Δ*setCD* mutant under the same conditions (Figure 1C and Dataset S1). Most genes that are not significantly affected by the expression of SetCD in SXT appear to be either inactive or constitutively expressed, and are mainly found in variable cargo DNA. Their functions are unknown or not directly tied to conjugative transfer, and include the antibiotic resistance genes, transposase genes, the *s027-s040* gene cluster, the diguanylate cyclase gene *dgcL* among others (Dataset S1).

2.2.7.2. Characterization of SetCD-dependent promoters

We carried out *de novo* motif discovery of DNA sequences bound by SetCD for each independent ICE ChIP-exo dataset, thereby generating three highly similar logo sequences (Figure 1D) in which subtle differences between the extracted motifs reflect ICE-specific polymorphisms in promoter regions. For each ICE, we next

determined the exact location of proposed SetCD-binding sites within the ChIP-exo peaks (Figure 1B) and observed a footprint likely corresponding to the SetCD and RNA polymerase holoenzyme complexes bound to the corresponding promoters (Figure 2A-F, and Table S1) (31,37). In some instances, we were able to identify two occurrences of a SetCD motif within the same peak. For example, the intergenic region *s063-s089* contains back-to-back SetCD-binding motifs, thereby revealing the presence of two SetCD-activated divergent promoters (Figure 2E). Genome-wide 5' rapid amplification of cDNA ends (5'-RACE) and primer extension analyses (Figure 1C and S2) allowed us to determine transcription start sites in SXT, revealing that SetCD-binding motifs are located immediately upstream of the -35 promoter box. This promoter structure was observed for all transcription start sites located between a SetCD-binding motif and a gene in the same orientation (Figure 2G). This organization is reminiscent of class II activation, in which the activator binds to a sequence that overlaps the promoter -35 element and usually contacts the RNA polymerase σ subunit (37).

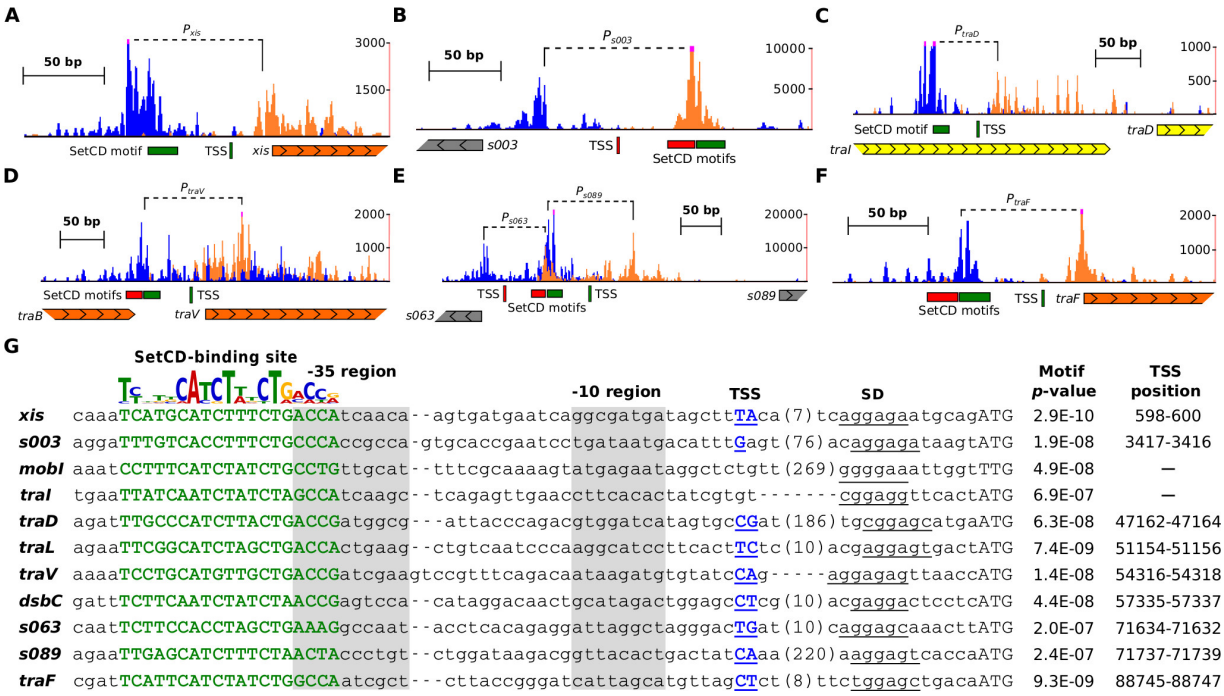

Figure 2. Organization of SetCD-dependent promoters. (A-F) SetCD-binding profile by ChIP-exo represented by the number of reads mapped on the promoter region upstream of *xis* (**A**), *s003* (**B**), *traD* (**C**), *traV* (**D**), *traF* (**F**) and in the intergenic region between *s063* and *s089* (**E**). The tracks plot ChIP-exo reads density (positive DNA strand in blue, negative DNA strand in orange) at single nucleotide resolution. SetCD-binding motifs and transcription start sites (TSS) are indicated (positive DNA strand in green, negative DNA strand in red). The TSS of *s089* was previously identified (35). Genes are represented by arrows and are color coded based on their function as in Figure 1. **(G)** Alignment of SetCD-dependent promoters in SXT. The SetCD-binding logo of SXT is reproduced from Figure 1D. The SetCD boxes are shown in bold green capital letters with their respective *p*-values. The position of the TSS in the promoter sequences is indicated when available (bold blue capital letters and underlined) along with the corresponding coordinates in SXT. Shine-Dalgarno sequences (SD) are underlined while start codons are in capital letters. The approximate position of the -35 and -10 regions is highlighted in gray. The numbers in bracket indicates the length in bp of spacers between the represented TSS and the Shine-Dalgarno sequence.

2.2.7.3. Validation of the SetCD operator sequences

We validated that the proposed SetCD-binding motifs promote the observed binding of SetCD and transcriptional activation by fusing the *lacZ* reporter gene to the promoters P_{s003} (DPL453) and P_{xis} (DPL393), respectively responsible for the expression of *int* and *xis*. For each promoter, two mutants were also generated. A first variant, Δ-*53*, lacked the sequence immediately upstream of the promoter-proximal SetCD motif (DPL400 and DPL394) while another variant, Δ-*36*, lacked the entire region upstream of the -35 promoter box, thus removing the first 17 bp of the proximal SetCD box (DPL490 and DPL489) (Figure 3A and 3B). β-galactosidase assays were then carried out upon *setCD* expression from the arabinose-inducible P_{BAD} promoter. Addition of the predicted SetCD boxes of P_{s003} and P_{xis} upstream of *lacZ* boosted the β-galactosidase activity by respectively ~520 and 2 200 fold, while the presence of additional upstream sequence made no difference (Figure 3B, compare WT with Δ-*53*). When the putative SetCD boxes were missing, the β-galactosidase activity dropped below the detection limit of our assay (Figure 3B, compare Δ-*53* with Δ-*36*).

To confirm that this motif alone, and not a hypothetical factor acting in *cis*, was sufficient to confer SetCD-dependent induction of gene expression, we substituted the native binding site of the cAMP receptor protein (CRP) of the P_{lac} promoter upstream of *lacZYA* in *E. coli* MG1655 by the SetCD box of P_{s003}. Two chimeric promoters containing the operator mutations *lacZo*^{CD1} and *lacZo*^{CD2} were constructed (Figure 3C). The *lacZo*^{CD1} mutant (DPL494) retained the -35 element of P_{lac}, whereas it was substituted by the -35 element of P_{s003} in the *lacZo*^{CD2} mutant (DPL501). The three operator sites o_1, o_2 and o_3 of the LacI repressor remained unaffected in both constructs. *setCD* was expressed under the control of P_{BAD} from pGG2B in the strains containing the constructions and in wild-type MG1655. The absence of significant

β-galactosidase activity observed on M9 glycerol medium supplemented with glucose for both *lacZo*^{CD1} or *lacZo*^{CD2} mutations confirmed the inability of the hybrid promoters to respond to the activation by CRP bound to cAMP regardless of the alleviation of LacI repression by IPTG and of the presence of repressed *setCD* (Figure 3D). In contrast, when glucose was replaced by arabinose, expression of *setCD* triggered a strong expression from the promoter containing *lacZo*^{CD1}, producing dark blue colonies, and weak expression from the one containing *lacZo*^{CD2} (Figure 3D and 3E). Strong expression from the latter was observed only upon concomitant alleviation of LacI repression by addition of IPTG (Figure 3D and 3E). These results indicate that expression of the *lacZYA* operon became SetCD-dependent when the CRP operator site of P_{lac} was replaced by either *lacZo*^{CD1} or *lacZo*^{CD2}. Interestingly, the variant *lacZo*^{CD2} seemed to remain strongly repressed by LacI upon *setCD* overexpression as shown by the lack of induction in the absence of IPTG, while the variant *lacZo*^{CD1} was not (Figure 3E). This phenotypical difference observed between the two mutants can be attributed to their respective -35 sequence. The -35 of *lacZo*^{CD2} (CACCGC) is very distant from the σ^{70} consensus, while *lacZo*^{CD1} harbors the more canonical -35 of P_{lac} (TTTACA). Taken together, these experiments confirm that the ChIP-exo derived SetCD motif alone is sufficient to confer SetCD-dependent activation of gene expression.

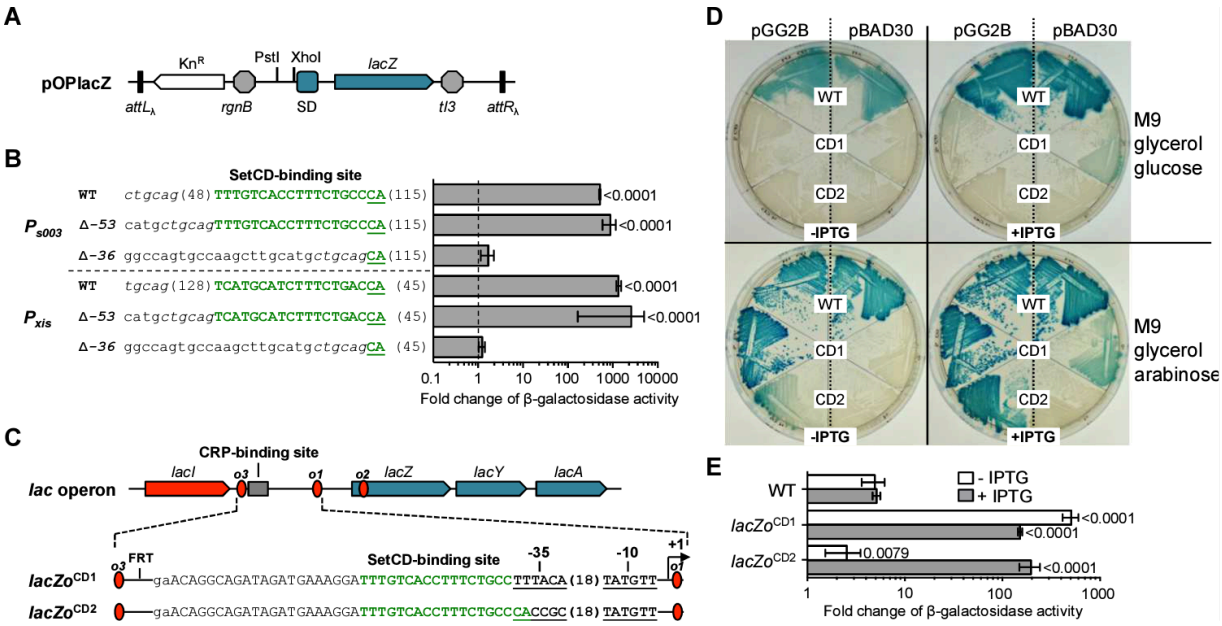

Figure 3. Validation of the DNA motif mediating the activation by SetCD. (A) Schematic representation of pOPlacZ, an integrative vector used as a promoter activity reporter system. Promoter regions were cloned between the PstI and XhoI restriction sites upstream of a promoterless *lacZ* gene transcriptionally isolated by the terminator sequences *rgnB* and *tI3*. SD, Shine-Dalgarno sequence. **(B)** β-galactosidase activity measured for P_{s003} and P_{xis} (WT) and derivatives Δ-*53* and Δ-*36* with SetCD expressed from an arabinose-inducible promoter on pGG2B. Nucleotides from the SetCD boxes are shown in green. Promoter variants were transcriptionally fused to *lacZ* in the pOPlacZ vector, and integrated in one copy at the $attB_{\lambda}$ site of *E. coli* BW25113 (Table 1). The right panel reports β-galactosidase activities expressed as the ratio between the Miller units in the arabinose-induced versus non-induced conditions. Results are the means and standard deviations of at least three independent biological replicates. The *p*-values from a two-way ANOVA with Tukey's multiple comparison test comparing the log of the means of the P_{s003} and P_{xis} WT or Δ-*53* variants relative to the corresponding Δ-*36* mutant are indicated. **(C)** Organization of two SetCD-dependent mutants of P_{lac}. In both mutants the CRP-binding site was replaced by the SetCD operator of P_{s003}. *lacZo*^{CD1} carries the -35 of P_{lac}, while *lacZo*^{CD2} has the -35 of P_{s003}. **(D)** Wild-type *E. coli* MG1655 (WT), *lacZo*^{CD1} mutant DPL494 (CD1) and *lacZo*^{CD2} mutant DPL501 (CD2) grown on M9 glycerol minimal medium supplemented with glucose or arabinose, with or without IPTG. The strains were carrying either pBAD30 or its *setCD*-expressing derivative pGG2B. **(E)** β-galactosidase activity measured for *E. coli* MG1655, DPL494 (*lacZo*^{CD1} mutant) and DPL501 (*lacZo*^{CD2} mutant) containing pGG2B grown in M9 glycerol minimal medium with glucose or arabinose, with or without IPTG. Ratios between the Miller units in the arabinose-induced versus glucose conditions are shown. Results are the means and standard deviations of four independent biological replicates. The *p*-values from a two-way ANOVA with Tukey's multiple comparison test comparing the log of the means of the *lacZo*^{CD1} and *lacZo*^{CD2} mutants relative to the wild-type in the corresponding conditions are indicated.

2.2.7.4. ChIP-exo assays reveal SetCD-regulated conserved genes in MGIs

The strain DPL493 used for ChIP-exo assays also contains MGI*Vfl*Ind1, an MGI originally detected in the same *V. fluvialis* strain that contains ICE*Vfl*Ind1 (Table 1), allowing us to monitor on this MGI the binding of SetCD provided in *trans* by ICE*Vfl*Ind1. Two major peaks were detected on MGI*Vfl*Ind1, both mapping upstream of two of the four conserved core genes (Figure 4A-C and Table S1). The first one was detected upstream of *rdfM* and the second one was found upstream of *cds4*, a gene of unknown function. To test whether *cds4* is regulated by SetCD, we measured its expression by quantitative real-time PCR (qRT-PCR) in *E. coli* containing SXT (AD72), SXT Δ*setCD* (AD133), or a plasmid expressing SetCD from P_{BAD} (AD132). While expression of *cds4* was induced by mitomycin C in the strain containing wild-type SXT, it was nearly abolished in the absence of *setCD* regardless of the presence of mitomycin C (Figure 4D). Overexpression of SetCD alone also dramatically increased the level of *cds4* transcript (~30,000-fold induction, Figure 4E), thereby confirming that *cds4* of MGIs is activated by SetCD. Motif Alignment and Search Tool (MAST) analysis on MGI*Vfl*Ind1 sequence revealed the presence of SetCD-binding motifs upstream of *cds4* and *rdfM* (Figure 4B, 4C and 4F). Mapping of the transcription start site of *rdfM* obtained by primer extension analysis (Figure S2) suggests that like in SXT, SetCD acts as a class II activator in MGIs (Figure 4C and 4F).

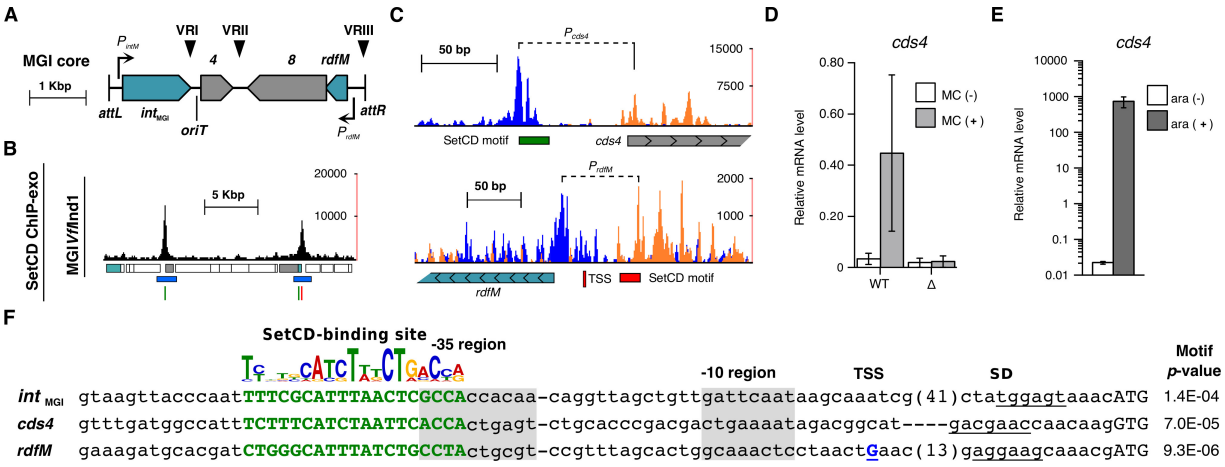

Figure 4. SetCD-dependent activation of MGI*Vfl*Ind1. (A) Representation of the conserved genes of MGIs. The color code is the same as in Figure 1. **(B)** SetCD ChIP-exo binding profile in MGI*Vfl*Ind1. Tracks are organized and color coded as in Figure 1B. SetCD motifs were derived from the SetCD-binding logo of ICE*Vfl*Ind1 presented in Figure 1D. **(C)** Close-up view of the peaks located in the intergenic region upstream of *cds4* and *rdfM* in MGI*Vfl*Ind1. Tracks are organized as in Figure 2A-F. Genes are represented by arrows and are color coded as in Figure 1B. **(D)** Quantification by qRT-PCR of the mRNA transcripts of *cds4* in *E. coli* strains containing wild-type SXT (WT; AD72) or SXT Δ*setCD* (Δ; AD133) in the presence or absence of mitomycin C (MC). **(E)** The effect of arabinose-induced SetCD expression from pGG2B on *cds4* mRNA in an *E. coli* strain devoid of ICE (AD132) as determined by qRT-PCR. **(F)** Alignment of SetCD-dependent promoters in MGI*Vfl*Ind1. The SetCD logo of ICE*Vfl*Ind1 is represented as in Figure 1D. Sequences are organized as in Figure 2G. The int_{MGI} SetCD-binding motif was found by FIMO while *cds4* and *rdfM* motifs were found by MAST. The transcription start site (TSS) of *rdfM* was determined by primer extension (Figure S2) and is located at position 18 284-18 285 on the negative DNA strand.

2.2.7.5. Establishment of SXT into a naive host requires *de novo* expression of *setCD*

Although it is known, and confirmed by our results, that SetCD activates the expression of SXT *int* in the donor strain (20,25), the importance of SetCD for the expression of *int* in recipient cells and consecutive integration of SXT in the recipient's chromosome is not clearly established as conflicting observations have been reported. Results from Beaber *et al*. (20) suggest that *setCD* is expendable in the recipient cells as Δ*setC* and Δ*setD* mutants of SXT have been shown to transfer and establish in recipient cells when the deletions were complemented in the donor cells exclusively. If SetC and SetD are not necessary for integration of SXT in the chromosome of the recipient cells, then expression of *int* likely occurs at a low constitutive level. Alternatively, SetCD could be produced in the donor cell and translocated through the mating pore into the recipient cell during conjugative transfer to stimulate *int* expression in the recipient. However, Burrus and Waldor reported that a suicide vector containing *attP* and *int* driven by its native promoter (P_{s003}) is unable to integrate in recipient cells lacking *setCD* (25), thereby suggesting that SetCD is required for *de novo* expression of *int* in recipient cells.

To clearly address the role of SetCD in the establishment of SXT in the recipient cells, we conducted mating assays using combinations of *E. coli* donor and recipient strains harboring SXT or its Δ*setCD* mutant with or without plasmids expressing either *int* or *setCD* under control of P_{BAD} (Figure 5A). Overexpression of *int* in recipient cells did not enhance the transfer of SXT (Figure 5A *a* and *b*), indicating that SXT integration into the recipient's chromosome is not a rate limiting step. This conclusion is also supported by *setCD* overexpression in the donor cell, which resulted in a ~3-log increase of transfer (Figure 5A *a* and *c*). Consequently, mating pore assembly or DNA

translocation is likely the rate-limiting step of SXT conjugative transfer. As expected, transfer of the Δ*setCD* mutant was abolished, even upon expression of *int* in recipient cells (Figure 5A *d* and *f*).

Intriguingly, transfer of SXT Δ*setCD* was only partially restored when *setCD* was overexpressed in donor cells, at a rate of only one fifth of wild-type SXT and ~4 logs lower than wild-type SXT upon *setCD* overexpression (Figure 5A, compare *a* vs *e*, and *c* vs *e*). Because in such a context, expression of the transfer genes is not compromised in donor cells, the low rate of transfer of mating *e* suggests that integration of SXT Δ*setCD* into the recipient's chromosome became the rate-limiting step of transfer. Assuming that this phenotype was attributable to weak or lack of expression of *int* in the recipient cells, we can rule out that the SetCD protein complex is translocated, at least in significant amounts, from the donor to recipient cells during SXT transfer. Indeed, stimulation of *int* expression mediated by translocated SetCD should have enabled normal integration of SXT Δ*setCD* at a wild-type rate despite the lack of *setCD* genes in the recipient cells. In fact, transfer of SXT Δ*setCD* was fully restored to wild-type level only upon concomitant overexpression of *setCD* in the donor and *int* in the recipient (Figure 5A *g* and *c* vs *e*). Altogether, these results suggest that the SetCD complex is not translocated from the donor to recipient cells during conjugation. Instead, *setCD* is expressed *de novo* upon entry of SXT in the recipient cells, allowing *int* expression to mediate SXT integration.

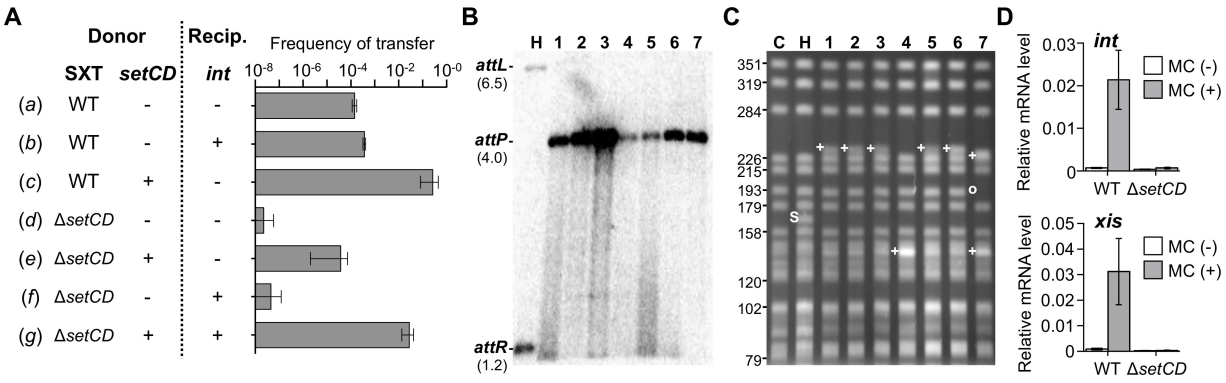

Figure 5. *De novo setCD* expression is required for normal establishment of SXT in a new host. **(A)** Conjugation assays using as donor *E. coli* CAG18439 derivatives harboring SXT (WT, HW220 or DPL2) or its Δ*setCD* mutant without (-) or with (+) *setCD* expressed from pGG2B (DPL3 or DPL5). The recipient strain is a Nx^r-derivative of *E. coli* MG1655 expressing (+) or not (-) *int* under control of P_{BAD} from pInt33 (DPL9 or VB111). Results are the mean values and standard deviations of three independent biological replicates. An unpaired t-test with Welch's correction was used to compare the log of the mean values. *p*-values are as follow: *a* vs *c*, < 0.0001; *a* vs *d*, 0.0916; *a* vs *g*, < 0.0001; *c* vs *g*, 0.0064. **(B)** Southern blot hybridization analysis of EcoRI/EcoRV-digested genomic DNAs of 7 randomly picked exconjugants isolated from the mating *e* (DPL5 × VB111) in panel A probed with the *attP* fragment of SXT (Lanes 1 to 7). Lane H, donor strain HW220. **(C)** Ethidium bromide-stained agarose gel of SpeI-digested genomic DNAs separated by contour-clamped homogeneous electric field pulsed field gel electrophoresis (CHEF-PFGE). Molecular sizes are in kilobases. (S), fragment containing SXT inserted in *prfC*; (+) and (o), respectively atypical and missing fragments compared to control lanes C and H. Lanes: C, CAG18439; H, HW220; 1 to 7, randomly picked exconjugants isolated from the mating *e* in panel A. **(D)** Quantification by qRT-PCR of the mRNA transcripts containing *int* and *xis* in CAG18439 containing wild-type SXT (WT; HW220) or SXT Δ*setCD* (DPL3) in the presence or absence of mitomycin C (MC).

2.2.7.6. A *setCD*-null mutant of SXT maintains atypically in exconjugant colonies

The Δ*setCD* mutation seemed to hinder the expression of *int* in the recipient cells, eventually leading to loss of the incoming SXT. Suboptimal *int* expression could

reduce integration or promote maintenance of the ICE by other means. We submitted a sample of 7 Δ*setCD* exconjugants randomly picked from mating *e* (Figure 5A) to profiling by Southern blot hybridization and PFGE analyses. Southern blot probing of the genomic DNA of these exconjugants with a fragment overlapping *attP* revealed atypical restriction patterns compared to the control donor strain containing SXT integrated as a single copy into *prfC* (HW220) (Figure 5B). All exhibited a signal for *attP* but lacked the characteristic *attL* and *attR* fragments normally present after correct SXT integration, suggesting that SXT Δ*setCD* failed to integrate site-specifically. This result was also supported by PCR amplification of *attB*, which confirmed that the 5' end of *prfC* was intact in all 7 exconjugants (Figure S3A and S3B). At least two possible mechanisms could explain the formation of such anomalous exconjugants. First, SXT Δ*setCD* could have integrated by homologous recombination or transposition potentially through one of the insertion sequence present in the variable region VRIII (Figure 1A). Alternatively, SXT Δ*setCD* could maintain as a circular replicative molecule. We tested both hypotheses by subjecting the genomic DNA of the same exconjugants to SpeI restriction, which does not cut SXT, and PFGE. None of the exconjugants exhibited the expected 166-kb SpeI fragment containing SXT as seen in the control donor strain HW220 (Figure 5C). Instead distinct restriction patterns were observed. In the most frequent pattern (5 out of 7 exconjugants), a new large fragment of ca. 230 kb was observed. This is inconsistent with the possibility of tandem integration of multiple copies of SXT at *prfC* since two copies of SXT would result in a larger 257-kb SpeI fragment (29) suggesting that SXT Δ*setCD* integrated into different chromosomal loci. None of these patterns exhibited a 99-kb band compatible with a replicative form of SXT. Altogether these results confirmed that SXT Δ*setCD* is unable to integrate into *prfC* in a site-specific fashion upon entry into a naive host, supporting that *int* expression was compromised or abolished in Δ*setCD* mutants.

2.2.7.7. Expression of *int* and *xis* requires activation by SetCD

The defective integration phenotype of SXT Δ*setCD* suggests that expression of *int* strictly depends upon activation by SetCD. To test this hypothesis, we monitored the expression of *int* by qRT-PCR on cDNA derived from *E. coli* strains containing SXT (HW220) or its Δ*setCD* derivative (DPL3). We also monitored the expression of *xis* in the same conditions as it is expected to be expressed only upon activation of SXT transfer to promote the site-specific excision reaction over integration. The DNA-damaging agent mitomycin C increased *int* and *xis* mRNA transcript levels by 31 and 39 fold respectively, in a *setCD*-dependent fashion (Figure 5D), thereby supporting our RNA-seq results showing that their expression is controlled by SetCD. Furthermore, both *int* and *xis* transcript levels were slightly above the limit of detection for wild-type SXT in non-inducing conditions (Figure 5D). Spontaneous induction of the SOS response in a subpopulation of the cell culture (38) likely accounts for this low basal level of expression and for the basal level of SXT transfer (~5×10^{-4} exconjugant/recipient in Figure 5A *a*). In the same conditions, expression of both *int* and *xis* dropped below the detection level for SXT Δ*setCD* regardless of the presence or absence of mitomycin C (Figure 5D). Together with the failure of the *setCD*-null mutant of SXT to integrate into *prfC* of recipient cells, these results demonstrate that *de novo* expression of *setCD* is required to trigger the expression of *int* in recipient cells and allow stable maintenance and inheritance of SXT/R391 ICEs.

2.2.8. Discussion

In many archetypical conjugative elements such as the IncFI F conjugative plasmid or the ICE Tn*916*, genes coding for the conjugative apparatus are organized as a single

polycistronic operon (39,40). In contrast, although the *tra* genes of SXT/R391 ICEs are syntenic with the *tra* genes of F-like conjugative plasmids (20,39), they are distributed in five distinct gene clusters, separated by variable DNA clusters (Figure 1A) (6). Such an organization requires an adaptable activation system that allows coherent expression of the diverse components essential for ICE conjugation. This includes activation of the site-specific recombination system mediating excision from the chromosome as a circular molecule (*xis* and *int*), assembly of the mating apparatus (*traLEKBVA*, *dsbC*/*traC*/*trhF*/*traWUN*, and *traFHG*), initiation of ICE DNA transfer (*mobI*/*oriT* and *traIDJ*), and finally integration into the recipient cell's chromosome (*int*) (Figure 6A).

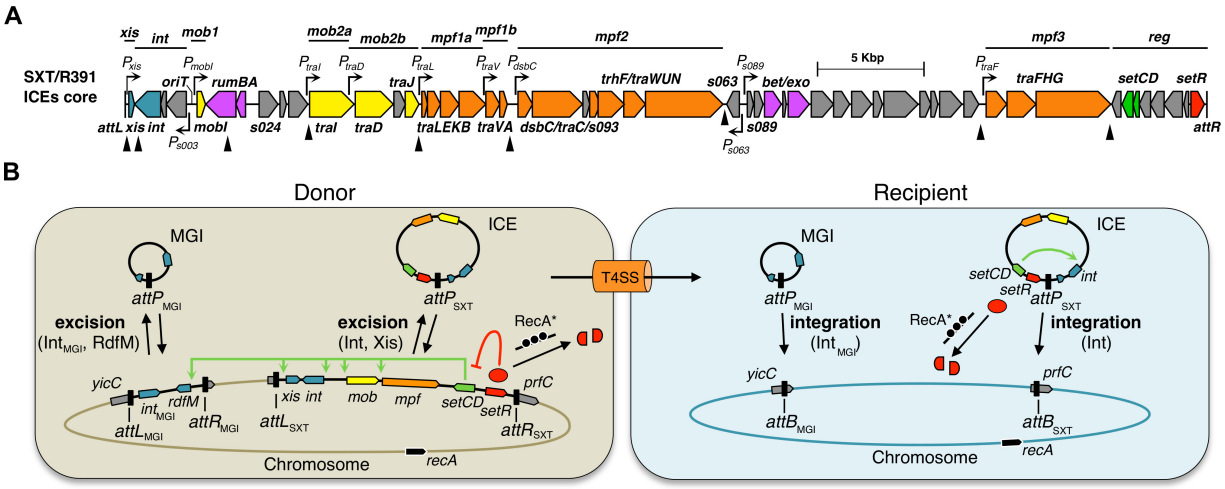

Figure 6. (A) Overview of the SetCD-regulated promoters on the conserved core of genes of SXT/R391 ICEs. Genes and annotations are color-coded as in figure 1A, and black triangles show the position of variable cargo DNA. (B) Summary of the main regulatory events occurring in the donor and recipient cells during conjugative transfer of SXT/R391 ICEs and MGIs.

In the present study we have established using ChIP-exo and RNA-seq experiments that SetCD coordinate the expression of many genes and operons by binding upstream of the transcription start sites of *xis*, *s003*, *mobI*, *traI*, *traD*, *traL*, *traV*, *dsbC*,

s063, *s089*, and *traF* (Figure 6A). This transcriptional regulation profile is similar to the one recently reported for the conjugative transfer functions encoded by the IncA/C conjugative plasmids that are activated by AcaCD, an activator complex distantly related to SetCD (34% and 23% identity between the C and D subunits, respectively) (31). In fact, activation of gene expression by SetCD and AcaCD seems to be reminiscent of the activation of transcription by the transcriptional master regulator of bacterial flagellum biogenesis FlhDC. The complex $FlhD_2C_2$ binds 30 to 40 bp upstream of the σ^{70}-dependent transcription start sites and activates the expression of class II flagellar operons, which encode components of the flagella basal body and export machinery (41). O'Halloran *et al* (42) previously reported a predicted potential SetCD binding site upstream of *xis* in R391 based on analogies with "$FlhD_2C_2$ box" arms (42). However, no such a binding motif exists upstream of *xis* in ICEs of the SXT/R391 family as the SetCD DNA recognition motif obtained by combining the SetCD-bound promoters of SXT, R391 and ICE*Vfl*Ind1 drastically differs from the $FlhD_2C_2$ (43-46) and AcaCD (31) binding sites. This is not surprising given the low similarity between SetCD, FlhCD and AcaCD (20,31). The -35 and -10 elements of all SetCD-dependent promoters studied here are poorly conserved relatively to the σ^{70} canonical -35 and -10 promoter boxes (47,48). In fact, the -35 region lacks a recognizable motif of the canonical -35 signal (TTGACA). In all the SetCD-dependent promoters, we found that the SetCD box overlaps or is located immediately upstream of the -35 element (Figure 2G and 4F). The close proximity of the SetCD box near the sequence usually recognized by σ factors suggest that binding of SetCD compensates for the lack of a recognizable -35 element, allowing recruitment of the RNA polymerase holoenzyme, in a manner similar to the class II CRP-, FNR- and $FlhD_2C_2$-dependent promoters (49-51). Biochemical characterization of SetCD is needed to clarify whether SetCD operates like $FlhD_2C_2$ by interacting with the RNA polymerase α subunit C-terminal domain (52).

Our data allowed us to deepen our understanding of the regulation of ICEs of the SXT/R391 family. To date both *traIDJ* and *traLEKBVA* were presumed to be two single polycistronic operons each regulated by a unique promoter located upstream of *traI* and *traL*, respectively (6,20). To our surprise, our data indicate that expression of these two gene clusters is much more complex. First, the promoter of *traI* differs between ICEs because the -10 element is part of the conserved core whereas the -35 region is brought by variable DNA found in the variable region HS5 (Figure 1A and S4B). Yet clearly, *traI* expression is dependent of SetCD in SXT as shown by our RNA-seq data (Dataset S1) and close examination of the ChIP-exo signal in the region upstream of *traI* suggests the presence of a potential weaker promoter footprint as well as a degenerate SetCD-binding motif for all three ICEs (Figure S4A and Table S1). This raises interesting questions about the selective pressure operating on sequences inserting near the -35 promoter element of *traI* for the conservation of a functional SetCD operator. Second, *traD* expression is driven from a SetCD-dependent promoter located within *traI* (Figure 2C and 2G). Finally, the *traLEKBVA* gene cluster likely corresponds to two independent operons, *traLEKB* and *traVA*, although we cannot rule out the existence of mRNA transcripts containing *traL* to *traA* (Figure 1A and 1C, Figure 2D and 2G). ChIP-exo data revealed a strong peak and a well-conserved SetCD motif within the 5' end of the coding sequence of *traV* in all three ICEs. Examination of the annotation of *traV* revealed that a much better suited ribosome binding site with a nearly canonical Shine-Dalgarno sequence is located 75 bp downstream of the original *traV* annotation. As a consequence, we propose to redefine the start codon of *traV* at this new location (Figure 2D and 2G). Wozniak and Waldor (28) reported that *mosAT*, which is located in HS2 and encodes a toxin-antitoxin system promoting the maintenance of SXT, was induced by SetCD. Clearly, our results show that there is no SetCD-binding site upstream of *mosA*, and that *mosAT* is not differentially expressed upon induction of SetCD by mitomycin C (Table S1 and Dataset S1). This suggests that the reported increased expression of

mosAT likely resulted from read-through of the mRNA transcript initiated by SetCD at the promoter P_{traV}.

The upstream region of the mutagenic DNA repair system *rumAB* of R391 was reported to contain a strong match to known LexA-binding sites at positions -44 to -25 relative to the *rumA* initiation codon (53). Although we have found a SetCD-binding site 530 bp upstream of *rumA* in the correct orientation to drive *rumA* expression in R391 only (Figure 1B), our RNA-seq data in SXT clearly shows that upon mitomycin C induction, *rumA* is not differentially expressed in a wild-type SXT compared to the corresponding *setCD*-null mutant (Dataset S1). Our data rather supports the idea that *rumAB* is part of the LexA regulon, not of the SetCD regulon (53).

The flexibility provided by SetCD has likely helped shaping the complex genetic structure and remarkable plasticity of SXT/R391 ICEs. Ironically, SetCD has also become a beacon signalling the presence of SXT/R391 ICEs to parasitic genomic islands that hijack the ICE transfer machinery. MGIs mimic SetCD-binding sites to activate their own excision in response to the presence of an ICE of the SXT/R391 family in the same cell. We have found here that SetCD binds upstream of *rdfM*, a gene that is known to be activated by SetCD (23), and upstream of *cds4*, a conserved gene of unknown function (Figure 4). Although int_{MGI} has been shown to be activated by SetCD (22,23), ChIP-exo failed to identify a statistically significant binding of SetCD upstream of this gene in MGI*Vfl*Ind1 but a more degenerate SetCD motif can be found using the Find Individual Motif Occurrences tool (FIMO) (Figure 4F). This is consistent with our previous report that int_{MGI} is induced only 300 fold by SetCD overexpression whereas *rdfM* is induced 2 000 fold in identical conditions (23). Nevertheless, int_{MGI} was also shown to be constitutively expressed at low level in the absence of SetCD, thereby allowing integration of the MGI into the chromosome of

the recipient cell independently of the cotransfer of an SXT/R391 ICE (23) (Figure 6B). This strategy likely favors MGI's "survival" and dissemination as less than 2% of recipient cells receiving an MGI have been shown to simultaneously receive a copy of the helper ICE (22). Our data confirm that MGI excision is strictly regulated by the activation of *rdfM* expression, not by int_{MGI} overexpression, and is consistent with the inability of an *rdfM*-null mutant to excise and transfer (23).

Unlike in MGIs, we demonstrated here that expression of SXT *int* and *xis* require the presence of SetCD in both donor and recipient cells. Our results indicate that an SXT Δ*setCD* mutant complemented in *trans* by SetCD only in the donor cells is incapable of integrating site-specifically into the 5' end of *prfC*. This observation led us to conclude that neither SetCD nor Int are translocated into the recipient strain during conjugation. Instead, *setCD* is expressed *de novo* in the recipient, allowing *int* expression to mediate SXT integration into *prfC* in a site-specific fashion (Figure 6B). This requirement contrasts with the regulation of MGI integration, which is independent of SetCD. As both *int* and *setCD* are physically linked on SXT/R391 ICEs, these mobile elements do not need to rely on a dual promoter regulation of their integrase gene to ensure their survival and dissemination.

Upon entry of an ICE of the SXT/R391 family into a naive host, the repressor SetR is initially absent, and concomitant expression of *setCD* and *setR* likely allow build-up of the SetCD and SetR pools. Entry by conjugation of the ICE as single-stranded DNA is known to activate the SOS response (54), which could transitorily contribute to maintain a low pool of SetR protein, thereby favoring the build-up of SetCD (Figure 6B). However, we showed that the transcript levels are similar for both *xis* and *int* in the presence of SetCD whereas the presence of the RDF Xis in the recipient cells reduces SXT transfer, likely by interfering during the recombination between *attP* and

attB (25). Therefore a yet uncharacterized mechanism likely maintains a low level of Xis protein or delays its expression in the recipient cell to favor chromosomal ICE integration. In the conjugative transposon Tn*916*, the RDF *xis* and integrase *int* genes are part of a single long tetracycline-inducible mRNA transcript reading through the *attP* attachment site and extending to all of the transfer genes (40). Unlike many mobile integrating elements, *xis* and *int* in SXT/R391 ICEs are not organized as an operon. Instead they are two convergent genes suggesting that other factors besides SetCD regulate their expression.

In summary, our study establishes the SetCD-dependent regulation of excision, transfer and integration the ICEs of the SXT/R931 family as well as the genomic islands they mobilize, and highlights the importance of SetCD in triggering the spread of antibiotic resistance conferring mobile elements in *V. cholerae* populations and in other *Gammaproteobacteria*.

2.2.9. Supplementary data

Supplementary Data are available at NAR online, including Supplementary References (60-73).

2.2.10. Acknowledgement

We are thankful to N. Carraro, F. Pelletier, M. Paquette-D'Avignon, S. Caron, A. Daccord, C. Déry, M. Belluau and the team of the Centre de calcul scientifique of the

Université de Sherbrooke for their technical assistance as well as P. Beauregard, A. Lavigueur, E. Bordeleau, and J.P. Brousseau for helpful discussions and critical reading of the manuscript.

2.2.11. Funding

This work was supported by the Fonds de recherche du Québec - Nature et technologies [PR-173580 to V.B. and S.R.], and by a Discovery Grant and Discovery Acceleration Supplement from the Natural Sciences and Engineering Council of Canada [326810 and 412288 to V.B.]. V.B. holds a Canada Research Chair in molecular bacterial genetics. D.P.L and D.M. were supported by scholarships from the Fonds de recherche du Québec. S.R. and P.É.J. are chercheur boursier junior 1 of the Fonds de recherche du Québec - Santé (FRQS).

2.2.12. References

1. Guglielmini, J., Quintais, L., Garcillan-Barcia, M.P., de la Cruz, F. and Rocha, E.P. (2011) The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet.*, **7**, e1002222.

2. Ghinet, M.G., Bordeleau, E., Beaudin, J., Brzezinski, R., Roy, S. and Burrus, V. (2011) Uncovering the prevalence and diversity of integrating conjugative elements in *Actinobacteria*. *PLoS One*, **6**, e27846.

3. Burrus, V., Pavlovic, G., Decaris, B. and Guedon, G. (2002) Conjugative transposons: the tip of the iceberg. *Mol. Microbiol.*, **46**, 601-610.

4. Burrus, V. and Waldor, M.K. (2004) Shaping bacterial genomes with integrative and conjugative elements. *Res. Microbiol.*, **155**, 376-386.

5. Wozniak, R.A. and Waldor, M.K. (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.*, **8**, 552-563.

6. Wozniak, R.A., Fouts, D.E., Spagnoletti, M., Colombo, M.M., Ceccarelli, D., Garriss, G., Dery, C., Burrus, V. and Waldor, M.K. (2009) Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet.*, **5**, e1000786.

7. Burrus, V., Marrero, J. and Waldor, M.K. (2006) The current ICE age: biology and evolution of SXT-related integrating conjugative elements. *Plasmid*, **55**, 173-183.

8. Spagnoletti, M., Ceccarelli, D., Rieux, A., Fondi, M., Taviani, E., Fani, R., Colombo, M.M., Colwell, R.R. and Balloux, F. (2014) Acquisition and evolution of SXT-R391 integrative conjugative elements in the seventh-pandemic *Vibrio cholerae* lineage. *mBio*, **5**, e01356-01314.

9. Chun, J., Grim, C.J., Hasan, N.A., Lee, J.H., Choi, S.Y., Haley, B.J., Taviani, E., Jeon, Y.S., Kim, D.W., Lee, J.H. *et al.* (2009) Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 15442-15447.

10. Garriss, G. and Burrus, V. (2013) In Roberts, A. P. and Mullany, P. (eds.), Bacterial integrative mobile genetic elements. Landes Biosciences, Austin, TX, pp. 217-234.

11. Kaper, J.B., Morris, J.G., Jr. and Levine, M.M. (1995) Cholera. *Clin. Microbiol. Rev.*, **8**, 48-86.

12. Chin, C.-S., Sorenson, J., Harris, J.B., Robins, W.P., Charles, R.C., Jean-Charles, R.R., Bullard, J., Webster, D.R., Kasarskis, A., Peluso, P. *et al.* (2011) The origin of the haitian cholera outbreak strain. *New Engl. J. Med.*, **364**, 33-42.

13. Ceccarelli, D., Spagnoletti, M., Cappuccinelli, P., Burrus, V. and Colombo, M.M. (2011) Origin of *Vibrio cholerae* in Haiti. *Lancet Infect. Dis.*, **11**, 262.

14. Katz, L.S., Petkau, A., Beaulaurier, J., Tyler, S., Antonova, E.S., Turnsek, M.A., Guo, Y., Wang, S., Paxinos, E.E., Orata, F. *et al.* (2013) Evolutionary dynamics of *Vibrio cholerae* O1 following a single-source introduction to Haiti. *MBio*, **4**.

15. Reimer, A.R., Van Domselaar, G., Stroika, S., Walker, M., Kent, H., Tarr, C., Talkington, D., Rowe, L., Olsen-Rasmussen, M., Frace, M. *et al.* (2011) Comparative genomics of *Vibrio cholerae* from Haiti, Asia, and Africa. *Emerg. Infect. Dis.*, **17**, 2113-2121.

16. Osorio, C.R., Marrero, J., Wozniak, R.A.F., Lemos, M.L., Burrus, V. and Waldor, M.K. (2008) Genomic and functional analysis of ICE*Pda*Spa1, a fish-pathogen-derived SXT-related integrating conjugative element that can mobilize a virulence plasmid. *J. Bacteriol.*, **190**, 3353-3361.

17. Harada, S., Ishii, Y., Saga, T., Tateda, K. and Yamaguchi, K. (2010) Chromosomally encoded bla_{CMY-2} located on a novel SXT/R391-related integrating conjugative element in a *Proteus mirabilis* clinical isolate. *Antimicrob. Agents Chemother.*, **54**, 3545-3550.

18. Badhai, J., Kumari, P., Krishnan, P., Ramamurthy, T. and Das, S.K. (2013) Presence of SXT integrating conjugative element in marine bacteria isolated from the mucus of the coral *Fungia echinata* from Andaman Sea. *FEMS Microbiol. Lett.*, **338**, 118-123.

19. Waldor, M.K., Tschape, H. and Mekalanos, J.J. (1996) A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. *J. Bacteriol.*, **178**, 4157-4165.

20. Beaber, J.W., Hochhut, B. and Waldor, M.K. (2002) Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *J. Bacteriol.*, **184**, 4259-4269.

21. Daccord, A., Ceccarelli, D., Rodrigue, S. and Burrus, V. (2013) Comparative analysis of mobilizable genomic islands. *J. Bacteriol.*, **195**, 606-614.

22. Daccord, A., Ceccarelli, D. and Burrus, V. (2010) 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.*, **78**, 576-588.

23. Daccord, A., Mursell, M., Poulin-Laprade, D. and Burrus, V. (2012) Dynamics of the SetCD-regulated integration and excision of genomic islands mobilized by integrating conjugative elements of the SXT/R391 family. *J. Bacteriol.*, **194**, 5794-5802.

24. Hochhut, B. and Waldor, M.K. (1999) Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. *Mol. Microbiol.*, **32**, 99-110.

25. Burrus, V. and Waldor, M.K. (2003) Control of SXT integration and excision. *J. Bacteriol.*, **185**, 5045-5054.

26. Beaber, J.W., Hochhut, B. and Waldor, M.K. (2004) SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature*, **427**, 72-74.

27. Beaber, J.W. and Waldor, M.K. (2004) Identification of operators and promoters that control SXT conjugative transfer. *J. Bacteriol.*, **186**, 5945-5949.

28. Wozniak, R.A.F. and Waldor, M.K. (2009) A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genet.*, **5**, e1000439.

29. Burrus, V. and Waldor, M.K. (2004) Formation of SXT tandem arrays and SXT-R391 hybrids. *J. Bacteriol.*, **186**, 2636-2645.

30. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 6640-6645.

31. Carraro, N., Matteau, D., Luo, P., Rodrigue, S. and Burrus, V. (2014) The master activator of IncA/C conjugative plasmids stimulates genomic islands and multidrug resistance dissemination. *PLoS Genet.*, **10**, e1004714.

32. Rhee, H.S. and Pugh, B.F. (2012) ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr Protoc Mol Biol*, **Chapter 21**, Unit 21 24.

33. Coetzee, J.N., Datta, N. and Hedges, R.W. (1972) R factors from *Proteus rettgeri*. *J. Gen. Microbiol.*, **72**, 543-552.

34. Ahmed, A.M., Shinoda, S. and Shimamoto, T. (2005) A variant type of *Vibrio cholerae* SXT element in a multidrug-resistant strain of *Vibrio fluvialis*. *FEMS Microbiol. Lett.*, **242**, 241-247.

35. Garriss, G., Poulin-Laprade, D. and Burrus, V. (2013) DNA-damaging agents induce the RecA-independent homologous recombination functions of integrating conjugative elements of the SXT/R391 family. *J. Bacteriol.*, **195**, 1991-2003.

36. Garriss, G., Waldor, M.K. and Burrus, V. (2009) Mobile antibiotic resistance encoding elements promote their own diversity. *PLoS Genet.*, **5**, e1000775.

37. Browning, D.F. and Busby, S.J. (2004) The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.*, **2**, 57-65.

38. McCool, J.D., Long, E., Petrosino, J.F., Sandler, H.A., Rosenberg, S.M. and Sandler, S.J. (2004) Measurement of SOS expression in individual *Escherichia coli* K-12 cells using fluorescence microscopy. *Mol. Microbiol.*, **53**, 1343-1357.

39. Lawley, T.D., Klimke, W.A., Gubbins, M.J. and Frost, L.S. (2003) F factor conjugation is a true type IV secretion system. *FEMS Microbiol. Lett.*, **224**, 1-15.

40. Celli, J. and Trieu-Cuot, P. (1998) Circularization of Tn*916* is required for expression of the transposon-encoded transfer functions: characterization of long tetracycline-inducible transcripts reading through the attachment site. *Mol. Microbiol.*, **28**, 103-117.

41. Chevance, F.F. and Hughes, K.T. (2008) Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.*, **6**, 455-465.

42. O'Halloran, J., McGrath, B. and Pembroke, J. (2007) The *orf4* gene of the enterobacterial ICE, R391, encodes a novel UV-inducible recombination directionality factor, Jef, involved in excision and transfer of the ICE. *FEMS Microbiol. Lett.*, **272**, 99-105.

43. Wozniak, C.E. and Hughes, K.T. (2008) Genetic dissection of the consensus sequence for the class 2 and class 3 flagellar promoters. *J. Mol. Biol.*, **379**, 936-952.

44. Lee, Y.-Y., Barker, C.S., Matsumura, P. and Belas, R. (2011) Refining the binding of the *Escherichia coli* flagellar master regulator, $FlhD_4C_2$, on a base-specific level. *J. Bacteriol.*, **193**, 4057-4068.

45. Stafford, G.P., Ogi, T. and Hughes, C. (2005) Binding and transcriptional activation of non-flagellar genes by the *Escherichia coli* flagellar master regulator $FlhD_2C_2$. *Microbiology*, **151**, 1779-1788.

46. Fitzgerald, D.M., Bonocora, R.P. and Wade, J.T. (2014) Comprehensive mapping of the *Escherichia coli* flagellar regulatory network. *PLoS Genet.*, **10**, e1004649.

47. Kumar, A., Malloch, R.A., Fujita, N., Smillie, D.A., Ishihama, A. and Hayward, R.S. (1993) The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. *J. Mol. Biol.*, **232**, 406-418.

48. Hawley, D.K. and McClure, W.R. (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res*, **11**, 2237-2255.

49. Lee, D.J., Minchin, S.D. and Busby, S.J. (2012) Activating transcription in bacteria. *Annu. Rev. Microbiol.*, **66**, 125-152.

50. Chilcott, G.S. and Hughes, K.T. (2000) Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar typhimurium and *Escherichia coli*. *Microbiol. Mol. Biol. Rev.*, **64**, 694-708.

51. van Hijum, S.A.F.T., Medema, M.H. and Kuipers, O.P. (2009) Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol. Mol. Biol. Rev.*, **73**, 481-509.

52. Liu, X., Fujita, N., Ishihama, A. and Matsumura, P. (1995) The C-terminal region of the alpha subunit of *Escherichia coli* RNA polymerase is required for transcriptional activation of the flagellar level II operons by the FlhD/FlhC complex. *J. Bacteriol.*, **177**, 5186-5188.

53. Kulaeva, O.I., Wootton, J.C., Levine, A.S. and Woodgate, R. (1995) Characterization of the *umu*-complementing operon from R391. *J. Bacteriol.*, **177**, 2737-2743.

54. Baharoglu, Z., Bikard, D. and Mazel, D. (2010) Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLoS Genet.*, **6**, e1001165.

55. Singer, M., Baker, T.A., Schnitzler, G., Deischel, S.M., Goel, M., Dove, W., Jaacks, K.J., Grossman, A.D., Erickson, J.W. and Gross, C.A. (1989) A collection of strains

containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*. *Microbiol. Rev.*, **53**, 1-24.

56. Grenier, F., Matteau, D., Baby, V. and Rodrigue, S. (2014) Complete genome sequence of *Escherichia coli* BW25113. *Genome Announc.*, **2**, e01038-01014.

57. Ceccarelli, D., Daccord, A., Rene, M. and Burrus, V. (2008) 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.*, **190**, 5328-5338.

58. Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.*, **177**, 4121-4130.

59. Demarre, G., Guerout, A.M., Matsumoto-Mashimo, C., Rowe-Magnus, D.A., Marliere, P. and Mazel, D. (2005) A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPa) conjugative machineries and their cognate *Escherichia coli* host strains. *Res. Microbiol.*, **156**, 245-255.

60. Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988) High efficiency transformation of *E.coli* by high voltage electroporation. *Nucleic Acids Res.*, **16**, 6127-6145.

61. Zeghouf, M., Li, J., Butland, G., Borkowska, A., Canadien, V., Richards, D., Beattie, B., Emili, A. and Greenblatt, J.F. (2004) Sequential peptide affinity (SPA) system for the identification of mammalian and bacterial protein complexes. *J. Proteome Res.*, **3**, 463-468.

62. Haldimann, A. and Wanner, B.L. (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.*, **183**, 6384-6393.

63. Heath, J.D., Perkins, J.D., Sharma, B. and Weinstock, G.M. (1992) NotI genomic cleavage map of *Escherichia coli* K-12 strain MG1655. *J. Bacteriol.*, **174**, 558-567.

64. Miller, J.F. (1992) *A short course in bacterial genetics*. Cold Spring Harbor Laboratory Press, Plainview, NY.

65. Sambrook, J., Fritsch, E. and Maniatis, T. (1989) *Molecular Cloning: A laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

66. Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, **2**, 28-36.

67. Bailey, T.L. and Gribskov, M. (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, **14**, 48-54.

68. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.

69. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, **9**, R137.

70. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol*, **11**, R106.

71. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017-1018.

72. Magoc, T., Wood, D. and Salzberg, S.L. (2013) EDGE-pro: Estimated degree of gene expression in prokaryotic genomes. *Evol Bioinform Online*, **9**, 127-136.

73. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841-842.

2.2.13. Supplementary data

2.2.13.1. Text S1 Supplementary Materials and Methods

2.2.13.1.1. Bacterial strains

The *E. coli* strains used in this study are described in Table 1. Antibiotics were used at the following concentrations: ampicillin (Ap), 100 µg/ml; chloramphenicol (Cm), 20 µg/ml; erythromycin (Em), 10 µg/ml; spectinomycin (Sp), 50 µg/ml; kanamycin (Kn), 50 µg/ml or 10 µg/ml for single copy integrants of pOPlacZ; nalidixic acid (Nx), 40 µg/ml; sulfamethoxazole (Su), 160 µg/ml; tetracycline (Tc), 12 µg/ml; trimethoprim (Tm), 32 µg/ml; mitomycin C (MC), 100 ng/ml.

2.2.13.1.2. Molecular biology methods

Genomic and plasmid DNA were prepared using the Wizard Genomic DNA Purification Kit (Promega) and EZ-10 Spin Column Plasmid DNA Minipreps Kit (Biobasic), respectively, according to manufacturer's instructions. All the enzymes used in this study were purchased from New England BioLabs or Enzymatics. PCR assays were performed with the primers described in Table S2. The PCR conditions were as follows: (i) 3 min at 94˚C; (ii) 30 cycles of 30 sec at 94˚C, 30 sec at the appropriate annealing temperature, and 1 minute/kb at 68˚C; and (iii) 5 min at 68˚C. When necessary, PCR products were purified using a EZ-10 Spin Column PCR Products Purification Kit (Biobasic) according to manufacturer's instructions. *E. coli*

was transformed by electroporation according to Dower *et al*. (1). Electroporation was carried out in a BioRad GenePulser Xcell apparatus set at 25 µF, 200 V and 1.8 kV using 1-mm gap electroporation cuvettes. Sequencing reactions were performed by the Plateforme de Séquençage et de Génotypage du Centre de Recherche du CHUL (Québec, QC, Canada).

2.2.13.1.3. Plasmid and strain constructions

Plasmids and primers used in this study are described in Table 1 and S2, respectively. The template plasmid pDPL458 was constructed by site-directed mutagenesis of pKD3 to introduce a SetCD box using the primer pair Int-CDbox-F/Int-CDbox-R and the Q5 Site-Directed Mutagenesis Kit (New England Biolabs). The template plasmid pNC12 was constructed by introducing the SPA tag from pJL148 (2) amplified with the primer pair SPACtBamHI.for/SPACtBamHI.rev into the BamHI site of pVI36.

The transcriptional fusion vector pOPlacZ contains the promoterless *E. coli* K12 *lacZ* gene with its native Shine-Dalgarno sequence (3). PCR fragments containing -100, -52 or -35 to +81 with respect to the transcriptional +1 site of SXT *s003*, and -180, -52 or -35 to +11 relative to the +1 of SXT *xis* were cloned between the PstI and XhoI sites of pOPlacZ to produce pDPL440, pDPL382, pDPL465, pDPL384, pDPL385 and pDPL467, respectively.

The Δ*setCD* (DPL3) mutation was introduced in SXT using the one-step chromosomal gene inactivation technique (4) with the primer pair setD2WF/setC2WR

and pKD13 as the templates. Similarly, translational fusions of the 3xFLAG tag the C-terminal end of SetC in SXT (DPL492), Orf90 in R391 (DPL491) and SetC in ICE*Vfl*Ind1 (DPL493) were constructed using the same technique with primer pairs DATsetC3xFLAG-F/DATsetC3xFLAG-SXT-R (SXT) or DATsetC3xFLAG-F/DATsetC3xFLAG-R391-R (R391 and ICE*Vfl*Ind1) and pNC12 as the template. The CRP operator upstream of the *lacZYA* operon in *E. coli* MG1655 was substituted by a SetCD box (*lacZo*^{CD1} and *lacZo*^{CD2}) using the same technique with primer pairs DATcdlacZ-FO3/DATcdlacZ-R3 and DATcdlacZ-FO3/DATcdlacZ-R4, respectively, and pDPL458 as the template, yielding DPL494 and DPL501. The strains DPL453, DPL400, DPL490, DPL393, DPL394 and DPL489 were constructed by transforming *E. coli* BW25113 expressing λ Int from pINT-ts with plasmids pDPL440, pDPL382, pDPL465, pDPL384, pDPL385 and pDPL467, respectively, following the method of Haldimann and Wanner (5). pGG2B was subsequently introduced in these strains to express *setCD* under control of the arabinose-inducible promoter P_{BAD}. All mutations were verified by PCR amplification and sequenced by the Plateforme de séquençage et génotypage des génomes (CHUL, Québec).

2.2.13.1.4. β-galactosidase assays

The substrates used to determine the LacZ levels were either o-2-nitrophenyl-β-D-galactopyranoside (ONPG) or 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). The β-galactosidase assays using ONPG were carried out as described previously (6) in LB medium (for strains DPL453, DPL400, DPL490, DPL393, DPL394, and DPL 489) or in M9 minimal medium supplemented with 1% glycerol and 1% thiamine (for strains MG1655, DPL494, and DPL501), and 100 µg/ml ampicillin for maintenance of pGG2B. The induction of *setCD* expression, the alleviation of LacI repression, and catabolic repression were done by supplementing with 0.02%

arabinose, 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), or 0.2% glucose, respectively, to a refreshed culture grown to an OD_{600nm} of 0.2 followed by 2-h incubation at 37°C with shaking prior cell sampling. For the assays using X-gal as a substrate, a single colony of each strain was re-suspended in 50 μl of LB broth by vortexing and streaked onto M9 minimal medium supplemented with 1.5% agar, 20 mg/ml X-gal, 1% glycerol, 1% thiamine and 100 mg/ml ampicillin for maintenance of either pBAD30 or pGG2B. Derepression or activation of the P_{lac} derivatives was done by further supplementing the media with 0.02% arabinose, 0.2% glucose and 1 mM IPTG.

2.2.13.1.5. Southern blot hybridization

The fragment *attP* for Southern blot hybridization was amplified by PCR using genomic DNA of HW220 as a template and primers CattPF/CAttPR (Table S2) (7). Chromosomal DNA was prepared using the Wizard Genomic DNA purification kit (Promega) as described in the manufacturer's instructions, followed by a phenol:chloroform:isoamyl alcohol (25:24:1) extraction and ethanol precipitation. Purified total genomic DNA was digested with EcoRI and EcoRV restriction enzymes, and analyzed by Southern blot hybridization (8) using the 602-bp EcoRI fragment of pDPL189 containing *attP* from SXT as a probe. The probe was radiolabeled using T4 polynucleotide kinase and dATP (γ-^{32}P) 3000 Ci/mmol and purified with an illustra MicroSpin G-25 column (GE Healthcare). The purified radiolabeled probe was added to the prehybridization solution (6X SSC, 5X Denhardt's reagent, 0.5% (w/v) SDS and 100 μg/ml sonicated salmon sperm DNA) with a specific activity of 2×10^6 cpm/μg. Radioactive signal was detected using a Storm 860 Molecular Imager (GMI) and brightness and contrasts were adjusted on the entire image using Quantity One 1-D analysis software (Bio-Rad).

2.2.13.1.6. CHEF-PFGE

E. coli chromosomal DNA used for contour-clamped homogeneous electric field pulsed field gel electrophoresis (CHEF-PFGE) was prepared as previously described (9). After digestion by the SpeI restriction enzyme, large *E. coli* restriction fragments were separated with a CHEF-DR II system (Bio-Rad) using the conditions described in Burrus and Waldor (10).

2.2.13.1.7. Real-Time quantivative PCR analysis

Total RNAs for reverse transcription real-time quantitative PCR (qRT-PCR) were extracted from *E. coli* strains HW220, DPL3, AD72, AD132 and AD133, and cDNA were synthesized as described elsewhere (11). The RNA quality was further verified with a 2100 Bioanalyzer instrument (Agilent Technologies). qRT-PCR was performed by measuring the increase of fluorescence using the Quantifast SYBR Green mix (QIAGEN) in an Eppendorf RealPlex (Eppendorf). Quantitative amplification of *int*, *xis* and *cds4* was carried out using primer pairs RTint-F/RTint-R, RTxis-F/RTxis-R and RTgene4Vf-F/RTgene4Vf-R respectively. For normalization, *rpoZ*, the gene coding for ω subunit of the RNA polymerase, was amplified using primers RTrpoZcoli-F and RTrpoZcoli-R and results were expressed as a relative expression based on the ΔCt calculation method. Experiments were carried out three times in triplicate and combined. Melting curves were carried out on the final reaction products (156-165 bp) to verify that amplification was specific to targets. All primer pairs exhibited efficiencies comprised between 91% and 94%. The detection limit was fixed at 0.0004, the mean of all results obtained with SXT Δ*setCD*.

2.2.13.1.8. Primer extension analysis

Primer extension experiment to determine the transcription start site of *s003* and *rdfM* was carried out as described elsewhere (12) using total RNA extracted respectively from HW220 and AD72 induced with 100 ng/ml mitomycin C. The primers used for the primer extension reactions using the Primer Extension System-AMV Reverse Transcriptase kit (Promega) were RRintR (*s003*) and PE-rdfM-2 (*rdfM*) (Table S2). A Sanger sequencing ladder was generated using the Sequenase 2.0 DNA Sequencing Kit (Affymetrix-USB).

2.2.13.1.9. SetCD ChIP-exo experiments and ChIP-exo libraries preparation

Chromatin immunoprecipitation coupled with exonuclease digestion (ChIP-exo) experiments and ChIP-exo libraries preparation were performed as described previously (3) with the following modifications. ChIP-exo experiments were performed in biological triplicates, each using 10 ml of *E. coli* strains DPL491, DPL492, or DPL493 bearing *setC*^{3xFLAG} (*orf90*^{3xFLAG} for R391) grown to an OD_{600nm} of ~0.6 with agitation in an orbital shaker at 37°C. Cultures were then induced with 200 ng/ml mitomycin C and incubated for an additional 2 hours at 37°C. Induced cultures were next subjected to 1% formaldehyde crosslinking for 20 min at 25°C followed by quenching with 0.125 M glycine. Immunoprecipitation procedure was then performed as described previously (3) using the Anti-FLAG M2 antibody (Sigma). ChIP-exo libraries were prepared as described previously (3) using the following primers (Table S2): TrueSeq-B-dT24-VN, 3'-T-IGA-A0 adaptor (IGA-A0-up annealed with IGA-A0-down-T/A; 30 nM adaptor final concentration), IGA-PCR-PE-F, and TrueSeq-MPEX-R with a specific barcode for each sample. The resulting libraries were purified with

Agencourt AMPure XP SPRI beads (Beckman Coulter) with a ratio of 1.0, before evaluating the library size and concentration using a 2100 Bioanalyzer instrument (Agilent Technologies). Barcoded libraries were pooled together in equal quantity to obtain approximately the same amount of reads per sample (Table S3).

2.2.13.1.10. Total RNA extractions and RNA-seq libraries preparation

Total RNA extractions were done with three biological replicates. 10 ml of *E. coli* DPL492 (SXT *setC*^{3xFLAG}) and 5 ml of *E. coli* DPL3 (SXT Δ*setCD*) were grown to an OD_{600nm} of ~0.3 as described for ChIP-exo experiments. DPL492 culture was split in two, and one half was induced with 200 ng/ml mitomycin C for two hours along with the entire DPL3 culture. All cultures reached a final OD_{600nm} of ~0.6 and total RNA was then extracted using the Direct-zol RNA MiniPrep kit (Zymo Research) according to the manufacturer's specifications. Purified total RNA was treated with 1.5 U of DNAse I in 1X DNAse I Reaction Buffer (Zymo Research) at 37°C for 20 min. DNAse-treated RNA samples were then purified using the RNA Clean & Concentrator-5 kit (Zymo Research) according to the manufacturer's specifications. The quality and concentration of total RNA extractions were evaluated using a 2100 Bioanalyzer instrument (Agilent Technologies). A total of nine RNA sequencing (RNA-seq) libraries were constructed in biological triplicates per experimental condition as described previously (3) using the following primers (Table S2): TrueSeq-3'-hybrid-B0, 5'-hybrid-A0, DSN-TrueSeq-F, DSN-TrueSeq-R, IGA-PCR-PE-F, and TrueSeq-MPEX-R with a specific barcode for each sample. Indexed libraries were pooled together in equal quantity to obtain approximately the same amount of reads per sample (Table S3).

2.2.13.1.11. Genome-wide 5'-RACE RNA libraries preparation

Equal quantities of purified total RNA from non-induced DPL492, induced DPL492, and induced DPL3 replicates were pooled to obtain a final amount of 2 µg for each experimental condition. 5' rapid amplification of cDNA ends (5'-RACE RNA) libraries were then prepared as described previously (3) using the following primers (Table S2) : TrueSeq-3'-hybrid-B0, 5'-hybrid-A0, IGA-PCR-PE-F, and TrueSeq-MPEX-R with a specific barcode for each sample. The resulting libraries were purified with Agencourt AMPure XP SPRI beads (Beckman Coulter) with a ratio of 0.9, before evaluating the library size and concentration using a 2100 Bioanalyzer instrument (Agilent Technologies). Indexed libraries were pooled together in equal quantity to obtain approximately the same amount of reads per sample (Table S3).

2.2.13.1.12. Illumina sequencing and data analysis

Illumina sequencing was performed on a Illumina HiSeq 2000 Sequencing system at the Plateau de biologie moléculaire et génomique fonctionnelle of the Institut de Recherche Cliniques de Montréal. Approximately 3 to 6 million paired-end reads of 50 bp were obtained for each library, each represented in triplicates in RNA-seq and ChIP-exo experiments (Table S3). Sample multiplexing relied on a 6-bp barcode located within the TrueSeq-MPEX-R primer (Table S2).

2.2.13.1.13. ChIP-exo data analysis

Only the forward reads of the Illumina paired-end reads obtained from SetCD ChIP-exo experiments (Table S3) were aligned using the Burrows-Wheeler Aligner (BWA) (13) since reverse reads do not contain interesting information using this type of library (3). The reference genomes used for the alignments are *E. coli* MG1655 genome (GenBank: U00096.3) as well as the appropriate ICE sequence: SXT (strain DPL492), GenBank: AY055428.1; R391 (strain DPL491), GenBank: AY090559.1; and ICE*Vfl*Ind1/MGI*Vfl*Ind1 (strain DPL493), GenBank: KM213605.1 (ICE*Vfl*Ind1) along with KC117176.1 (MGI*Vfl*Ind1). Alignment from biological triplicates were filtered, combined and sorted as described elsewhere (3). Normalized read density (Fig. 1B and Fig. 4B) as well as peak enrichments were calculated by the Model-based Analysis of ChIP-Seq tool (MACS) (14) as described elsewhere (3). Subpeaks (referred as “peaks” in the paper; Fig. 1B, Fig. 4B and Table S1) with a summit height below 40X of the theoretical bp coverage (obtained by calculating the ratio between the number of reads used by MACS and the combined *E. coli* MG1655 and respective ICE/MGI genome size) were discarded to eliminate the background signal. SetCD-binding logos (Fig. 1D) were created separately for each ICE independently by providing to the motif-based sequence analysis tool (MEME) (15) the nucleotide sequences corresponding to subpeaks summits ±100 bp with the “any number of repetitions” option and a maximum motif length of 20 bp (note that only the subpeaks found within the ICE sequences were provided to MEME for motifs elicitation). Precise locations of SetCD-binding motifs within each ICE were determined using the Motif Alignment and Search Tool (MAST) (16) and their respective matrix from MEME, and intersected with MACS subpeaks. The MAST option to process the negative DNA strand as a separated sequence was used, and only motifs with *p*-value below 1.0e-04 positioned completely within MACS peaks were considered significant hits (Fig. 1B and Fig. 2). If two or more motifs in the same orientation were

found within one peak, only the most significant *p*-value was conserved. The combined SetCD-binding logo was generated by submitting to MEME all significant SetCD motif sequences found by MAST in all three ICEs with the “One motif per sequence” option (Fig. 1D). Positions of SetCD binding motifs inside MGI*Vfl*Ind1 were identified by MAST with the ICE*Vfl*Ind1 SetCD binding-site matrix using the same parameters as for motifs searching in ICEs (Fig. 4B, 4C, and 4F). The degenerate SetCD-binding motifs located within the intergenic regions upstream *traI* were identified by providing the SetCD-binding matrices found for each ICE respectively to the Find Individual Motif Occurrences tool (FIMO) (17). Only motifs correctly oriented relative to the downstream gene with a *p*-value below 1.0e-03 were considered significant (Fig. S3). The degenerate SetCD-binding motif located within the intergenic region upstream int_{MGI} was identified by the same method using the ICE*Vfl*Ind1 SetCD binding-site matrix. High-resolution ChIP-exo data was generated by processing the filtered and sorted BWA aligned sequences to conserve only the first nucleotide of forward reads. Finally, near-single nucleotide resolution ChIP-exo reads density (Fig. 2A-F, 4C and S3A), as well as full reads ChIP-exo density, were calculated with Bedtools' genomecov function (18) for each DNA strand seperately.

2.2.13.1.14. RNA-seq data analysis

Paired-end Illumina reads obtained from RNA-seq libraries (Table S3) were trimmed and filtered as described elsewhere (3). Filtered reads were then aligned on the *E. coli* MG1655 genome (GenBank U00096.3) as well as the SXT sequence (GenBank AY055428.1) using the Estimated Degree of Gene Expression in Prokaryotes software (EDGE-pro) (19), which also generated reads per kilobase of transcript per million mapped reads values (RPKM) for each gene (Fig. 1C and Dataset S1). Differential expression analysis between RNA-seq experimental conditions (Dataset

S1 and Table S3) was then conducted using the DESeq R package (20) with the following options to calculate data dispersion: pooled method, fit-only sharing mode, and local fit Type. Statistically, genes between two given conditions were considered differentially expressed if the calculated *p*-value was equal or below 0.1 and the expression fold change was equal or above 5 (Dataset S1). To generate strand-specific RNA-seq density files for data visualization using a UCSC trackhub, filtered reads were also aligned by BWA on the *E. coli* MG1655 genome (GenBank U00096.3) as well as SXT (GenBank AY055428.1). Aligned reads were filtered with SAMtools view to discard those with an alignment quality score below 10, and combined per experimental condition using SAMtools merge. Pooled as well as individual filtered reads alignments were then sorted with SAMtools. RNA-seq reads density calculated per dataset with Bedtools genomecov for each DNA strand separately.

2.2.13.1.15. Genome-wide 5'-RACE data analysis and SetCD-regulated promoters characterization

Forward reads of Illumina paired-end sequences, which correspond to the start of transcribed RNAs, obtained from the 5'-RACE libraries (Table S3) were trimmed, aligned by BWA, filtered and sorted by SAMtools as described elsewhere (3). The reads were then processed to conserve only the first nucleotide, corresponding to potential transcription start sites. Single-bp resolution 5'-RACE reads density was calculated with Bedtools genomecov for each DNA strand separately (Fig. 1C). Properly oriented transcription start sites from the mitomycin C induced wild-type SXT that were located less than 125 bp downstream of a SetCD-binding motif were considered significant and reported in the SetCD-regulated promoter alignment (Fig. 2G).

2.2.13.2. Supplementary References

1. Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res*, **16**, 6127-6145.

2. Zeghouf, M., Li, J., Butland, G., Borkowska, A., Canadien, V., Richards, D., Beattie, B., Emili, A. and Greenblatt, J.F. (2004) Sequential peptide affinity (SPA) system for the identification of mammalian and bacterial protein complexes. *J. Proteome Res.*, **3**, 463-468.

3. Carraro, N., Matteau, D., Luo, P., Rodrigue, S. and Burrus, V. (2014) The master activator of IncA/C conjugative plasmids stimulates genomic islands and multidrug resistance dissemination. *PLoS Genet.*, **10**, e1004714.

4. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 6640-6645.

5. Haldimann, A. and Wanner, B.L. (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.*, **183**, 6384-6393.

6. Miller, J.F. (1992) *A short course in bacterial genetics*. Cold Spring Harbor Laboratory Press, Plainview, NY.

7. Burrus, V. and Waldor, M.K. (2003) Control of SXT integration and excision. *J. Bacteriol.*, **185**, 5045-5054.

8. Sambrook, J., Fritsch, E. and Maniatis, T. (1989) *Molecular Cloning: A laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

9. Heath, J.D., Perkins, J.D., Sharma, B. and Weinstock, G.M. (1992) NotI genomic cleavage map of *Escherichia coli* K-12 strain MG1655. *J. Bacteriol.*, **174**, 558-567.

10. Burrus, V. and Waldor, M.K. (2004) Formation of SXT tandem arrays and SXT-R391 hybrids. *J. Bacteriol.*, **186**, 2636-2645.

11. Daccord, A., Mursell, M., Poulin-Laprade, D. and Burrus, V. (2012) Dynamics of the SetCD-regulated integration and excision of genomic islands mobilized by integrating conjugative elements of the SXT/R391 family. *J. Bacteriol.*, **194**, 5794-5802.

12. Garriss, G., Poulin-Laprade, D. and Burrus, V. (2013) DNA-damaging agents induce the RecA-independent homologous recombination functions of integrating conjugative elements of the SXT/R391 family. *J. Bacteriol.*, **195**, 1991-2003.

13. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.

14. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, **9**, R137.

15. Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, **2**, 28-36.

16. Bailey, T.L. and Gribskov, M. (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, **14**, 48-54.

17. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017-1018.

18. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841-842.

19. Magoc, T., Wood, D. and Salzberg, S.L. (2013) EDGE-pro: Estimated degree of gene expression in prokaryotic genomes. *Evol Bioinform Online*, **9**, 127-136.

20. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol*, **11**, R106.

Table S1. Detailed description of SetCD ChIP-exo enrichment peaks and SetCD-binding motifs for all three ICEs and MGI*Vfl*Ind1.

ICE or MGI	MACS ChIP peak calls				SetCD-binding motifs			
	Shared	Peak	Summit	Peak	Position	Orientation	*p*-value	Oriented
SXT	I	251-1311	615	12017	547-566	Positive	2.9E-10	*xis*
	II	3001-4181	3425	16467	3450-3469	Negative	1.9E-08	*s003*
					3470-3489	Positive	4.9E-08	*mobI*
	-	20851-21580	21150	2930	-	-	-	-
	III	23901-25131	24485	6683	-	-	-	-
	-	-	-	-	45100-45119[a]	Positive[a]	6.9E-07[a]	*traI*
	IV	46801-47711	47165	6615	47112-47131	Positive	6.3E-08	*traD*
	V	50641-51651	51155	2807	51103-51122	Positive	7.4E-09	*traL*
	VI	53911-55451	54335	15008	54243-54262	Negative	1.0E-06	-
					54263-54282	Positive	1.4E-08	*traV*
	VII	56951-57701	57305	3826	57284-57303	Positive	4.4E-08	*dsbC*
	VIII	71311-72581	71735	28440	71666-71685	Negative	2.0E-07	*s063*
					71686-71705	Positive	2.4E-07	*s089*
	IX	88291-89121	88745	8118	88695-88714	Positive	9.3E-09	*traF*
R391	I	2261-2751	2525	5772	2457-2476	Positive	1.5E-09	*xis*
	II	4991-6441	5455	13262	5360-5379	Negative	2.1E-08	*s003*
					5380-5399	Positive	2.6E-08	*mobI*
	III	7881-8291	8105	4496	-	-	-	-
		8301-8971	8495	3410	8409-8428	Negative	1.3E-06	*rumA*
	-	-	-	-	32275-32294[a]	Positive[a]	9.5E-07[a]	*traI*
	IV	34171-34891	34435	4033	34287-34306	Positive	8.8E-05	*traD*
	V	37901-39661	38295	4569	38376-38395	Positive	2.1E-09	*traL*
	VI	41211-42671	41695	15622	41536-41555	Positive	2.0E-07	*traV*
		42681-43221	42795	3211	-	-	-	-
	VII	49121-50141	49565	3062	49489-49508	Negative	1.1E-06	*cds55*
					49509-49528	Positive	6.4E-08	*dsbC*
	VIII	60491-62331	61465	27717	61366-61385	Negative	8.9E-07	*s063*
					61386-61405	Positive	8.1E-09	*s089*
	IX	74121-74881	74325	7398	74271-74290	Negative	1.6E-06	-
					74291-74310	Positive	3.7E-09	*traF*
ICE*Vfl*Ind1	I	71-1180	455	12511	404-423	Positive	4.1E-10	*xis*
	II	5311-6081	5565	12265	5577-5596	Negative	3.9E-08	*s003*
					5597-5616	Positive	6.5E-08	*mobI*
	III	25781-27201	26585	4534	-	-	-	-
	-	42781-43361	43115	2627	43063-43082[a]	Positive[a]	4.8E-04[a]	*traI*
					43338-43357	Positive	9.3E-06	
	IV	44901-45860	45115	3791	45074-45093	Positive	6.0E-08	*traD*
	V	52911-53771	53500	2674	53438-53457	Positive	2.3E-09	*traL*
	VI	56281-57711	56655	8473	56598-56617	Positive	3.0E-08	*traV*
	VII	59381-59901	59680	1944	59616-59635	Positive	6.5E-08	*dsbC*
	VIII	71021-72250	71495	20690	71473-71492	Negative	4.7E-07	*s063*
					71493-71512	Positive	9.4E-09	*s089*
	IX	91581-92650	92155	5248	92076-92095	Negative	4.0E-07	*s075*
					92096-92115	Positive	4.3E-09	*traF*
MGI*Vfl*Ind1	-	-	-	-	229-248[b]	Positive[b]	1.4E-04[b]	int_{MGI}
	-	4991-6771	5685	11440	5653-5672	Positive	7.0E-05	*cds4*
	-	17531-19201	18335	8408	17958-17977	Positive	1.6E-05	*cds15*
					18318-18337	Negative	9.3E-06	*rdfM*

[a] Motifs obtained by FIMO using the corresponding ICE SetCD-binding logo depicted in Fig. 1D.
[b] Motif obtained by FIMO using the ICE*Vfl*Ind1 SetCD-binding logo depicted in Fig. 1D.

Table S2. DNA sequences of the primers used in this study.

Primer name	Nucleotide sequence (5' to 3')	Use in the study
Int-CDboxF	CGAACAGGCAGATAGATGAAAGGATTTGTCACCTTTCTGCCCACC	Cloning of the SetCD box of P_{s003}
Int-CDboxR	CGGGTGGGCAGAAAGGTGACAAATCCTTTCATCTATCTGCCTGTT	Cloning of the SetCD box of P_{s003}
DATcdlacZ-FO3	GACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATCATATGAATATCCTCCTTAG	Replacement of the CRP box and -35 element of P_{lac} promoter by the SetCD box of P_{s003}
DATcdlacZ-R3	ACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGGCAGAAAGGTGACAA	Replacement of the CRP box of P_{lac} promoter by the SetCD box of P_{s003}
DATcdlacZ-R4	ACAATTCCACACAACATACGAGCCGGAAGCATAAAGGCGGTGGGCAGAAAGGTGACAA	Replacement of the CRP box and -35 of P_{lac} promoter by the SetCD box and -35 of P_{s003}
DATsetC3xflag-F	ATGGCCGAGCAAACCACTAGAGCACTTTTTGAGAACGGTGAGCTCGACTACAAAGA	C-terminal translational fusion of SPA tag with *setC* in SXT, R391 and ICE*Vfl*Ind1
DATsetC3xflag-SXT-R	GCACGGGCGGTGCACAATCAAATCATGTATCAGCATGTGTAGGCTGGAGCTGCTTC	C-terminal translational fusion of SPA tag with *setC* in SXT
DATsetC3xflag-R391-R	GCACGGGCGGTGCACAATCAAATCATGCATCAGCATGTGTAGGCTGGAGCTGCTTC	C-terminal translational fusion of SPA tag with *setC* in R391 and ICE*Vfl*Ind1
SPACtBamHI.for	GGGATCCTCCATGGAAAAGAGAAGATGGA	Construction of pNC12
SPACtBamHI.rev	GGGATCCCTACTTGTCATCGTCATCCTTG	Construction of pNC12
RRintR	ACGACGTTTGGCGTCTCGAT	Identification +1 of *s003*
PE-rdfM-2	TGGATGATTGTGTAGTTGGGT	Identification of +1 of *rdfM*
RTrpoZcoli-F	GCTCGTCAGATGCAGGTAGG	Amplification of *rpoZ*
RTrpoZcoli-R	GCTTGTAATTCAGCGGCTTC	Amplification of *rpoZ*
RTint-F	AGTTTTATGACTGGCTGATT	Amplification of *int*
RTint-R	TTTTTGACTCATGCAAATAG	Amplification of *int*
RTxis-F	TAACCCCTTTTCTAACATCT	Amplification of *xis*
RTxis-R	TTCTCATACTCCTCCAAATC	Amplification of *xis*
RTgene4Vf-F	GCAATGGCCAAACCTGTATT	Amplification of *cds4*
RTgene4Vf-R	GGCGTAATGCACCAAAAGAT	Amplification of *cds4*
CattPF	TTCGAAGGTTTAGCCACAGTTGTTTATGAGTG	Amplification of $attP_{SXT}$
CattPR	TTCGAATATTCCGCTTTTGTAATGTCGAAA	Amplification of $attP_{SXT}$
setD2WF	CTATCGCAATATTTTACATGAACTGGAGGTCCGTCAATTCCGGGGATCCGTCGACC	Deletion of *setCD* in SXT
setC2WR	CACGGGCGGTGCACAATCAAATCATGTATCAGCATGGTGTAGGCTGGAGCTGCTTCG	Deletion of *setCD* in SXT
VISRF	CTCTCATTAACTGGGTTCAGG	Amplification of *attR*
VISLR3	GCATTCTCCTGAAAATCAATG	Amplification of *attL*
EattBF	GCCGCACTTTTGCCATTATT	Amplification of *attB* and *attL*
prfCQR	TTGTTGGAGATTGGACGGAAC	Amplification of *attB*
EattBR	AGCAGCACCTTCTCGGTGAT	Amplification of *attR*
TrueSeq-B-dT24-VN[a]	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTTTVN	Second strand synthesis in ChIP-exo libraries preparation
IGA-A0-up[b,c]	/5AmMC6/ACACTCTTTCCCTACACGACGCTCTTCCGATCT	ChIP-exo libraries preparation
IGA-A0-down-T/A[c]	GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGAT	ChIP-exo libraries preparation
IGA-PCR-PE-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	ChIP-exo and RNA-seq libraries amplification
TrueSeq-MPEX-R[d]	CAAGCAGAAGACGGCATACGAGAT-index-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	ChIP-exo and RNA-seq libraries amplification, contains 6 bp barcode
TrueSeq-3'-hybrid-B0[b,e]	/5Phos/rArGrArUrCrGrGrArArGAGCACACGTCT/3AmMO/	RNA-seq libraries preparation
5'-hybrid-A0[e]	ACACGACGrCrUrCrUrUrCrCrGrArUrCrU	RNA-seq libraries preparation
DSN-TrueSeq-F	CACGACGCTCTTCCGATCT	RNA-seq libraries amplification
DSN-TrueSeq-R	AGACGTGTGCTCTTCCGATCT	RNA-seq libraries amplification

[a] Mixed bases: V: A, C or G; N: A, T, C or G.

[b] Nucleotide modifications: /5AmMC6/ : 5' amino modifier C6; /5Phos/ : 5' phosphorylation; /3AmMO/ : 3' amino modifier.

[c] Form the 3'-T-IGA-A0 adaptor when annealed together.

[d] Index corresponds to an unique combination of 6 nucleotides and serve as a barcode for sample multiplexing.

[e] r letter corresponds to RNA bases.

Table S3. Illumina libraries sequenced in this study.

Sequencing line	Library type	Experimental condition	Replicate	Index in TrueSeq-MPEX-R oligo	Index read by instrument	Sequenced paired-end reads (millions)
1	ChIP-exo	DPL492 (SXT *setC*^{3xFLAG}) +MC	1	AGTACC	GGTACT	3.1
			2	CAGACC	GGTCTG	3.0
			3	GCAACC	GGTTGC	2.8
		DPL491 (R391 *orf90*^{3xFLAG}) +MC	1	CCTTAC	GTAAGG	3.2
			2	GGCGAC	GTCGCC	4.2
			3	ATAGAC	GTCTAT	3.0
		DPL493 (ICE*Vfl*Ind1 *setC*^{3xFLAG} MGI*Vfl*Ind1) +MC	1	GATAAC	GTTATC	3.1
			2	ACGAAC	GTTCGT	3.3
			3	CTCAAC	GTTGAG	3.1
2	RNA-seq	DPL492 (SXT *setC*^{3xFLAG})	1	TGACCG	CGGTCA	4.5
			2	CCGCTC	GAGCGG	5.0
			3	GAGGTC	GACCTC	5.4
		DPL492 (SXT *setC*^{3xFLAG}) +MC	1	AGCCGC	GCGGCT	4.0
			2	CGGTTG	CAACCG	4.5
			3	GTAATT	AATTAC	4.9
		DPL3 (SXT Δ*setCD*) +MC	1	ATGACT	AGTCAT	5.3
			2	TCCAGC	GCTGGA	6.0
			3	CCTACT	AGTAGG	5.4
	5'-RACE	DPL492 (SXT *setC*^{3xFLAG})	1	GTTGAT	ATCAAC	9.0
		DPL492 (SXT *setC*^{3xFLAG}) +MC	1	GCGGCT	AGCCGC	7.1
		DPL3 (SXT Δ*setCD*) +MC	1	ACTCTT	AAGAGT	6.9

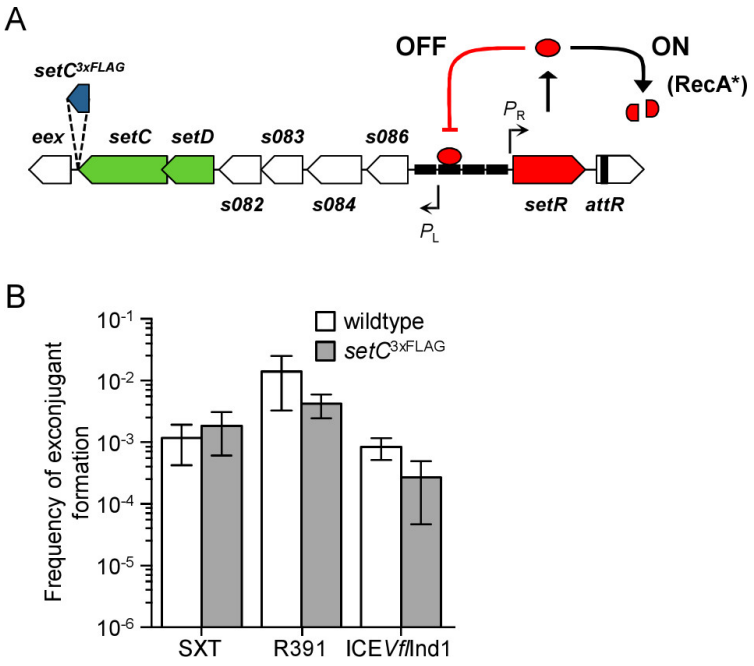

Figure S1. The bimodal switch governing the conjugative transfer of SXT/R391 ICEs and MGIs. **(A)** Schematic representation of the regulatory region of SXT/R931 ICEs. The fate of SetR proteins (red ellipse) governs the state of transfer (ON/OFF). Genes are represented by arrows and are color coded as follow: red, transcriptional repressor SetR; green, transcriptional activators SetC and SetD; white, other or unknown functions. 3xFLAG-labelling of the C-terminal coding sequence of *setC* is indicated in blue. **(B)** SetC^{3xFLAG} remains functional to activate the conjugative transfer of SXT/R391 ICEs. Conjugation assays using *E. coli* VB17 (BW25113 SXT), DPL513 (BW25113 R391), AD57 (CAG18439 ICE*Vfl*Ind1 MGI*Vfl*Ind1), and their *setC*^{3xFLAG} derivatives (DPL492, DPL491 and DPL493, respectively) as donors upon mitomycin C induction. The recipient strains were either *E. coli* CAG18439 or VB111 (MG1655 Nx^{r}). Results are the means and standard deviations of three independent biological replicates.

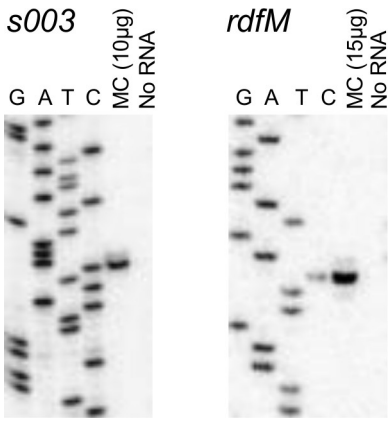

Figure S2. Primer extension analysis of the promoters of *s003* in SXT and *rdfM* in MGI*Vfl*Ind1. G, A, T and C, Sanger sequencing reaction lanes; MC, mitomycin C-induced samples; No RNA, negative control. Concentrations of RNA used for the primer extension reaction are indicated. Each well was loaded with equal reaction volumes.

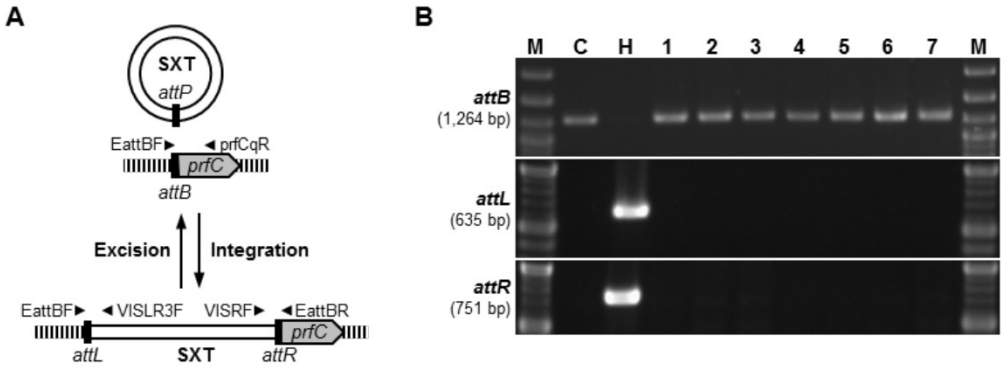

Figure S3. Atypical maintenance of SXT Δ*setCD* in fresh exconjugants. (A) Model of SXT integrating into and excising from the 5' end of *prfC* in a site-specific fashion. **(B)** PCR amplification of *attB*, *attL* and *attR* junction fragments. Ethidium bromide-stained 1% agarose gels of PCR products amplified using primer pairs EattBF/prfCqR (*attB*), EattBF/VISLR3 (*attL*) and VISRF/EattBR (*attR*) (Table S2). The position of the primers is indicated in panel A. Lanes: M, 2-Log DNA Ladder molecular size marker; C, CAG18439 (SXT- strain, negative control); H, HW220 (CAG18439 containing SXT integrated into the 5' end of *prfC*); 1 to 7, Nx^r Su^r Tm^r exconjugants resulting from a mating experiment between *E. coli* DPL5 and *E. coli* VB111 (mating *e* in Figure 5A).

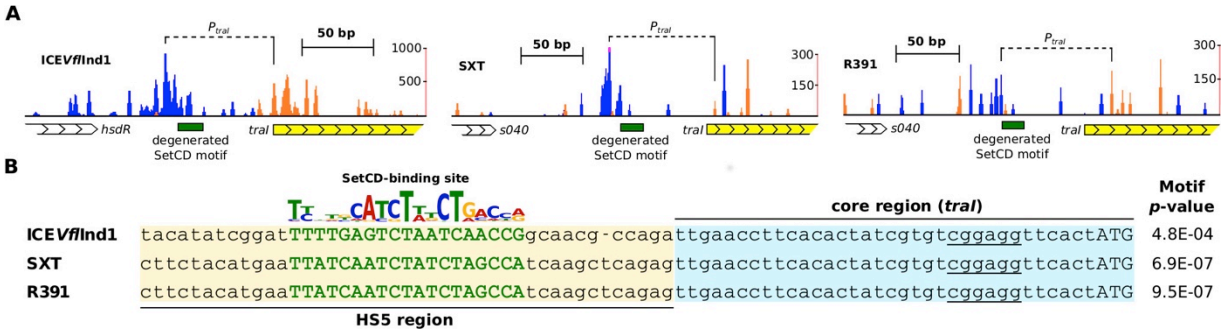

Figure S4. The *tral* promoter is disrupted by the insertion of variable DNA. (A) ChIP-exo binding profile of SetCD in the intergenic region upstream of *tral* in ICE*Vfl*Ind1, SXT and R391, respectively. Tracks are organized as in Figure 2A-F. The green boxes correspond to degenerate SetCD binding-motifs located on the positive DNA strand that were detected within ChIP-exo peaks by FIMO but not by MAST analysis. Genes are represented by arrows and are color coded as in Figure 1. **(B)** Alignment of the *tral* upstream sequences showing the proposed SetCD-binding motifs in all three ICEs. The blue shaded sequences correspond to identical sequences shared by the three ICEs (core). The orange shaded sequences correspond to variable DNA (HS5) that is shared by SXT and R391 but different in ICE*Vfl*Ind1. The degenerate SetCD-binding motifs found in panel A are represented in bold green capital letters with their respective *p*-value. The combined SetCD logo is as in Figure 1D. Shine-Dalgarno sequences are underlined while start codons are in capital letters.

CHAPITRE 3

DISCUSSION GÉNÉRALE ET CONCLUSION

Ce projet doctoral a permis l'identification de CroS, un régulateur transcriptionnel encodé par les ICE SXT/R391 qui, avec SetR, constituent un interrupteur génétique gouvernant la transition de l'état quiescent vers le transfert conjugatif de ces éléments. De plus, un cinquième site opérateur de SetR (et de CroS) a été mis en évidence et participe à la répression du transcrit contenant les gènes *croS*, *setC* et *setD*. Ces deux derniers codent pour l'activateur principal des ICE SXT/R391, SetCD, dont le régulon a été établi à l'échelle génomique. Nos travaux ont également montré une boucle de rétroaction positive par SetCD qui implique la réplication en cercle-roulant de l'ICE lors du transfert. Finalement, ces travaux ont permis d'éclaircir des évidences conflictuelles concernant l'activation de l'expression de l'intégrase par SetCD dans la cellule réceptrice. Ce projet a donc grandement amélioré notre compréhension du système de régulation régissant cette famille d'ICE, renommée comme étant d'importants vecteurs de dissémination de résistance aux antibiotiques.

3.2. Répresseur SetR

3.2.1. Reconnaissance de l'ADN par SetR

Le répresseur SetR maintient les ICE SXT/R391 dans leur état quiescent et intégré dans le chromosome en liant ses sites opérateurs situés dans le module de

régulation. Ses fonctions d'autorégulation et de répression de *setCD* ont fait l'objet d'études (Beaber and Waldor, 2004; Beaber *et al.*, 2002, 2004), mais les mécanismes précis de régulation par SetR sont toujours inconnus. La recherche approfondie sur les mécanismes de régulation utilisés par les répresseurs de type CI permet d'avancer certaines hypothèses concernant la structure (Figure 17A) et les mécanismes utilisés par SetR. Ces déductions permettent de mieux cerner le rôle de CroS et du site opérateur *O4*, des composantes additionnelles de régulation mise en évidence par les travaux de ce projet de doctorat.

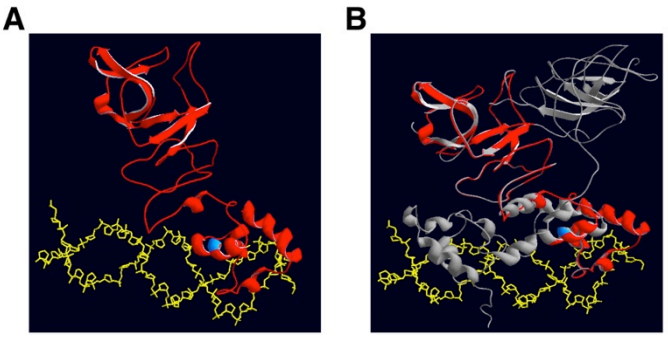

Figure 17. Structure putative du répresseur SetR. (A) La structure en ruban 3D de SetR a été modélisée par Phyre2 (Kelley and Sternberg, 2009) à partir du gabarit c3bdnB (CI du bactériophage λ) avec 100 % de confiance, 96 % de couverture et 31 % d'identité. L'ADN est en jaune et le monomère de SetR est en rouge. Le site de clivage G49 est en bleu. Dimensions du modèle (Å): X,38.526; Y,42.839; Z,75.645. **(B)** Superposition des surfaces des PBD de SetR (rouge) et λ CI (gris). La représentation montre un monomère de SetR et un dimère de λ CI tel que cristallisé lorsque lié à l'ADN (Stayrook *et al.*, 2008).

La liaison de CI à ses cibles est médiée par le positionnement des hélices de reconnaissance d'un dimère de CI (une par sous-unité monomérique) dans deux sillons majeurs consécutifs de l'ADN. Cette configuration permet l'interaction des chaînes latérales des hélices avec les 4 à 5 pb bordant l'opérateur (Figure 17B)

(Aggarwal *et al.*, 1988; Anderson *et al.*, 1987; Harrison and Aggarwal, 1990). L'interface N-terminale des dimères de CI détermine la préférence pour les bases centrales, ce qui définit l'affinité spécifique du répresseur pour chacun de ses sites opérateurs (Bell and Koudelka, 1995; Harrison and Aggarwal, 1990; Koudelka, 1998; Koudelka and Carlson, 1992; Mauro *et al.*, 2003). L'affinité différentielle entre les sites opérateurs est cruciale pour la régulation exercée par les répresseurs de type CI (Koudelka and Lam, 1993; Little *et al.*, 1999; Ptashne, 2004). L'affinité relative observée de SetR pour ses opérateurs est *O1>O4>O3>OL>O2* (Beaber and Waldor, 2004; Poulin-Laprade and Burrus, 2015). Les opérateurs de SetR sont partiellement symétriques et sont séparés par des espaceurs riches en adénosine et en thymine (Figure 18A) (Beaber and Waldor, 2004).

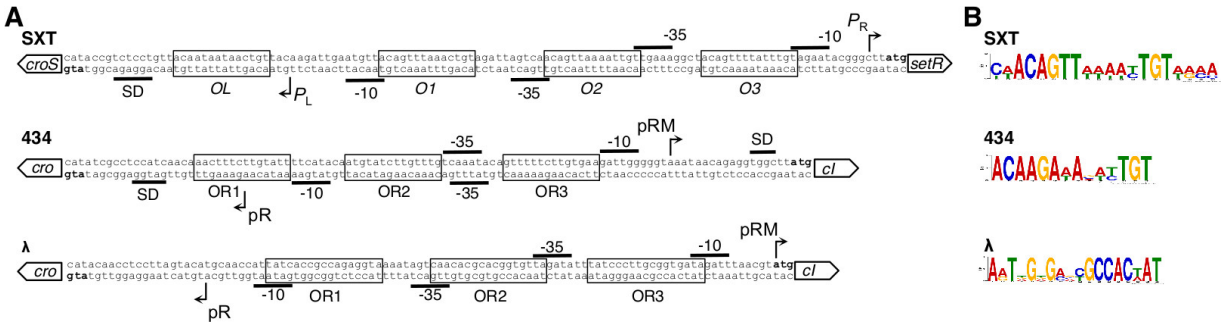

Figure 18. Comparaison de la région régulatrice de SXT avec les opérateurs OR des bactériophages λ et 434. (A) Séquences nucléotidiques de la région régulatrice de SXT et des opérateurs OR des bactériophages 434 et λ (Beaber and Waldor, 2004; Bushman, 1993; Schubert *et al.*, 2007). Les sites opérateurs sont indiqués par des rectangles. Les séquences Shine-Dalgarno (SD), les promoteurs et les boîtes -10 et -35 sont indiqués en aval des gènes *setR*, *croS*, *cI* et *cro* qui sont représentés par des flèches. Les codons d'initiation sont en gras. **(B)** Séquences logo générées par l'outil *Multiple Em for Motif Elicitation* (Bailey and Elkan, 1994) pour les opérateurs de SetR et CroS chez SXT, et ceux de CI et Cro chez 434 et λ.

Le motif d'ADN reconnu par SetR est plus similaire à celui de 434 CI que celui reconnu par λ CI (Figure 18B). Ceci s'explique par l'homologie plus importante entre

les séquences protéiques de SetR et 434 CI que celle entre SetR et λ CI (Aggarwal *et al.*, 1988; Beaber and Waldor, 2004; Bushman, 1993; Pirrotta, 1979).

Les espacements entre les sites opérateurs consécutifs de SetR *O1*-*O2* et *O2*-*O3* sont respectivement de 24 et 23 pb (Figure 18A). Cette distance est idéale pour la liaison coopérative entre deux répresseurs liés à des sites adjacents étant donné qu'ils sont localisés sur la même face de l'hélice d'ADN (Mao *et al.*, 1994). Il est tentant d'assumer que *O1*-*O2* est l'organisation favorisée lorsque le niveau cellulaire de SetR est bas puisque SetR a une affinité plus forte pour *O1* que pour *O3* (Beaber and Waldor, 2004; Poulin-Laprade and Burrus, 2015). L'espacement entre *OL* et *O1* est de 30 pb, ce qui suggère que la liaison de SetR à *O1* ne peut pas mener à la liaison coopérative à *OL*, l'orientation des dimères autour de l'hélice ne le permettant pas. Les avantages conférés par la coopérativité entre dimères de CI situés sur des sites adjacents sont l'augmentation de la sensibilité et de la stabilité de l'interrupteur génique qui détermine la voie de développement du bactériophage (Ptashne, 2004).

3.2.2. Mécanisme de régulation par SetR

Malgré les similitudes entre les systèmes de régulation des bactériophages lambdoïdes, les détails des mécanismes utilisés par les répresseurs de type CI diffèrent parfois entre les phages. Par exemple, la position relative des éléments des promoteurs pR et pRM par rapport aux sites de l'opérateur droit (OR) est différente chez λ et 434 (Figure 18A). Ces différences entraînent des divergences dans l'encombrement stérique entre CI et l'holoenzyme, et dans les mécanismes d'interférence entre promoteurs adjacents. La position relative des boîtes -35 des promoteurs P_L et P_R par rapport à l'opérateur *O2* est plus apparentée aux équivalents

de λ qu'à ceux de 434 dans lequel les boîtes -35 de pR et pRM se chevauchent et sont positionnées entre OR2 et OR3 (Figure 18A). Ceci suggère que les mécanismes de régulation encourus par SetR sont plus similaires à ceux de λ CI qu'à ceux de 434 CI, bien que la séquence protéique de SetR soit plus proche de celle de 434 CI que de celle de λ CI.

3.2.2.1. Régulation du promoteur P_R

Le facteur σ qui médie la spécificité de la RNAP envers les promoteurs P_L et P_R de SXT est σ^{70}, le facteur σ permettant la reconnaissance des promoteurs de gènes communs chez *E. coli*. Les boîtes -10 et -35 du promoteur P_L sont plus similaires aux motifs consensuels reconnus par le facteur σ^{70} qu'à celles du promoteur P_R. L'autoactivation par SetR est donc importante pour contrebalancer la force du promoteur P_L et pour assurer la répression du système en absence d'induction. Un dimère de λ CI lié au site OR2 interagit directement avec le facteur σ^{70} de la RNAP lié au promoteur pRM (Li *et al.*, 1994) et stimule l'isomérisation du complexe promoteur-RNAP vers un complexe ouvert compétent pour la transcription de *cI* (Ptashne, 2004; Ptashne *et al.*, 1976). L'autoactivation par SetR a été suggérée, mais n'a jamais été observée. L'interaction éventuelle de SetR avec la RNAP n'a pas non plus été investiguée.

L'autorépression par SetR, quant à elle, a été observée (Beaber and Waldor, 2004) et est probablement médiée par la liaison de SetR à *O3*, comme ce qui est observé chez les équivalents de λ et 434 (Koudelka and Lam, 1993; Maurer *et al.*, 1980). Lorsque le niveau cellulaire de SetR est bas, SetR serait lié seulement à ses sites de plus hautes affinités, *O1*-*O2*, ce qui stimulerait son autoactivation en recrutant

l'holoenzyme. Lorsque le niveau de SetR excède une certaine limite, SetR se lierait plus fréquemment à son site de moindre affinité *O3*, ce qui mènerait a un encombrement stérique empêchant la liaison de la RNAP-σ^{70} à P_R.

3.2.2.2. Répression de P_L

La répression de pR (l'équivalent de P_L chez SXT) procède différemment chez λ et 434 (Xu and Koudelka, 2001). La liaison de λ CI aux opérateurs OR1 et OR2 exclut la liaison de l'holoenzyme à pR (Meyer *et al.*, 1980), tandis que la liaison de 434 CI aux opérateurs OR1 et OR2 n'exclut pas la liaison de l'holoenzyme à pR, mais bloque la transition du complexe holoenzyme-pR fermé vers un complexe ouvert compétent pour la transcription (Xu and Koudelka, 2001). Ceci suggère que la répression de *setCD* par SetR est médiée en bloquant l'accès de l'holoenzyme au promoteur P_L, probablement en étant lié coopérativement aux opérateurs *O1*-*O2*, comme ce qui a été établi pour λ (Meyer *et al.*, 1980). Pour confirmer que SetR agit de façon similaire à λ CI, une expérience montrant que SetR et l'holoenzyme s'excluent serait nécessaire. La liaison de SetR aux opérateurs *OL* et *O4* pourrait aussi contribuer au blocage de l'élongation du transcrit contenant *croS* et *setCD*.

3.2.2.3. *O4* participe à la répression de P_L

L'addition d'une boucle dans l'ADN peut conférer plusieurs avantages à des réseaux de régulation, par exemple l'augmentation de la stabilité du système, de son inductibilité ainsi que l'augmentation de la concentration locale en répresseurs (Cournac and Plumbridge, 2013; Morelli *et al.*, 2009). L'article présenté à la section

2.1. montre, entre autres, que la présence du site *O4*, localisé dans le cadre de lecture de *s084*, contribue à la répression du transcrit qui débute à P_L et qui contient *croS* et *setCD* (Poulin-Laprade and Burrus, 2015). Ceci est en accord avec la régulation de λ si l'on considère que *O4* de SXT est l'équivalent de l'opérateur gauche (OL) de λ qui est composé de trois sites distincts (OL1, OL2 et OL3). Il est intéressant de noter que la boucle ne serait pas dans la même orientation chez λ et SXT (section 2.1. Figure 4A). Deux hypothèses non exclusives pourraient expliquer l'effet répresseur de la disponibilité du site *O4*. Premièrement, la liaison de SetR ou CroS à *O4* pourrait créer un obstacle à l'élongation du transcrit P_L par l'holoenzyme. Deuxièmement, SetR pourrait s'associer en multimère pour lier simultanément *O4* et *O1-O2*, les sites de hautes affinités localisés 800 pb plus loin. La délétion de l'opérateur gauche chez λ (OL) mène à une répression plus forte de λ *cro* que l'effet sur l'expression de *croS* de la délétion de *O4* chez SXT, suggérant que le complexe OL1-OL2-CI_8-OR1-OR2 est plus stable que le complexe *O4*-SetR-*O1-O2*. Ceci s'explique facilement par la liaison coopérative de quatre dimères de λ CI, la configuration optimale pour les répresseurs de type CI (Dodd *et al.*, 2001). L'absence d'un site opérateur adjacent à *O4* pourrait déstabiliser l'oligomérisation de SetR et la formation subséquente de la boucle. Cependant, l'espacement entre *O4* et *O1-O2* est plus court que celui observé entre les régions OL et OR de λ, ce qui pourrait contribuer à la formation d'une boucle soutenue par SetR (Priest *et al.*, 2014). CroS n'a pas le domaine C-terminal qui permet l'oligomérisation, la liaison coopérative de l'ADN et la compétence subséquente pour former une boucle (Bell and Lewis, 2001; Bell *et al.*, 2000).

L'implication du site *O4* dans la régulation du promoteur P_R n'a pas été vérifiée. Chez λ, la boucle générée par la liaison d'un octamère de CI aux opérateurs OL et OR mène à une stimulation du promoteur pRM de deux fois si le site OR3 est libre (Anderson and Yang, 2008). Une autre étude chez λ a montré que la sous-unité α de

la RNAP liée à pRM interagit avec un élément UP situé dans la région de l'opérateur gauche (OL), à quelque 2 kb plus loin (Cui *et al.*, 2013). Cette interaction à longue distance favorise la transcription à partir du promoteur pRM, ce qui rappelle les éléments *enhancers* rencontrés chez les eucaryotes. Aucune évidence d'un tel élément UP n'est observée à proximité du site *O4*, l'équivalent chez SXT de la région OL. La formation d'une boucle entre les opérateurs OL et OR augmente aussi la concentration locale de CI, ce qui favorise sa liaison à OR3 et mène à son autorépression à partir de pRM (Dodd, 2004; Dodd *et al.*, 2001).

3.3. Répresseur CroS

3.3.1. Répresseur de type Cro

L'homologue structurel le plus apparenté de CroS est Xfaso I, un répresseur de type Cro dont la structure a été déterminée (Roessler *et al.*, 2008). Les évidences expérimentales concernant la fonction ou l'expression de Xfaso I sont très limitées (Ahern *et al.*, 2014; de Mello Varani *et al.*, 2008; Summer *et al.*, 2010). Les prédictions de structures secondaires suggèrent que Xfaso I et CroS sont entièrement composés d'hélices α (Poulin-Laprade and Burrus, 2015). Cette structure est typique du répresseur Cro du bactériophage P22 qui a été initialement isolé chez *Salmonella* (Newlove *et al.*, 2004), mais différente de la structure de λ Cro qui montre un mélange de feuillets β et d'hélices α (Roessler *et al.*, 2008). Le réseau de régulation de P22 diffère de celui de λ, car il code pour des répresseurs additionnels qui contribuent au maintien de l'état lysogène (Susskind and Botstein, 1978). Malgré leurs différences structurelles, les protéines Cro de P22 et de λ ont des fonctions très similaires concernant la régulation des promoteurs pR et pRM (Poteete and Ptashne,

1982; Poteete *et al.*, 1986). En effet, une affinité opposée est observée chez les Cro (OR3 > OR2 > OR1) et leur répresseur CI (OR1 > OR2 > OR3) respectif (Ptashne, 2004). De plus, la variation de l'affinité des Cro pour leurs opérateurs est plus faible que celle des CI (Poteete *et al.*, 1986). Des essais de gel à retardement présentés à l'article de la section 2.1. ont montré que l'affinité relative de CroS pour ses opérateurs va comme suit : *O3* > *OL* > *O4* > *O1* ≈ *O2*. La confirmation de ces résultats par d'autres expériences est nécessaire.

L'étude de nombreuses versions du réseau de régulation lambdoïde a mené à la conclusion que la plupart des composantes du réseau sont optionnelles, mais qu'ensemble elles ont un effet synergique qui contribue à l'efficacité globale et à la stabilité du système (Little, 2010). Nous avons montré dans l'article présenté à la section 2.1. que CroS est particulièrement important pour la régulation de SXT étant donné que trois composantes importantes pour la réduction du bruit de fond et pour la stabilisation du système λ sont absentes chez SXT, soit l'activateur CII et les équivalents des sites OL2 et OL3 (Avlund *et al.*, 2010; Morelli *et al.*, 2009).

3.3.2. Mécanisme de régulation de CroS

En absence d'induction, en moyenne 1 cellule sur 10 000 agit comme donneuse, ce qui suggère que, dans la plupart des cellules de la population, la quantité d'ARNm de *croS* est plus faible que celle contenant *setR* (Poulin-Laprade and Burrus, 2015). CroS est présumément dominant dans la faible fraction de la population bactérienne qui est en transition entre l'état quiescent et l'état conjugatif de SXT. L'article de la section 2.1. a établi que CroS est essentiel à l'induction SOS de SXT, ce qui est en accord avec la régulation exercée par *cro* chez λ (Schubert *et al.*, 2007). Cette

fonction est médiée par la répression des promoteurs P_L et P_R lorsque la répression par SetR est levée. CroS retarderait la recrudescence du niveau de SetR et déplacerait l'interrupteur génétique vers l'expression à partir de P_L assez longtemps pour que le transfert conjugatif procède correctement. La répression de P_L par CroS préviendrait également un engouement de l'expression de *setCD*. Étant donné que les séquences consensuelles des boîtes -10 et -35 du promoteur P_L sont plus apparentées à σ^{70} qu'à celles de P_R, la levée de la répression par SetR serait suffisante pour la reconnaissance par l'holoenzyme. L'action d'un activateur ne serait donc pas requise pour l'isomérisation d'un complexe ouvert compétent pour la transcription à partir de P_L. Ce modèle s'apparente à la régulation des promoteurs pR et pRM de λ (Ptashne, 2004; Strainic *et al.*, 2000).

Un phénomène d'immunité prévenant la surinfection d'une bactérie lysogène a été observé dès les balbutiements de l'étude de la régulation de λ (Ptashne, 1967). Les travaux de Peng Luo, un chercheur invité au laboratoire en 2014, montrent que la présence de SetR dans une cellule réceptrice exerce une immunité contre un ICE entrant apparenté, malgré la compatibilité de leur groupe d'exclusion (Luo et Burrus, communication personnelle). L'immunité contre un ICE entrant est 60 fois plus importante dans une souche qui porte une copie chromosomique de $setR_{A056\text{-}2}$ exprimée à partir de son promoteur natif que dans une souche qui porte l'ICE complet ICE*Val*A056-2. Ces résultats montrent que SetR est le facteur responsable de cette immunité, mais qu'un ou des facteurs présents dans l'ICE complet inhibent partiellement la fonction d'immunité. CroS est un bon candidat pour cette inhibition de l'immunité étant donné que les résultats de ce projet de doctorat ont montré que CroS réprime P_R, le promoteur de *setR* (Poulin-Laprade and Burrus, 2015).

3.4. Activation par SetCD

Une autre partie du projet de doctorat consistait à raffiner notre compréhension du complexe activateur SetCD, principalement en identifiant l'ensemble de ses promoteurs cibles dans trois ICE SXT/R391, dans un MGI dépendant de cette famille d'ICE ainsi que dans le génome de *E. coli* (Poulin-Laprade *et al.*, 2015a). Ces résultats sont présentés dans l'article de la section 2.2. et confirment la liaison de SetCD aux promoteurs qui conduisent l'activation des gènes impliqués dans :

1) la recombinaison site-spécifique et la partition (P_{s003}, P_{xis});
2) la recombinaison homologue indépendante de RecA (P_{s089});
3) la machinerie associée au transfert conjugatif (P_{mobI}, $P_{traI,}$ P_{traD}, $P_{traL,}$ P_{traV}, $P_{dsbc,}$ P_{s063}, P_{traF});
4) et la mobilisation des MGI (P_{intM}, P_{rdfM}).

Ce profil de régulation est similaire à celui du complexe activateur AcaCD codé par les plasmides IncA/C (Carraro *et al.*, 2014; Poulin-Laprade *et al.*, 2015b). En effet, ce projet de doctorat a influencé l'étude parallèle du complexe activateur AcaCD par Nicolas Carraro, un stagiaire postdoctoral aussi sous la supervision de Vincent Burrus. Nos découvertes ont fait l'objet d'un article de revue où nous comparons extensivement les deux systèmes (Annexe 4). Le mécanisme d'activation employé par SetCD et AcaCD ressemble à celui de FlhCD (Carraro *et al.*, 2014; Poulin-Laprade *et al.*, 2015a, 2015b). Ces complexes activateurs reconnaissent des promoteurs de classe II qui sont définis par la position du site de liaison de l'activateur qui chevauche une boîte -35 dégénérée (Lee *et al.*, 2012). Notez que ce

mécanisme d'activation semble également employé par SetR pour l'activation de P_R (et par λ CI pour l'activation de pRM, Figure 18A).

L'équipe de O'Halloran *et al.* (2007) a suggéré un site de liaison de SetCD en amont de *xis* chez R391 qui ressemble grandement au motif reconnu par FlhCD. Cependant, ils ont introduit des erreurs dans le promoteur de *xis* chez R391. Premièrement, ils ont déplacé de 51 nucléotides en aval le codon d'initiation de *xis*, qui avait été établi en 2002 par l'équipe de Waldor (Beaber *et al*., 2002) (Annexe 6). Deuxièmement, ils ont introduit 5 nucléotides, dont la séquence ne se retrouve en amont d'aucun gène *xis* de cette famille d'ICE, y compris R391, qui correspondent exactement à une portion de la séquence consensus reconnue par FlhCD (Annexe 6). Le motif de reconnaissance de SetCD a été établi expérimentalement (Poulin-Laprade *et al.*, 2015a) et diffère drastiquement des motifs reconnus par AcaCD (Carraro *et al.*, 2014) et FlhCD (Fitzgerald *et al.*, 2014; Lee *et al.*, 2011; Stafford *et al.*, 2005; Wozniak and Hughes, 2008). Ceci n'est pas surprenant considérant la très faible similarité entre SetCD, FlhCD et AcaCD (Beaber *et al.*, 2002; Carraro *et al.*, 2014). Une étude a montré que l'activation par FlhCD est causée par son interaction avec la région C-terminale de la sous-unité α de la RNAP (Liu *et al.*, 1995). La caractérisation biochimique de SetCD et AcaCD est requise pour vérifier si leur rôle activateur est médié par un contact direct avec l'holoenzyme.

Le motif de reconnaissance de SetCD établi expérimentalement pourrait servir à trouver d'autres cibles de SetCD dans le génome de ses bactéries hôtes. Ceci a été effectué avec le motif reconnu par AcaCD et a permis l'identification de cibles chromosomiques potentielles dans le génome de nombreuses *Enterobacteriaceae* et *Vibrionaceae* (Carraro *et al.*, 2014, 2015b). Plusieurs sites ont été retrouvés chez SGI1 (*Salmonella genomic island 1*) qui confère la multirésistance à des souches

pathogènes de *S. enterica* ainsi que chez d'autres îlots génomiques qui peuvent être mobilisés par les plasmides IncA/C.

3.4.1. Structure du complexe SetCD

La structure de SetCD est toujours inconnue. De nombreuses évidences génétiques supportent l'hypothèse que les produits des gènes *setD-setC* et *acaD-acaC* s'associent pour former les complexes protéiques hétéromériques appelés respectivement SetCD et AcaCD (Beaber *et al.*, 2002; Carraro *et al.*, 2014). Les travaux de Carraro *et al.* (2014) ont montré que la sous-unité AcaD co-précipite avec une sous-unité AcaC marquée d'une étiquette 6xHis. Ce résultat supporte la formation de complexes hétéromériques comme celui observé pour FlhCD (Wang *et al.*, 2006). Les travaux de Wang *et al.* (2006) ont également mis en évidence la liaison d'un ion zinc par quatre résidus cystéines de FlhC (Cys137, Cys140, Cys157 et Cys160) qui sont conservés chez SetCD et AcaCD (Figure 14 et Figure 19). Ceci suggère que ces résidus et leurs interactions potentielles avec un ion zinc pourraient contribuer à la compétence de SetCD et AcaCD à se lier à leurs séquences cibles dans l'ADN.

```
SetC  RSQLASLHRCR-CGSLYLTVSQQRI-QLKCPVCEIMAEQTTR
AcaC  RSGDAMLEVCDNCKCTYFTSVNQRT-CVECPFCKEQGRHGGG
FlhC  ESGLLQLSSCNCCGGNFITHAHQPVGSFACSLCQPPSRAVKR
```

Figure 19. Alignement de SetC et ses homologues. Alignement généré avec Clustal Ω d'une portion de la région C-terminale des séquences protéique de SetC (de SXT), de AcaC (de pVCR94ΔX) et de FlhC (d'*E. coli* K12 MG1655). Les résidus en gras indiquent les cystéines présumément responsables de l'interaction avec des ions Zn (Wang *et al.*, 2006) qui sont conservés chez SetC et AcaC.

3.4.2. SetCD, un régulateur pléiotropique

3.4.2.1. Effets secondaires de la surexpression de SetCD

SetCD active plusieurs gènes dont les produits sont potentiellement délétères et/ou mutagènes, par exemple le gène *traV* qui code pour une composante du pore de conjugaison et les systèmes de recombinaison qui sont conservés chez les ICE SXT/R391 (Armshaw and Pembroke, 2015; Beaber *et al.*, 2002; Burrus and Waldor, 2003; Poulin-Laprade *et al.*, 2015a). L'activation par SetCD peut indirectement mener à l'expression de gènes des régions variables dont nombre d'entres eux peuvent entraîner des réarrangements chromosomiques comme les transposases. La surexpression de SetCD augmente aussi de manière artificielle l'expression de TraI, la relaxase qui prépare l'ADN pour le transfert et qui agit comme facteur d'initiation de la réplication (Carraro *et al.*, 2015a; Lee *et al.*, 2010a). La surexpression de SetCD augmente la proportion de la population chez qui l'ICE s'excise et se réplique, ce qui a pour effet d'augmenter le nombre de copies de l'ICE observées dans la population. L'article présenté à la section 2.1. montre qu'il y a donc plus d'ARNm qui portent la séquence codante des régulateurs, ce qui occasionne une rétroaction positive sur l'expression de SetCD et exacerbe sa surexpression (Poulin-Laprade and Burrus, 2015). Ce phénomène de rétroaction positive est particulièrement important chez R391 qui peut posséder jusqu'à 20 copies de l'ICE excisé (Carraro *et al.*, 2015a), ce qui est quatre fois plus important que chez SXT (Burrus and Waldor, 2004b; Poulin-Laprade *et al.*, 2015a).

3.4.2.2. Effets secondaires de la délétion de *setCD*

La délétion de *setCD* abolit le transfert de SXT (Beaber *et al.*, 2002). L'article présenté à la section 2.2. démontre que l'expression de l'intégrase et de l'excisionase de SXT requiert la présence de SetCD dans la cellule donneuse et dans la cellule réceptrice (Poulin-Laprade *et al.*, 2015a). En effet, un mutant SXT Δ*setCD* complémenté seulement dans la donneuse est incapable de s'intégrer de manière site-spécifique dans l'extrémité 5' du gène *prfC* de la réceptrice. Cette observation suggère fortement que ni SetCD ni l'intégrase ne sont transloqués dans la réceptrice durant la conjugaison. Alternativement, *setCD* est exprimé *de novo* dans la réceptrice, permettant l'expression de *int* et l'intégration site-spécifique subséquente. Ce besoin de SetCD dans la cellule réceptrice contraste avec la régulation de l'intégration des MGI dont la mobilisation est dépendante des ICE SXT/R391. En effet, l'excisionase de ceux-ci nécessite l'activation par SetCD dans la donneuse, mais leur intégrase est exprimée constitutivement, ce qui permet l'intégration site-spécifique d'un MGI dans une cellule réceptrice qui ne possède pas d'ICE mobilisateur. Étant donné que *int* et *setCD* sont physiquement liés sur les ICE SXT/R391, ces éléments n'ont pas besoin d'exprimer constitutivement leur intégrase pour assurer leur maintien et leur dissémination.

3.4.3. Cibles atypiques de SetCD

Les résultats de ChIP-exo présentés dans l'article de la section 2.2. ont montré des enrichissements en séquences liées par SetCD à proximité des boîtes -35 de nombreux promoteurs présents chez les ICE SXT/R391 et chez leurs MGI associés. Ces cibles dites typiques ont un enrichissement très localisé et retrouvé dans le

promoteur d'un gène en amont d'un site d'initiation de la transcription établi dans cette même étude par 5' RACE (*5' rapid amplification of cDNA ends*) ou par extension d'amorce (Poulin-Laprade *et al.*, 2015a).

Ces expériences à l'échelle génomique ont également permis d'observer trois occurrences d'enrichissement atypiques dans le génome de *E. coli* :

1) *yfiA* code une protéine qui s'associe à la sous-unité 30S du ribosome et qui est impliquée dans l'adaptation au stress environnemental. Ce gène est exprimé en phase stationnaire ou lors de choc froid (Agafonov and Spirin, 2004; Vila-Sanjurjo *et al.*, 2004);
2) *ssrS* code un ARN 6S qui interagit avec l'holoenzyme et inhibe son activité à certains promoteurs lorsqu'il y a une limitation en nutriments (Trotochaud and Wassarman, 2004; Wassarman and Saecker, 2006);
3) *ychH* code pour une protéine qui est impliquée dans la réponse à de nombreux stress cellulaires, par exemple la résistance au cadmium, au H_2O_2 et à l'acide. Cette protéine serait aussi importante pour la formation de biofilm et pour la dégradation du *cis*-1,2-dichloroéthylène (Lee *et al.*, 2010b).

Ces cibles dites atypiques ont leur ORF recouvert en entier par SetCD, ce qui suggère un rôle répresseur par blocage de l'élongation. Des expériences sont nécessaires pour confirmer cette fonction présumée de SetCD et son implication dans la biologie des bactéries hôtes. Il serait intéressant de répéter l'expérience de ChIP-exo en utilisant d'autres hôtes des ICE SXT/R391 comme *V. cholerae*, *Providencia rettgeri* et *Proteus mirabilis*. Il serait alors possible d'observer si les cibles dans le génome d'*E. coli* sont conservées chez les autres hôtes, ce qui pourrait

contribuer à évaluer l'interférence des ICE SXT/R391 sur la régulation des processus bactériens de réponse au stress.

3.4.4. Diversité des régulateurs de type FlhCD

La recherche de régulateurs de type FlhCD additionnels parmi d'autres éléments génétiques mobiles a été effectuée (Annexe 4) (Poulin-Laprade *et al.*, 2015b). Étant donné que l'homologie entre SetCD, AcaCD et FlhCD est très faible, de nouveaux profils HMM (*hidden Markov models*) basés sur l'alignement de la séquence primaire de SetC/AcaC et SetD/AcaD ont été générés. La base de données des séquences protéiques non redondante de Genbank a été criblée et a révélé plusieurs protéines homologues codées par plusieurs types d'éléments génétiques mobiles d'*Enterobacteriaceae* et de *Vibrionaceae*. Fait intéressant, ces orthologues sont retrouvés dans des contextes génétiques similaires (Annexe 4) (Poulin-Laprade *et al.*, 2015b). Des expériences sont nécessaires pour évaluer leur rôle dans la régulation de leurs éléments respectifs ainsi que leur impact potentiel sur d'autres éléments mobiles.

3.5. Comparaisons avec les systèmes de régulation d'autres familles d'ICE

Les ICE ont en commun leurs fonctions d'excision/intégration et de transferts conjugatifs (Burrus *et al.*, 2002). Cependant, les systèmes de régulation codés par les différentes familles d'ICE divergent autant par les régulateurs qu'ils codent et leurs mécanismes d'action que par leurs signaux inducteurs. La majorité des régulateurs sont des facteurs de transcription protéiques, mais des petits ARN ont été décrits, par

exemple l'ARN antisens dirigé contre *orf9* qui code pour un répresseur de Tn*916* (Figure 20) (Celli and Trieu-Cuot, 1998). Cette diversité des systèmes de régulation a été traitée dans plusieurs ouvrages (Carraro and Burrus, 2014; Roberts and Mullany, 2009; Waters and Salyers, 2013; Wozniak and Waldor, 2010). L'induction du transfert conjugatif est généralement étroitement liée à un état de stress chez l'hôte bactérien. En voici quelques exemples :

1) de l'ADN endommagé (ex. SXT/R391, ICE*Bs1*) (Auchtung *et al.*, 2005; Beaber *et al.*, 2004);
2) la carence en nutriments (ex. ICE*clc*, ICE*Bs1*) (Auchtung *et al.*, 2005; Sentchilo *et al.*, 2003a);
3) la phase stationnaire, l'exposition au 3-chlorobenzoate et des dommages cellulaires (ex. ICE*clc*) (Miyazaki *et al*., 2012; Reinhard and van der Meer, 2014; Sentchilo *et al*., 2003a);
4) une diminution du niveau de traduction (ex. Tn*916*) (Su *et al.*, 1992);
5) l'exposition à la tétracycline (ex. Tn*916*, CTnDOT) (Salyers *et al.*, 1995; Su *et al.*, 1992).

Certaines familles d'ICE, telles que SXT/R391, s'intègrent dans un site précis du chromosome, tandis que d'autres, telles que Tn*916*, montrent beaucoup moins de spécificité dans leur site d'insertion (Hochhut and Waldor, 1999). Les mécanismes de régulation et l'organisation des transcrits codants pour l'intégrase, l'excisionase et les gènes de transfert sont aussi très variables. Par exemple, la forme circulaire de Tn*916* arbore tous les gènes dans un même opéron tandis que leur expression est découplée dans la plupart des autres familles d'ICE (Figure 20).

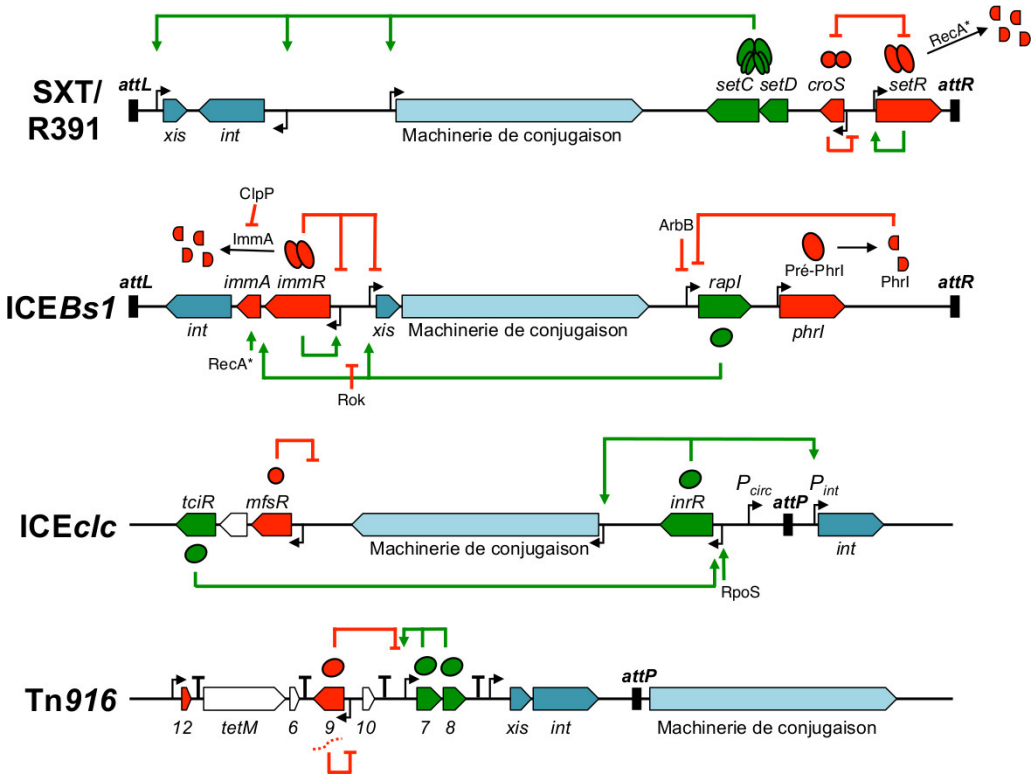

Figure 20. Régulation de l'excision/intégration et du transfert conjugatif chez les ICE des familles SXT/R391, ICE*Bs1*, ICE*clc* et Tn*916*. SetR (SXT/R391), ImmR et ImmA (ICE*Bs1*) sont sujet à la protéolyse aidée respectivement par RecA*, ImmA et ClpP. Le gène *phrI* (ICE*Bs1*) code pour un prépeptide (pré-PhrI) qui est clivé en sa forme mature (PhrI) et externalisé. Lorsque la concentration extracellulaire de PhrI atteint un certain niveau, il est importé dans la cellule et inhibe l'activité de RapI. Les activateurs (gènes et protéines) sont indiqués en vert tandis que les répresseurs sont en rouge. Les gènes ou groupes de gènes impliqués dans l'intégration/excision et conjugaison sont en bleu. Notez que ce schéma est simplifié et n'est pas à l'échelle. Inspiré de Carraro et Burrus (2014), Pradervand *et al.* (2014), Roberts et Mullany (2009), Wozniak et Waldor (2010).

3.5.1. Levée de la répression liée à la réponse SOS

Le transfert conjugatif des ICE des familles SXT/R391 et ICE*Bs1* est induit parallèlement à la réponse SOS bactérienne par différents mécanismes (Carraro and Burrus, 2014). L'induction des ICE SXT/R391 est médiée par l'autoprotéolyse de SetR catalysée par RecA* (Beaber *et al.*, 2004). Les gènes de transfert et de l'excisionase de ICE*Bs1* sont régulés par la paire ImmR et ImmA, respectivement un répresseur et sa protéase associée, ainsi que par le *quorum-sensing* médié par RapI et le pentapeptide PhrI (Figure 20) (Auchtung *et al.*, 2005, 2007; Bose and Grossman, 2011). La protéine ImmR réprime le promoteur qui conduit l'expression de *xis* et de la machinerie de conjugaison de ICE*Bs1* (Auchtung *et al.*, 2007). La protéase ImmA est activée par la réponse SOS selon un mécanisme inconnu qui implique RecA* puis clive ImmR, ce qui l'inactive et lève la répression exercée sur les gènes de transfert de ICE*Bs1* (Bose and Grossman, 2011). Les répresseurs CI, SetR et ImmR ont en commun des mécanismes d'autorégulation qui permettent de maintenir les niveaux de répresseurs assez hauts pour maintenir la quiescence, mais assez bas pour répondre efficacement à un signal inducteur (Auchtung *et al.*, 2007; Beaber and Waldor, 2004; Ptashne, 2004).

3.5.2. Effet de la circularisation

La circularisation a un effet sur la dynamique d'intégration des ICE de la famille ICE*clc* (Sentchilo *et al.*, 2003b). En effet, le gène *int* de ICE*clc* peut être exprimé à partir de deux promoteurs. Le premier, P_{int}, est faible et activé par l'activateur InrR et permet l'expression de l'intégrase lorsque l'élément est sous sa forme intégrée (Figure 20). Le deuxième, P_{circ}, est fort et constitutif, et sa position sur l'élément fait

en sorte qu'il peut conduire l'expression de l'intégrase seulement lorsque ICE*clc* est sous sa forme circulaire. Cet arrangement favorise donc l'expression de l'intégrase lorsque l'ICE est excisé et vulnérable.

Chez Tn*916*, ce n'est pas l'expression de l'intégrase, mais l'expression des gènes de transfert qui est conditionnelle à l'excision. En effet, les transcrits portant les gènes de transfert sont exprimés à partir de promoteurs situés à l'autre extrémité de l'élément, en amont des gènes *tetM*, *orf7* et *xis* (Figure 20). L'expression de l'intégrase et de l'excisionase, dont les gènes sont organisés en opéron, est faible et constitutive en absence de tétracycline (Celli and Trieu-Cuot, 1998). L'ajout de cet antibiotique ciblant les ribosomes permet la transcription au-delà d'une structure de terminaison en amont de *tetM*, ce qui favorise l'expression des gènes de transfert situés sur les Tn*916* circulaires (Roberts and Mullany, 2009).

Chez les ICE SXT/R391, la circularisation n'a été associée à aucun mécanisme de régulation. Certains des résultats obtenus non publiés supportent que *setR* et *xis* ne peuvent pas faire partie d'un même transcrit débuté à P_R (Annexe 7). Les gènes codants pour l'intégrase et l'excisionase sont convergents et chacun possède son propre promoteur dépendant de SetCD (Section 2.2, Figures 1 et 2) (Poulin-Laprade *et al.*, 2015a). Les résultats de l'article présenté à la section 2.2 montrent que les niveaux d'ARNm portant *int* et *xis* sont comparables en contexte d'induction. Cependant, une autre étude a montré que la présence de Xis dans la cellule réceptrice diminue la fréquence de transfert, probablement en défavorisant la recombinaison entre les sites *attP* et *attB* (Burrus and Waldor, 2003). Par conséquent, il y a sûrement un mécanisme limitant l'expression ou l'activité du produit de *xis* dans la cellule réceptrice pour favoriser l'intégration chromosomique médiée par Int. Le gène *xis* et son promoteur sont localisés à proximité de la séquence

d'attachement *attL* (Burrus and Waldor, 2003). Une hypothèse qui expliquerait la limitation en Xis dans la cellule réceptrice pour favoriser l'intégration à l'excision serait que la circularisation de l'élément créerait de l'interférence au promoteur P_{xis}. Par exemple, il pourrait y avoir formation d'une structure secondaire comprenant l'intégrase à promixité du site *attP* qui préviendrait la reconnaissance de P_{xis} par la RNAP ou par SetCD ou encore permettant l'expression d'un facteur qui réprimerait *xis* à un niveau post-transcriptionnel. Une autre hypothèse concernant la limitation en Xis dans la cellule réceptrice serait qu'elle soit sujette à une activité protéolytique médiée par un facteur encore inconnu codé par l'ICE ou l'hôte bactérien.

3.1. Biais fonctionnel dans l'identification des ICE SXT/R391

Le transfert horizontal de matériel génétique joue un rôle crucial dans l'évolution bactérienne, et les ICE en sont d'importants vecteurs. L'engouement lié à la problématique de la résistance aux antibiotiques a généré un biais dans la recherche et l'identification de nouveaux ICE. En effet, la majorité des ICE SXT/R391 décrits porte la résistance à de multiples antibiotiques. Tel est le cas de l'élément prototype SXT, découvert chez *Vibrio cholerae*, qui porte les gènes codants pour la résistance à quatre antibiotiques, dont deux sont utilisés pour traiter le choléra (Kaper *et al.*, 1995). Considérant le retour imminent de l'ère préantibiotique prédite par l'Organisation mondiale de la Santé et le *Centers for Disease Control and Prevention* américain, ce biais dans la découverte de nouveaux ICE va sûrement se prolonger. Pourtant, les ICE peuvent conférer de nombreux autres avantages à leurs hôtes bactériens, tant en contexte d'infection qu'au sein de l'environnement.

Tous les ICE de la famille SXT/R391 possèdent des gènes impliqués dans la modification, la réparation et la recombinaison de l'ADN. Certains de ces dispositifs participent à la protection des bactéries contre les phages et les plasmides qui sont omniscients en milieux aquatiques et dans le microbiome humain. Des études ont d'ailleurs montré l'importance de la prédation des *Vibrio* par les phages dans des contextes d'infection intrahumaine (Seed *et al.*, 2014), dans la persistance de pathogènes de poisson (Balado *et al.*, 2013) et dans la biorémédiation d'eau de mer contaminée par du diesel (Sauret *et al.*, 2015). Les composantes de réparation et de recombinaison de l'ADN peuvent être très utiles en cas d'atteintes à l'intégrité de l'ADN bactérien. Celui-ci peut être endommagé soit par des erreurs de réplication, soit par des facteurs environnementaux (par exemple les rayons UV), soit par des composés chimiques ciblant l'ADN ou des enzymes y étant associés (comme la mitomycine C et la ciprofloxacine). Bien que la famille d'ICE SXT/R391 soit parmi les plus étudiées, son influence sur la prospérité et l'évolution de ses hôtes bactériens est sûrement sous-évaluée.

Le déclin alarmant de nos ressources naturelles pousse les chercheurs à s'intéresser plus activement à la remédiation et à la biotransformation. Cet enthousiasme mènera peut-être à la découverte d'ICE SXT/R391 ou autre élément apparenté dont la cargaison pourrait contribuer à de tels processus. De telles cargaisons ont été identifiées dans d'autres familles d'ICE, par exemple dans l'ICE*clc* de *Pseudomonas knackmussii* B13 qui permet l'utilisation par l'hôte des composés 3-chlorobenzoate et 2-aminophénol comme seule source de carbone (Minoia *et al.*, 2008).

CONCLUSION

La dissémination rapide de la résistance aux antibiotiques parmi les populations bactériennes est principalement attribuée au transfert horizontal de gènes et a mené à un problème de santé publique alarmant. De plus en plus d'infections deviennent incurables, par exemple les infections à *Clostridium difficile* ou *Neisseria gonorrhoeae* mutlirésistantes ou encore à *Staphylococcus aureus* résistantes à la méthicilline et à d'autres antibiotiques. Une meilleure compréhension des systèmes de régulation régissant le transfert horizontal de matériel génétique génère des pistes pour révolutionner l'utilisation des traitements antibiotiques, idéalement pour les rendre plus durables. La caractérisation des systèmes de régulation des bactériophages lambdoïdes et des ICE SXT/R391 et ICE*Bs1* ont révélé que la propagation de ces vecteurs de matériel génétique est induite par des agents antimicrobiens ciblant l'ADN ou les enzymes y étant liés, par exemple les quinolones, la mitomycine C et le cotrimoxazole. Ce dernier est un traitement constitué de deux antibiotiques synergiques dont la résistance est portée par SXT et certains de ses variants. Les antibiotiques ciblant l'ADN, surtout le cotrimoxazole, sont donc à proscrire pour le traitement du choléra, étant donné que la majorité des souches épidémiques porte un ICE SXT/R391. Non seulement ces souches épidémiques ont de bonnes chances d'être résistantes, mais en plus, ces traitements induisent la propagation de ces ICE parmi les *Gammaproteobacteria* et la mobilisation de matériel génétique en *cis* et en *trans* qui y est associée.

Ce projet a permis d'identifier CroS, un régulateur des ICE SXT/R391 important pour l'induction de leur transfert par conjugaison. Il a également permis de peaufiner notre compréhension de la régulation par SetR en y associant un site opérateur additionnel

ainsi que la régulation par SetCD en établissant son régulon et en déchiffrant certaines évidences conflictuelles passées. Les résultats de ce projet de recherche ont donc permis de vérifier les hypothèses présentées à la section 1.6. et d'atteindre les objectifs de la section 1.7.

Bien que ce projet doctoral ait contribué grandement à la description du système de régulation des ICE SXT/R391, plusieurs avenues demandent encore à être investiguées. Pour mieux comprendre les mécanismes qui constituent l'interrupteur génétique, il serait important de confirmer l'affinité relative de CroS et de SetR pour chacun de leurs sites opérateurs par d'autres méthodes, par exemple la résonance de plasmon. Il faudrait également vérifier l'effet du site *O4* sur la régulation de P_R, le promoteur de SetR. De plus, il serait intéressant de mieux cerner l'apport des facteurs de l'hôte dans la régulation de la biologie de ces ICE. La protéine nucléoïde IHF est essentielle chez *V. cholerae* pour qu'elle puisse agir comme donneuse d'un ICE SXT/R391, tandis que les *E. coli* donneuses peuvent s'en passer (McLeod *et al.*, 2006). Le mécanisme impliquant IHF chez *V. cholerae* est encore inconnu. La même équipe a également montré que Fis n'influence pas le transfert de SXT vers ou à partir de *V. cholerae* (McLeod *et al.*, 2006). L'impact sur les ICE SXT/R391 des autres protéines nucléoïdes comme H-NS, HU et LRP n'a pas été vérifié chez *E.coli* et *V. cholerae*, et rien n'est connu concernant les facteurs potentiels des autres hôtes possibles. L'analyse par BPROM de SXT suggère plusieurs cibles putatives de CRP, un régulateur global impliqué dans la répression catabolique (Solovyev and Salamov, 2011). Bien que l'ajout de glucose lors d'expériences de conjugaison semble affecter la fréquence de transfert de SXT, le ou les mécanismes en cause sont encore inconnus. L'apport de protéases dans la régulation post-transcriptionnelle de *xis* et de *setCD* serait également à investiguer.

Une suite importante du projet est d'utiliser le motif de SetCD généré expérimentalement pour trouver des cibles potentielles dans le chromosome des différents hôtes des ICE SXT/R391. Ceci pourrait éclaircir l'interaction entre les ICE SXT/R391 et le génome de leurs hôtes, probablement en lien avec les réponses de stress cellulaire. Cette analyse pourrait également mener à la découverte de nouvelles classes d'îlots génomiques mobilisés par ces ICE par des mécanismes inédits, comme ça a été le cas pour la recherche de cibles à partir du motif reconnu par le complexe activateur AcaCD codé par les plasmides IncA/C (Carraro *et al.*, 2014). En effet, le motif de reconnaissance de AcaCD a permit d'identifier des sites putatifs en amont des gènes *xis*, *s004*/*rep*, *traN* et *trhH*/*trhG* chez SGI1, ainsi qu'en amont de *xis*, et *mobI* chez MGIVmi1 (Annexe 4). Ces découvertes laissent prévoir que notre compréhension des vecteurs de matériel génétique bactérien n'en est qu'à ces balbutiements.

ANNEXE 1

Table 1. Liste des ICE SXT/R391 cités dans cet ouvrage.

Hôtes[a]	ICE[b]	Localisation	Année	Cargaison[c]
Clinique				
Vibrio cholerae O139	SXT	Inde	1992	Su, Tm, Cm, Sm, système TA
V. cholerae O1	ICE*Vch*Ind5	Inde	1994-2005	Su, Tm, Cm, Sm
V. cholerae O2	ICE*Vch*Ind6	Inde	1998	Tm, système RM
V. cholerae O1	ICE*Vch*Ban9	Bangladesh	1994	Su, Tm, Cm, Sm, Tc
V. cholerae O1 B33	ICE*Vch*Moz10	Mozambique	2004	Su, Cm, Sm, Tc
V. cholerae O1	ICE*Vch*Hai2	Haïti	2010	diguanylate cyclase, endonucéase
V. fluvialis H-08942	ICE*Vfl*Ind1	Inde	2002	Su, Tm, Cm, Sm, système TA
V. alginolyticus	ICE*Val*Por1	Portugal	2005	Tc, Tm, Cm, Rf
V. splendidus	ICE*Vsp*Por2	Portugal	2009	Tm, Sp, Rf
V. scophthalmi	ICE*Vsc*Spa1	Espagne	2010	-
Providencia rettgeri	R391	Afrique du Sud	1967	Kn, Hg
Providencia alcalifaciens	ICE*Pal*Ban1	Bangladesh	1999	Su, Tm, Cm, Sm, système TA
Proteus mirabilis	ICE*Pmi*USA1	États-Unis	1986	Hélicase dépendante de l'ATP
Proteus mirabilis	ICE*Pmi*Jpn1	Japon	2006	Cp
Photobacterium damselae	ICE*Pda*Spa1	Espagne	2003	Tc
Shewanella haliotis	ICE*Sha*Por1	Portugal	2002	-
Enterovibrio nigricans	ICE*Eni*Spa1	Espagne	2008	Tc, Sm, Hg, BC
Environnementale				
V. cholerae non-O1/139	ICEVchMex1	Mexique	2001	système RM, diguanylate cyclase
V. cholerae non-O1/139	ICE*Vch*Hai2	Haïti	2010	diguanylate cyclase, endonucéase
V. fortis	ICE*Vfo*And1	Mer d'Andaman	2010	nd
S. putrefaciens	pMERPH	Royaume-Uni	1990	Hg
S. putrefaciens	ICE*Spu*PO1	Océan Pacifique	2000	pompe à efflux (Zn,Co,Cd), RM système
Alteromonas macleodii	ICE*Ama*AS1	Méditerranée	1998-2003	nd
Marinomonas ostreistagni	ICE*Msp*And1	Mer d'Andaman	2010	nd

[a] Pathogènes humain ou de poisson.

[b] Les articles de référence sont cités dans le texte.

[c] Su, sulfaméthoxazole; Tm, triméthoprime; Cm, chloramphénicol; Sm, streptomycine; Tc, tétracycline; Rf, rifampicine; Sp, spectinomycine; Kn, kanamycine; Hg, mercure; Cp, céphalosporines de 3e génération; BC, benzalkonium chloride; système TA, système toxine-antitoxine; système RM, système restriction-modification; nd, données non disponibles.

ANNEXE 2

Garriss, G., **Poulin-Laprade, D.**, Burrus, V. (2013). DNA-Damaging Agents Induce the RecA-Independent Homologous Recombination Functions of Integrating Conjugative Elements of the SXT/R391 Family. J. Bacteriol, *195*(9), 1991-2003.

L'article, formaté tel que publié dans *Journal of Bacteriology*, est reproduit avec la permission de *ASM Journals*.

JB
Journals.ASM.org

DNA-Damaging Agents Induce the RecA-Independent Homologous Recombination Functions of Integrating Conjugative Elements of the SXT/R391 Family

Geneviève Garriss, Dominic Poulin-Laprade, Vincent Burrus

Département de Biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada

Integrating conjugative elements (ICEs) of the SXT/R391 family are major contributors to the spread of antibiotic resistance genes. These elements also catalyze their own diversity by promoting inter-ICE recombination through the action of the RecA-independent homologous recombination system that they encode. Here, we report that expression of this recombination system, which consists of the single-stranded DNA annealing protein Bet and the exonuclease Exo, is induced by DNA-damaging agents via ICE-encoded transcriptional regulators. We show that the *bet* and *exo* genes are part of a large polycistronic transcript that contains many conserved ICE genes that are not involved in the main integration/excision and conjugative transfer processes. We show that although the recombination genes are highly transcribed, their translation is subject to additional strong regulatory mechanisms. We also show that an ICE-encoded putative single-stranded DNA binding protein (Ssb) limits hybrid ICE formation. Finally, a thorough *in silico* analysis reveals that orthologues of Bet and Exo are widely distributed in bacterial strains belonging to very distantly related bacterial species and are carried by various mobile genetic elements. Phylogenetic analyses suggest that the annealing proteins and exonucleases that compose these systems sometimes have different evolutionary origins, underscoring the strong selective pressure to maintain the functionality of these unrelated cooperating proteins.

Bacteria have the capacity to rapidly adapt to a changing environment due their ability to acquire and share genes, providing them with new selectable traits. One of the most striking examples of bacterial genome plasticity is the emergence and spread of antibiotic resistance genes. This exchange of genes occurs through horizontal gene transfer (HGT) events such as transformation, transduction, and conjugation. HGT events can be mediated by an array of mobile genetic elements (MGEs) such as conjugative plasmids, bacteriophages, and integrating conjugative elements (ICEs), which are also known as conjugative transposons (1–4). Stable integration of the features acquired by HGT, through site-specific recombination, transposition, or homologous recombination, allows their vertical transmission as well (5).

ICEs of the SXT/R391 family are mobile genetic elements that greatly contribute to the spread of antibiotic resistance genes in *Vibrio cholerae* and related gammaproteobacterial species (6, 7). These elements are found site-specifically integrated in the chromosome of their host. Under certain conditions, ICEs of the SXT/R391 family are excised from their host's chromosome as a circular, covalently closed molecule and transferred via conjugation to a new recipient cell in the form of a single-stranded DNA substrate (8). All members of this family share a conserved set of 52 core genes, of which 25 are required for their key functions of integration/excision, conjugative transfer and regulation (7). Newly identified ICEs are classified in this family based on the conservation of the scaffold of conserved genes, the presence of a conserved P4-like site-specific tyrosine recombinase and on their site-specific integration into the 5′ end of *prfC*, which encodes the peptide chain release factor RF3 (9). Interspersed in intergenic regions of this conserved backbone are variable DNA sequences that confer a wide selection of accessory functions, for instance, antibiotic or heavy metal resistances, and genes involved in modifying cell-signaling pathways (10–18). Similar to the induction of the lytic cycle of bacteriophage λ in lysogenic cells, the expression of the transfer genes of SXT/R391 ICEs is derepressed under conditions that trigger the host's SOS response. Under noninducing conditions, the master repressor SetR, which is encoded by these elements, prevents the expression of the transcriptional activators SetCD. DNA-damaging agents such as mitomycin C, UV light, and the antibiotic ciprofloxacin alleviate this repression, allowing SetCD-mediated activation of the conserved genes involved in excision and conjugative transfer (19).

SXT/R391 ICEs also encode a RecA-independent homologous recombination system similar to the bacteriophage λ Red system. In a previous study, we demonstrated that this recombination system promotes the formation of hybrid ICEs by catalyzing the recombination of two ICEs integrated in a tandem fashion (20). The recombination system carried by SXT/R391 ICEs comprises two genes: *s065*, encoding a single-stranded DNA annealing protein (21), and *s066*, encoding a 5′-3′ exonuclease (22). These proteins are, respectively, related to λBet and λExo, and we therefore refer to *s065* as *bet* and to *s066* as *exo*. Homologues of the SXT/R391 recombination genes are also found in conjugative plasmids of the A/C incompatibility group (IncA/C) (Fig. 1). These multidrug resistance plasmids are widespread in *Salmonella* and other enterobacteria from agricultural sources, and it was recently demonstrated

Received 8 November 2012 **Accepted** 19 February 2013
Published ahead of print 22 February 2013
Address correspondence to Vincent Burrus, vincent.burrus@usherbrooke.ca.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/JB.02090-12.

doi:10.1128/JB.02090-12

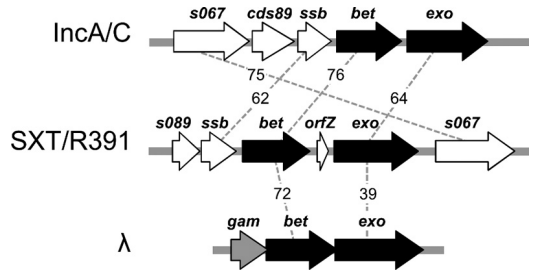

FIG 1 Schematic comparison of the recombination loci of SXT/R391 ICEs, IncA/C plasmids, and bacteriophage λ. SXT/R391 *bet* and *exo* and their orthologues are represented in black. Surrounding genes that have no known role in recombination are represented in white. λ*gam* has no orthologue in SXT/R391 ICEs and IncA/C plasmids and is presented in gray. Dotted lines and associated numbers indicate the percent similarity between orthologous genes.

that most of the conserved genes of SXT/R391 ICEs are also present in this family of conjugative plasmids (7).

Although the λ Red and SXT/R391 recombination genes are functionally similar, they differ in their genetic organization. λ*bet* and λ*exo* are located in the λ *pL* operon with other genes involved in homologous and site-specific recombination and are expressed early in the lytic cycle (23, 24). λ*gam*, a gene coding for an inhibitor of the host exonuclease RecBCD, is found directly upstream of λ*bet* and is transcribed with the recombination genes (Fig. 1). The product of λ*gam* protects the phage's linear double-stranded DNA (dsDNA) by inhibiting the ability of RecBCD to bind dsDNA ends (25). No homologue of λ*gam* exists in SXT/R391 ICEs or IncA/C plasmids; instead, a gene encoding a putative single-stranded DNA-binding protein (*ssb*) is found upstream of *bet*. In SXT/R391 ICEs, *bet* and *exo* are separated by a 288-bp stretch that contains a small 141-bp open reading frame (ORF) of unknown function, termed *orfZ* (Fig. 1). Upstream of *ssb* is found a gene of unknown function, *s089*. No homologues of *orfZ* or *s089* are found in the IncA/C plasmids.

Here, we report that the RecA-independent homologous recombination functions of SXT/R391 ICEs are induced by the DNA-damaging agent mitomycin C via the main transcriptional activators SetCD. Our results also indicate that the recombination functions are further regulated at the translational level. Furthermore, deletion of *ssb* significantly increased Bet-Exo-mediated hybrid ICE formation, suggesting that it could act as a modulator of the recombination activity. Our *in silico* analysis revealed that similar recombination systems are widely distributed in a large number of strains belonging to very diverse bacterial taxonomic orders and that they are present not only in ICEs and λ-like bacteriophages but also in conjugative plasmids from various incompatibility groups. Finally, our results also suggest different evolutionary origins for the Bet and Exo proteins that form the recombination system of SXT/R391 ICEs.

MATERIALS AND METHODS

Bacterial strains, plasmids, and media. The strains and plasmids used in this study are described in Table 1. These strains were routinely grown in Luria-Bertani (LB) broth at 37°C in an orbital shaker/incubator. Antibiotics were used at the following concentrations: ampicillin, 30 μg/ml (pGG2B only) or 100 μg/ml; tetracycline, 12 μg/ml; sulfamethoxazole, 32 μg/ml; trimethoprim, 4 μg/ml; spectinomycin, 50 μg/ml, kanamycin, 50 μg/ml. When required, bacterial cultures were supplemented with L-arabinose (0.2%, wt/vol) or mitomycin C (100 or 200 ng/ml).

Plasmid and strain construction. The oligonucleotides used in this study are described in Table S1 in the supplemental material. Deletion and fusion mutants were constructed by using the one-step chromosomal gene inactivation technique of Datsenko and Wanner (26). All deletions were designed to be nonpolar. The Δ*recA* mutation was introduced into *Escherichia coli* BW25113 using the primer pair recAWF/recAWR (20) and plasmid pVI36 as the template. The Δ*setDC* mutation was introduced into *E. coli* BW25113 SXT using the primer pair setD2wF/setC2wR and plasmid pKD13 as the template. The Δ*s089* and Δ*ssb* mutations were introduced into SXT (in strain HW220) using the primer pairs s089WF/s089WR and ssb1WF/ssb1WR, respectively, and template plasmid pVI36. The corresponding mutations were introduced into R391 (in strain JO99) using the primer pairs R66WF/s089WR and ssb2WF/ssb2WR, respectively, and pVI36 as the template. P_{lac}-*lacZ* and P_{lac}-*galK* were introduced into strains containing SXT and R391 deletion mutants, using generalized P1*vir* transduction and *E. coli* VB40 and GG13 as donor strains. Strains containing tandem arrays were constructed by successively transferring SXT::*lacZ* and R391::*galK* (or their corresponding deletion derivatives) into GG55, yielding strains GG66 to GG68. Plasmid pVI67 was introduced into GG55 prior to the transfer of the ICEs. *lacZ* translational fusions of *bet* and *exo* were constructed in BW25113 SXT using the primer pairs 65lacZ42BF/65lacZ42BR and 66lacZ42BF/66lacZ42BR, respectively, and plasmid pVI42B as the template. For both *bet* and *exo*, *lacZ* was fused to the 8th codon of the gene. All deletions and fusions were verified by PCR amplification using primers flanking the deletion, cloning, and sequencing. Plasmid pGG2B was constructed by amplifying *setDC* from strain HW220 using the primer pairs setDF/setC2R, subcloning into pCR2.1TOPO (Invitrogen), and recloning into pBAD30 (27). Plasmid pGG7 was constructed by amplifying *orfZ* from strain HW220 using the primer pair orfZTOPOF/corfZR and cloning into pBAD-TOPO (Invitrogen). Plasmid pGG32 was constructed by amplifying the intergenic region located between *s063* and *s089* in SXT from *E. coli* HW220 using the primer pair 6389F/6389R and cloning into pCR2.1TOPO. Plasmids pDPL373, pDPL374, and pDPL375 were constructed by cloning into pBAD-TOPO the *bet′-′lacZ* translational fusion and various lengths of its 5′ untranslated region (UTR) that were amplified using the primers betTAlacZF1, betTAlacZF2, and betTAlacZF3, respectively, with primer lacZTAcommonR and genomic DNA of GG209 as the template. Plasmids pDPL376, pDPL377, and pDPL378 were similarly constructed for the *exo′-′lacZ* translational fusion and its 5′ UTR using primers exoTAlacZF1, exoTAlacZF2, and exoTAlacZF3, respectively, with primer lacZTAcommonR and genomic DNA of GG215 as the template. The resulting six plasmids were introduced into *E. coli* BW27784. All plasmid inserts were verified by sequencing by Centre d'Innovation Génome Québec (McGill University, Montreal, QC, Canada). *E. coli* was transformed by electroporation in 1-mm-gap cuvettes in a GenePulser Xcell apparatus (Bio-Rad) set at 2.5 μF, 200 Ω, and 1.8 kV.

RNA extraction and cDNA synthesis. Briefly, RNA extractions were performed as follows. Cultures were grown for 14 to 16 h in LB broth containing the appropriate antibiotics, diluted 1:100 in fresh medium in the absence of antibiotics (except when ampicillin was added for the maintenance of the pGG2B vector), and grown to an optical density at 600 nm (OD_{600}) of 0.2. They were then diluted 1:100 a second time and grown to an OD_{600} of 0.2. Each culture was then split between two flasks, of which one was induced with 100 ng/ml mitomycin C (MMC) or 0.02% L-arabinose (when using pGG2B), and both were incubated for 2 h at 37°C. A 1-ml sample of each culture was used for total RNA extraction using an RNeasy minikit and RNAprotect bacterial reagent (Qiagen) following the manufacturer's instructions. Once purified, the RNA samples were subjected to gDNA digestion using Turbo DNase (Ambion) and following the manufacturer's instructions. cDNA was synthesized using 1 μg of RNA, 50 ng of random hexamers, or 2 pmol of the gene-specific primer s073RT (Integrated DNA Technologies) (see Table S1 in the supplemental material), and the reverse transcriptase SuperScript II (Invitrogen), following the manufacturer's instructions. Control reactions in the

TABLE 1 *E. coli* strains and plasmids used in this study

Strain or plasmid	Relevant genotype or phenotype[a]	Reference or source
E. coli strains		
BW25113	F$^-$ Δ(*araD-araB*)*567* Δ*lacZ4787*(::*rrnB3*) λ$^-$ *rph-1* Δ(*rhaD-rhaB*)*568 hsdR514*	26
BW27784	F$^-$ Δ(*araD-araB*)*567* Δ*lacZ4787*(::*rrnB-3*) λ$^-$ Δ(*araH-araF*)*570*(::FRT) Δ*araEp-532*::FRT ϕ*Pcp18 araE533* Δ(*rhaD-rhaB*)*568 hsdR514*	58
CAG18439	MG1655 *lacZU118 lacI42*::Tn*10* (Tc^r)	59
VB112	MG1655 Rf^r	60
HW220	CAG18439 *prfC*::SXT (Tc^r Su^r Tm^r)	61
JO99	CAG18439 *prfC*::R391 (Tc^r Kn^r)	62
VB40	CAG18439 Δ*lacZ prfC*::SXT::*lacZ* (Tc^r Su^r Tm^r)	20
GG13	CAG18439 Δ*galK prfC*::R391::*galK* (Tc^r Kn^r)	20
VB17	BW25113 *prfC*::SXT (Su^r Tm^r)	V. Burrus
GG247	BW25113 Δ*recA prfC*::SXT (Su^r Tm^r)	This study
GG242	BW25113 *prfC*::SXT Δ*setDC* (Su^r Tm^r)	This study
GG55	VB112 Δ*recA* (Rf^r)	20
GG66	VB112 *prfC*::[SXT::*lacZ*]-[R391::*galK*] (Rf^r Su^r Tm^r Kn^r)	20
GG67	VB112 *prfC*::[SXT::*lacZ* Δ*s089*]-[R391::*galK* Δ*s089*] (Rf^r Su^r Tm^r Kn^r)	This study
GG68	VB112 *prfC*::[SXT::*lacZ* Δ*ssb*]-[R391::*galK* Δ*ssb*] (Rf^r Su^r Tm^r Kn^r)	This study
VB47	CAG18439 Δ*galK* Δ*recA* (Tc^r)	20
GG209	BW25113 *prfC*::SXT *bet*′-′*lacZ* (Su^r Tm^r)	This study
GG215	BW25113 *prfC*::SXT *exo*′-′*lacZ* (Su^r Tm^r)	This study
Plasmids		
pBAD30	ori_{p15A} *araC* P_{BAD}	27
pGG2B	pBAD30::*setDC*	This study
pGG7	pBAD::*orfZ*	This study
pGG32	pCR2.1::Igr(*s063-s089*)$_{SXT}$	This study
pVI67	pAH57 Δ(xis_λ-int_λ)::*setDC*(Ts)	20
pVI36	Sp^r template for one-step chromosomal gene inactivation	60
pKD13	Kn^r template for one-step chromosomal gene inactivation	21
pVI42B	pVI36 BamHI::P_{lac}-*lacZ*	20
pDPL373	pBAD::TA$_{-10}$-*bet*′-′*lacZ*	This study
pDPL374	pBAD::TA$_{-24}$-*bet*′-′*lacZ*	This study
pDPL375	pBAD::TA$_{-53}$-*bet*′-′*lacZ*	This study
pDPL376	pBAD::TA$_{-11}$-*exo*′-′*lacZ*	This study
pDPL377	pBAD::TA$_{-34}$-*exo*′-′*lacZ*	This study
pDPL378	pBAD::TA$_{-41}$-*exo*′-′*lacZ*	This study

[a] Ap^r, ampicillin resistant; Cm^r, chloramphenicol resistant; Kn^r, kanamycin resistant; Rf^r, rifampin resistant; Su^r, sulfamethoxazole resistant; Sm^r, streptomycin resistant; Sp^r, spectinomycin resistant; Tc^r, tetracycline resistant; Tm^r, trimethoprim resistant; Ts, thermosensitive.

absence of reverse transcriptase (no-RT reactions) were performed for each sample.

Real-time quantitative PCR analysis. qRT-PCR was performed by measuring the increase of fluorescence using the Quantifast SYBR green mix (Qiagen) in an Eppendorf RealPlex (Eppendorf). Primers were designed to amplify an internal fragment of ca. 180 bp of the reference genes (*rpoZ* and *gyrA*) (28) or of each gene of interest (see Table S1 in the supplemental material). The expression ratios were calculated using the $2^{-\Delta\Delta CT}$ method, and the difference in the transcript levels of each gene of interest was calculated compared to the transcript level of *rpoZ* and *gyrA* using the $2^{-\Delta CT}$ method. The experiments were carried out with a minimum of three biological replicates, each with three technical replicates.

Primer extension analysis. Primer extension analysis was carried out using 1 and 5 μg of total RNA from BW25113 SXT cultured in the absence (CTL) or presence of mitomycin C (MMC) at 100 ng/ml, with radiolabeled TSS89up4 primer ([γ-^{32}P]ATP, 3000 Ci/mmol, 10 mCi/ml [PerkinElmer]) and a primer extension system with Primer Extension System-AMV Reverse Transcriptase (Promega) following the manufacturer's instructions. A control reaction in which water replaced the RNA was performed. A 3-μl portion of each primer extension reaction was directly mixed with an equal volume of loading dye, and the primer extension products were migrated in a 0.8% acrylamide–bis-acrylamide (19:1) gel in a SequiGenGT nucleic acid sequencing cell (Bio-Rad) alongside the sequencing reaction of the intergenic region located between *s063* and *s089* in SXT. Sanger sequencing of the 367-bp intergenic region located between *s063* and *s089* in SXT was performed using 5 μg of plasmid pGG32 as the template, [α-^{32}P]dATP (800 Ci/mmol, 10 mCi/ml) (PerkinElmer), primer TSS89up4, and the Sequenase version 2.0 kit (Affymetrix). Signal strength was detected using a Storm 860 molecular imager (GMI).

Semiquantitative PCR. PCRs aimed at amplifying *s063*, *s989*, and *bet* were carried out using cDNA prepared with the gene-specific primer s073RT (see Table S1 in the supplemental material). Each reaction mixture contained a 1 μM concentration of each primer, a 100 μM concentration of each deoxynucleoside triphosphate (dNTP), and 1 U *Taq* DNA polymerase (New England BioLabs); 1 μl of a 1:10 dilution of the cDNA synthesis reaction, 1 μl of a 1:10 dilution of the no-RT reactions (negative control), and 500 pg BW25113 SXT gDNA (positive control) were used as the templates for PCR amplification. PCR conditions were as follows: (i) 3 min at 95°C, (ii) 28 cycles of 30 s at 95°C, 30 s at the appropriate annealing temperature, and 30 s at 72°C, and (iii) 5 min at 72°C. After PCR, the samples were mixed with loading dye (orange G in 30% glycerol [vol/vol]) at a final concentration of 1×, and 15 μl of each reaction mixture was migrated in a 2% agarose gel in 1× TAE (Tris-acetate-EDTA) buffer,

alongside 1 μg 2-log DNA ladder (New England BioLabs). Migrated gels were colored in an ethidium bromide bath (3 μg/ml) for 30 min and then soaked for 30 min in distilled water. Gels were visualized under UV light in a GelDocXR system (Bio-Rad) and analyzed using Quantity One v6.3 software (Bio-Rad).

β-Galactosidase assays. β-Galactosidase assays were carried out using translational fusions of *lacZ* to *bet* or *exo* in BW25113 SXT or strain BW27784 containing pDPL plasmids, under induced or control conditions. Briefly, strains were grown for 14 to 16 h in LB with the appropriated antibiotics, then refreshed at 1:100, and grown to an OD_{600} of 0.2 in fresh LB broth in the absence of antibiotics, except when ampicillin was used to maintain pGG2B or pDPL plasmids. Cultures were split into 4-ml samples and grown in the absence or presence of MMC at 100 ng/ml or 200 ng/ml (for chromosomal fusions) for 2 or 16 h or in the presence or absence of L-arabinose 0.2% vol/vol (for assays performed with pDPL plasmids). β-Galactosidase activity was assessed using 100- or 500-μl samples as described elsewhere (29).

Bioinformatics analyses. The predicted proteomes of 2,714 plasmids, 4,093 viruses, and 1,706 complete microbial genomes available in March 2012 were obtained from the RefSeq database (30) and analyzed to identify putative recombination systems similar to the SXT/R391 system. We screened this data set using profile hidden Markov models of the RecT, YqaJ, and Gam families with HMMsearch from the HMMER v3.0 software package (31). The HMM profiles of Bet, Exo, and λGam were recovered from the Pfam 25.0 database, as follows: for Bet, RecT and PF03837; for Exo, YqaJ and PF09588; for λGam, Gam and PF06064. The sequences corresponding to each identified protein were downloaded and aligned using MUSCLE multiple sequence alignment software (32). Maximum-likelihood phylogenies of proteins belonging to the RecT and YqaJ families were generated using the PhyML v3.0 program (33) with the HKY85 substitution. Tree topologies were optimized by PhyML using the NNI and SPR methods model, and the starting trees were estimated using BioNJ. Branch support of the phylogenies was estimated using nonparametric bootstrap (100 replicates). Phylogenetic analyses were computed using amino acid alignment generated by MUSCLE, and the poorly aligned regions were removed with trimAl v1.2 software using the automated heuristic approach (34) prior to phylogenetic analyses. Phylogenetic trees were viewed in iTOL v2 (35) and are available as a shared project on the iTOL website (http://itol.embl.de/shared/GGarriss). Sequence alignments were performed using ClustalW (36).

RESULTS

DNA-damaging agents induce the expression of ICE recombination genes. In order to gain a better understanding of the conditions that favor hybrid ICE formation, we investigated the conditions that induce the transcription of *bet* and *exo*. Since all the conserved genes of SXT/R391 ICEs involved in integration/excision and conjugative transfer are induced under conditions that trigger their host's SOS response, we decided to verify whether the DNA-damaging agent mitomycin C (MMC) would also promote the transcription of the recombination genes. We measured by reverse transcription-quantitative PCR (RT-qPCR) the effect of the exposure to MMC on the transcript levels of *bet* and *exo*, using *E. coli* BW25113 harboring SXT (VB17). Upon exposure to MMC, the relative expression levels ($2^{-\Delta\Delta CT}$) of *bet* and *exo* were ~50 times higher than under the control condition (Fig. 2A). However, when an isogenic *recA*-null mutant (GG247) was used, the induction of *bet* and *exo* expression was abolished (expression ratio of ≈1), indicating that the pathway by which these genes are induced requires *recA*. In order to determine if this increase relies on the ICE's main activators, SetC and SetD, and not on the alleviation of repression by the host-encoded SOS response repressor LexA, we carried out the same experiments using a strain harboring an SXT

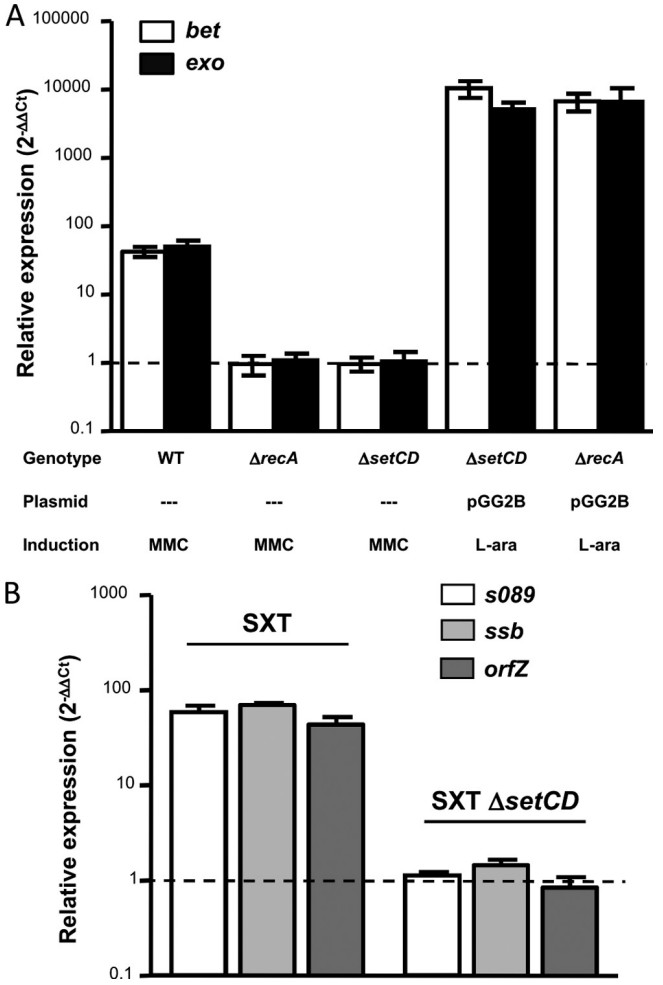

FIG 2 Expression of *bet* and *exo* is induced by mitomycin C. (A) Relative expression of *bet* and *exo*. The genotype of the ICE or strain, the presence of the *setCD* expression vector pGG2B, and induction with mitomycin C (MMC) or L-arabinose for pGG2B (L-ara) are indicated below the graph. (B) Relative expression of *s089*, *ssb*, and *orfZ* in an SXT or SXT Δ*setCD* background upon induction with MMC. Bars represent the ratio ($2^{-\Delta\Delta CT}$) between an induced and the noninduced conditions, using *rpoZ* as an internal reference as determined by RT-qPCR, and are means from a least three independent biological replicates. Standard deviations are indicated. Dotted lines indicate an expression ratio of 1 (no difference between the control and induced conditions).

Δ*setCD* mutant (GG242). Our results showed that in the absence of *setCD*, induction of *bet* and *exo* transcription with MMC was unachievable (expression ratio of ≈1). This deletion was more than successfully complemented using plasmid pGG2B, which carries *setCD* under the control of an arabinose-inducible promoter (expression ratio > 5,000). Furthermore, the absence of *recA*, which is required for derepression of *setCD* expression, could be bypassed by providing SetCD from pGG2B, which resulted in a >6,000-fold increase of *bet* and *exo* transcript levels. Our results show that agents that trigger the host's SOS response induce the transcription of ICE-encoded recombination functions, through the action of *recA* and the ICE-encoded transcriptional activators *setCD*.

We decided to verify whether the transcription of the two genes located upstream the recombination system, *s089* and *ssb*, as well *orfZ*, located between *bet* and *exo*, responded to the same induc-

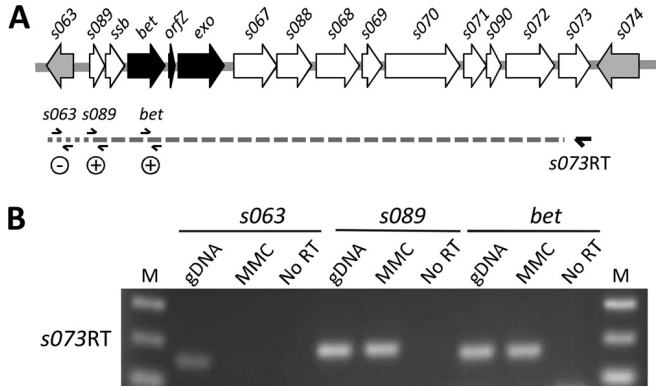

FIG 3 Operon structure of the SXT/R391 homologous recombination system. (A) Schematic representation of the *s089-s073* region of the SXT/R391 backbone. Genes are represented as follows: black, recombination genes; white, genes of other or unknown function; gray, immediate borders of the *s089-s073* locus. The relative positions of the reverse transcription primer *s073*RT as well as PCR primers used to amplify *bet*, *s063*, and *s089* are indicated. Dotted lines shows the reverse transcription product, and results of PCR are indicated as positive or negative. (B) A 2% agarose gel from an assay to amplify *s063*, *s089*, and *bet* on the s073RT product. VB17 genomic DNA (gDNA) and reverse transcription samples in the absence of reverse transcriptase (No RT) were used, respectively, as positive and negative PCR controls. MMC, MMC-induced samples; M, molecular weight marker.

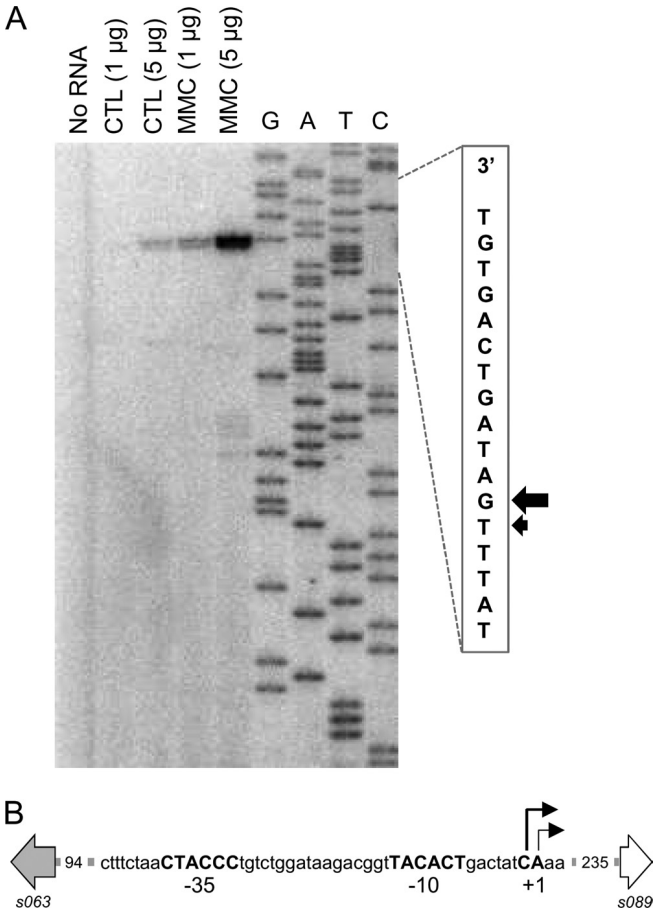

FIG 4 Promoter region of the SXT/R391 homologous recombination system. (A) Primer extension analysis of the transcribed strand of the *s063-s089* intergenic region. No RNA, negative control; CTL, noninduced control samples; MMC, MMC-induced samples; G, A, T and C, Sanger sequencing reaction lanes. Concentrations of RNA used for the primer extension reaction are indicated. Each well was loaded with equal reaction volumes. A partial read of the sequencing reaction is located to the right of the gel, with arrows indicating the transcription start sites on the transcribed strand. (B) Schematic representation of the promoter region. Positions of the two transcription start sites are indicated by angled arrows. Deduced −35 (5′-TTGACA-3′) and −10 (5′-TATAAT-3′) promoter elements recognized by *E. coli* σ^{70} are shown in bold capital letters. *s063* and *s089* are represented by gray and white arrows, respectively. Lengths (in base pairs) of the spacers between represented regions are indicated.

tion conditions as *bet* and *exo* using the same RT-qPCR assays. Our results show that the exposure to MMC causes a SetCD-dependent increase of *s089*, *ssb*, and *orfZ* transcripts comparable to that obtained for *bet* and *exo* (Fig. 2B).

The recombination genes are part of a polycistronic transcript that spans 11 kb of conserved ICE DNA and is driven from a MMC-induced promoter located upstream *s089*. The recombination genes are located within an ~11-kb conserved region of the SXT/R391 ICEs positioned between two inversely transcribed genes, *s063* (SXT/R391 conserved core gene) and *s074* (nonconserved gene present in SXT) (Fig. 3A) (7). Eight nearly identical and perfectly syntenic genes are located downstream of *exo* in this region. These genes play no role in conjugative transfer, and their respective functions remains obscure (7). The similar response of the genes between *s089* and *exo* to the exposure to MMC, combined with the fact that in λ the recombination genes are present in a single transcript (23), led us to hypothesize that a similar operon structure might also be found in SXT/R391 ICEs. The overall genetic organization of this locus led us to investigate whether all the genes between *s063* and *s074* are cotranscribed with *bet-exo*.

We carried out a reverse transcription experiment using a primer located at the 3′ end of *s073* (*s073*RT) and RNA obtained from MMC-treated VB17 cells (Fig. 3A). We then used PCR to amplify fragments located at the 5′ ends of *s063*, *s089*, and *bet* to determine if any of these genes are cotranscribed with *s073*. Our results show that *s089* and *bet* are part of a transcript that encompasses ca. 11 kb of conserved ICE DNA (Fig. 3B) and that expression of all of the genes from *s089* to *s073* is induced by MMC. Moreover, the impossibility of amplifying *s063* from the same cDNA sample indicates that this transcript originates within the intergenic region lying between divergently transcribed *s063* and *s089* (Fig. 3).

In order to determine the precise location of the promoter driving the transcription of the recombination genes, we analyzed the 367-bp intergenic region located between *s063* and *s089* by primer extension, using total RNA from *E. coli* VB17 prepared in noninduced and MMC-induced conditions. Our results show that the transcription start site is located 235 bp from the start codon of *s089* and corresponds to a C in the nontranscribed DNA strand (Fig. 4). Despite the presence of a large 5′ UTR, we found no evidence of another ORF in the region located between the transcriptional and translation start sites of the recombination system. As expected, the intensity of the band is dependent on the presence of MMC and also on the concentration of RNA used for the primer extension assay. A second band of lesser intensity is observed at A^{+2}, but this might be due to imprecise transcription initiation (Fig. 4A). Further analysis of the sequence upstream of C^{+1} revealed likely putative −35 (CTACCC) and −10 (TACACT)

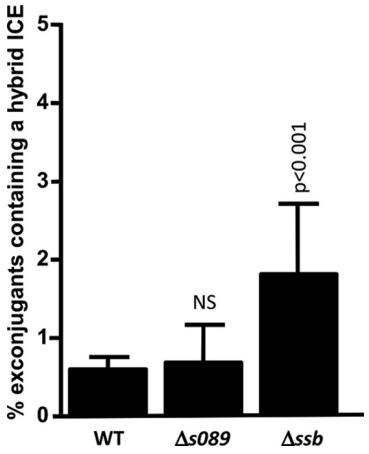

FIG 5 Impact of *s089* and *ssb* on hybrid ICE formation. *recA* donor strains containing wild-type SXT-R391 tandem arrays or tandem arrays of Δ*s089* and Δ*ssb* mutants of SXT and R391 were used in mating experiments with a *recA* recipient strain. Values are means and standard deviations from at least three independent experiments. One-way ANOVA with Dunnett's multiple comparison test was used to compare the means for hybrid-ICE-containing exconjugant colonies of each deletion mutant with the WT. The confidence intervals (p) for the comparisons are indicated. NS, not statistically different from the WT.

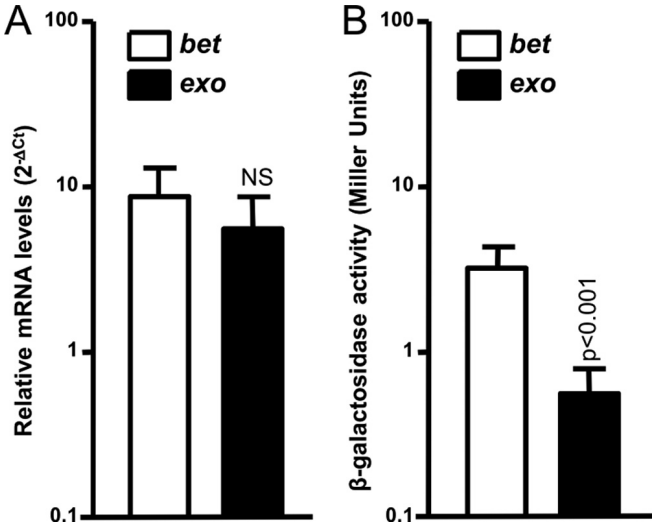

FIG 6 Relative quantification of the mRNA level and translation of *bet* and *exo* in the presence of MMC. (A) mRNA levels are expressed as the relative quantification by RT-qPCR ($2^{-\Delta CT}$) of each gene relative to *rpoZ*, in a WT background upon induction with MMC. (B) Translation results are expressed as Miller units obtained from β-galactosidase assays with *bet'-'lacZ* and *exo'-'lacZ* translational fusions (*lacZ* fused to the 8th codon of each gene) upon induction with MMC. Values are means and standard deviations from at least three independent experiments. One-way ANOVA with Tukey's multiple comparison test was used to compare the means between *bet* and *exo* transcription and translation. The mean number of Miller units obtained for *exo* was statistically significantly different from the values for *bet*, with a confidence interval of $P < 0.001$. NS, not statistically different.

promoter elements (Fig. 4B). No other DNA elements that contribute to promoter recognition by *E. coli* σ^{70} (37) could be identified in this region.

A gene encoding a single-stranded DNA binding protein modulates hybrid ICE formation. The presence of *s089* and *ssb* in the same transcript as *bet* and *exo* led us to investigate their possible involvement in Bet-Exo-mediated hybrid ICE formation. We constructed *recA* donor strains harboring tandem arrays of Δ*s089* and Δ*ssb* mutants of SXT and R391 and used them in conjugation assays with a *recA* recipient strain (VB47), using a quantitative detection assay designed to study hybrid ICE formation (20). These experiments were carried out in a Δ*recA* background to prevent RecA-mediated recombination and ensure that all hybrid ICEs are formed by Bet-Exo-mediated recombination (20). We bypassed the requirement for RecA for the activation of expression of the conjugative transfer genes (19, 20) as well as for expression of *bet* and *exo* by providing *setCD* in *trans* from plasmid pVI67. We found that although *s089* does not appear to be involved in hybrid ICE formation, the deletion of *ssb* caused a 3-fold increase of the percentage of exconjugants containing a hybrid ICE ($P < 0.001$) (Fig. 5). These results show clearly that the absence of Ssb allows Bet-Exo to more efficiently catalyze inter-ICE recombination.

Expression of *bet* and *exo* is subject to strong translational regulation. To determine whether the translation of *bet* and *exo* is consistent with the strong accumulation of their mRNA transcript upon MMC exposure, we used single-copy chromosomal *lacZ* translational fusions of the two genes in *E. coli* BW25113 derivatives (GG209 and GG215) to perform β-galactosidase assays. Interestingly, detectable translation of *bet'-'lacZ* and *exo'-'lacZ* occurred only after extensive exposure to MMC. Indeed, no differences in β-galactosidase activity between the induced and control samples were observed when the cultures were induced under conditions similar to those used for our RT-qPCR analysis (2 h with 100 ng/ml MMC). We confirmed by RT-qPCR with primers targeting *lacZ* that the mRNA transcript level of *lacZ* for each fusion was comparable to that obtained for the corresponding wild-type genes (see Fig. S1 in the supplemental material). We found that a 16-h induction with 200 ng/ml MMC was required for the detection of a significant induction of β-galactosidase activity for the *bet* and *exo* fusions. As expected, the translation of *bet* and *exo* was increased by MMC in a *recA*$^+$ strain: β-galactosidase activities were 3.56 ± 0.93 and 0.52 ± 0.23 Miller units (means ± standard deviations) for the *bet* and *exo* fusions, respectively, after MMC induction, compared with 0.00 ± 0.01 Miller units for both fusions without induction.

However, the comparison of RT-qPCR and β-galactosidase data revealed important differences in the transcriptional and translational profiles of *bet* and *exo*. Indeed, upon induction with MMC, *bet* and *exo* conserved the same relative mRNA amounts relative to the reference gene *rpoZ* ($2^{-\Delta CT}$) (Fig. 6A). However, β-galactosidase assay results demonstrate that *exo* is significantly less translated (nearly a 5-fold difference) than *bet* ($P < 0.001$) (Fig. 6B). Therefore, although *bet* and *exo* are part of the same transcript, the production of the Bet and Exo proteins is differentially regulated. As RT-qPCR assays ruled out the possibility that the difference observed between transcriptional and translational data was due to possible polar effects resulting from the *lacZ* fusions (see Fig. S1 in the supplemental material), our results suggest that expression of *bet* and *exo* is subject to an additional posttranscriptional or translational regulatory mechanism.

The SXT/R391 recombination system differs from other recombination systems composed of proteins related to Bet and Exo by the presence of a gene of unknown function, termed *orfZ*, located between the *bet* and *exo* genes (Fig. 1). Although OrfZ does

not have any known homologues, its small size suggests that it could act as a transcriptional or translational regulator. To test this hypothesis, we cloned *orfZ* under the control of an arabinose-inducible promoter (pGG7) and expressed it in *trans* in *E. coli* VB17. We used an RT-qPCR assay to measure the impact of its overexpression on the expression of *bet* and *exo*. Relative expression ($2^{-\Delta\Delta CT}$) of *bet* and *exo* remained unaffected by *orfZ* overexpression (expression ratio of ≈1). We also tested its impact on the translation of the single-copy chromosomal *bet'-'lacZ* and *exo'-'lacZ* fusions in the presence and absence of arabinose to induce the expression of *orfZ* from pGG7. No change in β-galactosidase activity was observed for these fusions upon overexpression of *orfZ* (expression ratio of ≈1). Taken together, these results suggest that *orfZ* does not participate in Bet-Exo-mediated recombination by regulating their expression. Furthermore, our analysis of the intergenic regions on each side of *orfZ* shows that they share some sequence similarity (see Fig. S4 in the supplemental material). The presence of *orfZ* only in SXT/R391 ICEs suggests that it results from an insertion event that led to the duplication of the *bet-exo* intergenic region and that it does not have a significant role in recombination (see below).

Conserved putative translational attenuators are found upstream of *bet* and *exo*. We used the RibEx server (38) to assess the presence of putative regulatory elements that could affect *bet* and *exo* translation. We identified two putative translational attenuators located immediately upstream of *bet* (TA*bet*) and *exo* (TA*exo*) (Fig. 7A). Moreover, the attenuator located upstream of *exo* has a higher predicted stability ($\Delta G = -25.50$ kcal mol^{-1}) than the one found upstream of *bet* ($\Delta G = -21.60$ kcal mol^{-1}), which could also explain the stronger translational attenuation observed for the production of Exo. These structures are highly conserved in all SXT/R391 ICEs (see Fig. S2 in the supplemental material), and RibEx analyses place similar structures upstream of the *bet* and *exo* genes found in IncA/C plasmids (see Fig. S3 in the supplemental material). The presence of these putative translational attenuators is consistent with our experimental finding that the expression of *bet* and *exo* is subject to a strong translational regulation.

In order to determine if these putative translational attenuators are indeed functional, we cloned *bet'-'lacZ* and *exo'-'lacZ*, conserving different lengths of the 5′ UTR upstream of each fusion, under the control of an arabinose-inducible promoter. This yielded three constructions for each fusion: one devoid of the putative attenuator sequence (TA_{-10} and TA_{-11}), one containing a truncated attenuator sequence (TA_{-24} and TA_{-34}), and one containing the full-length putative attenuator sequence (TA_{-53} and TA_{-41}) (Fig. 7A). We then assayed the translation levels of these constructions using β-galactosidase assays. In agreement with our prediction, under arabinose-inducing conditions, the absence of TA*bet* (TA_{-10}) allowed a higher translation level, which corresponds to an almost-8-fold increase compared to the level of translation obtained with the chromosomal fusion (Fig. 7B). In the presence of the full-length TA*bet* (TA_{-53}), a translation level comparable to that obtained with the chromosomal fusion was obtained. The truncated construction (TA_{-24}) gave intermediate results, consistent with the destabilization of the first stem-loop of the predicted structure (Fig. 7). Taken together, these results confirm that the predicted TA*bet* sequence acts as a translational attenuator. Results obtained with TA*exo* did not allow us to confirm that it acts as a translational attenuator, since similar translation levels were obtained regardless of the length of the 5′ UTR (Fig. 7B). However, the high levels of translation obtained from these constructions confirm that the chromosomal translational fusion is indeed functional and suggest that other regulatory mechanisms might be in play to limit translation of *exo*.

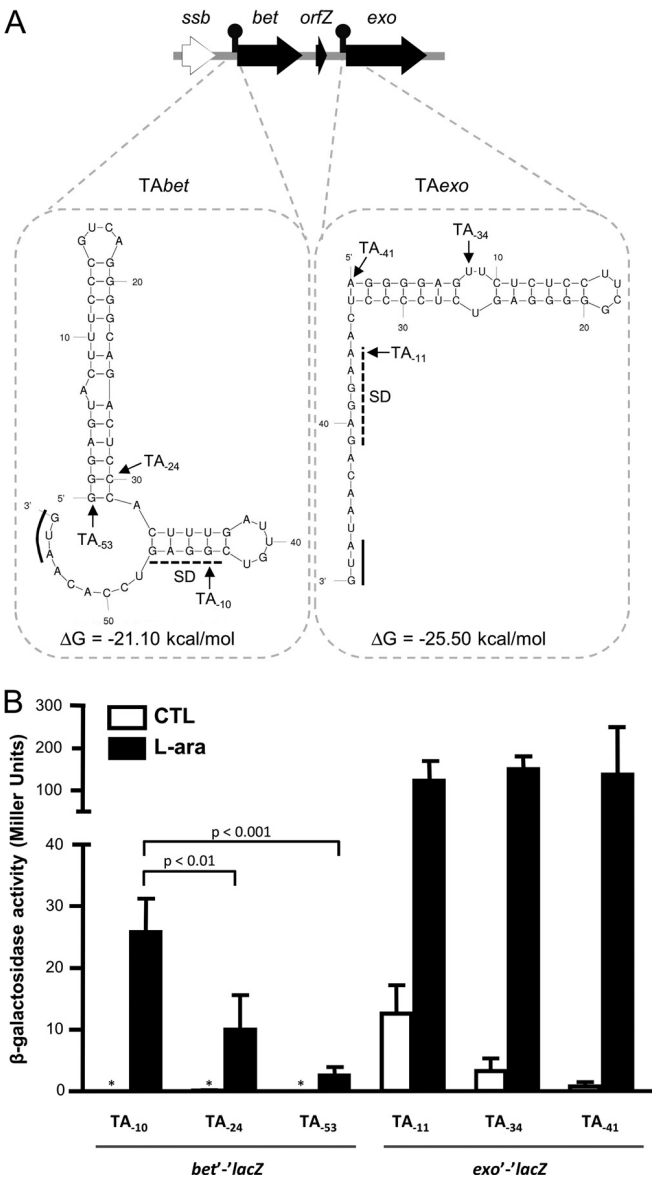

FIG 7 The recombination locus contains two predicted translational attenuators. (A) The presence of attenuators was determined with the RibEx server (38) and is shown by black lollipops upstream of *bet* and *exo*. Dotted rectangles contain folding predictions made with Mfold v4.6 (63) for each attenuator, along with the sequence located between the attenuator and the translation start sites of *bet* or *exo*. Positions of the Shine-Dalgarno (SD) sequences (dashed line), translation start sites (underlining), and Δ*G* values are indicated for each attenuator. Positions of the first nucleotide included in the construction of the pDPL *lacZ* translational fusion plasmids are indicated as TA sites and are numbered from the AUG start codon for each structure. (B) β-Galactosidase activity of *bet'-'lacZ* and *exo'-'lacZ* translational fusions (*lacZ* fused to the 8th codon of each gene) expressed in Miller units in the presence and absence (CTL) of L-arabinose. Values are means and standard deviations from three independent experiments. The asterisk indicates that activity was below the detection level. One-way ANOVA with Tukey's multiple comparison test was used to compare the means for the *bet'-'lacZ* fusion under arabinose-inducing conditions.

Similar recombination systems are widely distributed among various bacterial species and genetic elements. The similarity of the SXT/R391 recombination genes with the λ Red system and with genes found in the IncA/C family plasmids led us to investigate the prevalence of similar recombination systems in all sequenced bacterial species and mobile genetic elements. We performed a large-scale *in silico* analysis using the data extracted from the NCBI RefSeq database of fully sequenced viral, plasmid, and microbial genomes. We screened the proteomes of the 8,513 genomes of the database with the HMM profiles associated with the respective functional domains of Bet (RecT, PF03837) and Exo (YqaJ, PF09588) using HMMER3 software (31). Based on this analysis, we found 412 proteins containing a RecT domain and 632 proteins containing a YqaJ domain. A total of 163 recombination systems were identified based on the presence of a putative recombinase of the RecT family and a predicted exonuclease of the YqaJ family in the same locus (see Table S2 in the supplemental material).

The vast majority (125 of 163) of the recombination systems we identified were found in sequenced microbial genomes. Each system was further analyzed to determine whether it was part of an unidentified prophage-like element, integrating conjugative element, or integrated plasmid. We reclassified 96 of these systems as belonging to prophage-like elements on the basis of the presence of genes encoding a phage-like integrase and other unmistakable phage features (e.g., tail fiber protein, major capsid protein, and tail tape measure protein) in the vicinity of the recombination system. Four of the 125 systems identified in microbial genomes belonged to the SXT/R391 ICEs ICE*Spu*PO1 (CP000503), ICE*Pmi*USA1 (AM942759), ICE*Vch*Ban9 (CP001485) and to one previously unidentified putative SXT/R391 ICE in *V. cholerae* 2010EL-1786 (NC_016445). The remaining recombination systems found in microbial genomes that could not be reclassified were assumed to be located on genomic islands (GIs). Most of the plasmid-borne recombination systems (8 of 14) were found in IncA/C plasmids, the closest SXT/R391 relatives. Surprisingly, three systems were identified in conjugative plasmids belonging to the IncP-7 group (pCAR1, pCAR1.2, and pDK1), one in a plasmid of the IncT group (Rts1), as well as two in nonconjugative plasmids pMAQU02 and pALVIN02.

We analyzed the distribution of the identified recombination systems according to the taxonomic order of their host. As expected, more recombination systems (69 of 163) were identified in strains belonging to the order *Enterobacteriales*, the predominant hosts of lambdoid phages, than in strains of any other order. The remaining 94 systems were distributed among strains belonging to very diverse taxonomic orders, some of them very distant from SXT/R391 ICEs' hosts, such as the orders *Rhizobiales* and *Thermonoanaerobacterales* (Fig. 8; also, see Table S2 in the supplemental material).

SXT/R391 Bet and Exo have different evolutionary origins. We carried out a phylogenetic analysis to determine the evolutionary relationships between the 163 pairs of RecT domain- and YqaJ domain-containing proteins identified (Fig. 8). As SXT and R391 are found in genomes that are not fully assembled and are thus absent from the RefSeq database, we manually added their respective Bet and Exo sequences to our data set. Our analysis shows that the predicted recombinases and exonucleases identified in phages or GIs seem to cluster relative to their host's taxonomic order. However, proteins encoded by ICEs and plasmids tend to cluster relative to the family of mobile genetic elements to which they belong (Fig. 8). For instance, the Bet and Exo homologues of IncA/C plasmids form monophyletic groups, closely related to the corresponding ones of SXT/R391 ICEs. Similarly, plasmids belonging to the IncP-7 and IncT (Rts1) families also cluster together (Fig. 8).

Our analysis shows that partner recombinases and exonucleases do not always share a common origin. Indeed, the Bet proteins of SXT/R391 ICEs and IncA/C plasmids seem to have evolved from the ancestor of Bet proteins of λ-like phages (Fig. 8A). However, their cognate Exo proteins are very distantly related to those encoded by lambdoid phages (Fig. 8B). Moreover, while the YqaJ domain proteins encoded by all the plasmid- and ICE-borne recombination systems derive from the same common ancestor (Fig. 8B), their cognate RecT domain proteins have different phylogenetic relationships: RecT domain proteins from SXT/R391 ICEs and IncA/C plasmids form a monophyletic group, while those from IncP-7 plasmids (pCAR1, pCAR1.2, and pDK1), the IncT plasmid Rts1, and plasmid pMAQU02 form another distantly related group, more closely related to those derived from phages and prophage-like elements from *Enterobacteriales* and *Burkholderiales* (Fig. 8A). We also found that all plasmids but pALVIN02 harbor a gene encoding an Ssb in the vicinity of the predicted recombinase gene (see Table S2 in the supplemental material). Interestingly, pALVIN02 is the only plasmid that harbors a RecT orthologue unrelated to other plasmid-borne putative recombinase proteins. Although most bacteriophages from the order *Enterobacteriales* encode a homologue of λGam, none was found in the recombination systems carried by ICEs or plasmids (Fig. 8A). Finally, the only six occurrences of an ORF related to *orfZ* located between the recombinase- and exonuclease-encoding genes were found in SXT/R391 ICEs.

DISCUSSION

In previous studies, we demonstrated that ICEs of the SXT/R391 family generate their own genetic diversity by mediating recombination events between two ICEs arranged in a tandem array (20, 39). Two ICE-encoded proteins, Bet and Exo, which together function as a RecA-independent homologous recombination system, are able to catalyze a significant part of these events, generating functional ICEs with new features (20). Here, we investigated the conditions that regulate the expression of *bet* and *exo* to better understand which signals trigger the formation of hybrid ICEs. Like the host's SOS response, the expression of both genes is induced by DNA-damaging agents. RecA, as well as the ICE-encoded transcriptional regulators SetC and SetD, was found to be required for mitomycin C-induced expression of *bet* and *exo*. These results show that although they are induced during the SOS response, the recombination genes are not under the control of their host's SOS repressor, LexA. Instead, like the *tra* genes, *bet* and *exo* are under the control of the ICE-encoded main repressor SetR, which represses the expression of the SetCD transcriptional activator complex (11, 19, 40). SXT/R391 recombination functions are induced in response to DNA damages, the same conditions that trigger the conjugative transfer of these mobile elements to a new host (19), consequently providing the immediate possibility to segregate functional recombinants. Moreover, as we have previously shown that Bet-Exo and RecA homologous recombination pathways can cooperate to enhance hybrid ICE formation, their coinduction during SOS response highlights the extent to

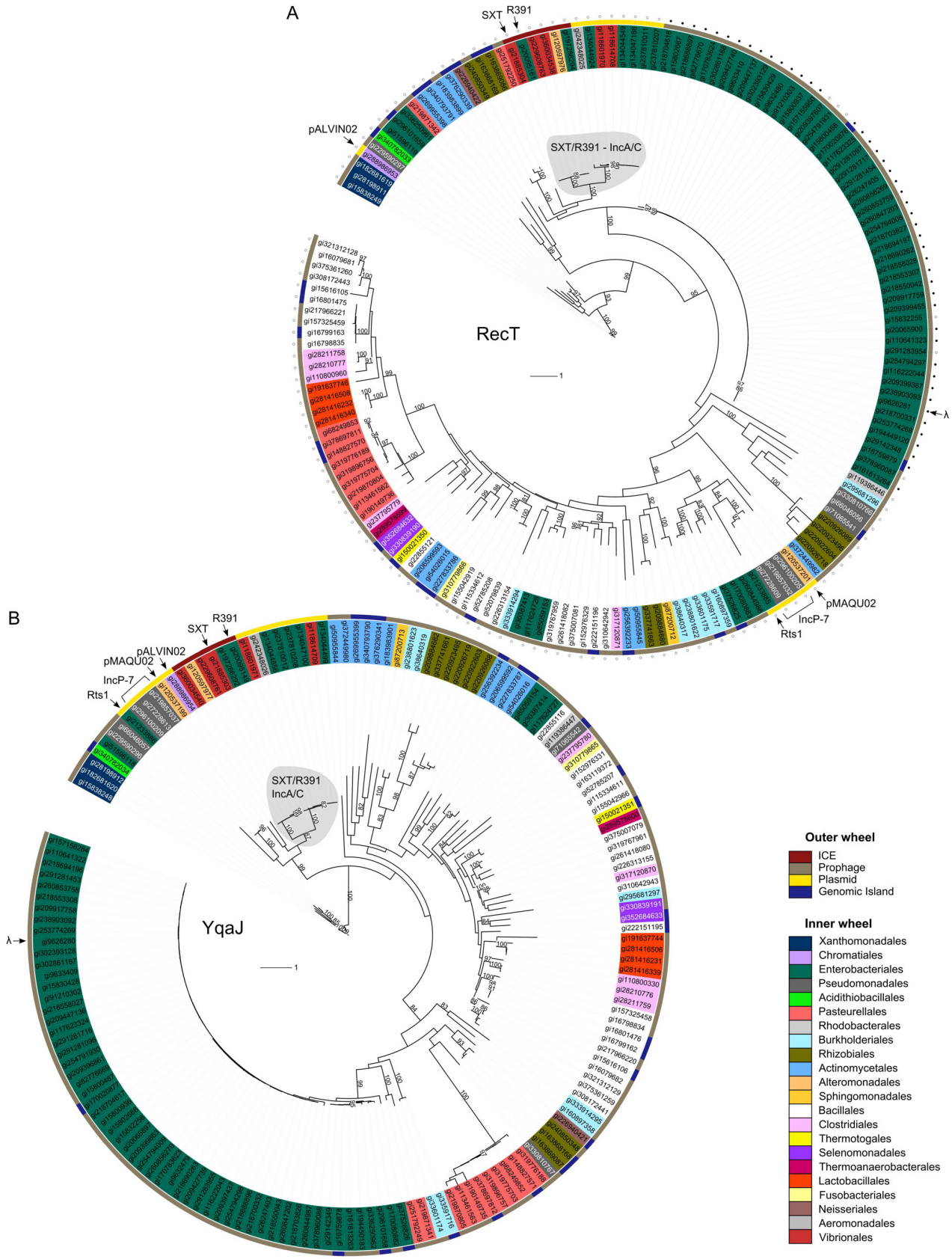

FIG 8 Phylogenetic trees of single-stranded DNA annealing proteins and cognate exonucleases encoded by gene pairs found in sequenced microbial, plasmid, and bacteriophage genomes. (A) Recombination proteins containing a RecT domain. (B) Exonucleases containing a YqaJ domain. Label colors indicate orders of bacterial strains in which each protein was found (inner wheel). Outer color strips indicate the element carrying each protein (outer wheel). Filled and open circles indicate the presence and absence, respectively, of Gam. Relevant elements are indicated by arrows. Bootstrap values of >80 are indicated.

which these elements take advantage of their host's cellular function to enhance their own diversity (20). Additionally, in the assay used to study hybrid ICE formation (20), derepression of SetCD in a *recA*$^+$ context probably depends upon the transient expression of the SOS response in a subpopulation of donor cells (41), since we did not provide the direct action of a DNA-damaging agent. Under those conditions, the total levels of RecA and Bet-Exo are expected to be much lower. Combined with the results we present here, this suggests that the percentage of hybrid ICEs obtained during those assays is possibly an ample underestimation of the real amount of hybrids that can be formed in the presence of DNA-damaging agents. Baharoglu and colleagues (42, 43) have recently demonstrated that, besides exposure to most antibiotics, conjugation itself induces the SOS response in *V. cholerae*, the natural host of SXT/R391 ICEs. Together with our results, this observation sheds new light on the major impact that these elements can have on their host's genome plasticity. Therefore, not only are *bet* and *exo* expressed in the induced donor cells, but they are also likely expressed in the fresh exconjugants, as the pool of SetR in the recipient cells has to build up to repress *setCD* expression. In addition, we anticipate that the accumulation of SetR could be slowed down by its simultaneous autocleavage, which would be promoted by the coprotease activity of RecA bound to the incoming single-stranded DNA.

In bacteriophage λ, λ*bet* and λ*exo* are part of the *pL* operon, which is also induced by DNA-damaging agents (23, 44). We, as well as others, have discussed in the past the possibility that these homologous recombination systems could act as a repair pathway for salvaging ICE or phage DNA broken by exposure to DNA-damaging agents (20, 24). Our finding that the SXT/R391 *bet* and *exo* genes are specifically expressed in the presence of DNA-damaging agents strengthens this hypothesis. We have found that the *bet* and *exo* genes are part of a polycistronic transcript that is driven from an SOS-inducible promoter located in the intergenic region between *s063* and *s089*. This transcript includes 12 other highly conserved and syntenic genes of the SXT/R391 backbone that have no apparent role in their transfer and is by far the largest set of contiguous genes of unknown function found in these elements (7). The conservation of these genes as well as their coexpression with the recombination and conjugative transfer genes suggests that at least some of them may have a role in recombination or in the fitness of these elements. Interestingly, 7 of the 14 genes of this locus, including *ssb*, *bet*, and *exo*, are also conserved in all the IncA/C conjugative plasmids, although their order differs slightly (7). To the best of our knowledge, the stimuli that trigger the conjugative transfer of IncA/C plasmids have not been identified. However, they lack a homologue of the ICE SetR repressor, suggesting that unless they rely on the host's LexA repressor, DNA damages are unlikely to activate their transfer. As a consequence, we are unable to predict which conditions trigger the expression of *bet* and *exo* orthologues in IncA/C plasmids.

The coexpression of *bet* and *exo* with genes located upstream of the recombination system (*s089* and *ssb*) led us to investigate their role in hybrid ICE formation. We found that the deletion of *s089* did not impact hybrid ICE formation, thus ruling out a role for the protein it encodes in recombination. However, we found that the deletion of *ssb*, a gene encoding a single-stranded DNA binding protein located immediately upstream *bet*, caused a significant (nearly 3-fold) increase in Bet-Exo-mediated hybrid ICE formation. Ssb proteins are known to be involved in a number of recombination, replication, and repair mechanisms, acting both cooperatively and competitively with a myriad of other proteins for binding to single-stranded DNA (ssDNA) (45, 46). Although we have not investigated its precise role in the recombination mechanism, it is possible that SXT Ssb limits the access of Bet to its substrate by competing for binding to ssDNA. This hypothesis is strengthened by recent findings that *in vivo* coexpression of SXT's Ssb-Bet-Exo diminishes by more than 2-fold the recombination frequency of a linear dsDNA substrate with the *E. coli* chromosome, compared with the coexpression of only Bet-Exo (22). Additionally, like λBet, SXT Bet is able to recombine ssDNA sequences that possess very short homologous sequences (36 bp) (21), and it has been demonstrated that λBet is able to promote strand exchange and displacement using short oligonucleotides containing several mismatches as substrates (47). In a context where the ICE recombination functions are induced by DNA-damaging agents, when large quantities of ssDNA would be available, it is possible that controlling the efficiency of recombination of sequences bearing only short homologous regions (36 bp [21]) would limit illegitimate recombination events, as they would lead to the formation of nonfunctional elements unable of transferring to a new host.

Our *in silico* analysis of the structure of several recombination systems shows that while a homologue of λGam is encoded by numerous phages, none is found in the plasmid or ICE-borne systems. λGam protects the phage linear dsDNA from degradation by inhibiting the host's RecBCD endonuclease (25, 48). Although we have not assessed the presence of other RecBCD inhibitors (46), the absence of λGam from plasmids and ICEs could be due to the fact that no step of these elements' normal "life cycle" involves a linear dsDNA molecule. Instead, all of these systems, with the exception of the one carried by the nonconjugative plasmid pALVIN02, contain a gene encoding an Ssb upstream of the gene encoding the Bet-like protein. Bacteriophage λ also carries a putative single-stranded DNA-binding protein, encoded by the gene *ea10* (49). Although the role of this precise putative Ssb in λ recombination has not been evaluated, other phage-encoded Ssb proteins have been recognized for their positive effects on phage homologous recombination and replication (for a review, see reference 46 and references therein). However, the specific role of Ssb proteins in the recombination catalyzed by systems found in conjugative elements has not been investigated yet. One of the possible roles for the ICE-encoded Ssb could be to protect the ssDNA substrate of conjugative transfer from degradation, aberrant recombination prior to its circularization and synthesis of its complementary strand, or even to limit RecA binding to prevent overactivation of the host's SOS response.

The role of *orfZ* located between *bet* and *exo* remains to be determined. However, our results suggest that it does not act as a transcriptional or translational regulator of *bet* and *exo* expression. In addition, no homologues of *orfZ* were found in any of the recombination systems we have analyzed *in silico*. The presence of long and imperfect repeats (66% identity) flanking this gene in SXT/R391 ICEs (see Fig. S4 in the supplemental material) suggests that it results from an insertion event and supports the idea that it is not relevant to the recombination mechanism of systems related to Bet-Exo.

The expression profiles and genetic organization of *bet* and *exo*

revealed that although mRNA transcripts of these genes accumulate at a high level in the presence of MMC, their translation remains weak. Our results are in accordance with other published data on the RecT-RecE recombination system of the *rac* prophage and on the λ Red system that show that overexpression of the recombinase with respect to its partner exonuclease favors recombination (50). We identified two putative translational attenuators, one located upstream of *bet* (TA*bet*) and the other upstream of *exo* (TA*exo*), that could account for the weak translation of the Bet- and Exo-LacZ fusion proteins. Our results suggest that TA*bet* is a genuine translational attenuator and when present causes a 10-fold reduction in the translation of the *bet′-′lacZ* fusion compared to when it is absent. On the other hand, the predicted translational attenuator located upstream of *exo* did not prove to effectively reduce translation. The high translation level of TA_{-11}-*exo′-′lacZ* compared to TA_{-10}-*bet′-′lacZ* (Fig. 7B) is in agreement with the predicted strength of the Shine-Dalgarno sequence upstream of *exo* (AAGGAG), which has a slightly better complementarity to the 3′ end of 16S RNA than the one found upstream of *bet* (CGGAGU) (Fig. 7A) (51). However, in its natural context, *exo* is much less translated than *bet* (Fig. 6B). Since TA*exo* does not seem to play a role as a translational attenuator in our artificial construction, another regulatory mechanism must be in play to reduce the efficient translation of Exo when it is part of the mRNA transcript initiated upstream of *s089*. Such translational regulation could depend upon the presence of an ICE-encoded small RNA, which would be present only in the chromosomal context of our translational fusions, and act by reducing the translation of Exo by binding to its 5′ UTR (for a review, see reference 52). Interestingly, while IncA/C plasmids lack an *orfZ* gene, an attenuator similar to TA*bet* is also found upstream the IncA/C *bet* gene, providing additional evidence for a common mode of translational regulation between these two recombination systems. Combined with our previously published results (20), our observations indicate that only very low levels of both Bet and Exo are sufficient to catalyze the formation of hybrid ICEs at high frequency.

The recombination proteins encoded by SXT and λ belong to the same families: λBet and SXT Bet belong to the RecT family (PF03837), and λExo and SXT Exo belong to the YqaJ family (PF0955), also sometimes referred to as the lambda exonuclease (LE) superfamily (53). Recombination systems composed of a putative recombinase of the RecT family and a putative exonuclease of the YqaJ family were identified in a variety of prophages, ICEs, and plasmids of different incompatibility groups as well as remnants of prophages or other genomic islands. Our analysis of the phylogenetic relationships between the predicted recombinase and exonuclease proteins from different origins suggests that those derived from phages cluster relative to the taxonomic order of their host, while those encoded by plasmids and ICEs cluster relative to the incompatibility group or family. This trend probably results from the narrow host range of phages, which would lead to fewer opportunities for interspecies gene exchange compared to plasmids and ICEs. Most mobile genetic elements are usually considered combinations of interchangeable functional modules that were originally defined as belonging to bacteriophages, conjugative plasmids, or transposons (6, 54–56). The widespread presence of recombination systems related to Bet-Exo in these mobile genetic elements might provide a suitable explanation for the exchange of such modules between different types of mobile elements, with a requirement for only very short homology regions on each side the different modules. The differences in the evolutionary relationships between the proteins of the RecT and YqaJ families suggest that many of the recombination systems identified have arisen by successive acquisition of the genes encoding these proteins rather than by the gain of a recombinase-exonuclease gene pair. While it is surprising that functionally cooperating proteins have been acquired from different origins, it is known that recombinases belonging to the RecT family are sometimes found associated with other types of exonucleases, such as RecE in the cryptic *rac* prophage (57), which has led to the assumption that the similar operon structures of different systems could arise independently due the selective pressure for the coexpression of cooperating proteins (57).

The prevalence of similar recombination genes in many mobile genetic elements found in nearly all bacterial orders suggests that mobile DNA may impact the plasticity and evolution of bacterial genomes far beyond the immediate benefit provided by selectable markers such as antibiotic resistance genes. Indeed, they could facilitate genome-wide rearrangement as well as incorporation of exogenous DNA into bacterial genomes by means of homologous recombination systems requiring much shorter identity regions than RecA-dependent pathways.

ACKNOWLEDGMENTS

We thank Eric Bordeleau for technical assistance with the HMM analysis. We are grateful to Nicolas Carraro, Eric Bordeleau, and Alain Lavigueur for insightful comments on the manuscript.

G.G. was the recipient a doctoral fellowship (Fonds de Recherche du Québec, Nature et Technologies, QC, Canada), and V.B. holds a Canada Research Chair in bacterial molecular genetics and is a member of the FRSQ-funded Centre de Recherche Clinique Étienne-Le Bel.

REFERENCES

1. **Frost LS, Leplae R, Summers AO, Toussaint A.** 2005. Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. **3**:722–732.
2. **Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW.** 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol. Rev. **33**:376–393.
3. **Ochman H, Lawrence JG, Groisman EA.** 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405**:299–304.
4. **Wiedenbeck J, Cohan FM.** 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol. Rev. **35**:957–976.
5. **Didelot X, Maiden MC.** 2010. Impact of recombination on bacterial evolution. Trends Microbiol. **18**:315–322.
6. **Wozniak RA, Waldor MK.** 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat. Rev. Microbiol. **8**:552–563.
7. **Wozniak RA, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, Garriss G, Dery C, Burrus V, Waldor MK.** 2009. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. PLoS Genet. **5**:e1000786. doi:10.1371/journal.pgen.1000786.
8. **Burrus V, Waldor MK.** 2003. Control of SXT integration and excision. J. Bacteriol. **185**:5045–5054.
9. **Burrus V, Marrero J, Waldor MK.** 2006. The current ICE age: biology and evolution of SXT-related integrating conjugative elements. Plasmid **55**:173–183.
10. **Beaber JW, Burrus V, Hochhut B, Waldor MK.** 2002. Comparison of SXT and R391, two conjugative integrating elements: definition of a genetic backbone for the mobilization of resistance determinants. Cell. Mol. Life Sci. **59**:2065–2070.
11. **Beaber JW, Hochhut B, Waldor MK.** 2002. Genomic and functional

analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. J. Bacteriol. **184**:4259–4269.
12. **Bordeleau E, Brouillette E, Robichaud N, Burrus V.** 2010. Beyond antibiotic resistance: integrating conjugative elements of the SXT/R391 family that encode novel diguanylate cyclases participate to c-di-GMP signalling in *Vibrio cholerae*. Environ. Microbiol. **12**:510–523.
13. **Harada S, Ishii Y, Saga T, Tateda K, Yamaguchi K.** 2010. Chromosomally encoded bla_{CMY-2} located on a novel SXT/R391-related integrating conjugative element in a *Proteus mirabilis* clinical isolate. Antimicrob. Agents Chemother. **54**:3545–3550.
14. **Iwanaga M, Toma C, Miyazato T, Insisiengmay S, Nakasone N, Ehara M.** 2004. Antibiotic resistance conferred by a class I integron and SXT constin in *Vibrio cholerae* O1 strains isolated in Laos. Antimicrob. Agents Chemother. **48**:2364–2369.
15. **Juiz-Rio S, Osorio CR, de Lorenzo V, Lemos ML.** 2005. Subtractive hybridization reveals a high genetic diversity in the fish pathogen *Photobacterium damselae subsp. piscicida*: evidence of a SXT-like element. Microbiology **151**:2659–2669.
16. **Ahmed AM, Shinoda S, Shimamoto T.** 2005. A variant type of *Vibrio cholerae* SXT element in a multidrug-resistant strain of *Vibrio fluvialis*. FEMS Microbiol. Lett. **242**:241–247.
17. **Bani S, Mastromarino PN, Ceccarelli D, Le Van A, Salvia AM, Ngo Viet QT, Hai DH, Bacciu D, Cappuccinelli P, Colombo MM.** 2007. Molecular characterization of ICE*Vch*Vie0 and its disappearance in *Vibrio cholerae* O1 strains isolated in 2003 in Vietnam. FEMS Microbiol. Lett. **266**:42–48.
18. **Osorio CR, Marrero J, Wozniak RA, Lemos ML, Burrus V, Waldor MK.** 2008. Genomic and functional analysis of ICE*Pda*SpaI, a fish-pathogen-derived SXT-related integrating conjugative element that can mobilize a virulence plasmid. J. Bacteriol. **190**:3353–3361.
19. **Beaber JW, Hochhut B, Waldor MK.** 2004. SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature **427**:72–74.
20. **Garriss G, Waldor MK, Burrus V.** 2009. Mobile antibiotic resistance encoding elements promote their own diversity. PLoS Genet. **5**:e1000775. doi:10.1371/journal.pgen.1000775.
21. **Datta S, Costantino N, Zhou X, Court DL.** 2008. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. Proc. Natl. Acad. Sci. U. S. A. **105**:1626–1631.
22. **Chen WY, Ho JW, Huang JD, Watt RM.** 2011. Functional characterization of an alkaline exonuclease and single strand annealing protein from the SXT genetic element of *Vibrio cholerae*. BMC Mol. Biol. **12**:16.
23. **Court DL, Oppenheim AB, Adhya SL.** 2007. A new look at bacteriophage lambda genetic networks. J. Bacteriol. **189**:298–304.
24. **Poteete AR.** 2001. What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. FEMS Microbiol. Lett. **201**:9–14.
25. **Murphy KC.** 2007. The lambda Gam protein inhibits RecBCD binding to dsDNA ends. J. Mol. Biol. **371**:19–24.
26. **Datsenko KA, Wanner BL.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. **97**:6640–6645.
27. **Guzman LM, Belin D, Carson MJ, Beckwith J.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. **177**:4121–4130.
28. **Daccord A, Mursell M, Poulin-Laprade D, Burrus V.** 2012. Dynamics of the SetCD-regulated integration and excision of genomic islands mobilized by integrating conjugative elements of the SXT/R391 family. J. Bacteriol. **194**:5794–5802.
29. **Miller JF.** 1992. A short course in bacterial genetics. Cold Spring Harbor Laboratory Press, Plainview, NY.
30. **Pruitt KD, Tatusova T, Klimke W, Maglott DR.** 2009. NCBI reference sequences: current status, policy and new initiatives. Nucleic Acids Res. **37**:D32–36.
31. **Eddy SR.** 2009. A new generation of homology search tools based on probabilistic inference. Genome Inform. **23**:205–211.
32. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–1797.
33. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. **59**:307–321.
34. **Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T.** 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**:1972–1973.
35. **Letunic I, Bork P.** 2011. Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. **39**:W475–478.
36. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**:2947–2948.
37. **Haugen SP, Ross W, Gourse RL.** 2008. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat. Rev. Microbiol. **6**:507–519.
38. **Abreu-Goodger C, Merino E.** 2005. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. Nucleic Acids Res. **33**:W690–W692.
39. **Burrus V, Waldor MK.** 2004. Formation of SXT tandem arrays and SXT-R391 hybrids. J. Bacteriol. **186**:2636–2645.
40. **Beaber JW, Waldor MK.** 2004. Identification of operators and promoters that control SXT conjugative transfer. J. Bacteriol. **186**:5945–5949.
41. **McCool JD, Long E, Petrosino JF, Sandler HA, Rosenberg SM, Sandler SJ.** 2004. Measurement of SOS expression in individual Escherichia coli K-12 cells using fluorescence microscopy. Mol. Microbiol. **53**:1343–1357.
42. **Baharoglu Z, Bikard D, Mazel D.** 2010. Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. PLoS Genet. **6**:e1001165. doi:10.1371/journal.pgen.1001165.
43. **Baharoglu Z, Mazel D.** 2011. *Vibrio cholerae* triggers SOS and mutagenesis in response to a wide range of antibiotics: a route towards multiresistance. Antimicrob. Agents Chemother. **55**:2438–2441.
44. **Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S.** 2005. Switches in bacteriophage lambda development. Annu. Rev. Genet. **39**:409–429.
45. **Kuzminov A.** 1999. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. Microbiol. Mol. Biol. Rev. **63**:751–813.
46. **Szczepanska AK.** 2009. Bacteriophage-encoded functions engaged in initiation of homologous recombination events. Crit. Rev. Microbiol. **35**:197–220.
47. **Li Z, Karakousis G, Chiu SK, Reddy G, Radding CM.** 1998. The beta protein of phage lambda promotes strand exchange. J. Mol. Biol. **276**:733–744.
48. **Marsic N, Roje S, Stojiljkovic I, Salaj-Smic E, Trgovcevic Z.** 1993. *In vivo* studies on the interaction of RecBCD enzyme and lambda Gam protein. J. Bacteriol. **175**:4738–4743.
49. **Ineichen K, Shepherd JC, Bickle TA.** 1981. The DNA sequence of the phage lambda genome between PL and the gene bet. Nucleic Acids Res. **9**:4639–4653.
50. **Muyrers JP, Zhang Y, Buchholz F, Stewart AF.** 2000. RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. Genes Dev. **14**:1971–1982.
51. **Komarova AV, Tchufistova LS, Supina EV, Boni IV.** 2002. Protein S1 counteracts the inhibitory effect of the extended Shine-Dalgarno sequence on translation. RNA **8**:1137–1147.
52. **Waters LS, Storz G.** 2009. Regulatory RNAs in bacteria. Cell **136**:615–628.
53. **Aravind L, Makarova KS, Koonin EV.** 2000. Survey and summary. Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. Nucleic Acids Res. **28**:3417–3432.
54. **Toleman MA, Walsh TR.** 2011. Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. FEMS Microbiol. Rev. **35**:912–935.
55. **Toussaint A, Merlin C.** 2002. Mobile elements as a combination of functional modules. Plasmid **47**:26–35.
56. **Burrus V, Pavlovic G, Decaris B, Guedon G.** 2002. The ICE*St*1 element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. Plasmid **48**:77–97.
57. **Iyer LM, Koonin EV, Aravind L.** 2002. Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. BMC Genomics **3**:8.
58. **Khlebnikov A, Datsenko KA, Skaug T, Wanner BL, Keasling JD.** 2001. Homogeneous expression of the P(BAD) promoter in Escherichia coli by

constitutive expression of the low-affinity high-capacity AraE transporter. Microbiology **147**:3241–3247.
59. **Singer M, Baker TA, Schnitzler G, Deischel SM, Goel M, Dove W, Jaacks KJ, Grossman AD, Erickson JW, Gross CA.** 1989. A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*. Microbiol. Rev. **53**:1–24.
60. **Ceccarelli D, Daccord A, Rene M, Burrus V.** 2008. Identification of the origin of transfer (*oriT*) and a new gene required for mobilization of the SXT/R391 family of ICEs. J. Bacteriol. **190**:5328–5338.
61. **Hochhut B, Waldor MK.** 1999. Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. Mol. Microbiol. **32**:99–110.
62. **Hochhut B, Beaber JW, Woodgate R, Waldor MK.** 2001. Formation of chromosomal tandem arrays of the SXT element and R391, two conjugative chromosomally integrating elements that share an attachment site. J. Bacteriol. **183**:1124–1132.
63. **Zuker M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31**:3406–3415.

ANNEXE 3

Daccord, A., Mursell, M., **Poulin-Laprade, D.**, and Burrus, V. (2012). Dynamics of the SetCD-Regulated Integration and Excision of Genomic Islands Mobilized by Integrating Conjugative Elements of the SXT/R391 Family. J. Bacteriol. *194*, 5794–5802.

L'article, formaté tel que publié dans *Journal of Bacteriology*, est reproduit avec la permission de *ASM Journals*.

JB Journals.ASM.org

Dynamics of the SetCD-Regulated Integration and Excision of Genomic Islands Mobilized by Integrating Conjugative Elements of the SXT/R391 Family

Aurélie Daccord, Mathias Mursell, Dominic Poulin-Laprade, and Vincent Burrus

Centre d'Étude et de Valorisation de la Diversité Microbienne (CEVDM), Département de Biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada

Mobilizable genomic islands (MGIs) are small genomic islands that are mobilizable by SXT/R391 integrating conjugative elements (ICEs) due to similar origins of transfer. Their site-specific integration and excision are catalyzed by the integrase that they encode, but their conjugative transfer entirely depends upon the conjugative machinery of SXT/R391 ICEs. In this study, we report the mechanisms that control the excision and integration processes of MGIs. We found that while the MGI-encoded integrase Int_{MGI} is sufficient to promote MGI integration, efficient excision from the host chromosome requires the combined action of Int_{MGI} and of a novel recombination directionality factor, RdfM. We determined that the genes encoding these proteins are activated by SetCD, the main transcriptional activators of SXT/R391 ICEs. Although they share the same regulators, we found that unlike *rdfM*, int_{MGI} has a basal level of expression in the absence of SetCD. These findings explain how an MGI can integrate into the chromosome of a new host in the absence of a coresident ICE and shed new light on the cross talk that can occur between mobilizable and mobilizing elements that mobilize them, helping us to understand part of the rules that dictate horizontal transfer mechanisms.

Horizontal gene transfer plays a fundamental role in bacterial evolution (22, 24, 28, 32, 33). By transferring from one bacterial genome to another, mobile genetic elements allow bacteria to acquire new DNA fragments encoding a wide array of new functions (16, 26). Genomic islands (GIs) are mobile genetic elements that play a fundamental role in horizontal gene transfer (26). GIs are DNA segments (10 to 550 kb) that are often associated with tRNA genes and exhibit a G+C content usually different from the surrounding chromosome (16, 26). Based upon the functions that they encode, GIs are also known as pathogenicity, symbiosis, metabolic, resistance, or fitness islands (26).

Integrating conjugative elements (ICEs) are self-transmissible GIs found in many Gram-positive and Gram-negative bacteria (8, 11, 36–38, 42). ICEs confer a variety of functions on their host, such as virulence factors, establishment of symbiosis, new metabolic traits, resistance to antibiotics, and factors that enhance bacterial fitness (11). ICEs transfer via conjugation in a conjugative plasmid-like manner, and like many temperate bacteriophages, they integrate into their host's chromosome, along with which they are replicated. The well-studied family of SXT/R391 ICEs includes more than 30 members that are found mostly in clinical and environmental *Vibrio* strains as well as in several other gammaproteobacterial species (7). SXT/R391 ICEs share a conserved set of 52 genes, with nearly half of them encoding proteins necessary for conjugation, integration/excision, and regulation (Fig. 1A) (41). They integrate by site-specific recombination into the 5′ end of *prfC*, a nonessential gene involved in the termination of translation (25). While integration and excision of SXT/R391 ICEs are catalyzed by the site-specific tyrosine recombinase Int_{SXT}, their excision from the chromosome is facilitated by the recombination directionality factor (RDF) Xis (10). Conjugative transfer of SXT/R391 ICEs is initiated at a *cis*-acting locus called the origin of transfer ($oriT_{SXT}$) by the putative relaxase TraI and the auxiliary mobilization protein MobI, which likely form together a nucleoprotein complex called the relaxosome (12). By analogy with conjugative plasmids, translocation of the ICE DNA through the membranes of the donor and the recipient cell is thought to occur as a linear single-stranded DNA molecule covalently bound to TraI (4). Once in the recipient cell, the ICE DNA is recircularized and its complementary strand is synthesized prior to integration into the chromosome. Regulation of excision and transfer of SXT/R391 ICEs is controlled by *setR*, which encodes a λ CI-like transcriptional repressor that represses the expression of *setCD* (4, 5). SetCD is a transcriptional activator complex that triggers the expression of all the genes involved in integration, excision, and conjugative transfer. SetR repression of *setCD* expression is alleviated by DNA damage (5), allowing SetCD to activate excision and transfer of the ICE.

Besides ICEs and bacteriophages, the vast majority of GIs do not have any known mechanisms of transfer and are therefore considered non-self-transmissible. However, such GIs typically harbor functional or cryptic genes that encode site-specific recombinases (integrases) or transposases. Their mechanisms of transfer likely involve the participation of mobilizing self-transmissible elements, such as generalized transducing phages, conjugative plasmids, or ICEs (6). We have recently identified in several genomes of *Vibrio* a new family of GIs that rely on a unique mechanism for gene transfer (13). These mobilizable genomic islands (MGIs) have a size of less than 25 kb and can be mobilized at high frequency by SXT/R391 ICEs using a *cis*-acting *oriT* sequence that mimics $oriT_{SXT}$. MGIs integrate into the 3′ end of *yicC*, a conserved gene encoding a putative stress-induced protein (13). MGI

Received 16 July 2012 Accepted 9 August 2012
Published ahead of print 24 August 2012
Address correspondence to Vincent Burrus, Vincent.Burrus@USherbrooke.ca.

doi:10.1128/JB.01093-12

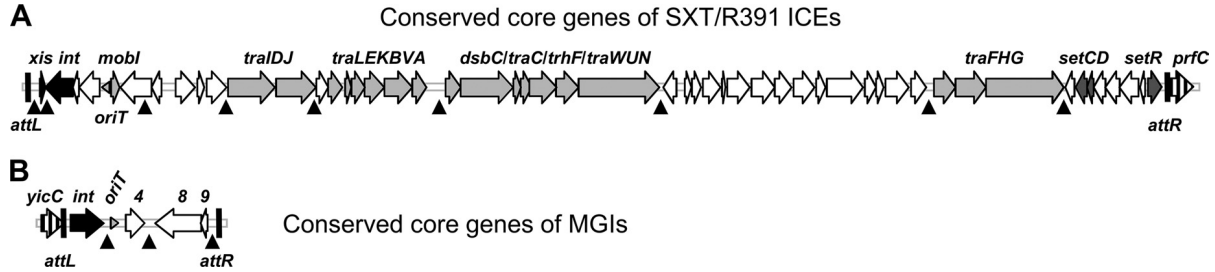

FIG 1 Schematic representation of the core sets of conserved genes of SXT/R391 ICEs (A) and MGIs (B). Vertically hatched open reading frames indicate the integration sites of the elements (*prfC* for SXT/R391 ICEs and *yicC* for MGIs). Black open reading frames represent genes involved in site-specific excision and integration. Light gray open reading frames represent genes encoding the conjugative transfer machinery. Dark gray open reading frames correspond to genes involved in regulation (*setCDR*), and white open reading frames represent genes of unknown function. *oriT*s are represented by horizontal gray arrowheads. Hot spots for insertion of variable DNA are indicated by black arrowheads pointing upward.

integration is catalyzed by the site-specific recombinase Int_{MGI}, a distant relative of Int_{SXT}. Besides int_{MGI} and the $oriT_{SXT}$-like $oriT_{MGI}$ sequence, all MGIs identified to date share only three conserved genes (Fig. 1B), none of which are predicted to encode components of a conjugative transfer machinery or an RDF. Interestingly, while MGI excision is independent of int_{SXT} and *xis*, it requires the presence of the ICE-encoded SetCD transcriptional activators (13).

In this study, we report the identification of the new RDF RdfM, which is required for MGI chromosomal excision. Like int_{MGI}, expression of *rdfM* is activated by SetCD. Comparison of the regulation of the integration/excision genes of SXT/R391 ICEs and that of those of MGIs revealed that they are similarly regulated by SetCD in the donor cells, yet int_{MGI} is expressed independently of SetCD in the recipient cells, allowing MGIs to integrate into the chromosome of a cell lacking an SXT/R391 ICE. To the best of our knowledge, this is the first report of such an intimate interaction between two unrelated families of mobile genetic elements.

MATERIALS AND METHODS

Bacterial strains and media. The bacterial strains and plasmids used in this study are described in Table 1. The strains were routinely grown in

TABLE 1 Strains and plasmids used in this study

Strain or plasmid	Relevant genotype or phenotype[a]	Reference
Escherichia coli strains		
β2163	(F⁻) RP4-2-Tc::Mu Δ*dapA*::(*erm-pir*) (Kn^r Em^r)	15
MC4100 λpir	F⁻ *araD139* Δ(*argF-lac*)*U169 rpsL150* (Sm^r) *relA1 flbB5301 deoC1 ptsF25 rbsR* λ*pir*	19
CAG18439	MG1655 *lacZU118 lacI42*::Tn*10* (Tc^r)	39
VB112	MG1655 Rf^r	12
AD57	CAG18439 *prfC*::ICE*Vfl*Ind1 *yicC*::MGI*Vfl*Ind1 (Tc^r Su^r Tm^r)	13
AD63	CAG18439 *prfC*::ICE*Vfl*Ind1 *yicC*::MGI*Vfl*Ind1::*aph* (Tc^r Su^r Tm^r Kn^r)	13
AD72	CAG18439 *prfC*::SXT *yicC*::MGI*Vfl*Ind1::*aph* (Tc^r Su^r Tm^r Kn^r)	13
AD81	CAG18439 *prfC*::ICE*Vfl*Ind1 *yicC*::MGI*Vfl*Ind1 Δ*int*::*aad7* (Tc^r Su^r Tm^r Sp^r)	13
AD130	CAG18439 *yicC*::MGI*Vfl*Ind1::*aph* (Tc^r Kn^r)	This study
AD132	CAG18439 *yicC*::MGI*Vfl*Ind1::*aph* pGG2B (Tc^r Kn^r Ap^r)	This study
AD133	CAG18439 *prfC*::SXT Δ*setCD yicC*::MGI*Vfl*Ind1::*aph* (Tc^r Su^r Tm^r Kn^r)	This study
AD167	CAG18439 *prfC*::ICE*Vfl*Ind1 *yicC*::MGI*Vfl*Ind1 Δ*cds9*::*aph* (Tc^r Su^r Tm^r Kn^r)	This study
AD169	CAG18439 *prfC*::ICE*Vfl*Ind1 *yicC*::MGI*Vfl*Ind1 Δ*cds4*::*aph* (Tc^r Su^r Tm^r Kn^r)	This study
AD192	CAG18439 *prfC*::ICE*Vfl*Ind1 *yicC*::MGI*Vfl*Ind1 Δ*cds8*::*aph* (Tc^r Su^r Tm^r Kn^r)	This study
AD207	CAG18439 *prfC*::R997 *yicC*::MGI*Vfl*Ind1::*aph* (Tc^r Ap^r Kn^r)	This study
AD208	CAG18439 pIntVvu (Tc^r Ap^r)	This study
AD210	β2163 pVB200 (Em^r Cm^r)	This study
AD217	CAG18439 *yicC*::pVB200 pIntVvu (Tc^r Cm^r Ap^r)	This study
AD232	β2163 pSW23T (Em^r Cm^r)	This study
Plasmids		
pIntVvu	pBAD-TOPO $int_{MGIVvuTai1}$ (Ap^r)	13
pGG2B	pBAD30 *setCD* (Ap^r)	G. Garriss
pSW23T	$oriT_{RP4}$; $oriV_{R6K\gamma}$ (Cm^r)	15
pVB200	pSW23T $attP_{MGIVflInd1}$ (Cm^r)	This study
p8	pBAD-TOPO $cds9_{MGIVvuTai1}$ (Ap^r)	This study
p9	pBAD-TOPO $cds9_{MGIVvuTai1}$ (Ap^r)	This study
p9-8	pBAD-TOPO $cds9\text{-}8_{MGIVflInd1}$ (Ap^r)	This study
pKD13	PCR template for one-step chromosomal gene inactivation (Kn^r)	14

[a] Ap^r, ampicillin resistant; Cm^r, chloramphenicol resistant; Kn^r, kanamycin resistant; Em^r, erythromycin resistant; Rf^r, rifampin resistant; Sm^r, streptomycin resistant; Sp^r, spectinomycin resistant; Su^r, sulfamethoxazole resistant; Tc^r, tetracycline resistant; Tm^r, trimethoprim resistant.

TABLE 2 DNA sequences of the primers used in this study

Primer name	Nucleotide sequence (5′ to 3′)	Use in the study
AD4-V-F	TAGCAGTGAGGAAGCAAACGATG	Amplification of $cds9_{MGIVvuTai1}$
AD4-R1	TTATCCACGGCCATAAGCAGC	Amplification of $cds9_{MGIVvuTai1}$
AD5-F	GCCGTGGATAAACCATCAGCA	Amplification of $cds8_{MGIVvuTai1}$
AD5-V-R1	TTAGTCATCCAAAATACTGCCTTT	Amplification of $cds8_{MGIVvuTai1}$
AD5-A-R1	TTAGTCATCCAAGATGCTGCCTTT	Amplification of $cds9\text{-}8_{MGIVflInd1}$
AD4-A-F	TAGCCGATTAGTACTGGCAAACTCC	Amplification of $cds9\text{-}8_{MGIVflInd1}$
AD11-WF	CAGCCCACGGCACGCGCACCAATACGAATGGAACGTGTGTAGGCTGGAGCTGCTTCG	Deletion of *cds9* in MGI*Vfl*Ind1
AD11-WR	ATGAACCCAACTACACAATCATCCACCACATCAACAATTCCGGGGATCCGTCGACC	Deletion of *cds9* in MGI*Vfl*Ind1
AD13-WF	AGTGCTAACGATTGGGATAGAGAATGGATACAGCAGGTGTAGGCTGGAGCTGCTTCG	Deletion of *cds4* in MGI*Vfl*Ind1
AD13-WR	CAGCGCCCTGTGAGGGGTTACTCTTTTTTCAGGCCTATTCCGGGGATCCGTCGACC	Deletion of *cds4* in MGI*Vfl*Ind1
attPAD-L1	TCGGCTTTGCTGTATGCAATA	Amplification of $attP_{MGI}$, 1st round
attPAD-R1-AC	TCTGCCATAGCAACAGCAAT	Amplification of $attP_{MGI}$, 1st round
attPAD-L2	GAGTTTTCCCATGTTTACTCCATA	Amplification of $attP_{MGI}$, 2nd round
attPAD-R2-AC	GTGACAGCTTTGCCTGCTT	Amplification of $attP_{MGI}$, 2nd round
Gene8-WF	TATGCCTAGCAACATGCCAAAATTACCAGCTGGTTTGTGTAGGCTGGAGCTGCTTCG	Deletion of *cds8* in MGI*Vfl*Ind1
Gene8-WR	TGACTTTGCGCTGTGGTCGGTTGCCATCGGGGATTAATTCCGGGGATCCGTCGACC	Deletion of *cds8* in MGI*Vfl*Ind1
Q-PCR-1F	AAGTGACAAACTCCGCCATC	Amplification of *attB* in *E. coli*
Q-PCR-1R	GCACGCAAAACAGAATTGAA	Amplification of *attB* in *E. coli*
Q-PCR-2F	GAAAACGGCAAGCTGAAAAC	Amplification of *rph* in *E. coli*
Q-PCR-2R	GTCCCCTGCACTTCAATGAT	Amplification of *rph* in *E. coli*
RTgene9-F b	TTATCCACGGCCATAAGCAG	Amplification of $cds9_{MGIVflInd1}$
RTgene9-R b	AGGTAAAGGCCAACTCAGGCTT	Amplification of $cds9_{MGIVflInd1}$
RTintVf-F	CCGTATCGGGTTTACACCAA	Amplification of $int_{MGIVflInd1}$
RTintVf-R	TTATCGCATGTCGAAACAGC	Amplification of $int_{MGIVflInd1}$
RTrpoZcoli-F	GCTCGTCAGATGCAGGTAGG	Amplification of *rpoZ*
RTrpoZcoli-R	GCTTGTAATTCAGCGGCTTC	Amplification of *rpoZ*
RTyicC-F	GAGTGACGAAGGGGAAATCA	Amplification of *yicC*
RTyicC-R	GCTTTCAGTGCCTGACCTTC	Amplification of *yicC*

Luria-Bertani (LB) broth at 37°C in an orbital shaker/incubator and were maintained at −80°C in LB broth containing 15% (vol/vol) glycerol. Antibiotics were used at the following concentrations: ampicillin (Ap), 100 μg/ml; kanamycin (Kn), 50 μg/ml; rifampin (Rf), 50 μg/ml; spectinomycin (Sp), 50 μg/ml; sulfamethoxazole (Su), 160 μg/ml; tetracycline (Tc), 12 μg/ml; and trimethoprim (Tm), 32 μg/ml. When required, bacterial cultures were supplemented with 0.3 mM DL-α,ε-diaminopimelic acid (DAP), 100 ng/ml mitomycin C, or 0.02% L-arabinose.

Plasmids and strain constructions. Plasmids and primers used in this study are described in Tables 1 and 2, respectively. Plasmid pVB200 was constructed by subcloning the XbaI-flanked digestion product $attP_{MGIVflInd1}$ into the XbaI site of pSW23T. Product $attP_{MGIVflInd1}$ was amplified using genomic DNA of *Vibrio fluvialis* H-08942 as a template and primer pair attPAD-L1/attPAD-R1-AC for the first round and primer pair attPAD-L2/attPAD-R2-AC for the second round and then cloned into vector pCR2.1-TOPO (Invitrogen). Plasmids p9, p8, and p9-8 were constructed by cloning $cds9_{MGIVvuTai1}$, $cds8_{MGIVvuTai1}$, or $cds9\text{-}8_{MGIVflInd1}$ into the TA cloning expression vector pBAD-TOPO (Invitrogen) according to the manufacturer's instructions. $cds9_{MGIVvuTai1}$, $cds8_{MGIVvuTai1}$, and $cds9\text{-}8_{MGIVflInd1}$ were amplified by PCR from genomic DNA of *Vibrio vulnificus* YJ016 or *V. fluvialis* H-08942 as a template using primer pairs AD4-V-F/AD4-R1, AD5-F/AD5-V-R1, and AD5-A-R1/AD4-A-F, respectively (Table 2).

All deletion mutants were constructed in *Escherichia coli* AD57 using the one-step chromosomal gene inactivation technique (14). All mutations were designed to be nonpolar. The Δ*cds4*, Δ*cds8*, and Δ*cds9* mutations were introduced in MGI*Vfl*Ind1 using primer pairs AD13-WF/AD13-WR, Gene8-WF/Gene8-WR, and AD11-WF/AD11-WR (Table 2), respectively, and pKD13 as the template. All deletion mutations were verified by PCR amplification using primers flanking the deletion.

Bacterial conjugation. Conjugation assays were used to transfer SXT, R997, MGI*Vfl*Ind1, or plasmids into *E. coli*. Mating assays were performed by mixing equal volumes of overnight cultures of donor and recipient strains. The cells were harvested by centrifugation and resuspended in a 1/20 volume of LB broth. Cell suspensions were poured onto LB agar plates and incubated at 37°C for 6 h. The cells were then resuspended in 1 ml of LB medium, and serial dilutions were plated onto appropriate selective media to determine the number of donors, recipients, and exconjugants. Frequency of transfer was expressed as the number of exconjugant cells per recipient cell in the mating mixture at the time of plating. *E. coli* CAG18439, MC4100 λpir, or VB112 was used as the recipient in conjugation experiments (Table 1). To induce expression of Int_{MGI} from pIntVvu, SetCD from pGG2B, protein 8 from p8, protein 9 (RdfM) from p9, or proteins 9 and 8 from p9-8 (Table 1) in complementation assays, mating experiments were carried out on LB agar plates supplemented with 0.02% L-arabinose.

Molecular biology techniques. All the enzymes were used according to the manufacturer's instructions (New England BioLabs). Plasmid DNA was prepared with a QIAprep Spin miniprep kit (Qiagen), and chromosomal DNA was prepared with a Wizard Genomic DNA purification kit (Promega) as described in the manufacturer's instructions.

PCR assays were carried out in 50-μl PCR mixtures with 1 U of *Taq* DNA polymerase (New England BioLabs). The PCR conditions were as follows: (i) 3 min at 94°C; (ii) 30 cycles of 30 s at 94°C, 30 s at a suitable annealing temperature, and 30 s to 60 s at 72°C; and (iii) 5 min at 72°C. When needed, PCR products were purified using a QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions. The purified PCR products or inserts of constructed plasmids were sequenced by Centre d'Innovation Génome Québec (McGill University, Montréal, Québec, Canada). DNA sequences were compared with the GenBank DNA sequence database using the BLASTN program (3). *E. coli* was transformed by electroporation in 1-mm-gap cuvettes according to the method of Dower et al. (18), using a GenePulser Xcell apparatus (Bio-Rad) set at 25 μF, 200 Ω, and 1.8 kV.

Real-time quantitative PCR assays for relative quantification of *attB* and *rph*. Real-time quantitative PCR assays were used to measure the percentages of cells in a culture that contained an unoccupied MGI *attB* site (the 3′ end of *yicC*) as described previously (13). Briefly, this corresponds to a comparison of the amounts of excised circularized MGI relative to the amounts of chromosome copies deduced from the amplification of *rph*, a gene located immediately 5′ of *yicC*. Primer pairs Q-PCR-1F/Q-PCR-1R and Q-PCR-2F/Q-PCR-2R were used for the amplification of *attB* and *rph*, respectively (Table 2).

RNA extraction and cDNA synthesis. Bacterial cultures were grown at 37°C to early exponential phase (optical density at 600 nm [OD_{600}], 0.2). Cultures were split in two, and induction was initiated by addition of 100 ng/ml mitomycin C or 0.02% L-arabinose. Two hours after induction, aliquots of bacterial cultures were directly mixed with RNA Protect bacterial reagent (Qiagen) and treated according to the manufacturer's instructions. Bacterial RNA was isolated after treating the cells with lysozyme (Sigma), using the RNeasy minikit (Qiagen). In addition, RNA samples were treated with DNase (RNase-free DNase set; Qiagen) during purification and Turbo DNase (Ambion) after purification. RNA purity and concentration were evaluated with an ND-1000 NanoDrop spectrophotometer (Thermo Fisher Scientific/NanoDrop Products). cDNA was prepared using SuperScript II (Invitrogen) according to the manufacturer's recommendations. Fifty nanograms of random hexamers (Integrated DNA Technologies) and 1 μg of total bacterial RNA were used in each reaction. After synthesis, cDNA sample mixtures were purified with the PCR purification kit (Qiagen) and stored at −20°C.

Reverse transcription quantitative PCR. The MasterCycler EP Realplex 4 sequence detection system (Eppendorf) was used to quantify the increase in fluorescence emission of SYBR green I during PCR. The Realplex software (version 1.5; Eppendorf) was used for data acquisition and analysis. Each 25-μl reaction mixture contained 12.5 μl of 2× SYBR green PCR Master Mix (Qiagen), 1 μM (each) primer, and 1 μl of cDNA template. Primer pairs RTgene9-F b/RTgene9-R b, RTintVf-F/RTintVf-R, RTrpoZcoli-F/RTrpoZcoli-R, and RTyicC-F/RTyicC-R were used for the amplification of $cds9_{MGIVflInd1}$, $int_{MGIVflInd1}$, *rpoZ*, and *yicC*, respectively (Table 2). The PCR conditions were (i) 5 min at 95°C, (ii) 45 cycles of 10 s at 95°C and 30 s at 60°C, (iii) 15 s at 95°C, (iv) 15 s at 60°C, (v) melting curve from 60°C to 95°C, and (vi) 15 s at 95°C. Three reactions were performed for each sample. For normalization, the *rpoZ* gene was used and results were expressed as relative expression based on the threshold cycle ($\Delta\Delta C_T$) calculation method. Experiments were carried out three times and combined.

RESULTS

Int_{MGI} is the only MGI-encoded protein necessary for MGI integration. In a previous study, we showed that Int_{MGI} is required for integration and excision of MGI*Vfl*Ind1 (13). To examine whether Int_{MGI} is the sole MGI-encoded protein necessary to mediate MGI's integration into the 3′ end of *yicC*, Int_{MGI}-mediated recombination between $attP_{MGI}$ and *attB* at *yicC* was monitored using pVB200, a derivative of the mobilizable Cm^r suicide vector pSW23T harboring $attP_{MGI}$ (Fig. 2A). Since pSW23T requires the product of *pir* to replicate, Cm^r exconjugants can be isolated after its conjugative transfer from a pir^+ host to a *pir* host only if it has integrated into the chromosome.

For a negative control, we first mobilized empty pSW23T from a mob^+ pir^+ donor strain to CAG18439 or CAG18439 harboring pIntVvu (AD208), a plasmid expressing Int_{MGI} under the control of an arabinose-inducible promoter. In both cases, the frequency of exconjugant formation was below 5×10^{-6} exconjugants/recipient, a value that we established as our baseline for subsequent experiments (Fig. 2B). The few recovered exconjugants can be attributed to random integration of the plasmid into the recipient chromosome. For a positive control, we also mobilized pVB200 from the same donor strain to a pir^+ strain (MC4100 λ*pir*) to verify that the constructed plasmid remained mobilizable. We found that up to 39% of the pir^+ recipient cells acquired and maintained the plasmid. We then mobilized pVB200 from the same donor strain to CAG18439 or AD208 (Fig. 2B). When Int_{MGI} was expressed in the *pir* recipient strain, the frequency of exconjugant formation was as high as that of the positive control, indicating that the plasmid was able to maintain itself by site-specific integration into the recipient's chromosome. Thus, we conclude that Int_{MGI} is the only MGI-encoded protein needed to mediate efficient integration of MGIs.

MGI integration does not require activation by SetCD. In our initial study of MGIs, we showed that the SXT/R391 ICE-encoded transcriptional activator SetCD is necessary for MGI excision from the chromosome, suggesting that it is required to activate the expression of Int_{MGI} (13). Surprisingly, we also observed that colonies harboring an MGI but devoid of any ICE were also recovered at high frequency in exconjugant populations. This observation is consistent with the natural occurrence of environmental and clinical isolates (*Vibrio cholerae* RC385 and *V. vulnificus* YJ016) having similar configurations (13) and suggests that while SetCD is necessary for excision, it is not required for *de novo* expression of Int_{MGI} in the recipient cells. To test this hypothesis, we mobilized pVB200 into CAG18439 harboring MGI*Vfl*Ind1 alone (AD130) or along with either R997 (AD207), an Ap^r-conferring ICE of the SXT/R391 family, or pGG2B (AD132), a plasmid expressing SetCD under the control of an arabinose-inducible promoter (Fig. 2C). Interestingly, exconjugants formed at high frequency in the sole presence of MGI*Vfl*Ind1 whereas the presence of R997 or expression of SetCD in the recipient cells did not significantly improve transfer and maintenance of pVB200, supporting the notion that SXT/R391 ICEs and SetCD are necessary for MGI's excision but not its integration.

Int_{MGI} alone does not promote efficient excision of MGIs. Next, we examined whether Int_{MGI} alone was able to promote efficient excision of pVB200 integrated into *yicC*. We used a semiquantitative PCR assay to detect unoccupied *attB* sites in the cell populations compared to *rph* as a reference target. The formation of an *attB* site was tested in CAG18439 as a positive control and in CAG18439 harboring *attB*::pVB200 along with pIntVvu (AD217), MGI*Vfl*Ind1 (AD130), MGI*Vfl*Ind1 and R997 (AD207), or MGI*Vfl*Ind1 and pGG2B (AD132). We found that Int_{MGI} alone did not mediate efficient excision, even when overexpressed (Fig. 2D, lanes 1 to 3). In fact, excision was detectable only in the presence of MGI*Vfl*Ind1 either along with R997 or upon expression of SetCD from pGG2B (Fig. 2D, lanes 4, 5, and 7). These results led us to suppose that an unidentified MGI-encoded factor likely helps Int_{MGI} to mediate efficient site-specific excision and that expression of this factor is likely activated by SetCD.

MGIs encode a putative RDF. Considering that Int_{MGI} is required but not sufficient to promote efficient excision, we looked at the genes conserved among sequenced MGIs to identify an RDF that could facilitate the Int_{MGI}-mediated excision of MGIs. RDFs control the directionality of tyrosine recombinase-mediated site-specific recombination events (30) and are usually small basic proteins (<100 amino acids) with or without a putative helix-turn-helix (HTH) DNA-binding motif. Besides int_{MGI}, only 3 genes are common to all MGIs identified and sequenced to date: *cds4* encodes a 214-amino-acid protein of unknown function, *cds8* encodes a 580-amino-acid putative helicase, and *cds9* encodes an

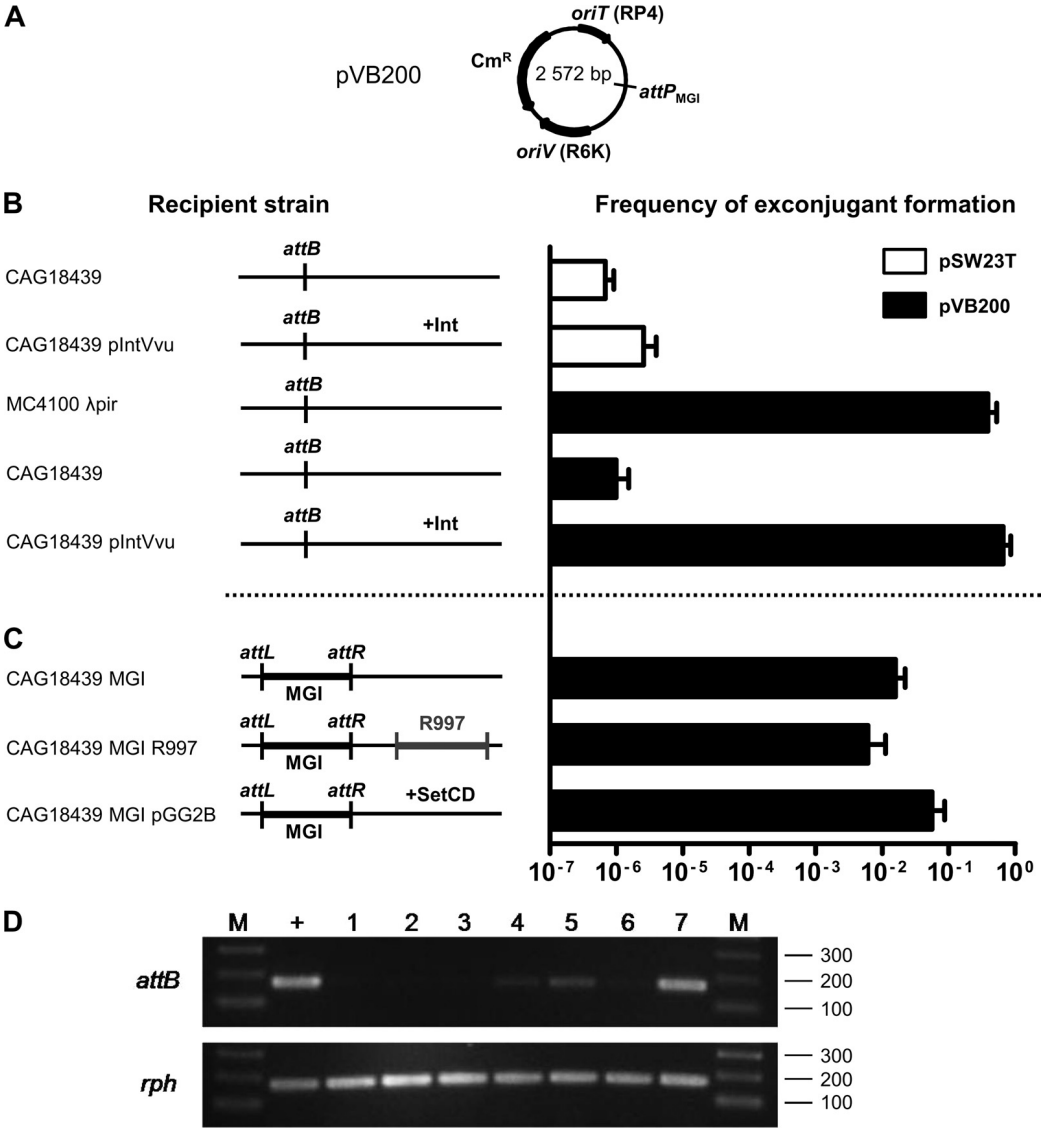

FIG 2 Genetic requirements for integration and excision of a replication-deficient plasmid containing the *attP* site of MGI*Vfl*Ind1. (A) Schematic map of pVB200. (B and C) Mobilization assays of pSW23T and pVB200 performed to assess plasmid integration into the 3′ end of *yicC* (*attB*). Conjugation assays were carried out using *E. coli* β2163 (*pir*$^+$) as a donor and MC4100 λpir (*pir*$^+$) or CAG18439 variants (*pir*) as recipient strains. The genetic background of each recipient strain is indicated on the left side of the panels. R997 is an Ap^r-conferring ICE of the SXT/R391 family. To induce expression of Int_{MGI} from pIntVvu or of SetCD from pGG2B, the conjugation assays were carried out on media supplemented with 0.02% arabinose. The frequency of exconjugant formation was obtained by dividing the number of exconjugants (Tc^r Cm^r CFU for CAG18439 or Sm^r Cm^r for MC4100 λpir) by the number of recipients (Tc^r or Sm^r CFU, respectively). The bars indicate the mean values and standard deviations obtained from three independent experiments. (D) Analysis of excision of pVB200 integrated into *yicC* (*attB*). Ethidium bromide-stained 2% agarose gel of *attB* and *rph* fragments amplified by semiquantitative PCR. Lanes: M, 2-log molecular size marker; +, CAG18439; 1 and 2, CAG18439 *yicC*::pVB200 pIntVvu; 3, CAG18439 *yicC*::pVB200-MGI*Vfl*Ind1; 4 and 5, CAG18439 *yicC*::pVB200-MGI*Vfl*Ind1 *prfC*::R997; 6 and 7, CAG18439 *yicC*::pVB200-MGI*Vfl*Ind1 pGG2B. Lanes 2 and 7, cultures were induced with 0.02% arabinose; lane 5, culture was induced with 100 ng/ml mitomycin C.

80-amino-acid predicted transcriptional regulator (Fig. 1B). Interestingly, the translation product of *cds9* shares 36% identity with Hef encoded by the high-pathogenicity island (HPI) of *Yersinia pseudotuberculosis* and 29% identity with AlpA encoded by *E. coli* prophage CP4-57 (Fig. 3). Hef has been reported to act as an RDF (29), whereas AlpA has been reported to activate the expression of its cognate integrase gene (27). Given the size of the predicted translation product of *cds9* and its similarity with Hef, we considered it to be a good candidate for a putative RDF. Yet, given its similarity with the transcriptional regulator AlpA, we could not rule out at this point the possibility that the product of *cds9* could activate the expression of int_{MGI}.

Deletion of *cds9* dramatically affects MGI excision and transfer. To assess whether one of the three MGI conserved genes could act as an RDF, we first constructed deletion mutants of each gene in MGI*Vfl*Ind1. We tested the ability of each mutant to be mobilized by ICE*Vfl*Ind1. While deletions of *cds4* or *cds8* had virtually no impact, deletion of *cds9* led to a dramatic reduction of the MGI frequency of transfer (Fig. 4). Mobilization of MGI*Vfl*Ind1 Δ*cds9* could be partially restored when *cds9* was expressed in *trans* from

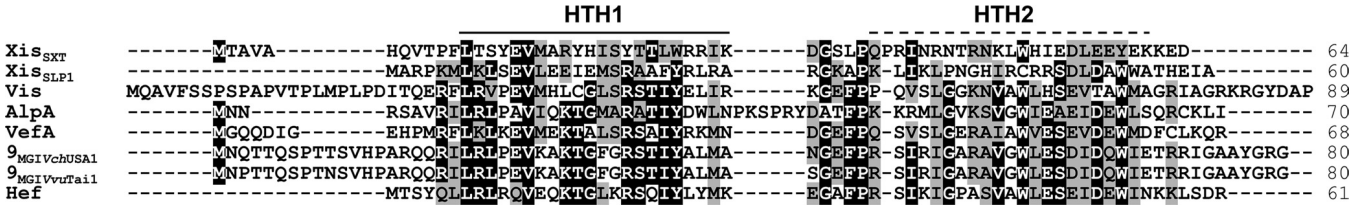

FIG 3 Sequence alignment of the translation products of *cds9*$_{MGIVvuTai1}$ and *cds9*$_{MGIVchUSA1}$ with related RDFs. The primary sequences of RDFs encoded by two sequenced MGIs were aligned using MUSCLE with the transcriptional regulator AlpA from CP4-57 prophage (NP_417113) and the RDFs Xis encoded by ICEs of the SXT/R391 family (ACV96240), Hef of HPI from *Y. pseudotuberculosis* (CAB46594), VefA of VPI-2 from *V. cholerae* N16961 (NP_231420), SLP1 of plasmid SLP1 from *Streptomyces coelicolor*, and Vis of the satellite bacteriophage P4 (NP_042041). Amino acid residues that are identical or similar (BLOSUM62 substitution matrix) in at least 60% of the sequences are indicated by a black or gray background, respectively. The solid bar indicates a helix-turn-helix (HTH) DNA-binding motif predicted in all proteins based on the Dodd-Egan method (17) (Dodd-Egan scores of 2.36 or higher), whereas the dashed bar highlights a secondary HTH motif exclusively predicted in RDFs encoded by MGIs (Dodd-Egan score of 3.07). The length of each protein is indicated in the right column.

an inducible promoter. Complementation with *cds8* did not restore MGI transfer, whereas complementation with *cds9-cds8* expressed from the same inducible promoter restored mobilization to the same level as that with *cds9* alone, ruling out the possible polar effects of the Δ*cds9* deletion on *cds8* that could have explained the partial complementation phenotype observed upon *cds9* overexpression.

To investigate whether the reduction of MGI*Vfl*Ind1 transfer was a consequence of reduced or abolished excision caused by deletion of *cds9*, we conducted real-time quantitative PCR assays to measure the percentage of cells in a culture containing unoccupied *attB* sites and found that excision of MGI*Vfl*Ind1 Δ*cds9* was undetectable (Fig. 4). In contrast, excision of the same mutant was dramatically enhanced (50-fold over wild-type level) when *cds9* was expressed in *trans*. Interestingly, the rate of excision of the mutant was restored to wild-type level when *cds9* and *cds8* were expressed together in *trans*, suggesting a possible regulatory activity of the protein encoded by *cds8*. These results indicate that the product of *cds9* plays an important role in MGI excision, either by acting as an RDF or by activating *int*$_{MGI}$ expression, or both.

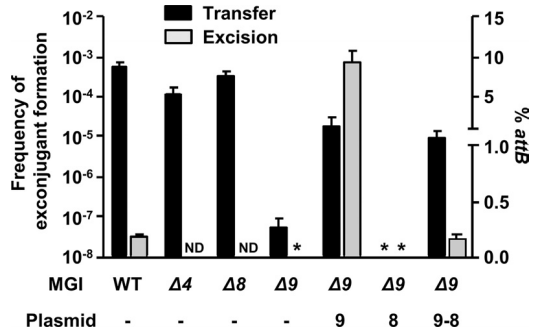

FIG 4 Genetic requirements for excision and transfer of MGI*Vfl*Ind1. Mobilization assays of MGI*Vfl*Ind1::*aph* or its Δ*cds4*, Δ*cds8*, or Δ*cds9* mutants by ICE*Vfl*Ind1 were carried out using *E. coli* CAG18439 containing ICE*Vfl*Ind1 and MGI*Vfl*Ind1 mutants as donors. When indicated, the donor expressed *cds9*, *cds8*, or *cds9* and *cds8* from p9, p8, or p9-8, respectively. *E. coli* VB112 (Rf^r) was used as the recipient strain. The frequency of exconjugant formation was calculated by dividing the number of exconjugants (Rf^r Km^r CFU) by the number of donors (Tc^r CFU). Real-time quantitative PCR was used to determine the percentage of unoccupied *attB* sites resulting from the circularization of MGI*Vfl*Ind1 or of its Δ*cds9* mutant in *E. coli* CAG18439 harboring ICE*Vfl*Ind1, p9, p8, or p9-8. The bars indicate the mean values and standard deviations obtained from three independent experiments. ND, not determined. Asterisks indicate that the frequency of exconjugant formation or the percentage of *attB* sites was below the limit of detection of the assays ($<1 \times 10^{-8}$ or 0.0004%, respectively).

The product of *cds9* acts as an RDF, not as a transcriptional activator. AlpA was shown to activate the expression of the integrase gene of the cryptic prophage CP4-57 in *E. coli* (27). This ability prompted us to investigate whether the product of *cds9* could act as a transcriptional activator of *int*$_{MGI}$. Expression of *int*$_{MGI}$ was measured by reverse transcription real-time quantitative PCR in the presence or absence of mitomycin C and in different genetic backgrounds, including cells devoid of ICE (AD130), cells containing an ICE and a wild-type (AD72) or Δ*cds9* (AD167) copy of MGI*Vfl*Ind1, or cells containing both ICE*Vfl*Ind1 and MGI*Vfl*Ind1 and expressing *cds9* from an inducible promoter (Fig. 5A). We found that *int*$_{MGI}$ expression was strongly stimulated by the addition of the DNA-damaging agent mitomycin C, but only in the presence of an SXT/R391 ICE. The absence of *cds9* had no effect on the mitomycin C-induced activation of *int*$_{MGI}$ expression. Overexpression of *cds9* in the absence of mitomycin C induction led to a slight yet nonsignificant activation of *int*$_{MGI}$ expression. This barely detectable level of activation is probably not dependent upon *cds9* expression, but rather the result of the expression of SetCD in a subpopulation of cells inherently expressing the SOS response, as reported by McCool et al. (31). This phenomenon also explains the constitutive basal level of transfer of SXT/R391 ICE in the absence of DNA-damaging agents. These results combined with our previous observations on SetCD-mediated activation of MGI excision indicate that the product of *cds9* acts as an RDF rather than an activator of *int*$_{MGI}$ expression. From now on, *cds9* will therefore be referred to as *rdfM* for recombination directionality factor of MGI.

Expression of both *int*$_{MGI}$ and *rdfM* is activated by SetCD. Given that SetCD activates MGI excision (13) and that DNA-damaging agents stimulate *int*$_{MGI}$ expression, we hypothesized that SetCD acts as a transcriptional activator of both *int*$_{MGI}$ and *rdfM*. To verify this hypothesis, we measured the expression of *int*$_{MGI}$ and *rdfM* in *E. coli* cells carrying MGI*Vfl*Ind1 and SXT (AD72) or SXT Δ*setCD* (AD133) or expressing SetCD from pGG2B (AD132). We also measured the expression of *yicC* in the same cells since it is described in GenBank as a gene coding for a putative stress-induced protein and the relative positions and orientations of *yicC* and *int*$_{MGI}$ suggest that the two genes may be cotranscribed. Induction was carried out with either mitomycin C (AD72 and AD133) or L-arabinose (AD132). First, we observed

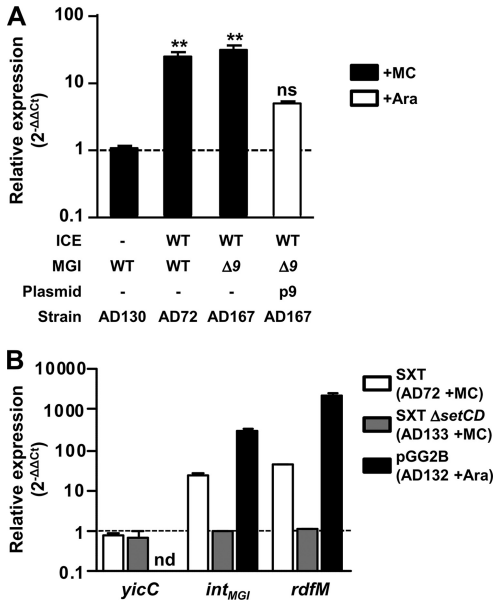

FIG 5 Regulation of the expression of integration and excision genes of MGIs and SXT/R391 ICEs. (A) Impact of protein 9 on the expression of int_{MGI}. The expression of int_{MGI} was measured by quantitative PCR upon induction of the SOS response (+MC) in AD72, AD130, and AD167 or overexpression of *cds9* from p9 in AD167 p9 (+Ara). One-way analysis of variance with a Dunnett posttest was used to compare the means of relative expression of int_{MGI} against the strain devoid of an ICE. The confidence interval for the comparisons was $P < 0.01$ (**). ns, nonsignificant. (B) Impact of SetCD on int_{MGI} and *rdfM* expression. The expression of *yicC*, int_{MGI}, and *rdfM* was measured upon induction of the SOS response (+MC) in AD72 and AD133 or upon overexpression of *setCD* from pGG2B (+Ara) in AD132.

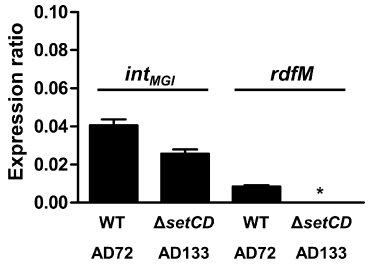

FIG 6 int_{MGI} has a basal level of expression in the absence of SetCD. The expression of int_{MGI} was measured by quantitative PCR in AD72 and AD133. The graph shows differential gene expression values (ΔC_T) compared with that of the housekeeping gene *rpoZ*. Results are expressed as the means of three independent biological replicates. The asterisk indicates that expression was below the detection limit. WT, wild type.

that neither SetCD nor mitomycin C modulates the expression of *yicC*, ruling out the possible expression of int_{MGI} from the promoter of *yicC* when the MGI is integrated into its 3′ end (Fig. 5B). In contrast, in the presence of wild-type SXT, mitomycin C was found to stimulate the expression of both int_{MGI} and *rdfM* (24- and 44-fold increase, respectively), whereas it had no effect in the presence of the SXT Δ*setCD* mutant. This stimulation of expression is attributable to the increased expression of SetCD from SXT, as the expression of SetCD from pGG2B in a strain lacking SXT resulted in increases of ~300- and ~2,000-fold in the transcript levels of int_{MGI} and *rdfM*, respectively.

int_{MGI} is constitutively expressed, allowing MGI integration in a strain devoid of an SXT/R391 ICE. We previously reported that upon mating with an *E. coli* donor strain harboring MGI*Vfl*Ind1 and ICE*Vfl*Ind1, ~98% of the isolated exconjugant colonies selected for the MGI were devoid of ICE*Vfl*Ind1, highlighting the independence of MGIs from ICEs for their integration into the chromosome (13). This result is supported by naturally occurring isolates containing MGIs but devoid of SXT/R391 ICEs (13). However, it contrasts with our abovementioned expression results indicating that SetCD activates the expression of int_{MGI}. To explain how MGIs integrate into *yicC* in the absence of ICE-encoded SetCD, we had a closer look at int_{MGI} expression data under non-induced conditions, which revealed that int_{MGI} has a low-level constitutive expression. In the presence of SXT (AD72), int_{MGI} and *rdfM* exhibit detectable $2^{-\Delta\Delta CT}$ values of 0.041 and 0.01 relative to *rpoZ*, respectively (Fig. 6). This level of expression is most likely a consequence of spontaneous induction of the SOS response (31). When the same experiment was carried out using SXT Δ*setCD* (AD133), *rdfM* expression was below the limit of detection whereas expression of int_{MGI} was reduced by only 36% (Fig. 6). This result, which is also supported by the high rate of exconjugant formation upon mobilization of pVB200 to a strain containing MGI*Vfl*Ind1 but lacking an ICE (Fig. 2C), indicates that int_{MGI} is constitutively expressed at a low level in the absence of SetCD. This basal level of expression is necessary and sufficient to promote integration of MGIs into the chromosome of a new host in the absence of a helper SXT/R391 ICE.

DISCUSSION

In this study, we investigated the integration and excision dynamics of MGIs. We found that while the MGI-encoded integrase alone is sufficient to promote efficient MGI integration into the chromosome, it also requires the MGI-encoded RDF RdfM to promote efficient MGI excision. We found that both int_{MGI} and *rdfM* are activated by the SXT/R391 ICE-encoded transcriptional regulator SetCD. However, the expression of int_{MGI} does not strictly require SetCD whereas the expression of *rdfM* does. These findings help to establish how an MGI cannot excise from the chromosome of a cell devoid of an SXT/R391 ICE but can integrate into the chromosome of such a cell. Accordingly, we propose a model of the regulation pathways responsible for the excision and integration processes of MGIs in the donor and recipient cells (Fig. 7).

Integration and excision are critical steps in the maintenance and dissemination of an integrative mobile genetic element, whether it is a temperate bacteriophage, an ICE, or a mobile genomic island. Site-specific integration typically requires the action of a single mobile element-encoded site-specific recombinase and may require the help of host-encoded nucleoid proteins, such as the integration host factor (IHF) and the factor for inversion stimulation Fis (reviewed in references 20 and 21). In contrast, the reverse recombination event, the site-specific excision, usually requires an additional genetic element-encoded protein, the RDF, also known as Xis/excisionase, although it usually lacks a catalytic activity *per se*. RDFs are usually small basic proteins that play architectural roles in the recombination events catalyzed by their cognate tyrosine or serine recombinase. Given their small size, RDF-encoding genes can be difficult to identify, and since a subset of RDFs harbor putative helix-turn-helix domains, they are often annotated as putative transcriptional regulators.

Lewis and Hatfull divided RDFs into 11 subgroups based on

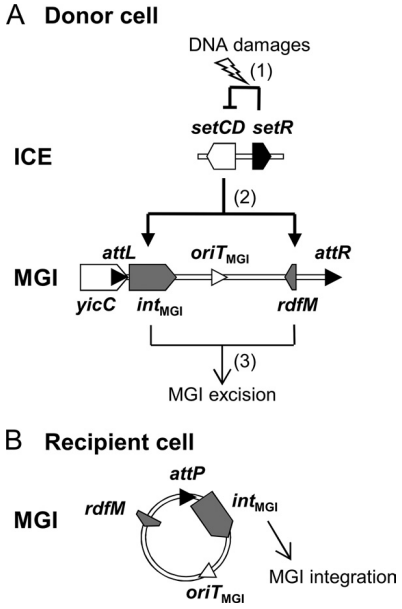

FIG 7 Integration and excision dynamics of MGIs in donor and recipient cells. (A) Excision from the donor cell chromosome. (1) SOS response is activated by DNA damages, alleviating the SetR-mediated repression of *setCD*. (2) The transcriptional activator SetCD activates the expression of int_{MGI} and *rdfM*. (3) Int_{MGI} and RdfM mediate the excision of the MGI. (B) Integration into the recipient cell chromosome. int_{MGI} is expressed at a low level and allows MGI integration regardless of the presence or absence of an SXT/R391 ICE.

sequence similarity (30). We showed here that RdfM belongs to the SLP1 subgroup of RDFs and as such has a putative conserved N-terminal HTH DNA-binding motif (HTH1) (Fig. 3). Peculiarly, unlike other members of this subgroup, RdfM is predicted to contain a second C-terminal HTH DNA-binding motif (HTH2) (Fig. 3), the role of which remains to be determined, as our results indicate that RdfM is not a transcriptional regulator of expression of int_{MGI}. This strongly contrasts with observations reported for AlpA of CP4-57, which activates the expression of *slpA*, the integrase gene of this cryptic prophage, and for which the role as an actual RDF remains unclear (27). Similarly, in addition to its RDF function, Vis of satellite prophage P4 has been shown to negatively regulate the expression of the P4 integrase by acting as an RNA-binding protein that posttranscriptionally regulates *int* expression (35). To date, all the RDFs belonging to the SLP1 subgroup have been found associated with P4-type integrases, as is also the case for RdfM (2, 10, 29, 30).

Given their importance for transfer and stability of integrating mobile elements, excision and integration must be tightly regulated. In many temperate bacteriophages and ICEs, the genes coding for the integration and excision functions are organized as a functional module in which the gene encoding the RDF precedes the gene coding for the recombinase, and both genes are organized in an operon-like structure (1, 9, 23, 40). In contrast, the organization of the genes coding for the integration and excision functions of MGIs is atypical: int_{MGI} is located immediately adjacent to *attL*, and *rdfM* is located on the opposite side, near the *attR* site (Fig. 1B). The integrase gene *intV2* (VC1758) and the RDF gene *vefA* (VC1809) of the *Vibrio* pathogenicity island 2 (VPI-2) are organized in a similar manner (2). This organization also resembles that of the KplE1 prophage, in which the *attL* site overlaps with the promoter of the *intS* gene, and the gene coding for the RDF TorI is remotely located, near the *attR* site (34). Yet while the recombination functions of MGIs and KplE1 seem to be organized alike, they are functionally unrelated. In KplE1, the *intS* gene is tightly regulated by its own product as well as by the TorI protein (34). In contrast, we showed that in MGIs, *rdfM* is not a transcriptional regulator of int_{MGI}. In fact, the uncoupled transcription of *rdfM* and int_{MGI} is likely a feature selected for by the opportunistic behavior of MGIs, as they rely on the SXT/R391 ICE-encoded transcriptional regulator SetCD to excise from their host chromosome. Unlike MGI excision, MGI integration does not require activation by SetCD. Once in a recipient cell, the MGI integrates site specifically through the action of Int_{MGI} expressed at a low level in the absence of SetCD. This process allows an MGI to establish itself in the host cell and be maintained in its progeny even in the absence of an SXT/R391 ICE.

Such a regulatory mechanism might have been selected to prevent MGI loss due to unproductive excision in the absence of a potentially mobilizing ICE of the SXT/R391 family. As a consequence, like for these ICEs, MGI excision and transfer are triggered by any physical or chemical agents that will damage DNA and stimulate the bacterial SOS response, including UV-light irradiation or exposure to mitomycin C and antibiotics such as ciprofloxacin (5, 13). Almagro-Moreno et al. (2) showed that *Vibrio* pathogenicity island 2 (VPI-2) excises from chromosome I of *V. cholerae* N16961 after sublethal UV-light irradiation of the cells. Increased excision of VPI-2 was correlated with increased expression of *intV2* and *vefA*. However, since *V. cholerae* N16961 is devoid of SXT/R391 ICE, induction of VPI-2 excision cannot be controlled by the SetR/SetCD pathway, suggesting that it may instead rely on the SOS response regulon repressor *lexA*.

This study brings a new understanding of the dynamics of excision, transfer, and integration of mobile genetic elements, giving a better insight into the rules that direct their mobility. We also show a novel interaction between two phylogenetically unrelated families of GIs and show how a small non-self-transmissible GI with no conjugative functions can take advantage of the conjugative machinery and regulatory pathways of a self-transmissible GI in order to transfer from one cell to another.

ACKNOWLEDGMENTS

This work was supported by a Discovery Grant and Discovery Acceleration Supplement from the Natural Sciences and Engineering Council of Canada (V.B.). D.P.-L. holds a Fonds de Recherche du Québec doctoral fellowship. M.M. was supported by the Deutscher Akademischer Austausch Dienst RISE program. V.B. holds a Canada Research Chair in molecular bacterial genetics and is a member of the FRSQ-funded Centre de Recherche Clinique Étienne-Le Bel.

We are grateful to D. Mazel for kindly providing us with *E. coli* β2163 and plasmid pSW23T. We thank E. Bordeleau, G. Garriss, N. Carraro, and A. Lavigueur for helpful comments and critical reading of the manuscript.

REFERENCES

1. **Abremski K, Gottesman S.** 1981. Site-specific recombination Xis-independent excisive recombination of bacteriophage lambda. J. Mol. Biol. **153**:67–78.
2. **Almagro-Moreno S, Napolitano MG, Boyd EF.** 2010. Excision dynamics of *Vibrio* pathogenicity island-2 from *Vibrio cholerae*: role of a recombination directionality factor VefA. BMC Microbiol. **10**:306. doi:10.1186/1471-2180-10-306.
3. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

4. **Beaber JW, Hochhut B, Waldor MK.** 2002. Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae.* J. Bacteriol. **184**:4259–4269.
5. **Beaber JW, Hochhut B, Waldor MK.** 2004. SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature **427**:72–74.
6. **Boyd EF, Almagro-Moreno S, Parent MA.** 2009. Genomic islands are dynamic, ancient integrative elements in bacterial evolution. Trends Microbiol. **17**:47–53.
7. **Burrus V, Marrero J, Waldor MK.** 2006. The current ICE age: biology and evolution of SXT-related integrating conjugative elements. Plasmid **55**:173–183.
8. **Burrus V, Pavlovic G, Decaris B, Guedon G.** 2002. Conjugative transposons: the tip of the iceberg. Mol. Microbiol. **46**:601–610.
9. **Burrus V, Pavlovic G, Decaris B, Guedon G.** 2002. 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 **48**:77–97.
10. **Burrus V, Waldor MK.** 2003. Control of SXT integration and excision. J. Bacteriol. **185**:5045–5054.
11. **Burrus V, Waldor MK.** 2004. Shaping bacterial genomes with integrative and conjugative elements. Res. Microbiol. **155**:376–386.
12. **Ceccarelli D, Daccord A, Rene M, Burrus V.** 2008. 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. **190**:5328–5338.
13. **Daccord A, Ceccarelli D, Burrus V.** 2010. 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. **78**:576–588.
14. **Datsenko KA, Wanner BL.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. **97**:6640–6645.
15. **Demarre G, et al.** 2005. 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. **156**:245–255.
16. **Dobrindt U, Hochhut B, Hentschel U, Hacker J.** 2004. Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. **2**:414–424.
17. **Dodd IB, Egan JB.** 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18**:5019–5026.
18. **Dower WJ, Miller JF, Ragsdale CW.** 1988. High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res. **16**:6127–6145.
19. **Dziejman M, Mekalanos JJ.** 1994. Analysis of membrane protein interaction: ToxR can dimerize the amino terminus of phage lambda repressor. Mol. Microbiol. **13**:485–494.
20. **Finkel SE, Johnson RC.** 1992. The Fis protein: it's not just for DNA inversion anymore. Mol. Microbiol. **6**:3257–3265.
21. **Friedman DI.** 1988. Integration host factor: a protein for all reasons. Cell **55**:545–554.
22. **Frost LS, Leplae R, Summers AO, Toussaint A.** 2005. Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. **3**:722–732.
23. **Ghinet MG, et al.** 2011. Uncovering the prevalence and diversity of integrating conjugative elements in actinobacteria. PLoS One **6**:e27846. doi:10.1371/journal.pone.0027846.
24. **Gogarten JP, Townsend JP.** 2005. Horizontal gene transfer, genome innovation and evolution. Nat. Rev. Microbiol. **3**:679–687.
25. **Hochhut B, Waldor MK.** 1999. Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC.* Mol. Microbiol. **32**:99–110.
26. **Juhas M, et al.** 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol. Rev. **33**:376–393.
27. **Kirby JE, Trempy JE, Gottesman S.** 1994. Excision of a P4-like cryptic prophage leads to Alp protease expression in *Escherichia coli.* J. Bacteriol. **176**:2068–2081.
28. **Lawrence JG, Hendrickson H.** 2005. Genome evolution in bacteria: order beneath chaos. Curr. Opin. Microbiol. **8**:572–578.
29. **Lesic B, et al.** 2004. Excision of the high-pathogenicity island of *Yersinia pseudotuberculosis* requires the combined actions of its cognate integrase and Hef, a new recombination directionality factor. Mol. Microbiol. **52**:1337–1348.
30. **Lewis JA, Hatfull GF.** 2001. Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. Nucleic Acids Res. **29**:2205–2216.
31. **McCool JD, et al.** 2004. Measurement of SOS expression in individual *Escherichia coli* K-12 cells using fluorescence microscopy. Mol. Microbiol. **53**:1343–1357.
32. **Ochman H, Lawrence JG, Groisman EA.** 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405**:299–304.
33. **Ochman H, Lerat E, Daubin V.** 2005. Examining bacterial species under the specter of gene transfer and exchange. Proc. Natl. Acad. Sci. U. S. A. **102**(Suppl 1):6595–6599.
34. **Panis G, et al.** 2010. Tight regulation of the intS gene of the KplE1 prophage: a new paradigm for integrase gene regulation. PLoS Genet. **6**:e1001149. doi:10.1371/journal.pgen.1001149.
35. **Piazzolla D, et al.** 2006. Expression of phage P4 integrase is regulated negatively by both Int and Vis. J. Gen. Virol. **87**:2423–2431.
36. **Salyers AA, Shoemaker NB, Stevens AM, Li LY.** 1995. Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. Microbiol. Rev. **59**:579–590.
37. **Scott JR, Churchward GG.** 1995. Conjugative transposition. Annu. Rev. Microbiol. **49**:367–397.
38. **Seth-Smith H, Croucher NJ.** 2009. Genome watch: breaking the ICE. Nat. Rev. Microbiol. **7**:328–329.
39. **Singer M, et al.** 1989. A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli.* Microbiol. Rev. **53**:1–24.
40. **Su YA, Clewell DB.** 1993. Characterization of the left 4 kb of conjugative transposon Tn*916*: determinants involved in excision. Plasmid **30**:234–250.
41. **Wozniak RA, et al.** 2009. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. PLoS Genet. **5**:e1000786. doi:10.1371/journal.pgen.1000786.
42. **Wozniak RA, Waldor MK.** 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat. Rev. Microbiol. **8**:552–563.

ANNEXE 4

Poulin-Laprade, D., Carraro, N., and Burrus, V. (2015b). The extended regulatory networks of SXT/R391 integrative and conjugative elements and IncA/C conjugative plasmids SXT/R391 and IncA/C conjugation regulation. Front. Microbiol. *6*, 837.

frontiers in Microbiology

REVIEW
published: 20 August 2015
doi: 10.3389/fmicb.2015.00837

The extended regulatory networks of SXT/R391 integrative and conjugative elements and IncA/C conjugative plasmids

Dominic Poulin-Laprade[†], *Nicolas Carraro*[†] *and Vincent Burrus*[*]

Laboratory of Bacterial Molecular Genetics, Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada

OPEN ACCESS

Edited by:
Yi-Cheng Sun, Chinese Academy of Medical Sciences and Peking Union Medical College, China

Reviewed by:
Lanming Chen, Shanghai Ocean University, China
Tony J. Pembroke, University of Limerick, Ireland
Jie Feng, Chinese Academy of Sciences, China

***Correspondence:**
Vincent Burrus, Laboratory of Bacterial Molecular Genetics, Département de Biologie, Faculté des Sciences, Université de Sherbrooke, 2500 Boul. Université, Sherbrooke, QC J1K 2R1, Canada
vincent.burrus@usherbrooke.ca

[†]*These authors have contributed equally to this work.*

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *28 May 2015*
Accepted: *31 July 2015*
Published: *20 August 2015*

Citation:
Poulin-Laprade D, Carraro N and Burrus V (2015) The extended regulatory networks of SXT/R391 integrative and conjugative elements and IncA/C conjugative plasmids. Front. Microbiol. 6:837. doi: 10.3389/fmicb.2015.00837

Nowadays, healthcare systems are challenged by a major worldwide drug resistance crisis caused by the massive and rapid dissemination of antibiotic resistance genes and associated emergence of multidrug resistant pathogenic bacteria, in both clinical and environmental settings. Conjugation is the main driving force of gene transfer among microorganisms. This mechanism of horizontal gene transfer mediates the translocation of large DNA fragments between two bacterial cells in direct contact. Integrative and conjugative elements (ICEs) of the SXT/R391 family (SRIs) and IncA/C conjugative plasmids (ACPs) are responsible for the dissemination of a broad spectrum of antibiotic resistance genes among diverse species of *Enterobacteriaceae* and *Vibrionaceae*. The biology, diversity, prevalence and distribution of these two families of conjugative elements have been the subject of extensive studies for the past 15 years. Recently, the transcriptional regulators that govern their dissemination through the expression of ICE- or plasmid-encoded transfer genes have been described. Unrelated repressors control the activation of conjugation by preventing the expression of two related master activator complexes in both types of elements, i.e., SetCD in SXT/R391 ICEs and AcaCD in IncA/C plasmids. Finally, in addition to activating ICE- or plasmid-borne genes, these master activators have been shown to specifically activate phylogenetically unrelated mobilizable genomic islands (MGIs) that also disseminate antibiotic resistance genes and other adaptive traits among a plethora of pathogens such as *Vibrio cholerae* and *Salmonella enterica*.

Keywords: SXT/R391, IncA/C, SGI1, regulation, integrative and conjugative elements, conjugative plasmids, genomic islands, pVCR94

Mobile Genetic Elements in the Modern World of Multiresistance

The discovery of penicillin by Alexander Fleming over 80 years ago marked the end of the pre-antibiotic era and revolutionized the prevention and treatment of many bacterial infections responsible for high morbidity and mortality. However, Sir Fleming himself warned the scientific community about antibiotic resistance and foresaw that inadequate usage of antibiotics could lead to "educated microbes." Since then, the use and misuse of antibiotics have led to the rapid and widespread emergence and selection of microorganisms resistant to a wide range of antimicrobial compounds. Today, multidrug resistance (MDR) has become one of the most alarming healthcare

issue on a global scale, so much so that in 2014 the World Health Organization (WHO) predicted a bleak short-term future: "A post-antibiotic era—in which common infections and minor injuries can kill—far from being an apocalyptic fantasy, is instead a very real possibility for the 21st Century" (World Health Organization, 2014).

Point mutations and/or gene amplification can allow bacteria to withstand hostile environments, such as the exposure to antimicrobial compounds (Gorgani et al., 2009; Davies and Davies, 2010; Toprak et al., 2012). Most often, MDR results from the acquisition by horizontal gene transfer of mobile genetic elements carrying multiple antibiotic resistance genes (Burrus et al., 2006; Mulvey et al., 2006; Welch et al., 2007; Escudero et al., 2014). Conjugation, which mediates DNA transfer between two bacterial cells in direct contact, is the most effective mechanism of horizontal gene transfer in terms of host range and quantity of genes translocated to a recipient cell per transfer event (Llosa et al., 2002; de la Cruz et al., 2010). Integrative and conjugative elements (ICEs) and conjugative plasmids of various incompatibility groups were shown to have a major impact on the global emergence of multidrug resistant pathogenic bacteria, in both clinical and environmental settings (Burrus et al., 2006; Fricke et al., 2009; Smillie et al., 2010; Wozniak and Waldor, 2010; Guglielmini et al., 2011; Walsh et al., 2011; Carattoli, 2013). Although both types of elements transfer from cell to cell by conjugation, their mechanism of persistence in the bacterial host cell genome is different. On the one hand, ICEs maintain themselves by integration into the chromosome of their host and excise prior to transfer as circular molecules (Burrus et al., 2002; Burrus and Waldor, 2004; Wozniak and Waldor, 2010). On the other hand, conjugative plasmids are maintained by replication as episomes, i.e., DNA molecules that are distinct from the chromosome.

This review focuses on the regulatory networks that govern the conjugative transfer of ICEs belonging to the SXT/R391 family (SRIs) and conjugative plasmids of the A/C incompatibility group (ACPs). Both classes of elements bear highly similar and nearly syntenic core sets of conserved genes and code for comparable transfer activator complexes (Wozniak et al., 2009; Carraro et al., 2014a; Poulin-Laprade et al., 2015). Recent investigations of the regulatory circuitries that activate SRIs and ACPs transfer have also contributed to the discovery of three classes of genomic islands (GIs) specifically mobilized by either SRIs or ACPs (Doublet et al., 2005; Daccord et al., 2010, 2012; Carraro et al., 2014a; Poulin-Laprade et al., 2015).

Diversity and Prevalence of SRIs and ACPs

SRIs and ACPs are major contributors to worldwide dissemination of adaptative traits such as antibiotic resistance among several species of *Enterobacteriaceae* and *Vibrionaceae* of clinical origin or isolated from the aquatic environment.

The SXT/R391 family is one of the largest, diverse and well-studied set of ICEs among Gram-negative bacteria. Extensive experimental and bioinformatic studies have led to a deeper understanding of their prevalence, diversity, and evolution (Boltner et al., 2002; Wozniak et al., 2009; Garriss and Burrus, 2013; Carraro and Burrus, 2014; Spagnoletti et al., 2014). SRIs are large conjugative elements (79 to 110 kb) found integrated into the 5′ end of *prfC* in the chromosome of several species of *Vibrio*, *Photobacterium*, *Providencia*, *Proteus*, *Alteromonas*, *Marinomonas*, and *Shewanella*, and are easily transferred to *E. coli* in the laboratory (Coetzee et al., 1972; Waldor et al., 1996; Hochhut and Waldor, 1999; Beaber et al., 2002a; Pembroke and Piterina, 2006; Osorio et al., 2008; Harada et al., 2010; Rodriguez-Blanco et al., 2012; Badhai et al., 2013; Lopez-Perez et al., 2013; Spagnoletti et al., 2014). Notably, SRIs played a key role in the dissemination of MDR in the seventh-pandemic lineage of *V. cholerae*, the etiological agent of the diarrhoeal disease cholera (Spagnoletti et al., 2014). *V. cholerae* is endemic in Asia, Africa, and Central America and epidemics of cholera are usually blooming in locations where the sanitation infrastructures and access to clean water are compromised. Indeed, cholera is considered by the WHO as an indicator of sanitation mismanagement and humanitarian crisis (e.g., refugee camps). Currently, most clinical isolates of *V. cholerae* carry an SRI and are multidrug resistant worldwide. Most SRIs found in epidemic strains of *V. cholerae* contain the genes *floR*, *strBA*, *sul2*, and *dfrA1* or *dfr18*, respectively conferring resistance to florfenicol/chloramphenicol, streptomycin, sulfamethoxazole and trimethoprim (Waldor et al., 1996; Hochhut et al., 2001; Wozniak et al., 2009). Sulfamethoxazole and trimethoprim have synergistic antibacterial activities and are often used in combination for the treatment of cholera (Kaper et al., 1995). Other SRIs from the aquatic environment and from diverse pathogens confer resistance to kanamycin (*aph*) or tetracycline (*tetAR*) (Coetzee et al., 1972; Osorio et al., 2008; Wozniak et al., 2009; Bi et al., 2012). In the countries where the sanitation infrastructures are appropriate, the domestic cases of cholera and other vibriosis caused by hosts of SRIs are widely associated with the consumption of raw or undercooked seafood (Morris, 2003; Song et al., 2013; Hara-Kudo and Kumagai, 2014; Robert-Pillot et al., 2014). For instance, a few cases of cholera acquired in the US are declared each year. These sporadic cholera cases are generally attributed to consumption of seafood gathered from the US Gulf coast (Loharikar et al., 2015). Antibiotic resistance genes carried by SRIs are also troublesome for aquaculture as resistance genes can hinder the treatment of diseased fish and enter the food chain (Osorio et al., 2008; Rodriguez-Blanco et al., 2012; Nonaka et al., 2014). Indeed, consumption of raw fish and shellfish contaminated by live bacteria bearing SRIs could facilitate the dissemination of MDR among *Gammaproteobacteria* of the human host microbiome.

ACPs are large (>110 kb) circular plasmids grouped as a family based on the high percentage of sequence conservation of their *repA* gene, which codes for their replication intiator protein (Llanes et al., 1994, 1996; Carattoli et al., 2005; Fricke et al., 2009). Multidrug resistant ACPs are found worldwide in pathogens associated with human infections such as *Citrobacter freundii*, *V. cholerae*, *Salmonella enterica*, *Proteus mirabilis*, *E. coli*, *Yersinia pestis* and *ruckeri*, *Klebsiella pneumoniae*, and *Providencia stuartii* (Bauernfeind et al., 1996; Galimand et al., 1997; Giles et al., 2004; Welch et al., 2007; Ding et al., 2008; Fricke et al., 2009; Call et al., 2010; Fernandez-Alarcon et al., 2011; Lindsey et al., 2011; Walsh et al., 2011; Carattoli, 2013;

Carraro et al., 2014b; Rahman et al., 2014). ACPs carrying MDR are also increasingly encountered in enteropathogenic bacteria recovered from food-producing animals and food products, mainly *S. enterica* and *E. coli* (Glenn et al., 2011; Lindsey et al., 2011; Randall et al., 2011; Folster et al., 2012; Del Castillo et al., 2013; Guo et al., 2014). Disturbingly, recent studies identified multiple extended-spectrum β-lactamases (ESBLs)-encoding ACPs conferring resistance to a wide range of β-lactam antimicrobials (Fernandez-Alarcon et al., 2011; Folster et al., 2011, 2012; Walsh et al., 2011; Harmer and Hall, 2015). Carbapenems were the last effective β-lactams for the treatment of infectious bacteria carrying ESBLs. Unfortunately, several recently isolated ACPs propagate the infamous New Delhi metallo-β-lactamase bla_{NDM-1} gene and its variants, which code for zinc metallo-β-lactamases that hydrolyze all penicillins, cephalosporins and carbapenems (Walsh et al., 2005, 2011; Yong et al., 2009; Nordmann et al., 2011; Tijet et al., 2015).

ACPs and SRIs are a threat to antibiotic therapies due to the large variety of antibiotic resistance genes that they bear on dynamic genetic structures such as integrons and transposons, further promoting the exchange and capture of resistance genes from other mobile genetic elements (Hochhut et al., 2001; Mazel, 2006; Welch et al., 2007; Fricke et al., 2009; Wozniak et al., 2009; Lindsey et al., 2011; Carraro et al., 2014b). Acquisition and exchange of antibiotic resistance genes are strongly enhanced by the broad host range of these elements, which can easily spread across several genera and species of *Gammaproteobacteria*. This phenomenon is likely further exacerbated by their mechanism of transfer as single-stranded DNA molecules have been shown to stimulate the SOS response in recipient cells, thereby promoting the intra- and inter-integrons movement of resistance cassettes (Guerin et al., 2009; Baharoglu et al., 2010, 2012; Cambray et al., 2011; Escudero et al., 2014).

Modular Organization of SRIs and ACPs

All SRIs share 47 kb of DNA corresponding to a highly conserved core set of 52 genes with over 95% identity at the nucleotide level (Wozniak et al., 2009). About half of these genes have been shown to be essential to ensure the basic maintenance, transfer and regulatory functions of SRIs. These essential genes are clustered in four main modules (**Figure 1**), i.e., the *int* module which codes for the integrase and excisionase and ensures intracellular mobility, the *mob* and *mpf* modules which code for a type IV secretion system (T4SS) and is responsible of the intercellular mobility (DNA processing and mating pore formation), and the *reg* module coding for the regulatory network governing the expression of the other modules. Each module can contain one to several transcriptional unit(s) (**Figure 1**; Poulin-Laprade et al., 2015). The *reg* module of SRIs is the most highly conserved locus amongst members of this family of ICEs (Wozniak et al., 2009).

ACPs are characterized by ∼110 kb of conserved core genes with over 98% nucleotide sequence identity (Fricke et al., 2009; Fernandez-Alarcon et al., 2011; Del Castillo et al., 2013; Carraro et al., 2014b; Harmer and Hall, 2014, 2015). Although ACP conserved core is larger than the one shared by SRIs, their organization is highly similar and syntenic (Welch et al., 2007; Wozniak et al., 2009). In particular, the *tra* genes of the *mob* and *mpf* modules of ACPs and SRIs are reminiscent of the IncFI F and IncHI1 R27 plasmids suggesting a common ancestry (Lawley et al., 2003). One of the most striking differences between SRIs and ACPs reflects their respective biology. The *int* module, which ensures chromosomal integration and excision of SRIs, is replaced by the *rep* module driving the replication of the episomal ACPs. The conserved core of ACPs also contains several genes of unknown functions beyond those also found in SRIs.

Distinctive features of the individual members of SRI and ACP families are provided by insertions of variable cargo DNA in hotposts dispersed in their respective conserved core. These insertions vary in size (from ∼60 to 20,000 bp) and encode adaptative traits that may provide a selective advantage to the bacterial host in specific conditions, such as resistance to antibiotics, heavy metals or phage infection, or synthesis of the second messenger c-di-GMP (Welch et al., 2007; Fricke et al., 2009; Wozniak et al., 2009; Bordeleau et al., 2010; Carraro et al., 2014a).

Control of the Conjugative Functions of SRIs and ACPs

Control of SRI and ACP conjugative transfer is a key attribute for their propagation and stability. Excessive repression would impair their dissemination, while overactivation would be a burden for the bacterial host causing reduced fitness, and ultimately their instability in the cell population (Lundquist and Levin, 1986; Scott et al., 1988; Beaber et al., 2002b; Ramsay et al., 2006; Bellanger et al., 2009; Haft et al., 2009). Moreover, SRIs and ACPs not only drive their self-transfer, but also the transfer of phylogenetically unrelated mobilizable genomic islands (MGIs). Additionally, SRIs in association with MGIs can mobilize up to 1.5 Mb of chromosomal DNA each in Hfr-like conjugal events initiated prior to their excision (Hochhut et al., 2000; Daccord et al., 2010). Hence, these elements can potentially mobilize more than 60% of *V. cholerae* chromosome I in a single conjugal event.

Transcriptional repressors encoded by SRIs and ACPs repress the expression of master activator genes, maintaining these elements in a quiescent state in most cells of the bacterial population. Both SRIs and ACPs thrive in a large array of *Enterobacteriaceae* and *Vibrionaceae*, which implies that their regulatory networks are likely autonomous and orthogonal, i.e., they allow the activation/repression of the element while avoiding crosstalks with regulatory networks of the host cell.

The Regulation Module of SRIs and ACPs

SRIs and ACPs bear distinct regulatory modules that govern their self-transmissibility (**Figure 2**). These regulatory modules code for unrelated repressors: SetR for SRIs and Acr1 and Acr2 for ACPs (Beaber et al., 2004; Carraro et al., 2014a). In contrast, the regulatory module of SRIs and ACPs code for related transcriptional activator complexes, respectively SetCD and AcaCD, that drive the expression of the conjugative genes and other functions (Beaber et al., 2002b; Carraro et al., 2014a; Poulin-Laprade et al., 2015). SetCD and AcaCD are distant relatives of FlhCD, the master activator of flagellum biosynthesis in many

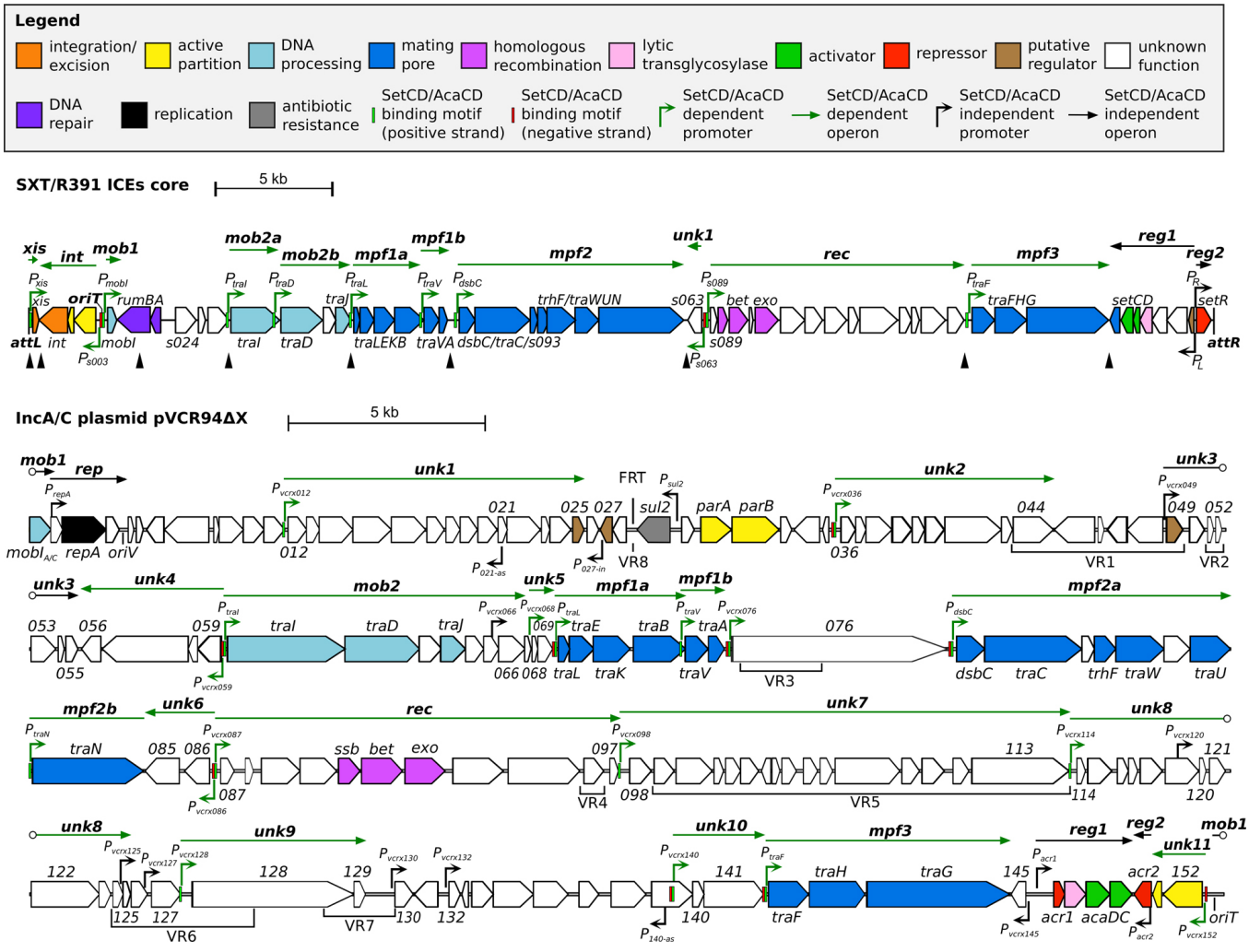

FIGURE 1 | Schematic representation of the genetic organization and transcriptional units of the conserved core of SXT/R391 ICEs (integrated in *prfC*) and pVCR94ΔX (circular map linearized at the start position of gene *mobI*) adapted from Poulin-Laprade et al. (2015) and Carraro et al. (2014a, 2015a). Genes are represented by arrows and color coded according to their function as indicated in the legend. For clarity, ORF names *vcrxXXX* were shortened as *XXX* for pVCR94ΔX. SetCD- and AcaCD-binding motifs located on positive and negative DNA strands are represented by light green and red narrow boxes, respectively. Operons are indicated by arrows positioned above represented genes. SetCD- and AcaCD-regulated promoters and operons are colored in green. Open circles mark operons interruptions generated by the map format. *mob1-2*, DNA processing; *rep*, replication; *unk1-11*, unknown; *mpf1-3*, mating pore formation; *rec*, recombination; *reg1-2*, regulation. P_{021-as} and P_{140-as}: *vcrx021* and *vcrx140* antisens promoters, respectively. P_{027-in}: *vcrx027* internal promoter. Black triangles show the position of variable cargo DNA in SRIs, while variable DNA regions inserted in the conserved core of ACPs are indicated below genes (VR1 to VR8). The origin of replication (*oriV*) and the origin of transfer (*oriT*) are indicated. The position of the FRT site resulting from the deletion of the antibiotic resistance gene cluster in pVCR94 is also shown.

Gram-negative bacteria (Chevance and Hughes, 2008; Fitzgerald et al., 2014). Recent studies established the AcaCD and SetCD regulons and refined the models of transcriptional organization of the functional core of both types of elements (**Figure 1**; Carraro et al., 2014a; Poulin-Laprade et al., 2015).

Repression of SRIs Dissemination

The SetR Repressor

The dominant regulatory state of SRIs is the quiescent state in which the element is integrated into the chromosome and the genes associated with recombination and transfer are silent (Beaber et al., 2004; Poulin-Laprade et al., 2015). In this dormant state, very few genes are transcribed, including genes independently regulated belonging to cargo DNA (e.g., antibiotic resistance genes) and *setR*. The *setR* gene is located at the rightmost end of the integrated ICE (**Figure 1**). SetR is an acronym for SXT excision and transfer repressor. *setR* mRNA transcript is leaderless, expressed from the P_R promoter, and codes for a 215-amino acid residue protein with a DNA binding helix-turn-helix motif (HTH_3, PF01381) in its N-terminal moiety and a C-terminal LexA-like autoproteolysis motif (Peptidase_S24, PF00717). SetR shares homology with λ CI-like repressors encoded by lambdoid bacteriophages (Beaber et al., 2002b).

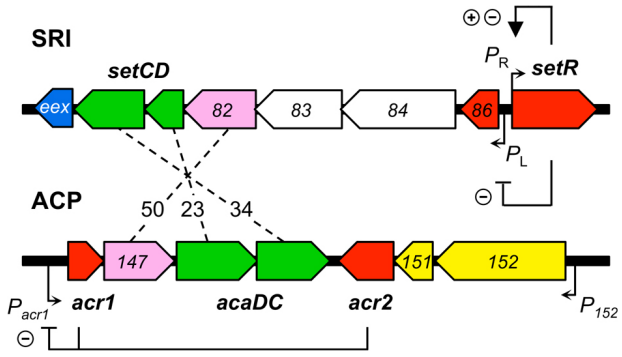

FIGURE 2 | Comparison of the regulatory modules of SXT/R391 ICEs (SRI) and IncA/C plasmids (ACP). The genes are color-coded as indicated in **Figure 1** legend. Numbers between the elements represent the percentage of identity between orthologous proteins. The regulation exerted by SetR, Acr1, and Acr2 is indicated (minus sign for repression, plus sign for activation). For clarity, ORF names *s0XX* were shortened as *XX* for SRI and *vcrxXXX* as *XXX* for ACP.

The pivotal role of SetR in SRIs regulation is reflected by the inability to generate a *setR* mutant of SXT without simultaneous *setR trans*-complementation or a preexisting *setCD* inactivation (Beaber et al., 2002b, 2004).

SetR Regulation of the P_L and P_R Early Promoters

SetR maintains the quiescent integrated state of SRIs by binding to four operator sites (*OL*, *O1*, *O2*, and *O3*) distributed in the intergenic region between *s086* and *setR* (**Figure 2**; Beaber and Waldor, 2004). Footprint assays revealed that the relative affinity of SetR for its operators is $O1 > O2 \approx O3 > OL$ (Beaber and Waldor, 2004). SetR operator sites bear partial dyad symmetry and are separated by AT-rich spacers. An additional site located 800 bp downstream of the P_L promoter was suggested but never assessed (Beaber and Waldor, 2004). It has been proposed that binding of SetR to the four operators between *s086* and *setR* leads to SetR's autoregulation of the P_R promoter (Beaber et al., 2004). Binding of SetR to *O1* is thought to lead to activation of the P_R promoter. When the cellular pool of SetR exceeds a threshold, SetR is thought to repress its own expression by further binding to the low affinity *O3* operator, concealing the -10 element of the P_R promoter (Beaber and Waldor, 2004). Beaber and Waldor (2004) observed the repressive effect of SetR on P_R by monitoring the β-galactosidase activity of a P_R-*lacZ* transcriptional fusion in strains containing or lacking SXT, or its Δ*setCD* or Δ*setCD* Δ*setR* mutants. Quantification of the β-galactosidase activity in these strains showed that the presence of SXT lowered the activity of P_R by 30% (SXT^- versus SXT^+) (Beaber and Waldor, 2004). Deletion of *setCD* did not significantly alter P_R activity compared to the SXT^+ background, whereas in cells containing SXT Δ*setCD* Δ*setR*, P_R activity was comparable to cells lacking SXT, thereby confirming that SetR represses P_R (Beaber and Waldor, 2004). SetR binding to *O1* and *OL* obstructs the P_L promoter which drives *setCD* expression and subsequent activation of conjugative functions.

The mRNA transcript starting at P_L codes for seven proteins including a predicted λCro-like repressor, the two subunits of the activator complex SetCD, and the entry exclusion determinant Eex (**Figure 2**; Beaber et al., 2002b; Beaber and Waldor, 2004; Marrero and Waldor, 2005; Poulin-Laprade et al., 2015).

Alleviation of SetR Repression

In donor cells, the inductive cue triggering SRI propagation is linked to the SOS response (Waldor et al., 1996; Beaber et al., 2004). Using the energy of ATP, RecA polymerizes onto single-stranded DNA, generating RecA-ssDNA filaments (RecA*) that are competent for homologous recombination and are also allosteric effectors unleashing the latent proteolytic activity of LexA and λ CI-like repressors (Little, 1984; Chen et al., 2008). Thus, RecA is the central factor linking DNA damages (sometimes caused by antibiotics) to the cellular SOS stress response (DNA mutagenesis and repair), and to the induction of conjugative transfer of SRIs that are major vectors of MDR (Beaber et al., 2004; Baharoglu et al., 2010). Inspired by the extensive work done on the λ CI repressor, the link between RecA* and SetR was drawn with the mutant $setR^{G49E}$ in which the Ala-Gly cleavage site activated by RecA is disrupted (Gimble and Sauer, 1985; Beaber et al., 2004). As expected, the $setR^{G49E}$ mutant of SXT is unresponsive to mitomycin C, a DNA damaging agent known to trigger the bacterial SOS response.

Upon DNA damage, SetR becomes a substrate for RecA*-mediated self-cleavage, thereby alleviating SetR's repression on P_L and allowing *setCD* expression. The -10 and -35 promoter elements of P_L are more similar to the recognition motif of σ^{70}-bound RNA polymerase (RNAP) than those of P_R, likely leading to a quicker isomerization into an open complex competent for transcription initiation. Alleviation of SetR repression would then be sufficient for recognition of P_L by RNAP, without the need of a transcriptional activator. This model is reminiscent of the regulation of λ P_R and P_{RM} early promoters (Strainic et al., 2000; Ptashne, 2004).

SetR acts as a sentinel "sensing" DNA damages and triggering the "escape" of SRIs to recipient cells. For an optimal responsiveness and avoidance of cellular resources misallocation, SetR expression is tightly regulated and maintained at low levels (Beaber and Waldor, 2004). The *setR* transcript is a leaderless mRNA; the absence of a Shine-Dalgarno sequence is a post-transcriptional mechanism that likely contributes to a low intracellular level of SetR protein (Van Etten and Janssen, 1998; Beaber and Waldor, 2004). Spontaneous induction of the SOS response in a subpopulation of cells is thought to account for the low basal transfer of SRIs, which varies between individual SRIs for reasons that remain unknown (Beaber and Waldor, 2004; McCool et al., 2004; McGrath et al., 2005; Poulin-Laprade et al., 2015).

Repression of ACPs

While no SetR homolog has been found in ACPs, their regulatory module codes for two repressors named Acr1 and Acr2 (IncA/C repressor 1 and 2; Carraro et al., 2014a). *acr1* codes for a 90-amino acid Ner-like protein that is mainly composed of a helix-turn-helix DNA binding domain (HTH_35, PF13693). Acr1 directly represses its own expression from the constitutive promoter

P_{acr1} (**Figure 2**). This promoter drives the expression of *acr1* and also the expression of *acaC* and *acaD*, which code for the activator complex AcaCD. Acr2 is a 139-amino acid H-NS-like repressor (Histone_HNS, PF00816) that also directly represses P_{acr1} (Carraro et al., 2014a,b). H-NS proteins are known to globally repress expression of horizontally acquired DNA by binding AT-rich sequences (Dorman, 2004, 2014; Navarre et al., 2006). Besides P_{acr1}, Acr2 might also repress other plasmid- or host-borne promoters, potentially having a wider impact on the biology of ACPs and their interaction with host cells.

The frequency of transfer of ACPs varies widely from non detectable to very high (1 in 10 cells for pVCR94; Welch et al., 2007; Fricke et al., 2009; Carraro et al., 2014b). Inducing factors triggering the conjugative transfer of ACPs have yet to be identified (Carraro et al., 2014a,b). Consistent with the absence of SOS-dependent repressors such as λ CI or ImmR, conjugative transfer of ACPs is independent of *recA* and the SOS response (Auchtung et al., 2005; Carraro et al., 2014b).

The Heteromeric Complexes SetCD and AcaCD

It was previously established that individual deletion of either *setC* or *setD* abolished the excision and transfer of the prototypical SRI SXT (Beaber et al., 2002b). These deletions were complemented in *trans* with plasmids expressing the individual genes, thereby confirming the central role of SetC and SetD in the biology of SRIs. Transcriptional *lacZ* fusions with promoters driving the expression of *int*, *traL* and *traG* demonstrated that SetCD is a transcriptional activator of the site-specific recombination and conjugative transfer genes (Beaber et al., 2002b; Poulin-Laprade et al., 2015).

A similar characterization was recently carried out for *acaC* and *acaD*, which code for the master activator of ACP conjugative transfer (Carraro et al., 2014a). For both sets of transcriptional activators, genetic assays strongly suggest that the products of the *setD-setC* and *acaD-acaC* genes assemble into higher order protein complexes designated SetCD and AcaCD, respectively. While no direct experimental evidence support the oligomerization of SetCD, AcaD was shown to copurify with 6xHis-tagged AcaC subunit, supporting the formation of heteromeric complexes as observed for the flagellar gene activator complex FlhCD (Wang et al., 2006; Carraro et al., 2014a).

Conflicting evidence suggest a possible autoregulation of SetCD expression. On the one hand, overexpression of SetCD was reported to result in a 40-fold activation of expression of a chromosomal *setD::lacZ* fusion in SXT (Beaber et al., 2002b). On the other hand, expression from P_L, which drives *setCD* expression, remained unaffected by deletion of *setCD* regardless of the presence of mitomycin C (Beaber et al., 2004). An exhaustive list of the promoters targeted by SetCD was recently established for three representative members of the SRI family (SXT, R391 and ICE*Vfl*Ind1) using chromatin immunoprecipitation coupled with exonuclease digestion (ChIP-exo) and RNA sequencing (RNA-seq; Poulin-Laprade et al., 2015). No SetCD binding site was found upstream of P_L or elsewhere in the regulatory module.

A similar experimental approach also allowed to establish the list of the promoters targeted by AcaCD in pVCR94ΔX, a prototypical ACP lacking most of its resistance genes (Carraro et al., 2014a,b). The DNA motifs recognized by SetCD and AcaCD were deduced from the multiple targets that were experimentally determined. Operator sites for SetCD and AcaCD fixation greatly differ from each other, and from the DNA motif recognized by *E. coli* FlhCD (**Figure 3A**; Carraro et al., 2014a; Fitzgerald et al., 2014; Poulin-Laprade et al., 2015). Despite their functional homology, SetCD, AcaCD, and FlhCD exhibit a high degree of divergence, which is reflected in their respective DNA target preference and specificity (Liu and Matsumura, 1994; Carraro et al., 2014a; Fitzgerald et al., 2014; Poulin-Laprade et al., 2015).

SetCD and AcaCD are Pleiotropic Transcriptional Activators

In many mobile genetic elements, genes involved in a given biological function are often arranged in an operon structure within a single module expressed from a single promoter (Celli and Trieu-Cuot, 1998; Toussaint and Merlin, 2002; Auchtung et al., 2005; Carraro et al., 2011). The genes coding for the conjugative machinery of the *E. coli* IncF1 F plasmid or the *Enterococcus faecalis* ICE Tn*916* are good examples of such an organization (Celli and Trieu-Cuot, 1998; Lawley et al., 2003). In contrast, the conjugation modules of SRIs and ACPs are fragmented in multiple and distinct operons (**Figures 1** and **3B,C**). This fragmentation of functional modules is most often attributed to insertions of variable cargo DNA, insertion sequences (IS) and transposons (Fricke et al., 2009; Wozniak et al., 2009; Fernandez-Alarcon et al., 2011; Meinersmann et al., 2013). These insertions occur in sites most likely selected because of their minimal impact on genes essential for transfer and subsequent maintenance of SRIs and ACPs in bacterial populations. Discontinuity of the functional modules complexifies the genetic regulation in terms of timing and gene dosage for coordinated expression of their functions allowing the dissemination of SRIs and ACPs. The efficient activation of the machinery for DNA processing and mating pore assembly relies on the flexibility and accuracy of DNA binding by the activator complexes SetCD and AcaCD. For instance, SetCD can be tolerant to insertion of cargo DNA in the promoter driving expression of *traI* in SXT, an essential component of conjugal transfer (Poulin-Laprade et al., 2015).

Mechanism of Activation by SetCD and AcaCD

ChIP-exo experiments have revealed 11 SetCD-dependent promoters in SRIs and 19 AcaCD-dependant promoters in ACPs (Carraro et al., 2014a; Poulin-Laprade et al., 2015). SetCD- and AcaCD-dependent promoters have poorly conserved -10 and non-conserved -35 boxes, compared to the canonical σ^{70} promoter elements (Hawley and McClure, 1983; Kumar et al., 1993). In each promoter, the DNA motif recognized by the activator complex partially overlaps the -35 element, which is usually bound by the σ^{70} subunit of RNAP (Carraro et al., 2014a; Poulin-Laprade et al., 2015). This suggests that, as observed for FlhCD, SetCD and AcaCD compensate for the lack of a recognizable -35 elements by binding in the -35 region, facilitating the recruitment of σ^{70}-bound RNAP to the promoters. As such, FlhCD, SetCD and AcaCD act as typical class II transcriptional activators (Browning and Busby, 2004). FlhCD

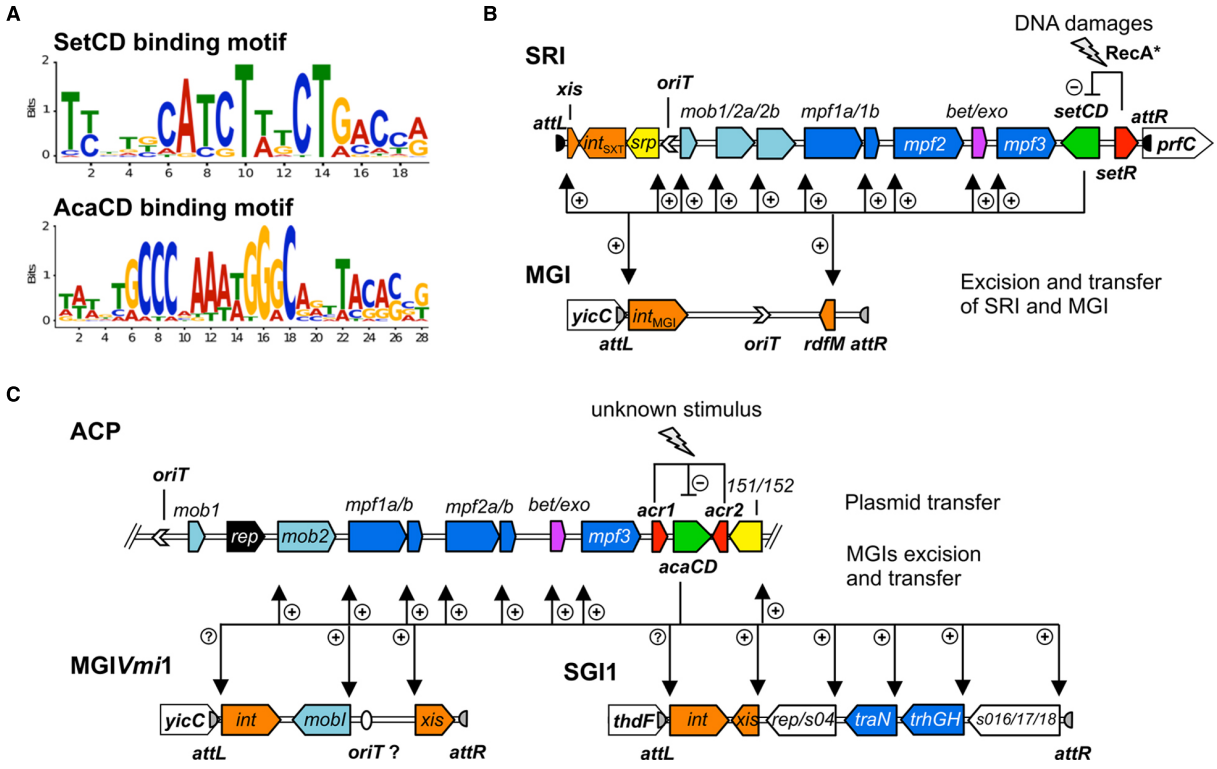

FIGURE 3 | Activation by heteromeric complexes SetCD and AcaCD. **(A)** Experimentally determined recognition motifs of SetCD and AcaCD (Carraro et al., 2014a; Poulin-Laprade et al., 2015). **(B)** Representation of SetCD targets in SRIs and in MGIs they mobilize. The arrows indicate transcriptional repression by SetR (minus sign) and transcriptional activation by SetCD (plus signs). **(C)** Representation of AcaCD targets in ACPs and in MGIs from MGI*Vmi*1 and SGI1 families. The arrows indicate transcriptional repression by Acr1 and Acr2 (minus sign) and transcriptional activation by AcaCD (plus signs). In both B and C panels, the modules of DNA processing (*mob*) and mating pair formation (*mpf*) are indicated and color-coded as described in **Figure 1** legend.

was shown to interact with the C-terminal domain of RNAP (Liu et al., 1995). Biochemical characterizations are needed to establish whether AcaCD and SetCD directly interact with RNAP.

Activation of Integration, Excision, and Stability Functions

As SRIs maintain by integration in the host cell chromosome, the major contributor to their maintenance in a bacterial lineage is the integration/excision module (Hochhut and Waldor, 1999; Burrus and Waldor, 2003). This module contains the *int* gene coding for the integrase, a site-specific tyrosine recombinase, the *xis* gene coding for a recombination directionality factor, as well as their cognate attachment sites, i.e., *attP* on the circular form, or *attL* and *attR* at both ends of the chromosomally integrated SRI. Expression of both *int* and *xis* is SetCD-dependent, yet driven from two separate promoters (Burrus and Waldor, 2003; Poulin-Laprade et al., 2015). Stability of SRIs is also provided by toxin-antitoxin systems (TA) and a type II active partition system named *srpRMC* (SXT/R391 partition; Dziewit et al., 2007; Wozniak and Waldor, 2009; Carraro et al., 2015b). The *srpRM* genes code for the proteins driving the active partition of the excised element in daughter cells, while *srpC* is a centromere-like sequence bound by SrpR (Baxter and Funnell, 2014; Carraro et al., 2015b). Regulation of integration, excision and active partition of SRIs are interconnected as *srpRM* and *int* are cotranscribed from the same SetCD-dependent promoter (Poulin-Laprade et al., 2015).

As plasmids, ACPs maintain in bacterial lineages by autonomous replication, which is mediated by the *repA*/*oriV* locus in an AcaCD-independent fashion (Llanes et al., 1996; Carraro et al., 2014a). Orthologs of the SRI's *srpRM* genes are also found in ACPs (*vcrx151*/*vcrx152* in pVCR94). Reminiscent of SRIs, expression of these *srpRM* orthologs is AcaCD-dependent (Carraro et al., 2014a). Interestingly, ACPs also carry genes coding for a type I ParABC-like partitioning system (Walker-type ATPase; *vcrx031*/*vcrx032* in pVCR94), whose regulation is likely independent of AcaCD (Baxter and Funnell, 2014; Carraro et al., 2014a, 2015b).

Activation of the Conjugative Machinery

SRIs and ACPs code for very similar conjugative machineries, as reflected by the syntenic organization of their transfer genes and the closely related proteins they encode (Fricke et al., 2009; Wozniak et al., 2009). The mobilization modules (*mob*) code for key factors involved in DNA molecule preparation (DNA processing functions) that will be translocated to the recipient cell through the type IV secretion system encoded by the *tra* modules. The relaxase TraI, with the help of the auxiliary protein MobI, is thought to recognize the origin of transfer (*oriT*) located immediately upstream of *mobI* in both SRIs and ACPs (**Figures 1**

and **3B,C**; Ceccarelli et al., 2008; Carraro et al., 2014b). By analogy with other better characterized conjugative systems such as F, the resulting nucleoprotein complex, aka relaxosome, is thought to nick one DNA strand within *oriT* (Llosa et al., 2002). This DNA strand is delivered to the mating pore linking the donor and recipient cells. Based on the mechanism of single-stranded conjugative transfer of the F plasmid, it is assumed that SRIs and ACPs replicate using the rolling-circle mechanism during translocation of the transferred DNA strand. Several studies on ICEs from both Gram-negative and Gram-positive bacteria showed that ICEs are capable of intracellular rolling-circle plasmid-like replication (Kiewitz et al., 2000; Pembroke and Murphy, 2000; Dimopoulou et al., 2002; Grohmann, 2010; Lee et al., 2010; Carraro et al., 2011, 2015b; Sitkiewicz et al., 2011). This replication only occurs in a subpopulation of cells as it is conditional on element activation. Mechanistically, it does not strikingly differ from rolling-circle replication used for the stable maintenance of plasmids, uses *oriT* as an origin of replication and the relaxase TraI as a replication initiator protein. In fact, the replication module is part of the mobilization module (Grohmann, 2010; Lee et al., 2010; Carraro and Burrus, 2014). Although the exact mechanism remains to be elucidated, SRIs have been shown to replicate in an *oriT*, TraI and SetCD-dependent manner (Pembroke and Murphy, 2000; Carraro et al., 2015b).

Genome-wide footprinting of SetCD and AcaCD DNA binding coupled with transcriptomic analyses revealed that the syntenic *mob* modules of SRIs and ACPs are divided into different transcriptional units (Carraro et al., 2014a; Poulin-Laprade et al., 2015). In SRIs, SetCD binds upstream of *mobI* (*mob1*), *traI* (*mob2a*), and *traDJ* (*mob2b*). Interestingly, the canonic promoter driving the expression of *traI* is disrupted by an insertion into hotspot 5 (Poulin-Laprade et al., 2015). The -10 element of P_{traI} is part of the conserved core and retained, while the -35 element is variable and provided by inserted cargo DNA. Alteration of P_{traI} is associated with a poorer affinity for SetCD as determined by ChIP-exo, which could contribute to the lower transfer and replication of SXT compared with R391 (Carraro et al., 2015b; Poulin-Laprade et al., 2015). In ACPs, AcaCD activates the expression of *traIDJ* (*mob2*) from a unique promoter (**Figure 1**; Carraro et al., 2014a, 2015a). No ACPs available to date in the Genbank database harbor a P_{traI} promoter altered by insertion of cargo DNA (Carraro et al., 2014a). Surprisingly, no AcaCD binding-site was detected upstream of $mobI_{A/C}$ (formerly known as *vcrx001*, Carraro et al., 2014a,b). As for MobI of SRIs, $MobI_{A/C}$ is essential for conjugative transfer of ACPs (Carraro et al., 2014b). The impact of such subtle differences on the regulation of conjugative transfer of SRIs and ACPs need to be experimentally addressed. Altered regulation of the *mob* functions can have drastic effects on the dynamics of these elements since initiation of transfer (*oriT* recognition and nicking by the relaxosome) was shown to be the rate limiting step of SRIs dissemination (Carraro et al., 2015b).

Other essential components for conjugative transfer of SRIs and ACPs are the pilus, which stabilizes the initial contact between cells, and the type IV secretion system (mating pore) through which DNA is translocated to recipient cells. This conjugative machinery is encoded by three mating pair formation modules (*mpf*) which are, as the *mob* modules, syntenic between SRIs, ACPs and the F plasmid (**Figure 1**; Lawley et al., 2003; Fricke et al., 2009; Wozniak et al., 2009). In both SRIs and ACPs, the *mpf1a* module contains the *traLEKB* genes, while *traAV* are found in *mpf1b* (**Figure 1**; Armshaw and Pembroke, 2013; Carraro et al., 2014a, 2015a; Poulin-Laprade et al., 2015). The *mpf2* modules are organized differently in SRIs (*mpf2*: *dsbC-traC-s093-trhF-traWUN*) and ACPs (*mpf2a*: *dsbC-traC-vcrx079-trhF-traW-vcrx082-traU* and *mpf2b*: *traN* in pVCR94ΔX), the latters expressing *traN* from its own AcaCD-dependent promoter (**Figures 3B,C**). Finally, the *mpf3* module (*traFHG*) has the same operon structure in both types of elements.

SetCD targets were exclusively found in the conserved backbone of SRIs (Poulin-Laprade et al., 2015). In contrast, AcaCD binding sites were also detected upstream of operons containing genes of unknown functions, as well as in regions that are not conserved (Carraro et al., 2014a, 2015a). The relevance of these AcaCD-regulated genes for the biology of ACPs remains to be determined.

Activation of RecA-independent Homologous Recombination Functions

In addition to conjugative transfer functions, SRIs and ACPs code for diverse mutagenic and recombination functions. Both types of elements include the well-conserved *bet* and *exo* genes, which code for a λ Red-like RecA-independent homologous recombination system (Garriss et al., 2009). This system contributes to the formation of hybrid ICEs by recombineering elements inserted in tandem in the chromosome, generating new patterns of antibiotic resistance genes. In both SRIs and in ACPs, the expression of *bet* and *exo* is under the control of the SetCD- and AcaCD-dependent P_{s089} and $P_{vcrx087}$ promoters, respectively (Garriss et al., 2013; Carraro et al., 2014a; Poulin-Laprade et al., 2015). In both cases, the promoter driving their expression exhibits the highest ChIP-exo enrichment peaks. Although *bet* and *exo* are highly transcribed, their expression is hindered by a strong translational attenuator located upstream of *bet* in SXT (Garriss et al., 2013). This translational attenuator is also present in ACPs, but its functionality remains to be investigated.

SetCD and AcaCD Trigger the Expression of Genomic Island-bound Genes

Several autonomous conjugative elements were shown to mobilize non-autonomous GIs using various mechanisms (Bellanger et al., 2014). For instance, the conjugative transposon Tn*916* *trans*-mobilizes the 1.7 kb-GI mTnSAG1 from *Streptococcus agalactiae* by recognition of a cryptic *oriT* located within the *lnu*(C) gene, which confers resistance to lincomycin (Achard and Leclercq, 2007). ICEs from *Streptococcus thermophilus* were shown to *cis*-mobilize elements called CIMEs (*cis*-mobilizable elements) by a mechanism designated as accretion-mobilization (Pavlovic et al., 2004; Bellanger et al., 2011). SRIs and ACPs can also *trans*-mobilize diverse GIs using distinct strategies for their dissemination. Interestingly, these strategies are all coupled to the

regulatory network of their cognate helper element (Daccord et al., 2010, 2012, 2013; Douard et al., 2010; Carraro et al., 2014a, 2015a; Poulin-Laprade et al., 2015).

SRIs-dependent Mobilization of Genomic Islands

Characterization of the *oriT* sequence of SRIs allowed identification of chromosomal *oriT*-like sequences that were more than 63% identical (Ceccarelli et al., 2008; Daccord et al., 2010). Further investigations revealed that these cryptic *oriT* sequences belong to MGIs integrated at the 3′ end of *yicC* in the chromosome of *Vibrio*, *Alteromonas*, *Pseudoalteromonas*, and *Methylophaga* species (Daccord et al., 2010, 2013). The size of these MGIs ranges from 18 to 33 kb with a conserved core sequence of ∼5.5 kb encompassing four genes (int_{MGI}, *cds4*, *cds8*, *rdfM*) and their cognate regulatory sequences. int_{MGI} and *rdfM* code for the integrase and recombination directionality factor that allow MGIs to excise from and integrate into the host cell chromosome. Function of *cds4* and *cds8* remains unknown. The conserved backbone of MGIs is disrupted by DNA fragments that vary in size and gene content. Most of these genes code for adaptive functions such as type I and type III restriction-modification system that may confer resistance to bacteriophage infection (Daccord et al., 2013).

The initial step of an MGI mobilization by SRIs is its excision from the chromosome. Excision requires the transcriptional activation of int_{MGI} and *rdfM* by the SRIs-encoded master activator SetCD (**Figure 3B**), and the subsequent recombination between the $attL_{MGI}$ and $attR_{MGI}$ attachment sites flanking the MGI (Daccord et al., 2012; Poulin-Laprade et al., 2015). The resulting circular extrachromosomal MGI carries $oriT_{MGI}$, which acts as a *cis*-acting sequence that mimics the *oriT* of SRIs and hijacks the relaxosome encoded by SRIs. Ultimately, the MGI is translocated to the recipient cell throught the mating apparatus encoded by SRIs. Once in the recipient cell, the MGI becomes completely independent of the helper SRI and its transcriptional activator SetCD to establish itself in the new host. The MGI constitutively expresses int_{MGI} at a low level, thereby allowing its own integration into the 3′ end of *yicC* (Daccord et al., 2012). MGI*Vfl*Ind1, initially isolated from *Vibrio vulnificus*, was used as a prototype to study MGIs and was reported to be mobilized at high frequency between *E. coli* strains by both ICE*Vfl*Ind1 and SXT (10^{-3} transconjugants per donor cell). This frequency rose to 10^{-1} transconjugants per donor cell upon overexpression of *setCD* or induction with mitomycin C treatment (Daccord et al., 2010, 2012). MGI*Vfl*Ind1 is also able to *cis*-mobilize over 1 Mb of chromosomal DNA located 5′ of *yicC* in an Hfr-like manner (Daccord et al., 2010). Chromosomal DNA mobilization by MGI*Vfl*Ind1 involves initiation of transfer at $oriT_{MGI}$ by the relaxosome of a SRI prior to excision of the MGI from the chromosome.

ACP-dependent Mobilization of Genomic Islands

Discovery of the sequences targeted by AcaCD in ACPs was the cornerstone for the identification of potential chromosomal targets in genomes of several *Enterobacteriaceae* and *Vibrionaceae* (Carraro et al., 2014a, 2015a). Notably, multiple AcaCD binding sites were detected in the *Salmonella* genomic island 1 (SGI1), which confers MDR to pathogenic *S. enterica* and was reported to be mobilized in *trans* by ACPs by an unknown mechanism (Doublet et al., 2005; Douard et al., 2010; Carraro et al., 2014a). AcaCD sites were also detected in other GIs that are or could be mobilized by ACPs (Carraro et al., 2014a, 2015a).

SGI1-like Elements

SGI1 is a 43-kb chromosomal mobile element carrying a class 1 integron that confers resistance to ampicillin, chloramphenicol, streptomycin, sulfonamides and tetracycline (ACSSuT phenotype; Mulvey et al., 2006). SGI1 and related MDR-conferring GIs have been found integrated at the 3′ end of *thdF* (*trmE*) in a large variety of multidrug resistant *S. enterica* serovars and in *P. mirabilis* (Boyd et al., 2008; Wiesner et al., 2009; Hall, 2010; Girlich et al., 2015). All SGI1-like elements share a highly conserved ∼27 kb core region, which mostly contain genes of unknown function (Mulvey et al., 2006; Boyd et al., 2008). The conserved genes *int* and *xis* mediate SGI1's excision from and integration into the chromosome (Doublet et al., 2005). Three conserved *tra* genes code for putative mating pore components sharing 40, 60, and 78% identity with ACP's TraN, TraG and TraH, respectively. In most SGI1-like elements, the core region is disrupted between the resolvase-encoding gene *res* and *s044* by complex integrons conferring MDR (Boyd et al., 2008). Interestingly, a similar variable region is inserted in *s023* in the SGI1-variant SGI2 (formerly SGI1-J; Levings et al., 2008).

Currently, the genetic determinants and the nature of the interactions allowing the specific mobilization of SGI1 by ACPs remain largely unknown. Recent work revealed that the transcriptional activator AcaCD encoded by ACPs triggers the excision of SGI1 (Carraro et al., 2014a). Indeed, AcaCD-binding motifs were identified in the promoter region of the recombination directionality factor-encoding gene *xis* as well as upstream of putative regulatory and conjugation proteins (**Figure 3C**). SGI1 was reported to be highly stable, even after 350 generations without selective pressure (Kiss et al., 2012). Nevertheless, these assays were carried out in cells lacking an ACP. This is a major bias since SGI1 cannot excise in the absence of AcaCD (Carraro et al., 2014a). Interspecies mobilization of SGI1 between *S. enterica* and *E. coli* was reported to be highly variable (10^{-5} to 10^{-2} transconjugants per donor cell after overnight matings), depending on the *Salmonella* strain, the SGI1 variant and the IncA/C helper plasmid (Doublet et al., 2005; Kiss et al., 2012). In contrast, the frequency of transfer of SGI1 mobilized by pVCR94ΔX between *E. coli* strains was so high that virtually all recipient cells harbored SGI1 after mating (Carraro et al., 2014a).

MGI*Vmi*1-like Elements

ACPs also *trans*-mobilize MGI*Vmi*1, a 16.5 kb element that belongs to a family of MGIs unrelated to SGI1 and to the MGIs mobilized by SRIs (Carraro et al., 2014a, 2015a). AcaCD binding sites were detected upstream of *490* and *xis* (formally *420*), which code respectively for a 195-amino acids MobI-like protein and a 94-amino acid residue putative recombination directionality factor (**Figure 3C**). Similar characteristics are found in several GIs inserted in *Vibrio mimicus*, *Vibrio parahaemolyticus*, and

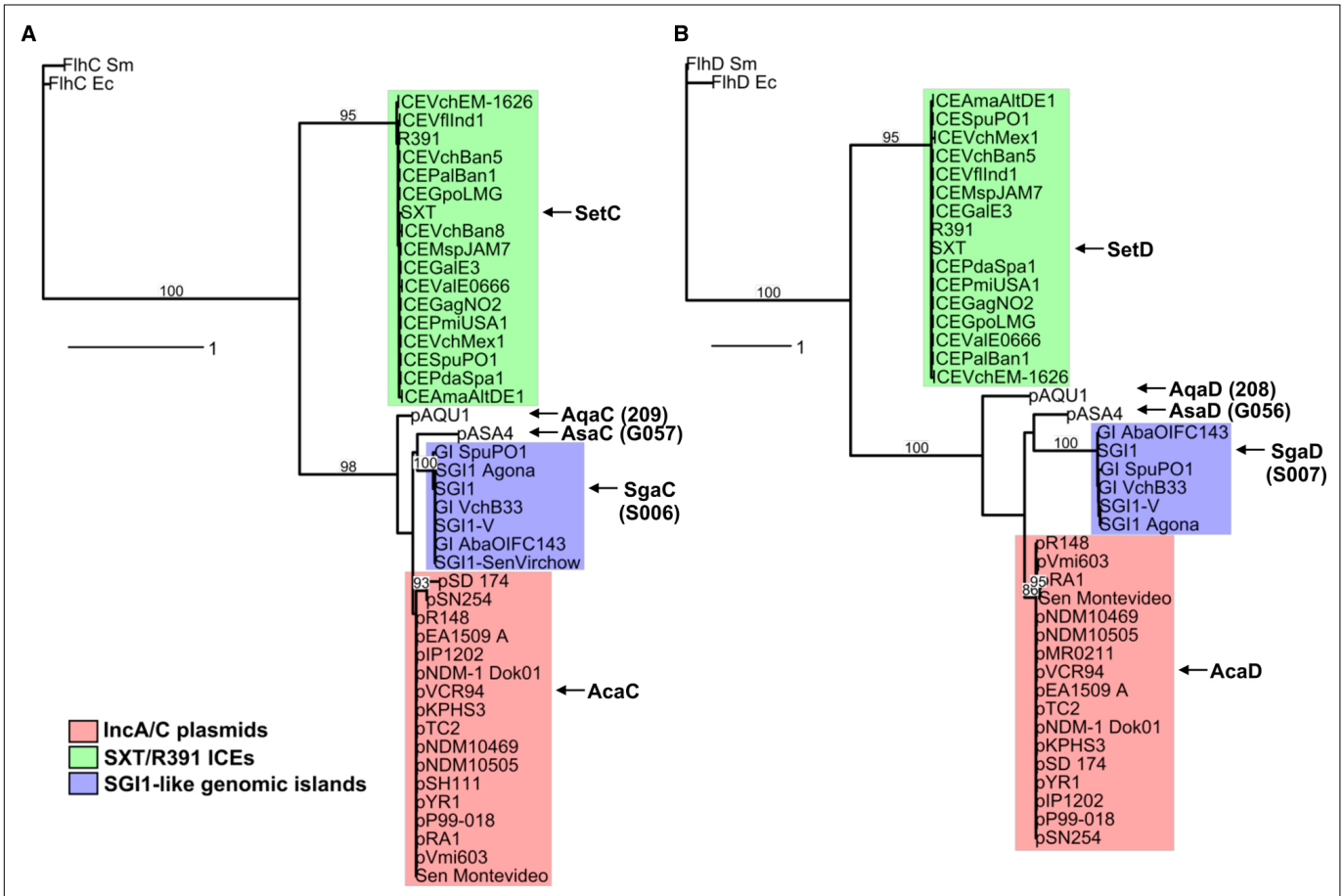

FIGURE 4 | Phylogenetic trees based on alignments of amino acid sequences of SetC/AcaC (A) and SetD/AcaD (B) orthologs. The flagellar transcriptional activator proteins FlhC and FlhD of *E. coli* (Ec) and *Serratia marcescens* (Sm) were used as outgroups in phylogenetic analyses. Bootstrap values are indicated when over 80%. The individual scale bars represent genetic distances.

Shewanella putrefaciens genomes (Carraro et al., 2015a). Although these GIs are highly variable in size and content, their conserved key features strongly suggest that they are mobilizable by ACPs. Excision of MGI*Vmi*1 was found to be AcaCD-dependent and its transfer requires the conjugative machinery encoded by ACPs. Here again, the exact mechanism remains unknown and further investigation is needed. However, based on the homology with the structure of the *mob1* mobilization module of SRIs and ACPs, we hypothesized that the *oriT* locus of these GIs lies within the large intergenic region located upstream of the AcaCD-dependent gene coding for a MobI homolog (Carraro et al., 2015a).

Diversity of FlhCD-like Transcriptional Activators Amongst Conjugative Elements

A search for additional FlhCD-like regulators amongst other mobile genetic elements was carried out. Because homologies between SetCD and AcaCD, and the master flagellar activator FlhCD, are very weak, the Pfam HMM profiles for FlhC (PF05280) and FlhD (PF05247) domains are unsuitable to find functional orthologs of SetC/AcaC-like and SetD/AcaD-like activators in bacterial genomes. To solve this problem, we generated new HMM profiles based on alignments of the primary sequence of SetC/AcaC and SetD/AcaD protein orthologs. Screening of the Genbank non-redundant protein sequence database using these new profiles revealed a large number of homologous proteins encoded by diverse types of mobile genetic elements in *Enterobacteriaceae* and *Vibrionaceae*. Phylogenetic reconstructions using a representative subset of SetC/AcaC and SetD/AcaD orthologs revealed identical clustering in three distinct families distantly related to FlhC and FlhD: (i) SetC and SetD encoded by SRIs, (ii) AcaC and AcaD encoded by ACPs, (iii) putative proteins encoded by SGI1-like elements, S006 and S007, here renamed SgaC and SgaD (SGI1 activator subunits C and D), (iv) putative proteins encoded by pAQU1-like conjugative plasmids, 208 and 209 that we named AqaD and AqaC (pAQU1 activator subunits C and D), (v) putative proteins encoded by pAsa4, G057 and G056, here renamed AsaC and AsaD (pAsa4 activator subunits C and D; **Figure 4**). Interestingly, the genes coding for these putative transcriptional regulators are found in a similar genetic context in all cases (**Figure 5**). They are found in close proximity to a gene coding for a TraG homolog, a component of the conjugative apparatus, a gene

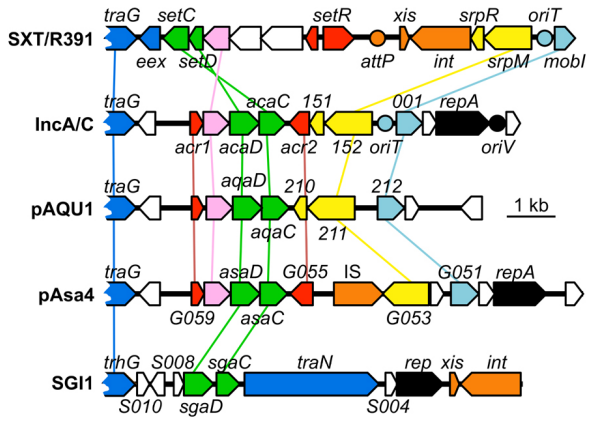

FIGURE 5 | Comparison of the genetic context of genes coding for SetCD/AcaCD orthologs in diverse mobile genetic elements. Schematic representation of demonstrated or putative regulatory regions of SXT from *V. cholerae* O139 (AY055428.1), pVCR94 from *V. cholerae* O1 El Tor (NC_023291), pAQU1 from *P. damselae* subsp. *damselae* (NC_016983.1), pAsa4 from *Aeromonas salmonicida* (CP000645.1) and *Salmonella* genomic island 1 (SGI1) from *S. enterica* subsp. *enterica* serovar Typhimurium DT104 (AF261825.2). Arrows of similar color represent genes predicted to have similar functions and are color-coded as in **Figure 1** legend. Circles indicate the position of origins of transfer (pale blue, *oriT*), of the origin of replication (black, *oriV*) of pVCR94 based on identity with pRA1, and the position of the *attP* site for chromosomal integration of SXT by site-specific recombination (orange).

coding for putative lysozyme-like protein. and genes coding for homologs of the SrpRM partition system and MobI protein. Further investigation are needed to confirm the functionality of these putative activator complexes regarding the activation of their cognate mobile GIs, and their potential impact on other genetic elements.

Concluding Remarks and Perspectives

In the current context of rapid emergence and spread of MDR, it has become essential to get a better understanding of the dynamics and the mechanisms of transfer and regulation of mobile GIs that promote such resistance. A plethora of studies have been aimed at characterizing the determinants of transfer of various model conjugative genetic elements such as ICE*Bs1*, Tn*916*, CTnDOT, R388, pESBL, pSL20, R27 (Celli and Trieu-Cuot, 1998; Marra and Scott, 1999; Auchtung et al., 2005; Gibert et al., 2013; Waters and Salyers, 2013; Fernandez-Lopez et al., 2014; Yamaichi et al., 2015). Extensive research on SRIs and recent work on ACPs have refined our grasp of the biology and regulation of these major players of MDR propagation. Classical genetics and modern molecular methods have facilitated the complete characterization of the regulons of SetCD and AcaCD, which greatly improved our understanding of SRIs, ACPs, and the elements they mobilize.

Mobilization of GIs requires the SetCD- or AcaCD-dependent activation of their excision, and involves three distinct mechanisms of *oriT* recognition and DNA processing. First, the MGI-SRI model is based on the recognition of a MGI-borne sequence mimicking the *oriT* of the self-transmissible SRI. This *oriT* counterfeit is likely recognized as a genuine *oriT* by the relaxosome of SRIs, thereby leading to MGI transfer through the SRI-encoded mating pore. Second, the MGI*Vmi*1-ACP model likely relies on *oriT* recognition of MGI by its cognate MobI-like protein ($MobI_{MGI}$), whose expression depends on AcaCD. The subsequent DNA processing at *oriT* of the MGI likely results from the interaction of $MobI_{MGI}$ with the $MobI_{A/C}$-less relaxosome encoded by ACPs. Finally, the SGI1-ACP model remains the most elusive mechanism of mobilization as no $oriT_{IncA/C}$ sequence or MobI-like encoding gene has been identified so far in the sequence of SGI1-like elements. Further experiments are ongoing to precisely determine the mechanisms leading to the mobiliation of such GIs, and potentially of newly identified GIs.

Exploitation of experimentally determined SetCD and AcaCD recognition motifs to use them as specific tags for *in silico* analyses of bacterial genomes is a powerful approach to identify new mobile GIs mobilized by either SRIs or ACPs (Carraro et al., 2014a, 2015a). Moreover, additional FlhCD-like regulators were identified, which given their degree of divergence with AcaCD and SetCD, likely recognized unrelated DNA motifs. We anticipate that characterization of these motifs will facilitate the discovery of additional families of MGIs in sequenced bacterial genomes.

Mobile genetic elements and their bacterial host are inherently connected. Horizontally acquired DNA is often silenced by bacterial host factors such as H-NS-like proteins, most likely to limit the impact of exogenous DNA on endogenous regulatory networks and metabolic pathways (Dorman, 2004, 2014; Navarre et al., 2006; Singh et al., 2014). In the case of SRIs, IHF was shown to be mandatory for *V. cholerae* to act as a donor of SXT, but dispensable in *E. coli* donors (McLeod et al., 2006). The host factor Fis does not influence SXT transfer to or from *V. cholerae*. Further studies will be required to evaluate the influence of host factors on the dynamics of ACPs. Reciprocally, SRIs and ACPs, could interfere with the regulation of the host cell cellular processes. Besides the targets identified in MGIs, no clear interactions of SetCD and AcaCD with chromosomal loci in *E. coli* were detected (Carraro et al., 2014a; Poulin-Laprade et al., 2015). Nevertheless, considering the limitations of experimental settings and the broad host range of SRIs and ACPs, interactions with host metabolic pathways or other bacterial responses cannot be excluded. SetCD and AcaCD could also target plasmids of other incompatibility groups or other mobile GIs. Finally, other conjugative elements that code for FlhCD-like regulators could have a significant impact on their host biology.

Further investigations on the regulation of self-transmissible and mobilizable genetic elements will ultimately unravel the interconnections and the mechanism by which MDR and other adaptive traits spread among bacterial populations. In the war against the rampant emergence of untreatable pathogenic bacteria, fundamental knowledge regarding the key determinants at play for the dissemination of MDR will be an undeniable asset. To prevent a possible and imminent post-antibiotic era, mankind could find its salvation in the development of a new generation of molecules targeting key regulators aimed at confining or halting the exchange of MDR-conferring mobile genetic elements in patients, or even cure them from bacterial populations.

Acknowledgments

We are thankful to A. Lavigueur for critical reading of the manuscript. We are thankful to D. Matteau for its contribution to **Figure 1**. This work was supported by a Discovery Grant and Discovery Acceleration Supplement from the Natural Sciences and Engineering Council of Canada (326810 and 412288 to VB). VB holds a Canada Research Chair in molecular bacterial genetics. DPL was supported by a scholarship from the Fonds de Recherche du Québec–Nature et Technologies (FRQNT). After acceptance of this review, our results demonstrating the involvement of AcaCD in the excision of SGI1 (Carraro et al., 2014a) were confirmed by Kiss et al. (2005). This paper also explores the role of SgaCD (**Figures 4** and **5** of this review) and strongly suggests that similarly to its IncA/C-encoded counterpart AcaCD, SgaCD of SGI1 activates the promoter of *xis* and the subsequent excision of this genomic island.

References

Achard, A., and Leclercq, R. (2007). Characterization of a small mobilizable transposon, MTnSag1, in *Streptococcus agalactiae*. *J. Bacteriol.* 189, 4328–4331. doi: 10.1128/JB.00213-07

Armshaw, P., and Pembroke, J. T. (2013). Control of expression of the ICE R391 encoded UV-inducible cell-sensitising function. *BMC Microbiol.* 13:195. doi: 10.1186/1471-2180-13-195

Auchtung, J. M., Lee, C. A., Monson, R. E., Lehman, A. P., and Grossman, A. D. (2005). Regulation of a Bacillus subtilis mobile genetic element by intercellular signaling and the global DNA damage response. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12554–12559. doi: 10.1073/pnas.0505835102

Badhai, J., Kumari, P., Krishnan, P., Ramamurthy, T., and Das, S. K. (2013). Presence of SXT integrating conjugative element in marine bacteria isolated from the mucus of the coral *Fungia echinata* from Andaman Sea. *FEMS Microbiol. Lett.* 338, 118–123. doi: 10.1111/1574-6968.12033

Baharoglu, Z., Bikard, D., and Mazel, D. (2010). Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLoS Genet.* 6:e1001165. doi: 10.1371/journal.pgen.1001165

Baharoglu, Z., Krin, E., and Mazel, D. (2012). Connecting environment and genome plasticity in the characterization of transformation-induced SOS regulation and carbon catabolite control of the *Vibrio cholerae* integron integrase. *J. Bacteriol.* 194, 1659–1667. doi: 10.1128/JB.05982-11

Bauernfeind, A., Stemplinger, I., Jungwirth, R., and Giamarellou, H. (1996). Characterization of the plasmidic beta-lactamase CMY-2, which is responsible for cephamycin resistance. *Antimicrob. Agents Chemother.* 40, 221–224.

Baxter, J. C., and Funnell, B. E. (2014). Plasmid partition mechanisms. *Microbiol. Spectr.* 2, 1–20. doi: 10.1128/microbiolspec.PLAS-0023-2014

Beaber, J. W., Burrus, V., Hochhut, B., and Waldor, M. K. (2002a). Comparison of SXT and R391, two conjugative integrating elements: definition of a genetic backbone for the mobilization of resistance determinants. *Cell. Mol. Life Sci.* 59, 2065–2070. doi: 10.1007/s000180200006

Beaber, J. W., Hochhut, B., and Waldor, M. K. (2002b). Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *J. Bacteriol.* 184, 4259–4269. doi: 10.1128/JB.184.15.4259-4269.2002

Beaber, J. W., Hochhut, B., and Waldor, M. K. (2004). SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 427, 72–74. doi: 10.1038/nature02241

Beaber, J. W., and Waldor, M. K. (2004). Identification of operators and promoters that control SXT conjugative transfer. *J. Bacteriol.* 186, 5945–5949. doi: 10.1128/JB.186.17.5945-5949.2004

Bellanger, X., Morel, C., Gonot, F., Puymege, A., Decaris, B., and Guedon, G. (2011). Site-specific accretion of an integrative conjugative element together with a related genomic island leads to *cis* mobilization and gene capture. *Mol. Microbiol.* 81, 912–925. doi: 10.1111/j.1365-2958.2011.07737.x

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

Bellanger, X., Roberts, A. P., Morel, C., Choulet, F., Pavlovic, G., Mullany, P., et al. (2009). Conjugative transfer of the integrative conjugative elements ICESt1 and ICESt3 from *Streptococcus thermophilus*. *J. Bacteriol.* 191, 2764–2775. doi: 10.1128/JB.01412-08

Bi, D., Xu, Z., Harrison, E. M., Tai, C., Wei, Y., He, X., et al. (2012). ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Res.* 40, D621–D626. doi: 10.1093/nar/gkr846

Boltner, D., Macmahon, C., Pembroke, J. T., Strike, P., and Osborn, A. M. (2002). R391: a conjugative integrating mosaic comprised of phage, plasmid, and transposon elements. *J. Bacteriol.* 184, 5158–5169. doi: 10.1128/JB.184.18.5158-5169.2002

Bordeleau, E., Brouillette, E., Robichaud, N., and Burrus, V. (2010). Beyond antibiotic resistance: integrating conjugative elements of the SXT/R391 family that encode novel diguanylate cyclases participate to c-di-GMP signalling in *Vibrio cholerae*. *Environ. Microbiol.* 12, 510–523. doi: 10.1111/j.1462-2920.2009.02094.x

Boyd, D. A., Shi, X., Hu, Q. H., Ng, L. K., Doublet, B., Cloeckaert, A., et al. (2008). *Salmonella* genomic island 1 (SGI1), variant SGI1-I, and new variant SGI1-O in *Proteus mirabilis* clinical and food isolates from China. *Antimicrob. Agents Chemother.* 52, 340–344. doi: 10.1128/AAC.00902-07

Browning, D. F., and Busby, S. J. (2004). The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* 2, 57–65. doi: 10.1038/nrmicro787

Burrus, V., Marrero, J., and Waldor, M. K. (2006). The current ICE age: biology and evolution of SXT-related integrating conjugative elements. *Plasmid* 55, 173–183. doi: 10.1016/j.plasmid.2006.01.001

Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. (2002). Conjugative transposons: the tip of the iceberg. *Mol. Microbiol.* 46, 601–610. doi: 10.1046/j.1365-2958.2002.03191.x

Burrus, V., and Waldor, M. K. (2003). Control of SXT integration and excision. *J. Bacteriol.* 185, 5045–5054. doi: 10.1128/JB.185.17.5045-5054.2003

Burrus, V., and Waldor, M. K. (2004). Shaping bacterial genomes with integrative and conjugative elements. *Res. Microbiol.* 155, 376–386. doi: 10.1016/j.resmic.2004.01.012

Call, D. R., Singer, R. S., Meng, D., Broschat, S. L., Orfe, L. H., Anderson, J. M., et al. (2010). blaCMY-2-positive IncA/C plasmids from *Escherichia coli* and *Salmonella enterica* are a distinct component of a larger lineage of plasmids. *Antimicrob. Agents Chemother.* 54, 590–596. doi: 10.1128/AAC.00055-09

Cambray, G., Sanchez-Alberola, N., Campoy, S., Guerin, E., Da Re, S., Gonzalez-Zorn, B., et al. (2011). Prevalence of SOS-mediated control of integron integrase expression as an adaptive trait of chromosomal and mobile integrons. *Mob. DNA* 2, 6. doi: 10.1186/1759-8753-2-6

Carattoli, A. (2013). Plasmids and the spread of resistance. *Int. J. Med. Microbiol.* 303, 298–304. doi: 10.1016/j.ijmm.2013.02.001

Carattoli, A., Bertini, A., Villa, L., Falbo, V., Hopkins, K. L., and Threlfall, E. J. (2005). Identification of plasmids by PCR-based replicon typing. *J. Microbiol. Methods* 63, 219–228. doi: 10.1016/j.mimet.2005.03.018

Carraro, N., and Burrus, V. (2014). Biology of Three ICE Families: SXT/R391, ICEBs1, and ICESt1/ICESt3. *Microbiol. Spectr.* 2, 1–20. doi: 10.1128/microbiolspec.MDNA3-0008-2014

Carraro, N., Libante, V., Morel, C., Decaris, B., Charron-Bourgoin, F., Leblond, P., et al. (2011). Differential regulation of two closely related integrative and conjugative elements from *Streptococcus thermophilus*. *BMC Microbiol.* 11:238. doi: 10.1186/1471-2180-11-238

Carraro, N., Matteau, D., Burrus, V., and Rodrigue, S. (2015a). Unraveling the regulatory network of IncA/C plasmid mobilization: when genomic islands hijack conjugative elements. *Mob. Genet. Elements* 5, 1–5. doi: 10.1080/2159256x.2015.1045116

Carraro, N., Poulin, D., and Burrus, V. (2015b). Replication and active partition of integrative and conjugative elements (ICEs) of the SXT/R391 family: the line between ICEs and conjugative plasmids is getting thinner. *PLoS Genet.* 11:e1005298. doi: 10.1371/journal.pgen.1005298

Carraro, N., Matteau, D., Luo, P., Rodrigue, S., and Burrus, V. (2014a). The master activator of IncA/C conjugative plasmids stimulates genomic

islands and multidrug resistance dissemination. *PLoS Genet.* 10:e1004714. doi: 10.1371/journal.pgen.1004714

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

Ceccarelli, D., Daccord, A., Rene, M., and Burrus, V. (2008). 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.* 190, 5328–5338. doi: 10.1128/JB.00150-08

Celli, J., and Trieu-Cuot, P. (1998). Circularization of Tn916 is required for expression of the transposon-encoded transfer functions: characterization of long tetracycline-inducible transcripts reading through the attachment site. *Mol. Microbiol.* 28, 103–117.

Chen, Z., Yang, H., and Pavletich, N. P. (2008). Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453, 489–484. doi: 10.1038/nature06971

Chevance, F. F., and Hughes, K. T. (2008). Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* 6, 455–465. doi: 10.1038/nrmicro1887

Coetzee, J. N., Datta, N., and Hedges, R. W. (1972). R factors from *Proteus rettgeri*. *J. Gen. Microbiol.* 72, 543–552.

Daccord, A., Ceccarelli, D., and Burrus, V. (2010). 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.* 78, 576–588. doi: 10.1111/j.1365-2958.2010.07364.x

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

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

Davies, J., and Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74, 417–433. doi: 10.1128/MMBR.00016-10

de la Cruz, F., Frost, L. S., Meyer, R. J., and Zechner, E. L. (2010). Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol. Rev.* 34, 18–40. doi: 10.1111/j.1574-6976.2009.00195.x

Del Castillo, C. S., Hikima, J., Jang, H. B., Nho, S. W., Jung, T. S., Wongtavatchai, J., et al. (2013). Comparative sequence analysis of a multidrug-resistant plasmid from *Aeromonas hydrophila*. *Antimicrob. Agents Chemother.* 57, 120–129. doi: 10.1128/AAC.01239-12

Dimopoulou, I. D., Russell, J. E., Mohd-Zain, Z., Herbert, R., and Crook, D. W. (2002). Site-specific recombination with the chromosomal tRNA(Leu) gene by the large conjugative Haemophilus resistance plasmid. *Antimicrob. Agents Chemother.* 46, 1602–1603. doi: 10.1128/AAC.46.5.1602-1603.2002

Ding, H., Yang, Y., Lu, Q., Wang, Y., Chen, Y., Deng, L., et al. (2008). The prevalence of plasmid-mediated AmpC beta-lactamases among clinical isolates of *Escherichia coli* and *Klebsiella pneumoniae* from five children's hospitals in China. *Eur. J. Clin. Microbiol. Infect. Dis.* 27, 915–921. doi: 10.1007/s10096-008-0532-4

Dorman, C. J. (2004). H-NS: a universal regulator for a dynamic genome. *Nat. Rev. Microbiol.* 2, 391–400. doi: 10.1038/nrmicro883

Dorman, C. J. (2014). H-NS-like nucleoid-associated proteins, mobile genetic elements and horizontal gene transfer in bacteria. *Plasmid* 75, 1–11. doi: 10.1016/j.plasmid.2014.06.004

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

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

Dziewit, L., Jazurek, M., Drewniak, L., Baj, J., and Bartosik, D. (2007). The SXT conjugative element and linear prophage N15 encode toxin-antitoxin-stabilizing systems homologous to the tad-ata module of the *Paracoccus aminophilus* plasmid pAMI2. *J. Bacteriol.* 189, 1983–1997. doi: 10.1128/JB.01610-06

Escudero, J., Loot, C., Nivina, A., and Mazel, D. (2014). The integron: adaptation on demand. *Microbiol. Spectr.* 3, 1–22. doi: 10.1128/microbiolspec.MDNA3-0019-2014

Fernandez-Alarcon, C., Singer, R. S., and Johnson, T. J. (2011). Comparative genomics of multidrug resistance-encoding IncA/C plasmids from commensal and pathogenic *Escherichia coli* from multiple animal sources. *PLoS ONE* 6:e23415. doi: 10.1371/journal.pone.0023415

Fernandez-Lopez, R., Del Campo, I., Revilla, C., Cuevas, A., and De La Cruz, F. (2014). Negative feedback and transcriptional overshooting in a regulatory network for horizontal gene transfer. *PLoS Genet.* 10:e1004171. doi: 10.1371/journal.pgen.1004171

Fitzgerald, D. M., Bonocora, R. P., and Wade, J. T. (2014). Comprehensive mapping of the *Escherichia coli* flagellar regulatory network. *PLoS Genet.* 10:e1004649. doi: 10.1371/journal.pgen.1004649

Folster, J. P., Pecic, G., Mccullough, A., Rickert, R., and Whichard, J. M. (2011). Characterization of bla(CMY)-encoding plasmids among *Salmonella isolated* in the United States in 2007. *Foodborne Pathog. Dis.* 8, 1289–1294. doi: 10.1089/fpd.2011.0944

Folster, J. P., Pecic, G., Singh, A., Duval, B., Rickert, R., Ayers, S., et al. (2012). Characterization of extended-spectrum cephalosporin-resistant *Salmonella enterica* serovar Heidelberg isolated from food animals, retail meat, and humans in the United States 2009. *Foodborne Pathog. Dis.* 9, 638–645. doi: 10.1089/fpd.2012.1130

Fricke, W. F., Welch, T. J., Mcdermott, P. F., Mammel, M. K., Leclerc, J. E., White, D. G., et al. (2009). Comparative genomics of the IncA/C multidrug resistance plasmid family. *J. Bacteriol.* 191, 4750–4757. doi: 10.1128/JB.00189-09

Galimand, M., Guiyoule, A., Gerbaud, G., Rasoamanana, B., Chanteau, S., Carniel, E., et al. (1997). Multidrug resistance in *Yersinia pestis* mediated by a transferable plasmid. *N. Engl. J. Med.* 337, 677–680. doi: 10.1056/NEJM199709043371004

Garriss, G., and Burrus, V. (2013). "Integrating conjugative elements of the SXT/R391 Family," in *Bacterial Integrative Mobile Genetic Elements*, eds A. P. Roberts and P. Mullany (Austin, TX: Landes Biosciences), 217–234.

Garriss, G., Poulin-Laprade, D., and Burrus, V. (2013). DNA-damaging agents induce the RecA-independent homologous recombination functions of integrating conjugative elements of the SXT/R391 family. *J. Bacteriol.* 195, 1991–2003. doi: 10.1128/JB.02090-12

Garriss, G., Waldor, M. K., and Burrus, V. (2009). Mobile antibiotic resistance encoding elements promote their own diversity. *PLoS Genet.* 5:e1000775. doi: 10.1371/journal.pgen.1000775

Gibert, M., Juarez, A., Madrid, C., and Balsalobre, C. (2013). New insights in the role of HtdA in the regulation of R27 conjugation. *Plasmid* 70, 61–68. doi: 10.1016/j.plasmid.2013.01.009

Giles, W. P., Benson, A. K., Olson, M. E., Hutkins, R. W., Whichard, J. M., Winokur, P. L., et al. (2004). DNA sequence analysis of regions surrounding blaCMY-2 from multiple *Salmonella* plasmid backbones. *Antimicrob. Agents Chemother.* 48, 2845–2852. doi: 10.1128/AAC.48.8.2845-2852.2004

Gimble, F. S., and Sauer, R. T. (1985). Mutations in bacteriophage lambda repressor that prevent RecA-mediated cleavage. *J. Bacteriol.* 162, 147–154.

Girlich, D., Dortet, L., Poirel, L., and Nordmann, P. (2015). Integration of the blaNDM-1 carbapenemase gene into *Proteus* genomic island 1 (PGI1-PmPEL) in a *Proteus mirabilis* clinical isolate. *J. Antimicrob. Chemother.* 70, 98–102. doi: 10.1093/jac/dku371

Glenn, L. M., Lindsey, R. L., Frank, J. F., Meinersmann, R. J., Englen, M. D., Fedorka-Cray, P. J., et al. (2011). Analysis of antimicrobial resistance genes detected in multidrug-resistant *Salmonella enterica* serovar Typhimurium isolated from food animals. *Microb. Drug Resist.* 17, 407–418. doi: 10.1089/mdr.2010.0189

Gorgani, N., Ahlbrand, S., Patterson, A., and Pourmand, N. (2009). Detection of point mutations associated with antibiotic resistance in *Pseudomonas aeruginosa*. *Int. J. Antimicrob. Agents* 34, 414–418. doi: 10.1016/j.ijantimicag.2009.05.013

Grohmann, E. (2010). Autonomous plasmid-like replication of Bacillus ICEBs1: a general feature of integrative conjugative elements? *Mol. Microbiol.* 75, 261–263. doi: 10.1111/j.1365-2958.2009.06978.x

Guerin, E., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Da Re, S., et al. (2009). The SOS response controls integron recombination. *Science* 324, 1034. doi: 10.1126/science.1172914

Guglielmini, J., Quintais, L., Garcillan-Barcia, M. P., De La Cruz, F., and Rocha, E. P. (2011). The repertoire of ICE in prokaryotes underscores the

unity, diversity, and ubiquity of conjugation. *PLoS Genet.* 7:e1002222. doi: 10.1371/journal.pgen.1002222

Guo, Y. F., Zhang, W. H., Ren, S. Q., Yang, L., Lu, D. H., Zeng, Z. L., et al. (2014). IncA/C plasmid-mediated spread of CMY-2 in multidrug-resistant *Escherichia coli* from food animals in China. *PLoS ONE* 9:e96738. doi: 10.1371/journal.pone.0096738

Haft, R. J., Mittler, J. E., and Traxler, B. (2009). Competition favours reduced cost of plasmids to host bacteria. *ISME J.* 3, 761–769. doi: 10.1038/ismej.2009.22

Hall, R. M. (2010). *Salmonella* genomic islands and antibiotic resistance in *Salmonella enterica*. *Future Microbiol.* 5, 1525–1538. doi: 10.2217/fmb.10.122

Harada, S., Ishii, Y., Saga, T., Tateda, K., and Yamaguchi, K. (2010). Chromosomally encoded blaCMY-2 located on a novel SXT/R391-related integrating conjugative element in a *Proteus mirabilis* clinical isolate. *Antimicrob. Agents Chemother.* 54, 3545–3550. doi: 10.1128/AAC.00111-10

Hara-Kudo, Y., and Kumagai, S. (2014). Impact of seafood regulations for *Vibrio parahaemolyticus* infection and verification by analyses of seafood contamination and infection. *Epidemiol. Infect.* 142, 2237–2247. doi: 10.1017/S0950268814001897

Harmer, C. J., and Hall, R. M. (2014). pRMH760, a precursor of A/C(2) plasmids carrying blaCMY and blaNDM genes. *Microb. Drug Resist.* 20, 416–423. doi: 10.1089/mdr.2014.0012

Harmer, C. J., and Hall, R. M. (2015). The A to Z of A/C plasmids. *Plasmid* doi: 10.1016/j.plasmid.2015.04.003 [Epub ahead of print].

Hawley, D. K., and McClure, W. R. (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res.* 11, 2237–2255.

Hochhut, B., Lotfi, Y., Mazel, D., Faruque, S. M., Woodgate, R., and Waldor, M. K. (2001). Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins. *Antimicrob. Agents Chemother.* 45, 2991–3000. doi: 10.1128/AAC.45.11.2991-3000.2001

Hochhut, B., Marrero, J., and Waldor, M. K. (2000). Mobilization of plasmids and chromosomal DNA mediated by the SXT element, a constin found in *Vibrio cholerae* O139. *J. Bacteriol.* 182, 2043–2047. doi: 10.1128/JB.182.7.2043-2047.2000

Hochhut, B., and Waldor, M. K. (1999). Site-specific integration of the conjugal *Vibrio cholerae* SXT element into prfC. *Mol. Microbiol.* 32, 99–110.

Kaper, J. B., Morris, J. G. Jr., and Levine, M. M. (1995). Cholera. *Clin. Microbiol. Rev.* 8, 48–86.

Kiewitz, C., Larbig, K., Klockgether, J., Weinel, C., and Tummler, B. (2000). Monitoring genome evolution ex vivo: reversible chromosomal integration of a 106 kb plasmid at two tRNA(Lys) gene loci in sequential *Pseudomonas aeruginosa* airway isolates. *Microbiology* 146, 2365–2373. doi: 10.1099/00221287-146-10-2365

Kiss, J., Papp, P. P., Szabó, M., Farkas, T., Murányi, G., Szakálla, E. et al. (2005). 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.* doi: 10.1093/nar/gkv758 [Epub ahead of print].

Kiss, J., Nagy, B., and Olasz, F. (2012). Stability, entrapment and variant formation of *Salmonella* genomic island 1. *PLoS ONE* 7:e32497. doi: 10.1371/journal.pone.0032497

Kumar, A., Malloch, R. A., Fujita, N., Smillie, D. A., Ishihama, A., and Hayward, R. S. (1993). The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. *J. Mol. Biol.* 232, 406–418. doi: 10.1006/jmbi.1993.1400

Lawley, T. D., Klimke, W. A., Gubbins, M. J., and Frost, L. S. (2003). F factor conjugation is a true type IV secretion system. *FEMS Microbiol. Lett.* 224, 1–15. doi: 10.1016/S0378-1097(03)00430-0

Lee, C. A., Babic, A., and Grossman, A. D. (2010). Autonomous plasmid-like replication of a conjugative transposon. *Mol. Microbiol.* 75, 268–279. doi: 10.1111/j.1365-2958.2009.06985.x

Levings, R. S., Djordjevic, S. P., and Hall, R. M. (2008). SGI2, a relative of *Salmonella* genomic island SGI1 with an independent origin. *Antimicrob. Agents Chemother.* 52, 2529–2537. doi: 10.1128/AAC.00189-08

Lindsey, R. L., Frye, J. G., Fedorka-Cray, P. J., and Meinersmann, R. J. (2011). Microarray-based analysis of IncA/C plasmid-associated genes from multidrug-resistant *Salmonella enterica*. *Appl. Environ. Microbiol.* 77, 6991–6999. doi: 10.1128/AEM.00567-11

Little, J. W. (1984). Autodigestion of lexA and phage lambda repressors. *Proc. Natl. Acad. Sci. U.S.A.* 81, 1375–1379.

Liu, X., Fujita, N., Ishihama, A., and Matsumura, P. (1995). The C-terminal region of the alpha subunit of *Escherichia coli* RNA polymerase is required for transcriptional activation of the flagellar level II operons by the FlhD/FlhC complex. *J. Bacteriol.* 177, 5186–5188.

Liu, X., and Matsumura, P. (1994). The FlhD/FlhC complex, a transcriptional activator of the *Escherichia coli* flagellar class II operons. *J. Bacteriol.* 176, 7345–7351.

Llanes, C., Gabant, P., Couturier, M., Bayer, L., and Plesiat, P. (1996). Molecular analysis of the replication elements of the broad-host-range RepA/C replicon. *Plasmid* 36, 26–35. doi: 10.1006/plas.1996.0028

Llanes, C., Gabant, P., Couturier, M., and Michel-Briand, Y. (1994). Cloning and characterization of the Inc A/C plasmid RA1 replicon. *J. Bacteriol.* 176, 3403–3407.

Llosa, M., Gomis-Ruth, F. X., Coll, M., and De La Cruz Fd, F. (2002). Bacterial conjugation: a two-step mechanism for DNA transport. *Mol. Microbiol.* 45, 1–8. doi: 10.1046/j.1365-2958.2002.03014.x

Loharikar, A., Newton, A. E., Stroika, S., Freeman, M., Greene, K. D., Parsons, M. B., et al. (2015). Cholera in the United States, 2001–2011: a reflection of patterns of global epidemiology and travel. *Epidemiol. Infect.* 143, 695–703. doi: 10.1017/S0950268814001186

Lopez-Perez, M., Gonzaga, A., and Rodriguez-Valera, F. (2013). Genomic diversity of "deep ecotype" *Alteromonas macleodii* isolates: evidence for Pan-Mediterranean clonal frames. *Genome Biol. Evol.* 5, 1220–1232. doi: 10.1093/gbe/evt089

Lundquist, P. D., and Levin, B. R. (1986). Transitory derepression and the maintenance of conjugative plasmids. *Genetics* 113, 483–497.

Marra, D., and Scott, J. R. (1999). Regulation of excision of the conjugative transposon Tn916. *Mol. Microbiol.* 31, 609–621.

Marrero, J., and Waldor, M. K. (2005). Interactions between inner membrane proteins in donor and recipient cells limit conjugal DNA transfer. *Dev. Cell* 8, 963–970. doi: 10.1016/j.devcel.2005.05.004

Mazel, D. (2006). Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.* 4, 608–620. doi: 10.1038/nrmicro1462

McCool, J. D., Long, E., Petrosino, J. F., Sandler, H. A., Rosenberg, S. M., and Sandler, S. J. (2004). Measurement of SOS expression in individual *Escherichia coli* K-12 cells using fluorescence microscopy. *Mol. Microbiol.* 53, 1343–1357. doi: 10.1111/j.1365-2958.2004.04225.x

McGrath, B. M., O'halloran, J. A., and Pembroke, J. T. (2005). Pre-exposure to UV irradiation increases the transfer frequency of the IncJ conjugative transposon-like elements R391, R392, R705, R706, R997, and pMERPH and is recA+ dependent. *FEMS Microbiol. Lett.* 243, 461–465. doi: 10.1016/j.femsle.2005.01.013

McLeod, S. M., Burrus, V., and Waldor, M. K. (2006). Requirement for *Vibrio cholerae* integration host factor in conjugative DNA transfer. *J. Bacteriol.* 188, 5704–5711. doi: 10.1128/JB.00564-06

Meinersmann, R. J., Lindsey, R. L., Bono, J. L., Smith, T. P., and Oakley, B. B. (2013). Proposed model for the high rate of rearrangement and rapid migration observed in some IncA/C plasmid lineages. *Appl. Environ. Microbiol.* 79, 4806–4814. doi: 10.1128/AEM.01259-13

Morris, J. G. Jr. (2003). Cholera and other types of vibriosis: a story of human pandemics and oysters on the half shell. *Clin. Infect. Dis.* 37, 272–280. doi: 10.1086/375600

Mulvey, M. R., Boyd, D. A., Olson, A. B., Doublet, B., and Cloeckaert, A. (2006). The genetics of *Salmonella* genomic island 1. *Microbes Infect.* 8, 1915–1922. doi: 10.1016/j.micinf.2005.12.028

Navarre, W. W., Porwollik, S., Wang, Y., Mcclelland, M., Rosen, H., Libby, S. J., et al. (2006). Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* 313, 236–238. doi: 10.1126/science.1128794

Nonaka, L., Maruyama, F., Onishi, Y., Kobayashi, T., Ogura, Y., Hayashi, T., et al. (2014). Various pAQU plasmids possibly contribute to disseminate tetracycline resistance gene tet(M) among marine bacterial community. *Front. Microbiol.* 5:152. doi: 10.3389/fmicb.2014.00152

Nordmann, P., Poirel, L., Walsh, T. R., and Livermore, D. M. (2011). The emerging NDM carbapenemases. *Trends Microbiol.* 19, 588–595. doi: 10.1016/j.tim.2011.09.005

Osorio, C. R., Marrero, J., Wozniak, R. A., Lemos, M. L., Burrus, V., and Waldor, M. K. (2008). Genomic and functional analysis of ICEPdaSpa1, a fish-pathogen-derived SXT-related integrating conjugative element that can mobilize a virulence plasmid. *J. Bacteriol.* 190, 3353–3361. doi: 10.1128/JB.00109-08

Pavlovic, G., Burrus, V., Gintz, B., Decaris, B., and Guedon, G. (2004). Evolution of genomic islands by deletion and tandem accretion by site-specific recombination: ICESt1-related elements from *Streptococcus thermophilus*. *Microbiology* 150, 759–774. doi: 10.1099/mic.0.26883-0

Pembroke, J. T., and Murphy, D. B. (2000). Isolation and analysis of a circular form of the IncJ conjugative transposon-like elements, R391 and R997: implications for IncJ incompatibility. *FEMS Microbiol. Lett.* 187, 133–138. doi: 10.1111/j.1574-6968.2000.tb09149.x

Pembroke, J. T., and Piterina, A. V. (2006). A novel ICE in the genome of *Shewanella putrefaciens* W3-18-1: comparison with the SXT/R391 ICE-like elements. *FEMS Microbiol. Lett.* 264, 80–88. doi: 10.1111/j.1574-6968.2006.00452.x

Poulin-Laprade, D., Matteau, D., Jacques, P. E., Rodrigue, S., and Burrus, V. (2015). Transfer activation of SXT/R391 integrative and conjugative elements: unraveling the SetCD regulon. *Nucleic Acids Res.* 43, 2045–2056. doi: 10.1093/nar/gkv071

Ptashne, M. (2004). *A Genetic Switch: Phage Lambda Revisited*, 3rd Edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Rahman, M., Shukla, S. K., Prasad, K. N., Ovejero, C. M., Pati, B. K., Tripathi, A., et al. (2014). Prevalence and molecular characterisation of New Delhi metallo-beta-lactamases NDM-1, NDM-5, NDM-6, and NDM-7 in multidrug-resistant Enterobacteriaceae from India. *Int. J. Antimicrob. Agents* 44, 30–37. doi: 10.1016/j.ijantimicag.2014.03.003

Ramsay, J. P., Sullivan, J. T., Stuart, G. S., Lamont, I. L., and Ronson, C. W. (2006). Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. *Mol. Microbiol.* 62, 723–734. doi: 10.1111/j.1365-2958.2006.05396.x

Randall, L. P., Clouting, C., Horton, R. A., Coldham, N. G., Wu, G., Clifton-Hadley, F. A., et al. (2011). Prevalence of *Escherichia coli* carrying extended-spectrum beta-lactamases (CTX-M and TEM-52) from broiler chickens and turkeys in Great Britain between 2006 and 2009. *J. Antimicrob. Chemother.* 66, 86–95. doi: 10.1093/jac/dkq396

Robert-Pillot, A., Copin, S., Himber, C., Gay, M., and Quilici, M. L. (2014). Occurrence of the three major *Vibrio* species pathogenic for human in seafood products consumed in France using real-time PCR. *Int. J. Food Microbiol.* 189, 75–81. doi: 10.1016/j.ijfoodmicro.2014.07.014

Rodriguez-Blanco, A., Lemos, M. L., and Osorio, C. R. (2012). Integrating conjugative elements as vectors of antibiotic, mercury, and quaternary ammonium compound resistance in marine aquaculture environments. *Antimicrob. Agents Chemother.* 56, 2619–2626. doi: 10.1128/AAC.05997-11

Scott, J. R., Kirchman, P. A., and Caparon, M. G. (1988). An intermediate in transposition of the conjugative transposon Tn916. *Proc. Natl. Acad. Sci. U.S.A.* 85, 4809–4813.

Singh, S. S., Singh, N., Bonocora, R. P., Fitzgerald, D. M., Wade, J. T., and Grainger, D. C. (2014). Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev.* 28, 214–219. doi: 10.1101/gad.234336.113

Sitkiewicz, I., Green, N. M., Guo, N., Mereghetti, L., and Musser, J. M. (2011). Lateral gene transfer of streptococcal ICE element RD2 (region of difference 2) encoding secreted proteins. *BMC Microbiol.* 11:65. doi: 10.1186/1471-2180-11-65

Smillie, C., Garcillan-Barcia, M. P., Francia, M. V., Rocha, E. P., and De La Cruz, F. (2010). Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* 74, 434–452. doi: 10.1128/MMBR.00020-10

Song, Y., Yu, P., Li, B., Pan, Y., Zhang, X., Cong, J., et al. (2013). The mosaic accessory gene structures of the SXT/R391-like integrative and conjugative elements derived from *Vibrio* spp. isolated from aquatic products and environment in the Yangtze River Estuary, China. *BMC Microbiol.* 13:214. doi: 10.1186/1471-2180-13-214

Spagnoletti, M., Ceccarelli, D., Rieux, A., Fondi, M., Taviani, E., Fani, R., et al. (2014). Acquisition and evolution of SXT-R391 integrative conjugative elements in the seventh-pandemic *Vibrio cholerae* lineage. *MBio* 5:e01356–14. doi: 10.1128/mBio.01356-14

Strainic, M. G. Jr., Sullivan, J. J., Collado-Vides, J., and Dehaseth, P. L. (2000). Promoter interference in a bacteriophage lambda control region: effects of a range of interpromoter distances. *J. Bacteriol.* 182, 216–220. doi: 10.1128/JB.182.1.216-220.2000

Tijet, N., Richardson, D., Macmullin, G., Patel, S. N., and Melano, R. G. (2015). Characterization of multiple NDM-1-producing enterobacteriaceae isolates from the same patient. *Antimicrob. Agents Chemother.* 59, 3648–3651. doi: 10.1128/AAC.04862-14

Toprak, E., Veres, A., Michel, J. B., Chait, R., Hartl, D. L., and Kishony, R. (2012). Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* 44, 101–105. doi: 10.1038/ng.1034

Toussaint, A., and Merlin, C. (2002). Mobile elements as a combination of functional modules. *Plasmid* 47, 26–35. doi: 10.1006/plas.2001.1552

Van Etten, W. J., and Janssen, G. R. (1998). An AUG initiation codon, not codon-anticodon complementarity, is required for the translation of unleadered mRNA in *Escherichia coli*. *Mol. Microbiol.* 27, 987–1001.

Waldor, M. K., Tschape, H., and Mekalanos, J. J. (1996). A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. *J. Bacteriol.* 178, 4157–4165.

Walsh, T. R., Toleman, M. A., Poirel, L., and Nordmann, P. (2005). Metallo-beta-lactamases: the quiet before the storm? Clin. *Microbiol. Rev.* 18, 306–325. doi: 10.1128/CMR.18.2.306-325.2005

Walsh, T. R., Weeks, J., Livermore, D. M., and Toleman, M. A. (2011). Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: an environmental point prevalence study. *Lancet Infect. Dis.* 11, 355–362. doi: 10.1016/S1473-3099(11)70059-7

Wang, S., Fleming, R. T., Westbrook, E. M., Matsumura, P., and Mckay, D. B. (2006). Structure of the *Escherichia coli* FlhDC complex, a prokaryotic heteromeric regulator of transcription. *J. Mol. Biol.* 355, 798–808. doi: 10.1016/j.jmb.2005.11.020

Waters, J. L., and Salyers, A. A. (2013). Regulation of CTnDOT conjugative transfer is a complex and highly coordinated series of events. *MBio* 4, e00569-00513. doi: 10.1128/mBio.00569-13

Welch, T. J., Fricke, W. F., Mcdermott, P. F., White, D. G., Rosso, M. L., Rasko, D. A., et al. (2007). Multiple antimicrobial resistance in plague: an emerging public health risk. *PLoS ONE* 2:e309. doi: 10.1371/journal.pone.0000309

Wiesner, M., Zaidi, M. B., Calva, E., Fernandez-Mora, M., Calva, J. J., and Silva, C. (2009). Association of virulence plasmid and antibiotic resistance determinants with chromosomal multilocus genotypes in Mexican *Salmonella enterica* serovar Typhimurium strains. *BMC Microbiol.* 9:131. doi: 10.1186/1471-2180-9-131

World Health Organization. (2014). Antimicrobial resistance: global report on surveillance. Available at: http://www.who.int/drugresistance/documents/surveillancereport/en/

Wozniak, R. A., Fouts, D. E., Spagnoletti, M., Colombo, M. M., Ceccarelli, D., Garriss, G., et al. (2009). Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet.* 5:e1000786. doi: 10.1371/journal.pgen.1000786

Wozniak, R. A., and Waldor, M. K. (2009). A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genet.* 5:e1000439. doi: 10.1371/journal.pgen.1000439

Wozniak, R. A., and Waldor, M. K. (2010). Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* 8, 552–563. doi: 10.1038/nrmicro2382

Yamaichi, Y., Chao, M. C., Sasabe, J., Clark, L., Davis, B. M., Yamamoto, N., et al. (2015). High-resolution genetic analysis of the requirements for horizontal transmission of the ESBL plasmid from *Escherichia coli* O104:H4. *Nucleic Acids Res.* 43, 348–360. doi: 10.1093/nar/gku1262

Yong, D., Toleman, M. A., Giske, C. G., Cho, H. S., Sundman, K., Lee, K., et al. (2009). Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in *Klebsiella pneumoniae* sequence type 14 from India. *Antimicrob. Agents Chemother.* 53, 5046–5054. doi: 10.1128/AAC.00774-09

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ANNEXE 5

Poulin-Laprade, D. et Burrus, V. (2015). Electrophoretic Mobility Shift Assay Using Radiolabeled DNA Probes. *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology. 1334. Springer Science+Business Media New York. DOI 10.1007/978-1-4939-2877-4_1

Chapter 1

Electrophoretic Mobility Shift Assay Using Radiolabeled DNA Probes

Dominic Poulin-Laprade and Vincent Burrus

Abstract

Electrophoretic mobility shift assays (EMSA) have proven their usefulness for studying interactions between biological molecules. In the present protocol, a purified protein of interest is mixed with a 5′-end radiolabeled DNA probe. The bound complex(es) is separated by electrophoretic migration through a polyacrylamide gel and detected with a phosphorimager. The applications of EMSA are diverse, from thermodynamic and kinetic analyses to observation of bending and other conformational changes, stoichiometric inferences, or insights into cooperative protein binding.

[AU1] **Key words** EMSA, Electrophoretic mobility shift assay, Gel shift, Gel retardation, DNA probe labeling, Native polyacrylamide gel, Phosphore 32, Protein–DNA interactions

Abbreviations

APS	Ammonium persulfate
ATP	Adenosine triphosphate
Bis-Tris	1,3-bis(tris(hydroxymethyl)methylamino)propane
BSA	Bovine serum albumin
cAMP	3′-5′-cyclic adenosine monophosphate
CAP	*E. coli* cAMP receptor protein
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediamine tetraacetic acid
EMSA	Electrophoretic mobility shift assay
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
MOPS	3-(*N*-morpholino)propanesulfonic acid
PMSF	Phenylmethylsulfonyl fluoride
SELEX	Systematic evolution of ligands by exponential enrichment
SDS-PAGE	Polyacrylamide gel electrophoresis carried in presence of sodium dodecyl sulfate
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_1,

TE	Tris-EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane

1 Introduction

1.1 The Principle

The formation of complexes between biological molecules is fundamental for life development. Studies involving association and dissociation of such complexes or the rate at which they occur are essential in the fields of drug discovery, cellular biology, DNA processing, and genetic regulation. Electrophoretic mobility shift assays have been used since the early 1980s [1] to study the DNA-binding ability of proteins, and it is still a trusted approach to assess qualitative and quantitative parameters of protein–nucleic acid interactions. The basic principle of the method is that the structure, the size, and the charge of a molecule will affect its migration through a native polyacrylamide (or agarose) gel. Typically, the negative charge of the labeled nucleic acid will drive its migration in the gel matrix up to a given position according to the conditions (gel matrix, buffer composition, duration, voltage) of the electrophoresis. Under the same conditions, binding of a protein on the DNA probe will slow down the DNA migration and shorten the distance it travels through the gel (shift) as a function of the overall charge and increased size of the protein–DNA complex.

1.2 The Binding Entities

The protocol described in this chapter focuses on 5′-end radiolabeled PCR-amplified probes bound by purified DNA-binding proteins. From this framework, derived protocols can be optimized to suit specific experimental needs and limitations. Figure 1 describes some alternatives for starting materials, detection methods, and analysis. The probe can be either single-stranded, duplex, triplex, or quadruplex DNAs or RNAs, as well as small circular DNAs [2]. Radiolabeled species can be detected at the sub-picomolar range, and nanomoles of proteins are generally sufficient to generate a shift. This high sensitivity combined with the narrow pore size of acrylamide gels, which is around 5–20 nm (50–200 Å) for 10 to 4 % acrylamide gels [3], allows the observation of subtle changes in conformation or stoichiometry of intermediate complexes. If lower sensitivity is sufficient, for example, the qualitative binding of a large protein, other labeling methods and agarose gels (700–7000 Å) can be used (*see* **Note 1**).

The methodology to produce and purify proteins of interest goes beyond the scope of the current chapter. Since they are generally soluble, DNA-binding proteins are easy to produce in most cases. For quantitative assessments of bacterial transcriptional factors, we recommend the use of recombinant strategies to purify the protein of interest by affinity chromatography. Many companies

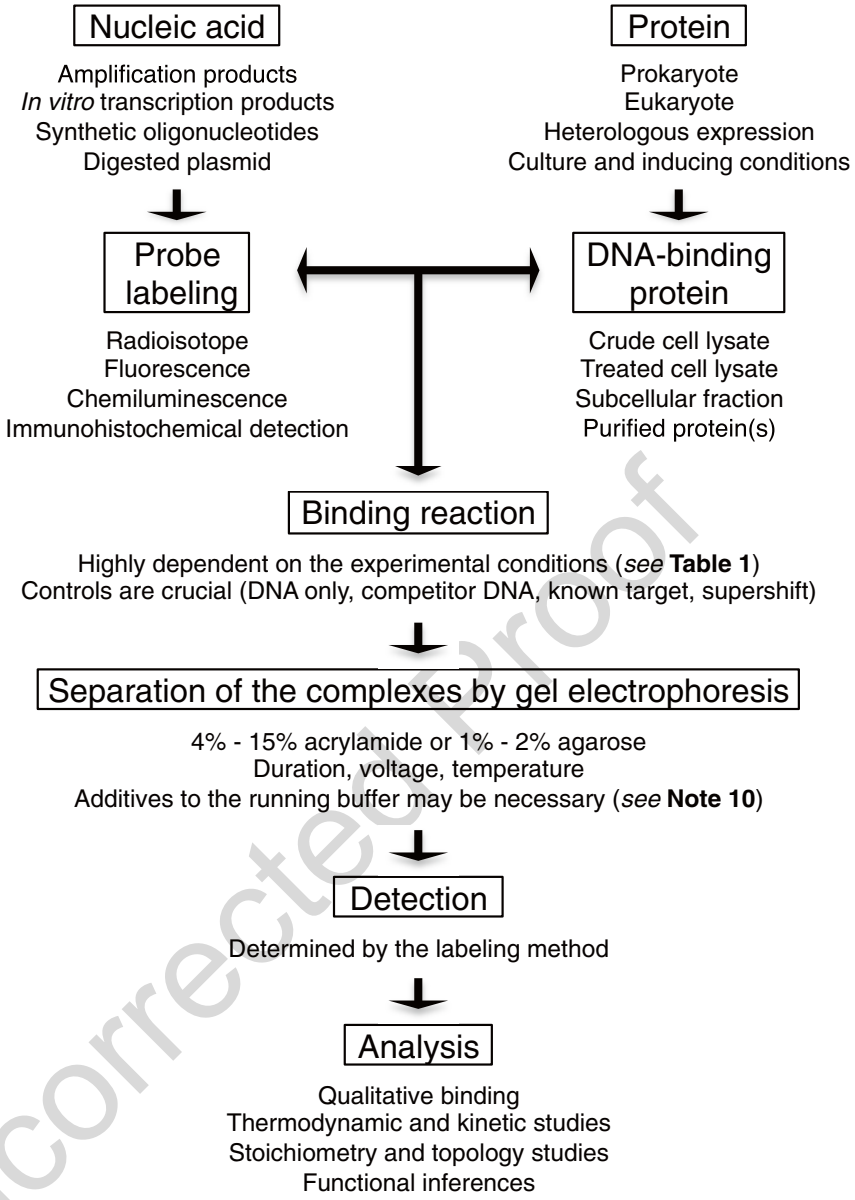

Fig. 1 Organizational chart of the EMSA procedure and the associated experience-specific possibilities

commercialize systems suitable for the production of recombinant proteins in *Escherichia coli*. The protein of interest should be purified to at least 95 % as observed on a Coomassie-stained SDS-PAGE. If the purification of the protein of interest is an issue or if elaborate multipartite DNA-binding complexes are studied, crude cell lysates or nuclear extracts containing the protein(s) of interest may be a convenient choice instead. When bacterial lysate is used, the probe may need to be labeled by incorporation because active exonucleases can promote degradation of the labeled moiety at the 5′ end of the probe. Bacterial lysates can be treated with streptomycin sulfate prior to use to precipitate nucleic acids and facilitate their

elimination by centrifugation [4]. The use of lysates is a faster and cheaper option when the DNA-binding properties have to be compared between multiple biological conditions and/or treatments. Finally, when lysates are used, small probes (oligonucleotides) are better suited to minimize false positives, and antibodies should be used to confirm the identity of a given DNA-binding protein.

1.3 The Binding Reaction

The physical and chemical conditions of the binding reactions are decisive for the observation of shifted probes. Table 1 shows a non-exhaustive list of parameters that require optimization for optimal binding. The buffer, protease inhibitors, and EDTA help protect the protein of interest from proteases, while salts and stabilizing and reducing agents contribute to maintain its folding and functionality. Optimal parameters greatly vary from one DNA-binding protein to another and may be difficult to predict when information regarding the biochemical properties of the protein is limited. Preliminary assays designed to screen various parameters can maximize the odds of observing a shift. Observation of distinct species and their binding behavior in various conditions may give appreciable biochemical cues for functional inferences. Most biological interactions are relatively weak since strong affinity could fix a state. An advantage of EMSA is that protein–DNA complexes are stabilized by the cage effect or molecular sequestration [5, 6]. In other words, during migration, dissociated complexes tend to reassociate because their partners are maintained close to each other by the mesh of the gel matrix. Coherently, smears are often associated with diffusion of low-affinity complexes. Controls are crucial to validate specific binding. A mandatory negative control is the migration of the probe in the absence of the protein. Other important negative controls can be biologically relevant DNA regions not thought to be bound or addition of competitor DNA (*see* Table 1). Known targets are great positive controls, as well as supershifts using antibodies raised against the protein of interest. EMSA can provide key information about protein–DNA interactions; however, other experiments such as DNAseI footprinting, chromatin immunoprecipitation, or reporter assays are necessary to confirm binding specificity and relevance in a given biological context.

1.4 Applications

Thermodynamics. The biological activity of a transcriptional regulator or any other DNA-binding protein usually correlates with its affinity for its recognition motif(s), which is quantified as a dissociation constant (K_D) or association constant (K_A) of a protein for a DNA target in defined conditions [7]. The K_D is usually determined by titration of the labeled probe upon addition of fold increments of the protein and subsequent quantification of the unbound and protein-bound DNA by densitometry.

Table 1
Physical and chemical parameters of the binding reaction

Parameter	Options[a]
Temperature	4 °C, 25 °C, 37 °C, or higher if proteins of heat-tolerant organisms are studied
Time	5 min to 2 h
Composition of the binding buffer	• Buffer (5–50 mM) – Tris, Glycine, HEPES, MOPS, Bis-Tris, Phosphate • Salts (1–50 mM) – NaCl, KCl, $MgCl_2$, $CaCl_2$, $ZnCl_2$ • Competitor DNA (0.5–50 µg/ml) – Bulk or sonicated chromosomal DNA Salmon sperm DNA *E. coli* chromosome Calf thymus – Synthetic DNA Poly d(I-C) or Poly d(A-T) Polysulfated carbohydrate heparin Unlabeled specific probes Designed oligonucleotides • Stabilizing agents – Glycerol, sucrose, BSA, triethylene glycol, Nonidet-P40 • Protease inhibitors – PMSF, protease inhibitor cocktails, EDTA • Reducing agent – DTT (*see* **Note 10**)

[a]These parameters, which vary depending on the nature of the protein and nucleic acid binding partners, usually require optimization for optimal binding. Further information can be found in [2, 3]

Stoichiometry. The use of a mix of wild-type and truncated or lengthened mutant proteins (which exhibit wild-type binding and multimerization properties) allows an estimation of the number of monomer subunits in a protein complex [8, 9]. The mutant proteins migrate differently than their wild-type counterparts given their altered size generating additional bands in which the DNA is bound either by the mutant proteins alone or by a mix of wild-type and mutant monomers. The number of additional bands allows the approximation of the number of subunits in the DNA-binding complex. The binding of more than one protein can also be studied by EMSA [10]. Additional protein partners alter the size, conformation, and charge of the complex, generating a discernible shift in the migration pattern of the labeled probe. An EMSA-based method was even designed to determine the molecular weight of complexes [11]. Relative affinity of multiple partners can be assessed by K_D determinations in defined conditions, preferably side-by-side in the same experiment since complex formation and migration are highly influenced by the experimental conditions [12].

The number of shifted bands observed represents the equilibrium state at the moment of loading and through the gel during migration. Generally, cooperative binding to multiple sites leads to a single shifted band because the equilibrium is rapidly shifted toward the all bound state. In contrast, in the case of noncooperative binding to adjacent sites, EMSA can reveal arrays of complexes of different sizes that form between the protein of interest and its target operators.

Topology. The sequence of the DNA probe itself can disturb migration patterns. For example, repetitions of 4–6 A•T tracks in phase with the DNA helix (every 10 bp) lead to significant bending of the DNA [13, 14]. Such a deflection of the helix axis is also often observed after binding of transcription factors or DNA processing enzymes. Succession of phased bends, either attributed to the sequence or protein binding, has been used in studies investigating the topology of DNA [15]. An approximation of the bending angle can be calculated by using permutated probes, i.e., probes of identical size partially overlapping each other and containing a given operator [16, 17]. Bending can lead to DNA loops often seen for strong repressors [18] or even be crucial for the functionality of activators [19].

SELEX. The EMSA principle was brought to the next level with the development of the systematic evolution of ligands by exponential enrichment procedure (SELEX) [20, 21]. Briefly, this method screens randomized nucleic acids for their affinity for a ligand that can be a protein or a small molecule. A clear advantage of this method when screening targets of DNA-binding proteins is that the results reflect a biologically relevant recognition motif, important for the establishment of an accurate consensus sequence [22]. The automation of the method largely contributed to the field of aptamer study and discovery, and several derivatives of the methods were developed [23].

2 Materials

Prepare all solutions using analytical grade reagents and molecular grade water for enzymatic reactions or deionized water for the preparation of buffers. Protein handling should always be on ice to maintain the activity. Diligently follow all waste disposal regulations, especially for the material contaminated with toxic reagents or radioisotopes. The following material marked by an asterisk (*) is hazardous, is harmful to the environment, and should be handled and discarded accordingly.

2.1 Binding Reaction Components

1. Purified protein of interest and the appropriate dilution buffer (*see* **Note 2**).
2. Reagent for protein quantification (*see* **Note 3**).

3. DNA template for amplification (genome, plasmid, or amplicon).
4. Primers to specifically amplify the probe of interest (*see* **Note 4**).
5. Taq DNA polymerase and the buffer provided by the manufacturer.
6. Commercial PCR Purification Kit or phenol–chloroform DNA extraction reagents [24].
7. T4 polynucleotide kinase and the buffer provided by the manufacturer.
8. *ATP γ-^{32}P 3000 Ci/mmol. Caution: ^{32}P emits harmful high-energy beta and secondary X radiations (*see* **Note 6**).
9. G-25 Sephadex columns.
10. Binding reaction buffer (2×): 20 mM Tris–HCl pH 8.0, 20 mM KCl, 1 mM $MgCl_2$, 5 mM DTT, 5 % glycerol. Mix 400 μl 1 M Tris–HCl pH 8.0, 1.6 ml 250 mM KCl, 10 μl 2 M $MgCl_2$, 15.42 mg DTT, and 0.5 ml glycerol. Complete the volume to 10 ml. Divide in aliquots and store at −20 °C (*see* Table 1).
11. Thermocycler.
12. Nanodrop (Thermo Scientific) or other device to quantify the DNA concentration.
13. Dry bath at 37 °C. The use of wet baths can mask a spill of radioactive material.
14. Radioactivity general equipment and permit.

2.2 Polyacrylamide Gel Components and Detection

1. 95 % or anhydrous ethanol.
2. Detergent powder for glassware.
3. Extra low-lint delicate task wipers.
4. Optional: *PlusOne Bind-Silane mixture: 3 μl γ-methacryloxypropyltrimethoxysilane (PlusOne Bind-Silane, GE Healthcare), 950 μl 95 % ethanol, 50 μl glacial acetic acid. Mix well and keep at room temperature for several months in a tightly sealed tube.
5. Tris borate EDTA (TBE; 10×): 890 mM Tris Base, 890 mM boric acid, 20 mM EDTA. 108 g Tris base, 55 g boric acid, 40 ml of 0.5 M EDTA pH 8.0. Filter to remove particles in the buffer that may affect migration. Sterilize by autoclaving and store at room temperature.
6. *Ammonium persulfate: 10 % (w/v) solution in water. Divide in single-use aliquots and store them at −20 °C.
7. *N,N,N,N-Tetramethylethylenediamine (TEMED). Store at 4 °C.
8. 30 % acrylamide/bis. Store at 4 °C.
9. Gel running buffer: 0.5× TBE. Dilute 50 ml of 10× TBE in 950 ml of deionized water. Keep at 4 °C (*see* **Note 9**).

10. Sample loading buffer: 0.25 % (w/v) bromophenol blue, 30 % (v/v) glycerol (*see* **Note 8**).
11. Fixing solution: 20 % (v/v) MeOH, 10 % (v/v) acetic acid in deionized water.
12. Filter paper.
13. Thin plastic food wrap (e.g., Saran wrap).
14. Vertical electrophoresis apparatus, ideally with a cooling system that helps stabilizing protein–DNA complexes. For example, the Emperor Penguin Water Cooled Dual Gel Electrophoresis System P9DS (Owl). Alternatively, gels can be run in a cold room.
15. Electrophoresis power supply (250 V, 200 mA capacity recommended).
16. Optional: Device for degassing the polyacrylamide gels [2].
17. Gel dryer. If you do not have access to a gel dryer, the gel can be frozen and exposed at −80 °C.
18. Phosphor screen and cassette (Kodak).
19. Phosphorimager instrument (e.g., Storm 860 Molecular Imager from GMI).
20. Image analysis software (e.g., QuantityOne from Bio-Rad).

3 Methods

All procedures must be carried on ice unless otherwise indicated.

3.1 Preparation of the DNA Probe

1. Prepare 6 PCR reactions of 50 µl to amplify the probe by following the Taq polymerase manufacturer's instructions. Optimize the PCR conditions to obtain a single amplicon as observed on an ethidium bromide-stained agarose gel. Caution: Ethidium bromide is a suspected mutagenic agent.
2. Pool and purify the PCR products.
3. Measure the DNA concentration of the purified probe and convert it to pmol/µl (*see* **Note 5**).

3.2 Probe Labeling

1. Prepare the labeling reaction mixture as described in Table 2. Incubate for 1 h at 37 °C. Shorter incubation time can be used, but the efficiency of the labeling will decrease accordingly. If nucleic acid material other than amplicons is used (e.g., oligonucleotides, RNAs), ensure that it is free of 5′ phosphate; otherwise treat with the Antarctic phosphatase prior to the labeling.
2. Purify the labeled probe with a G-25 column to eliminate the unused ATP γ-^{32}P 3000 Ci/mmol. This step lowers the exposure

t2.1 **Table 2**
t2.2 **DNA probe-labeling reaction mixture**

t2.3 Reagent	Volume (µl)
t2.4 H_2O	Up to 10
t2.5 10× polynucleotide kinase buffer	1
t2.6 10 pmol/µl labeled and purified probe	2
t2.7 ATP γ-^{32}P 3000 Ci/mmol	1–5 (*see* **Note 6**)
t2.8 T4 DNA polynucleotide kinase (10 U/µl)	1

t3.1 **Table 3**
t3.2 **Binding reaction mixture**

t3.3 Reagent	Volume (µl)
t3.4 2× reaction buffer	12
t3.5 H_2O	Up to 24 µl
t3.6 Diluted protein	0.25–8
t3.7 Competitor DNA[a]	0.25–8
t3.8 Diluted labeled probe	1

t3.9 [a]The concentration and the nature of the competitor DNA depend on the nucleic acid
t3.10 sample and the affinity of the protein of interest for the targets

to harmful radiations and prevents the camouflage of bands of low molecular weight by the strong signal generated by unused radioactive ATP. 275 276 277

3. Dilute the probe to an appropriate concentration in molecular grade water (*see* **Note 7**). 278 279

3.3 Binding Reaction

1. Just before use, dilute the protein sample with the appropriate buffer to concentrations ranging from 1 nM to 4 µM. Beware: Diluted proteins tend to rapidly loose activity over time. 280 281 282
2. Set the binding reactions on ice (without the labeled probe) and incubate for 10 min at room temperature (*see* Table 3). This step allows a specific binding of the protein to competitor DNA. Remember to always prepare a sample without protein as a negative control. 283 284 285 286 287
3. Add the labeled probe. Incubate for 30 min at 37 °C, then 5 min on ice to stabilize the complexes. 288 289
4. Add the loading buffer to the sample and keep on ice until gel loading. 290 291

3.4 Separation of Protein–DNA Complexes with a Native Polyacrylamide Gel

1. Carefully wash the glass plates, spacers, and combs with detergent and hot water. Rinse with deionized water, followed by 95 % ethanol. Dry the glass plates with delicate task wipes to retrieve any remaining dust and particles. Optional: Treat one glass plate with 50 µl of PlusOne Bind-Silane mixture only in the area that will contact the wells to prevent breaking of the gel when retrieving the comb.
2. Assemble the glass plates–spacers sandwich (*see* **Note 12**).
3. Prepare the acrylamide gel as depicted in Table 4 (*see* **Note 14**). Mix thoroughly the H_2O, 2× TBE, and acrylamide by swirling gently to avoid introducing bubbles. An optional degassing step of the gel mixture facilitates acrylamide polymerization and provides a more homogenous gel. Ensure to keep a small volume of gel mixture (≈1 ml) as a first polymerization control. Caution: Non-polymerized acrylamide is a potent neurotoxic agent.
4. Just before pouring, add 500 µl of 10 % APS and 28 µl of TEMED and mix thoroughly by swirling gently to avoid introducing bubbles. Note: Cold reagents or lower amounts of APS and TEMED will slow down the polymerization reaction, which might be useful for beginners.
5. Pour the acrylamide gel and insert the comb. Make sure to avoid bubbles in the gel which will lead to distorted migration patterns (*see* **Note 13**).
6. Once the acrylamide is polymerized, assemble the electrophoretic apparatus with the gel(s) and the migration buffer (0.5× TBE).
7. Precool the gel at 4 °C for 1 h prior to migration. The gel can be stored at 4 °C up to 1 week in 100 % humidity in a sealed bag with a wet paper.

Table 4
Preparation of native polyacrylamide gels of various concentrations

Reagent	4 %	6 %	8 %	10 %	15 %
H_2O (ml)	22.75	21	19.25	17.5	13.12
2× TBE[a] (ml)	8.75	8.75	8.75	8.75	8.75
40 % acrylamide (ml)	3.5	5.25	7	8.75	13.13
10 % APS[b] (µl)	500	500	500	500	500
TEMED[b] (µl)	28	28	28	28	28

[a]Volume for a final concentration of 0.5× TBE. Final concentration can range from 0.25× to 1×. Higher concentrations increase the speed of migration but also generate heat (*see* **Note 9**)
[b]APS and TEMED should be added just prior to pouring as they trigger polymerization of the acrylamide

8. Just before use, remove the comb and wash carefully the wells to remove non-polymerized acrylamide. It might be helpful to draw the edges of the wells on the glass plate if they are difficult to see after the removal of the comb to facilitate loading of the samples.

3.5 Migration and Detection

1. Pre-run the gel at 135 V (10 V/cm gel length). A second polymerization control can be carried out at this step by loading 2 μl of loading buffer (without sample) in every well. The migration of the bromophenol blue indicates if polymerization occurred properly. If not, discard the gel and prepare a new one.
2. Stop the power supply and quickly load the samples. After loading, run the gel at 120 V (*see* **Note 11**).
3. After migration, disassemble the apparatus and dry the gel plates assembly. Residual buffer can lead to vacuum formation when separating the plates, and contact of free buffer on the gel can elute radioactive nucleic acids potentially increasing the background signal in the autoradiography. Separate the glass plates, leaving the gel on one plate. Put the gel (and glass plate) in a large Pyrex dish containing the fixing solution for 20–30 min.
4. Remove the glass plate from the fixing solution and transfer the gel onto a filter paper. Put the paper on the gel and flip upside down. Delicately detach the gel paper from the glass plate. Place the gel paper on a larger filter paper onto the gel dryer surface, the gel facing up. Cover the gel with thin plastic food wrap while making sure to avoid bubbles and wrinkles. Beware: Transfer of gels with high acrylamide content might be challenging as they tend to not stick to the filter paper. A razor blade may be useful to detach the gel.
5. Dry the gel for 90 min with heat (70 °C).
6. Expose the thin plastic food wrap-covered gel on a phosphor screen in a cassette. Bands will appear after 30 min to several days of exposition depending on the signal strength. Beware of residual moisture as it can damage the phosphor screen. Also, phosphor screens should be handled under a dim light.
7. Detect the signal using a phosphorimager.
8. Quantitative measure of signal strength can be obtained by densitometry using image analysis software.

4 Notes

[AU3]

1. For concerns related to health, environment, or lack of experience, nonradioactive probe labeling can be achieved using high-sensitivity DNA detection dyes such as SYBR.

The downsides of this approach in comparison to radiolabeling are its lower sensitivity and the requirement for higher quantity of biologically active material. If this is not an issue, an elegant alternative to radiolabeling is the differential staining of both DNA and protein molecules using commercially available system such as the EMSA Kit with SYBR® Green & SYPRO® Ruby EMSA stains from Life Technologies. Keep in mind that the detection of these fluorescent dyes is only possible with adapted camera filters.

2. The dilution buffer depends on the nature of the starting protein material. It is generally the buffer used during the purification steps. To retain the DNA-binding activity, protein samples should always be handled on ice and used fresh, especially for crude cell lysates. Since the freezing/thawing procedure is inherently deleterious for protein activity, purified protein samples can be kept at 4 °C in a buffer containing protease inhibitors and a reducing agent (e.g., DTT) to prevent oxidation. For cell extracts and long-term conservation of purified proteins (>2 weeks), samples should be separated in single-use aliquots, quickly frozen in liquid nitrogen and stored at −80 °C. The addition of glycerol (50 % v/v final) may help retain the activity. However, dialysis will be mandatory to remove the glycerol prior to binding assays.
3. The easiest way to determine the protein concentration of a sample is with the Layne formula:

$$\text{Protein concentration (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260}).$$

For K_D determination or other quantitative assessments, we strongly advise the use of a more accurate method like the Bradford or BCA protein assays, respectively, commercialized by Bio-Rad and Thermo Scientific Pierce. The choice of the method depends on the composition of the protein sample as there are chemicals known to interfere with the reagents. Moreover, the composition in amino acids of the calibration protein should be considered to maximize the accuracy of the concentration measurement.

4. Ideally, the probe size should be between 50 and 600 bp. Labeled primers can also be used. Hybridized oligonucleotides are convenient to screen large libraries of mutants or single operators that are closely located in their native locus.
5. Example of conversion of double-stranded DNA concentration from ng/μl to pmol/μl:

Probe of 100 bp after purification: 30 μl at 200 ng/μl

$$\text{pmol/}\mu\text{l} = (0.2\ \mu\text{g/}\mu\text{l}) \times (1\ \text{pmol}/660\ \text{pg}) \times (10^6\ \text{pg}/1\ \mu\text{g}) \times (1/100\ \text{bp}) = 3\ \text{pmol/}\mu\text{l}.$$

6. The volume of ATP γ-^{32}P 3000 Ci/mmol added to the labeling reaction can vary depending on the availability of the isotope

and the specific activity of the probe needed (e.g., single complex formed, multiple complex formed). A higher concentration of radiolabeled ATP favors the labeling reaction by the T4 polynucleotide kinase. Always order and use a minimum of radioactive material because of its short-lived activity, its hazardousness, and its environmental implications. The lower tank of the electrophoresis apparatus may get contaminated with radioactive material. You should always verify with a Geiger counter and discard it accordingly.

7. The concentration of the DNA probe used can vary according to the requirements of the experiment. In the conditions indicated in the "Methods" section, a 1/50 dilution is used for <1 week after the calibration date of ^{32}P activity given by the manufacturer and 1/20 for 1–2 weeks old ^{32}P. It is better not to use material older than 2 weeks due to the radiolytic degradation of the sample.
8. In some instances, the bromophenol blue dye can disturb the binding reaction. To bypass this problem and still track the migration of the samples, the dye can be added to a well next to the undyed samples.
9. Low ionic strength buffers (e.g., TBE, TAE, TE) contribute to the detection of complexes by increasing the speed of migration and by generating less heat. Higher ionic strength conditions can disrupt ionic bonds which are important in complex stabilization [3]. Cofactors might be needed in the binding and running buffers for specific binding. For example, addition of cAMP is necessary for optimal DNA binding by CAP [25].
10. Dithiothreitol (DTT) is used at a concentration of 1 % (w/v) to reduce disulfide bonds between subunits of proteins during electrophoresis. At lower concentrations (1 mM), DTT is used to counteract the oxidation of proteins and preserve biological activity.
11. It may be useful to leave an empty well between two samples to show the position of your samples on the autoradiography. Avoid expelling bubbles while loading the samples into the wells.
12. The left, right, and bottom edges of the glass plates-separators sandwich can be secured with masking tape to prevent leakage of non-polymerized acrylamide gel.
13. Beware when putting your face close to the casting apparatus when inserting the comb. Splashes of non-polymerized acrylamide can occur. We find convenient to pour the gel with a 60 ml syringe topped with a 20G1 needle (BD PrecisionGlide). The volume of the syringe and the gauge of the needle vary according to the volume of gel needed and the space between the glass plates. The presence of bubbles in the gel affects the

migration because air does not conduct electric current well. Tilting the casting assembly at a 45° angle may help prevent bubbles formation. Gently tapping the assembly may help dislodge air bubbles.

14. A concentration of 8 % polyacrylamide is a good starting point when studying the binding of a small transcriptional regulator (>100 kDa) on a DNA probe of 80–300 bp. If larger proteins or DNAs are studied, the concentration can be lowered down to 4 %. A lower polyacrylamide content increases the speed of migration. Higher concentration (up to 15 %) can help separate multiple species and stabilizes complexes at a certain extent.

Acknowledgement

This work was supported by the Fonds Québécois de la recherche sur la nature et les technologies (D.P.L.) and a Discovery Grant and Discovery Acceleration Supplement from the Natural Sciences and Engineering Council of Canada (V.B.). V.B. holds a Canada Research Chair in molecular bacterial genetics.

References

1. Garner MM, Revzin A (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic Acids Res 9:3047–3060
2. Hellman LM, Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. Nat Protoc. doi:10.1038/nprot.2007.249
3. Lane D, Prentki P, Chandler M (1992) Use of gel retardation to analyze protein-nucleic acid interactions. Microbiol Rev 56:509–528
4. Oxenburgh MS, Snoswell AM (1965) Use of Streptomycin in the Separation of Nucleic Acids from Protein in a Bacterial Extract. Nature. doi:10.1038/2071416a0
5. Cann JR (1989) Phenomenological theory of gel electrophoresis of protein-nucleic acid complexes. J Biol Chem 264:17032–17040
6. Vossen KM, Fried MG (1997) Sequestration stabilizes lac repressor-DNA complexes during gel electrophoresis. Anal Biochem 245:85–92
7. Fried M, Crothers DM (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res 9:6505–6525
8. Hope IA, Struhl K (1987) GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. EMBO J 6:2781–2784
9. Kimsey HH, Waldor MK (2003) The CTX Repressor RstR Binds DNA Cooperatively to Form Tetrameric Repressor-Operator Complexes. J Biol Chem 279:2640–2647
10. Fried MG, Daugherty MA (1998) Electrophoretic analysis of multiple protein-DNA interactions. Electrophoresis 19:1247–1253
11. Orchard K, May GE (1993) An EMSA-based method for determining the molecular weight of a protein–DNA complex. Nucleic Acids Res 21:3335
12. Carey MF, Peterson CL, Smale ST (2013) Electrophoretic mobility-shift assays. Cold Spring Harb Protoc. doi:10.1101/pdb.prot075861
13. Haran TE, Mohanty U (2009) The unique structure of A-tracts and intrinsic DNA bending. Q Rev Biophys. doi:10.1017/S0033583509004752
14. Koo HS, Wu HM, Crothers DM (1986) DNA bending at adenine. thymine tracts. Nature 320:501–506

15. Zinkel SS, Crothers DM (1987) DNA bend direction by phase sensitive detection. Nature 328:178–181
16. Kim J, Zwieb C, Wu C, Adhya S (1989) Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene 85:15–23
17. Wu HM, Crothers DM (1984) The locus of sequence-directed and protein-induced DNA bending. Nature 308:509–513
18. Griffith J, Hochschild A, Ptashne M (1986) DNA loops induced by cooperative binding of lambda repressor. Nature 322:750–752
19. Liu-Johnson HN, Gartenberg MR, Crothers DM (1986) The DNA binding domain and bending angle of *E. coli* CAP protein. Cell 47:995–1005
20. Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346:818–822
21. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505–510
22. Lee Y-Y, Barker CS, Matsumura P et al (2011) Refining the Binding of the *Escherichia coli* Flagellar Master Regulator, $FlhD_4C_2$, on a Base-Specific Level. J Bacteriol 193: 4057–4068
23. Stoltenburg R, Reinemann C, Strehlitz B (2007) SELEX–a (r)evolutionary method to generate high-affinity nucleic acid ligands. Biomol Eng 24:381–403
24. Sambrook J, Russel DW (2001) Molecular cloning : A laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, New York
25. Fried MG, Crothers DM (1984) Equilibrium studies of the cyclic AMP receptor protein-DNA interaction. J Mol Biol 172:241–262

ANNEXE 6

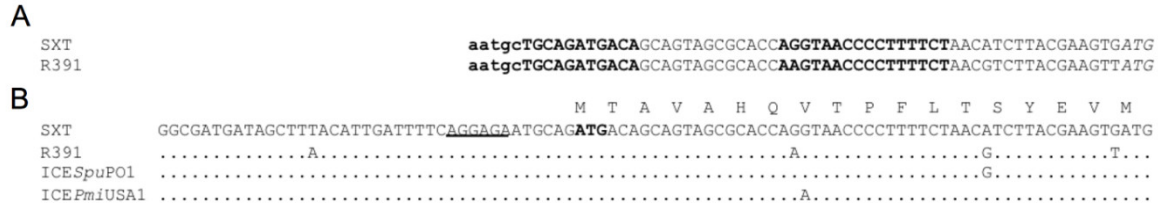

Figure 21. Prédiction erronée de l'opérateur de SetCD dans le promoteur de l'excisionase de R391. (A) Alignement du promoteur de *xis* (aussi appelé *jef*) tel que représenté par O'Halloran *et al.* (2007). Les lettres minuscules ne correspondent pas aux séquences nucléotidiques de SXT et R391 retrouvés dans la base de données Genbank. Le site de liaison de SetCD et le codon d'initiation de *xis* proposés par O'Halloran *et al.* (2007) sont respectivement indiqués en gras et en italique. Leur prédiction de l'opérateur de SetCD était basée sur la "boîte $FlhD_2C_2$" constituée de deux bras (AA(C/T)G(C/G)$N_{2/3}$AAATA(A/G)CG séparés par un espaceur non conservé de 10 à 12 nucléotides (Stafford *et al.*, 2005). **(B)** Alignement des régions 5' réelles de *xis* en provenance de SXT (AY055428), R391 (AY090559), ICE*Pmi*USA1 (AM942759) et ICE*Spu*PO1 (CP000503) telles que retrouvés dans la banque de données Genbank. La séquence de SXT est utilisée comme référence. Les nucléotides identiques dans les autres ICE sont représentés par des points tandis que les polymorphismes sont indiqués aux positions correspondantes. La séquence protéique de la portion N-terminale de Xis en provenance de SXT est inscrite au-dessus de sa séquence nucléotifique.

ANNEXE 7

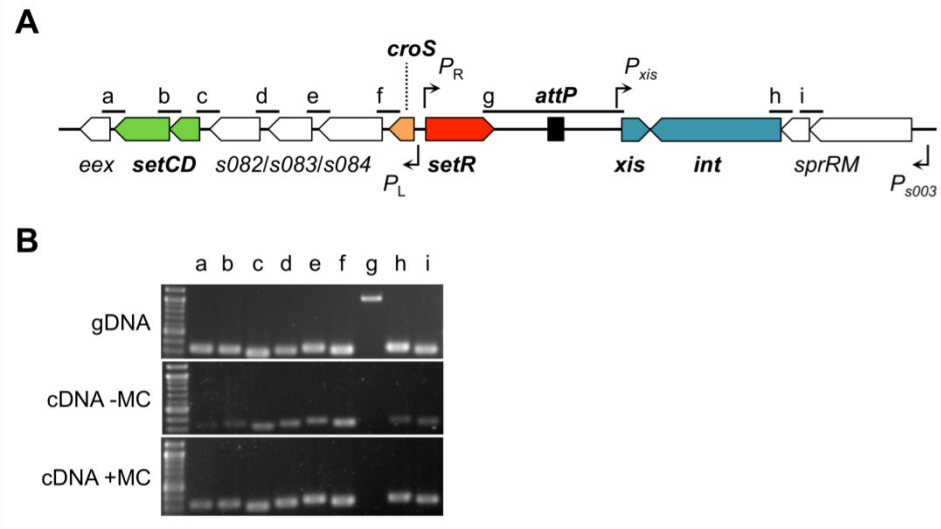

Figure 22. Organisation opéronique de la région régulatrice des ICE SXT/R391. (A) Représentation schématique des modules de régulation et d'intégration/excision des ICE SXT/R391 dans leur forme circulaire portant le site d'attachement *attP*. Les jonctions entre les gènes sont indiquées par les lettres a à i. **(B)** Amplification des jonctions entre les gènes de la région englobant *eex* et *sprM* à partir d'AND génomique, de cDNA d'E. coli BW25113 SXT traité ou non avec 100 ng/ml de mitomycine C (MC). Les produits d'amplification ont été séparés par migration dans un gel d'agarose coloré au bromure d'éthidium. Les contrôles d'amplification sur des échantillons non traités à la transcriptase inverse n'ont montré aucun produit.

BIBLIOGRAPHIE

Agafonov, D.E., and Spirin, A.S. (2004). The ribosome-associated inhibitor A reduces translation errors. Biochem. Biophys. Res. Commun. *320*, 354–358.

Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M., and Harrison, S.C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. Science *242*, 899–907.

Ahern, S.J., Das, M., Bhowmick, T.S., Young, R., and Gonzalez, C.F. (2014). Characterization of novel virulent broad-host-range phages of *Xylella fastidiosa* and *Xanthomonas.* J. Bacteriol. *196*, 459–471.

Ahmed, A.M., Shinoda, S., and Shimamoto, T. (2005). A variant type of *Vibrio cholerae* SXT element in a multidrug-resistant strain of *Vibrio fluvialis*. FEMS Microbiol. Lett. *242*, 241–247.

Alam, M.T., Weppelmann, T.A., Weber, C.D., Johnson, J.A., Rashid, M.H., Birch, C.S., Brumback, B.A., Beau de Rochars, V.E.M., Morris, J.G., and Ali, A. (2014). Monitoring water sources for environmental reservoirs of toxigenic *Vibrio cholerae* O1, Haiti. Emerg. Infect. Dis. *20*, 356–363.

Alam, M.T., Weppelmann, T.A., Longini, I., De Rochars, V.M.B., Morris, J.G., Jr, and Ali, A. (2015). Increased isolation frequency of toxigenic *Vibrio cholerae* O1 from environmental monitoring sites in Haiti. PLoS ONE *10*, e0124098.

Allison, C., Emödy, L., Coleman, N., and Hughes, C. (1994). The role of swarm cell differentiation and multicellular migration in the uropathogenicity of *Proteus mirabilis*. J. Infect. Dis. *169*, 1155–1158.

Alvarez-Martinez, C.E., and Christie, P.J. (2009). Biological diversity of prokaryotic type IV secretion systems. Microbiol. Mol. Biol. Rev. *73*, 775–808.

Anderson, L.M., and Yang, H. (2008). DNA looping can enhance lysogenic CI transcription in phage λ. Proc. Natl. Acad. Sci. *105*, 5827–5832.

Anderson, J.E., Ptashne, M., and Harrison, S.C. (1987). Structure of the repressor-

operator complex of bacteriophage 434. Nature *326*, 846–852.

Armshaw, P., and Pembroke, J.T. (2013). Generation and analysis of an ICE R391 deletion library identifies genes involved in the element encoded UV-inducible cell-sensitising function. FEMS Microbiol. Lett. *342*, 45–53.

Armshaw, P., and Pembroke, J.T. (2015). Examination of the cell sensitizing gene *orf43* of ICE R391 suggests a role in ICE transfer enhancement to recipient cells. FEMS Microbiol. Lett. *362*, 1–7.

Auchtung, J.M., Lee, C.A., Monson, R.E., Lehman, A.P., and Grossman, A.D. (2005). Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. Proc. Natl. Acad. Sci. U. S. A. *102*, 12554–12559.

Auchtung, J.M., Lee, C.A., Garrison, K.L., and Grossman, A.D. (2007). Identification and characterization of the immunity repressor (ImmR) that controls the mobile genetic element ICE*Bs1* of *Bacillus subtilis*: Immunity repressor from ICE*Bs1*. Mol. Microbiol. *64*, 1515–1528.

Avlund, M., Krishna, S., Semsey, S., Dodd, I.B., and Sneppen, K. (2010). Minimal gene regulatory circuits for a lysis-lysogeny choice in the presence of noise. PLoS ONE *5*, e15037.

Badhai, J., Kumari, P., Krishnan, P., Ramamurthy, T., and Das, S.K. (2013). Presence of SXT integrating conjugative element in marine bacteria isolated from the mucus of the coral *Fungia echinata* from Andaman Sea. FEMS Microbiol. Lett. *338*, 118–123.

Baharoglu, Z., Bikard, D., and Mazel, D. (2010). Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. PLoS Genet. *6*.

Bahl, M.I., Hansen, L.H., and Sørensen, S.J. (2009). Persistence mechanisms of conjugative plasmids. Methods Mol. Biol. Clifton NJ *532*, 73–102.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. ISMB Int. Conf. Intell. Syst. Mol. Biol. *2*, 28–36.

Bailone, A., Levine, A., and Devoret, R. (1979). Inactivation of prophage λ repressor *in vivo*. J. Mol. Biol. *131*, 553–572.

Balado, M., Lemos, M.L., and Osorio, C.R. (2013). Integrating conjugative elements of the SXT/R391 family from fish-isolated *Vibrios* encode restriction-modification systems that confer resistance to bacteriophages. FEMS Microbiol. Ecol. *83*, 457–467.

Banuett, F., Hoyt, M.A., McFarlane, L., Echols, H., and Herskowitz, I. (1986). *hflB*, a new *Escherichia coli* locus regulating lysogeny and the level of bacteriophage λ cII protein. J. Mol. Biol. *187*, 213–224.

Beaber, J.W., and Waldor, M.K. (2004). Identification of operators and promoters that control SXT conjugative transfer. J. Bacteriol. *186*, 5945–5949.

Beaber, J.W., Hochhut, B., and Waldor, M.K. (2002). Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. J. Bacteriol. *184*, 4259–4269.

Beaber, J.W., Hochhut, B., and Waldor, M.K. (2004). SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature *427*, 72–74.

Bell, A.C., and Koudelka, G.B. (1995). How 434 repressor discriminates between O1 and O3 : The influence of contacted and noncontacted base pairs. J. Biol. Chem. *270*, 1205–1212.

Bell, C.E., and Lewis, M. (2001). Crystal structure of the λ repressor C-terminal domain octamer. J. Mol. Biol. *314*, 1127–1136.

Bell, C.E., Frescura, P., Hochschild, A., and Lewis, M. (2000). Crystal structure of the λ repressor C-terminal domain provides a model for cooperative operator binding. Cell *101*, 801–811.

Bi, D., Xu, Z., Harrison, E.M., Tai, C., Wei, Y., He, X., Jia, S., Deng, Z., Rajakumar, K., and Ou, H.-Y. (2012). ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. Nucleic Acids Res. *40*, D621–D626.

Böer, S.I., Heinemeyer, E.-A., Luden, K., Erler, R., Gerdts, G., Janssen, F., and Brennholt, N. (2013). Temporal and spatial distribution patterns of potentially pathogenic *Vibrio spp*. at recreational beaches of the German north sea. Microb. Ecol. *65*, 1052–1067.

Bordeleau, E., Brouillette, E., Robichaud, N., and Burrus, V. (2010). Beyond antibiotic resistance: integrating conjugative elements of the SXT/R391 family that encode

novel diguanylate cyclases participate to c-di-GMP signalling in *Vibrio cholerae*. Environ. Microbiol. *12*, 510–523.

Bordeleau, E., Ghinet, M.G., and Burrus, V. (2012). Diversity of integrating conjugative elements in *Actinobacteria*. Mob. Genet. Elem. *2*, 119–124.

Bose, B., and Grossman, A.D. (2011). Regulation of horizontal gene transfer in *Bacillus subtilis* by activation of a conserved site-specific protease. J. Bacteriol. *193*, 22–29.

Burrus, V., and Waldor, M.K. (2003). Control of SXT integration and excision. J. Bacteriol. *185*, 5045–5054.

Burrus, V., and Waldor, M.K. (2004a). Shaping bacterial genomes with integrative and conjugative elements. Res. Microbiol. *155*, 376–386.

Burrus, V., and Waldor, M.K. (2004b). Formation of SXT tandem arrays and SXT-R391 hybrids. J. Bacteriol. *186*, 2636–2645.

Burrus, V., Pavlovic, G., Decaris, B., and Guédon, G. (2002). Conjugative transposons: the tip of the iceberg. Mol. Microbiol. *46*, 601–610.

Burrus, V., Marrero, J., and Waldor, M.K. (2006). The current ICE age: biology and evolution of SXT-related integrating conjugative elements. Plasmid *55*, 173–183.

Bushman, F.D. (1993). The bacteriophage 434 right operator. Roles of O(R)1, O(R)2 and O(R)3. J. Mol. Biol. *230*, 28–40.

Butler, S.M., and Camilli, A. (2005). Going against the grain: chemotaxis and infection in *Vibrio cholerae*. Nat. Rev. Microbiol. *3*, 611–620.

Calia, K.E., Murtagh, M., Ferraro, M.J., and Calderwood, S.B. (1994). Comparison of *Vibrio cholerae* O139 with *V. cholerae* O1 classical and El Tor biotypes. Infect. Immun. *62*, 1504–1506.

Cambray, G., Guerout, A.-M., and Mazel, D. (2010). Integrons. Annu. Rev. Genet. *44*, 141–166.

Cardinale, C.J., Washburn, R.S., Tadigotla, V.R., Brown, L.M., Gottesman, M.E., and Nudler, E. (2008). Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. Science *320*, 935–938.

Carraro, N., and Burrus, V. (2014). Biology of three ICE families: SXT/R391, ICE*Bs*1, and ICE*St*1/ICE*St*3. Microbiol. Spectr. *2*.

Carraro, N., Matteau, D., Luo, P., Rodrigue, S., and Burrus, V. (2014). The master activator of IncA/C conjugative plasmids stimulates genomic islands and multidrug resistance dissemination. PLoS Genet. *10*, e1004714.

Carraro, N., Poulin, D., and Burrus, V. (2015a). Replication and active partition of integrative and conjugative elements (ICEs) of the SXT/R391 Family: The line between ICEs and conjugative plasmids is getting thinner. PLoS Genet. *11*, e1005298.

Carraro, N., Matteau, D., Burrus, V., and Rodrigue, S. (2015b). Unraveling the regulatory network of IncA/C plasmid mobilization: When genomic islands hijack conjugative elements. Mob. Genet. Elem. *5*, 34–38.

Ceccarelli, D., Daccord, A., René, M., and Burrus, V. (2008). 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. *190*, 5328–5338.

Ceccarelli, D., Spagnoletti, M., Bacciu, D., Cappuccinelli, P., and Colombo, M.M. (2011a). New *V. cholerae* atypical El Tor variant emerged during the 2006 epidemic outbreak in Angola. BMC Microbiol. *11*, 130.

Ceccarelli, D., Spagnoletti, M., Bacciu, D., Danin-Poleg, Y., Mendiratta, D.K., Kashi, Y., Cappuccinelli, P., Burrus, V., and Colombo, M.M. (2011b). ICE*Vch*Ind5 is prevalent in epidemic *Vibrio cholerae* O1 El Tor strains isolated in India. Int. J. Med. Microbiol. IJMM *301*, 318–324.

Ceccarelli, D., Spagnoletti, M., Cappuccinelli, P., Burrus, V., and Colombo, M.M. (2011c). Origin of *Vibrio cholerae* in Haiti. Lancet Infect. Dis. *11*, 262.

Ceccarelli, D., Spagnoletti, M., Bacciu, D., Cappuccinelli, P., and Colombo, M.M. (2011d). New *V. cholerae* atypical El Tor variant emerged during the 2006 epidemic outbreak in Angola. BMC Microbiol. *11*, 130.

Ceccarelli, D., Spagnoletti, M., Hasan, N.A., Lansing, S., Huq, A., and Colwell, R.R. (2013). A new integrative conjugative element detected in Haitian isolates of *Vibrio cholerae* non-O1/non-O139. Res. Microbiol. *164*, 891–893.

Celli, J., and Trieu-Cuot, P. (1998). Circularization of Tn*916* is required for expression

of the transposon-encoded transfer functions: characterization of long tetracycline-inducible transcripts reading through the attachment site. Mol. Microbiol. *28*, 103–117.

Chilcott, G.S., and Hughes, K.T. (2000). Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar *typhimurium* and *Escherichia coli*. Microbiol. Mol. Biol. Rev. *64*, 694–708.

Cholera Working Group (1993). Large epidemic of cholera-like disease in Bangladesh caused by *Vibrio cholerae* O139 synonym Bengal. Lancet *342*, 387–390.

Chun, J., Grim, C.J., Hasan, N.A., Lee, J.H., Choi, S.Y., Haley, B.J., Taviani, E., Jeon, Y.-S., Kim, D.W., and Lee, J.-H. (2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. Proc. Natl. Acad. Sci. *106*, 15442–15447.

Claret, L., and Hughes, C. (2002). Interaction of the atypical prokaryotic transcription activator $FlhD_2C_2$ with early promoters of the flagellar gene hierarchy. J. Mol. Biol. *321*, 185–199.

Coetzee, J.N., Datta, N., and Hedges, R.W. (1972). R factors from *Proteus rettgeri*. J. Gen. Microbiol. *72*, 543–552.

Colwell, R.R. (1996). Global climate and infectious disease: the cholera paradigm. Science *274*, 2025–2031.

Cournac, A., and Plumbridge, J. (2013). DNA looping in prokaryotes: Experimental and theoretical approaches. J. Bacteriol. *195*, 1109–1119.

Court, D.L., Oppenheim, A.B., and Adhya, S.L. (2007). A new look at bacteriophage λ genetic networks. J. Bacteriol. *189*, 298–304.

Cui, L., Murchland, I., Shearwin, K.E., and Dodd, I.B. (2013). Enhancer-like long-range transcriptional activation by λ CI-mediated DNA looping. Proc. Natl. Acad. Sci. *110*, 2922–2927.

Daccord, A., Ceccarelli, D., and Burrus, V. (2010). Integrating conjugative elements of the SXT/R391 family trigger the excision and drive the mobilization of a new class of *Vibrio* genomic islands: ICE-mediated GI mobilization. Mol. Microbiol. *78*, 576–588.

Daccord, A., Mursell, M., Poulin-Laprade, D., and Burrus, V. (2012). Dynamics of the SetCD-regulated integration and excision of genomic islands mobilized by integrating

conjugative elements of the SXT/R391 family. J. Bacteriol. *194*, 5794–5802.

Daccord, A., Ceccarelli, D., Rodrigue, S., and Burrus, V. (2013). Comparative analysis of mobilizable genomic islands. J. Bacteriol. *195*, 606–614.

Darling, P.J., Holt, J.M., and Ackers, G.K. (2000). Coupled energetics of λ Cro repressor self-assembly and site-specific DNA operator binding II: cooperative interactions of Cro dimers. J. Mol. Biol. *302*, 625–638.

Das, A. (1993). Control of transcription termination by RNA-binding proteins. Annu. Rev. Biochem. *62*, 893–930.

Das, A., Ghosh, B., Barik, S., and Wolska, K. (1985). Evidence that ribosomal protein S10 itself is a cellular component necessary for transcription antitermination by phage λ N protein. Proc. Natl. Acad. Sci. U. S. A. *82*, 4070–4074.

Datta, A.B., Roy, S., and Parrack, P. (2005). Role of C-terminal residues in oligomerization and stability of λ CII: Implications for lysis-lysogeny decision of the phage. J. Mol. Biol. *345*, 315–324.

Dodd, I.B. (2004). Cooperativity in long-range gene regulation by the CI repressor. Genes Dev. *18*, 344–354.

Dodd, I.B., Perkins, A.J., Tsemitsidis, D., and Egan, J.B. (2001). Octamerization of lambda CI repressor is needed for effective repression of P_{RM} and efficient switching from lysogeny. Genes Dev. *15*, 3013–3022.

Dufour, A., Furness, R.B., and Hughes, C. (1998). Novel genes that upregulate the *Proteus mirabilis flhDC* master operon controlling flagellar biogenesis and swarming. Mol. Microbiol. *29*, 741–751.

Dziewit, L., Jazurek, M., Drewniak, L., Baj, J., and Bartosik, D. (2006). The SXT conjugative element and linear prophage N15 encode toxin-antitoxin-stabilizing systems homologous to the *tad-ata* module of the *Paracoccus aminophilus* plasmid pAMI2. J. Bacteriol. *189*, 1983–1997.

Eisen, H., Brachet, P., Pereira da Silva, L., and Jacob, F. (1970). Regulation of repressor expression in λ. Proc. Natl. Acad. Sci. U. S. A. *66*, 855–862.

Van Etten, W.J., and Janssen, G.R. (1998). An AUG initiation codon, not codon–anticodon complementarity, is required for the translation of unleadered mRNA in

Escherichia coli. Mol. Microbiol. *27*, 987–1001.

Faruque, S.M., Albert, M.J., and Mekalanos, J.J. (1998). Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*. Microbiol. Mol. Biol. Rev. *62*, 1301–1314.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., *et al*. (2014). Pfam: the protein families database. Nucleic Acids Res. *42*, D222–D230.

Fitzgerald, D.M., Bonocora, R.P., and Wade, J.T. (2014). Comprehensive mapping of the *Escherichia coli* flagellar regulatory network. PLoS Genet. *10*, e1004649.

Folkmanis, A., Maltzman, W., Mellon, P., Skalka, A., and Echols, H. (1977). The essential role of the *cro* gene in lytic development by bacteriophage λ. Virology *81*, 352–362.

Friedman, D.I., and Court, D.L. (1995). Transcription antitermination: the λ paradigm updated. Mol. Microbiol. *18*, 191–200.

Garcillán-Barcia, M.P., Francia, M.V., and de la Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol. Rev. *33*, 657–687.

Garriss, G., Waldor, M.K., and Burrus, V. (2009). Mobile antibiotic resistance encoding elements promote their own diversity. PLoS Genet. *5*, e1000775.

Garriss, G., Poulin-Laprade, D., and Burrus, V. (2013a). DNA-damaging agents induce the RecA-independent homologous recombination functions of integrating conjugative elements of the SXT/R391 family. J. Bacteriol. *195*, 1991–2003.

Garriss, G., Burrus, V., Roberts, A.P., and Mullany, P. (2013b). Integrating conjugative elements of the SXT/R391 family. Bact. Integr. Mob. Genet. Elem. 217.

Ghinet, M.G., Bordeleau, E., Beaudin, J., Brzezinski, R., Roy, S., and Burrus, V. (2011). Uncovering the prevalence and diversity of integrating conjugative elements in *Actinobacteria*. PLoS ONE *6*.

Gimble, F.S., and Sauer, R.T. (1985). Mutations in bacteriophage λ repressor that prevent RecA-mediated cleavage. J. Bacteriol. *162*, 147–154.

Guglielmini, J., Quintais, L., Garcillán-Barcia, M.P., de la Cruz, F., and Rocha, E.P.C. (2011). The repertoire of ICE in prokaryotes underscores the unity, diversity, and

ubiquity of conjugation. PLoS Genet *7*, e1002222.

Halary, S., Leigh, J.W., Cheaib, B., Lopez, P., and Bapteste, E. (2010). Network analyses structure genetic diversity in independent genetic worlds. Proc. Natl. Acad. Sci. U. S. A. *107*, 127–132.

Hall, R.H., and Sack, D.A. (2015). Introducing cholera vaccination in Asia, Africa and Haiti: A meeting report. Vaccine *33*, 487–492.

Hansen, S., Vulić, M., Min, J., Yen, T.-J., Schumacher, M.A., Brennan, R.G., and Lewis, K. (2012). Regulation of the *Escherichia coli* HipBA toxin-antitoxin system by proteolysis. PLoS ONE *7*, e39185.

Harada, S., Ishii, Y., Saga, T., Tateda, K., and Yamaguchi, K. (2010). Chromosomally encoded *blaCMY*-2 located on a novel SXT/R391-related integrating conjugative element in a *Proteus mirabilis* clinical isolate. Antimicrob. Agents Chemother. *54*, 3545–3550.

Harrison, S.C., and Aggarwal, A.K. (1990). DNA recognition by proteins with the helix-turn-helix motif. Annu. Rev. Biochem. *59*, 933–969.

Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., *et al*. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature *406*, 477–483.

Herrington, D.A., Hall, R.H., Losonsky, G., Mekalanos, J.J., Taylor, R.K., and Levine, M.M. (1988). Toxin, toxin-coregulated pili, and the *toxR* regulon are essential for *Vibrio cholerae* pathogenesis in humans. J. Exp. Med. *168*, 1487–1492.

Herskowitz, I., and Hagen, D. (1980). The lysis-lysogeny decision of phage λ: Explicit programming and responsiveness. Annu. Rev. Genet. *14*, 399–445.

Hochhut, B., and Waldor, M.K. (1999). Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. Mol. Microbiol. *32*, 99–110.

Hochhut, B., Marrero, J., and Waldor, M.K. (2000). Mobilization of plasmids and chromosomal DNA mediated by the SXT element, a constin found in *Vibrio cholerae* O139. J. Bacteriol. *182*, 2043–2047.

Hochhut, B., Lotfi, Y., Mazel, D., Faruque, S.M., Woodgate, R., and Waldor, M.K. (2001). Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae*

O139 and O1 SXT constins. Antimicrob. Agents Chemother. *45*, 2991–3000.

Hoyt, M.A., Knight, D.M., Das, A., Miller, H.I., and Echols, H. (1982). Control of phage λ development by stability and synthesis of cII protein: role of the viral cIII and host *hflA*, *himA* and *himD* genes. Cell *31*, 565–573.

Huq, A., West, P.A., Small, E.B., Huq, M.I., and Colwell, R.R. (1984). Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio cholerae* serovar O1 associated with live copepods in laboratory microcosms. Appl. Environ. Microbiol. *48*, 420–424.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

Johnson, A., Meyer, B.J., and Ptashne, M. (1978). Mechanism of action of the Cro protein of bacteriophage λ. Proc. Natl. Acad. Sci. U. S. A. *75*, 1783–1787.

Johnson, A.D., Poteete, A.R., Lauer, G., Sauer, R.T., Ackers, G.K., and Ptashne, M. (1981). λ repressor and cro: Components of an efficient molecular switch. Nature *294*, 217–223.

Jones, B.D., and Mobley, H.L. (1987). Genetic and biochemical diversity of ureases of *Proteus*, *Providencia*, and *Morganella* species isolated from urinary tract infection. Infect. Immun. *55*, 2198–2203.

Juhas, M., Van Der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., and Crook, D.W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution: Genomic islands and horizontal gene transfer. FEMS Microbiol. Rev. *33*, 376–393.

Kahler, A.M., Haley, B.J., Chen, A., Mull, B.J., Tarr, C.L., Turnsek, M., Katz, L.S., Humphrys, M.S., Derado, G., Freeman, N., *et al*. (2015). Environmental surveillance for toxigenic *Vibrio cholerae* in surface waters of Haiti. Am. J. Trop. Med. Hyg. *92*, 118–125.

Kaiser, A.D. (1957). Mutations in a temperate bacteriophage affecting its ability to lysogenize *Escherichia coli*. Virology *3*, 42–61.

Kaper, J.B., Morris, J.G., Jr, and Levine, M.M. (1995). Cholera. Clin. Microbiol. Rev. *8*, 48–86.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. *4*, 363–371.

Kloesges, T., Popa, O., Martin, W., and Dagan, T. (2011). Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. Mol. Biol. Evol. *28*, 1057–1074.

Kobiler, O., Koby, S., Teff, D., Court, D., and Oppenheim, A.B. (2002). The phage λ CII transcriptional activator carries a C-terminal domain signaling for rapid proteolysis. Proc. Natl. Acad. Sci. U. S. A. *99*, 14964–14969.

Kobiler, O., Rokney, A., Friedman, N., Court, D.L., Stavans, J., and Oppenheim, A.B. (2005). Quantitative kinetic analysis of the bacteriophage λ genetic network. Proc. Natl. Acad. Sci. U. S. A. *102*, 4470–4475.

Koudelka, G.B. (1998). Recognition of DNA structure by 434 repressor. Nucleic Acids Res. *26*, 669–675.

Koudelka, G.B., and Carlson, P. (1992). DNA twisting and the effects of non-contacted bases on affinity of 434 operator for 434 repressor. Nature *355*, 89–91.

Koudelka, G.B., and Lam, C.Y. (1993). Differential recognition of OR1 and OR3 by bacteriophage 434 repressor and Cro. J. Biol. Chem. *268*, 23812–23817.

Krauss, H., Weber, A., Appel, M., Enders, B., Isenberg, H.D., Schiefer, H.G., Slenczka, W., von Graevenitz, A., and Zahner, H. (2003). Zoonoses infectious diseases transmissible from animals to humans (Washington: ASM press).

Kulaeva, O.I., Wootton, J.C., Levine, A.S., and Woodgate, R. (1995). Characterization of the *umu*-complementing operon from R391. J. Bacteriol. *177*, 2737–2743.

Lanka, E., and Wilkins, B.M. (1995). DNA processing reactions in bacterial conjugation. Annu. Rev. Biochem. *64*, 141–169.

Lawley, T.D., Klimke, W.A., Gubbins, M.J., and Frost, L.S. (2003). F factor conjugation is a true type IV secretion system. FEMS Microbiol. Lett. *224*, 1–15.

De Lay, N., and Gottesman, S. (2012). A complex network of small non-coding RNAs regulate motility in *Escherichia coli*: sRNA regulation of motility in *Escherichia coli*. Mol. Microbiol. *86*, 524–538.

Lee, C.A., Babic, A., and Grossman, A.D. (2010a). Autonomous plasmid-like

replication of a conjugative transposon. Mol. Microbiol. *75*, 268–279.

Lee, D.J., Minchin, S.D., and Busby, S.J.W. (2012). Activating transcription in bacteria. Annu. Rev. Microbiol. *66*, 125–152.

Lee, J., Hiibel, S. r., Reardon, K. f., and Wood, T. k. (2010b). Identification of stress-related proteins in *Escherichia coli* using the pollutant *cis*-dichloroethylene. J. Appl. Microbiol. *108*, 2088–2102.

Lee, Y.-Y., Barker, C.S., Matsumura, P., and Belas, R. (2011). Refining the binding of the *Escherichia coli* flagellar master regulator, $FlhD_4C_2$, on a base-specific level. J. Bacteriol. *193*, 4057–4068.

Lewis, K. (2008). Multidrug tolerance of biofilms and persister cells. Curr. Top. Microbiol. Immunol. *322*, 107–131.

Li, M., Moyle, H., and Susskind, M.M. (1994). Target of the transcriptional activation function of phage λ CI protein. Science *263*, 75–77.

Little, J.W. (1984). Autodigestion of LexA and phage λ repressors. Proc. Natl. Acad. Sci. U. S. A. *81*, 1375–1379.

Little, J.W. (2010). Evolution of complex gene regulatory circuits by addition of refinements. Curr. Biol. *20*, R724–R734.

Little, J.W., Shepley, D.P., and Wert, D.W. (1999). Robustness of a gene regulatory circuit. EMBO J. *18*, 4299–4307.

Liu, X., and Matsumura, P. (1994). The FlhD/FlhC complex, a transcriptional activator of the *Escherichia coli* flagellar class II operons. J. Bacteriol. *176*, 7345–7351.

Liu, X., Fujita, N., Ishihama, A., and Matsumura, P. (1995). The C-terminal region of the α subunit of *Escherichia coli* RNA polymerase is required for transcriptional activation of the flagellar level II operons by the FlhD/FlhC complex. J. Bacteriol. *177*, 5186–5188.

Llosa, M., Gomis-Rüth, F.X., Coll, M., and de la Cruz, F. (2002). Bacterial conjugation: a two-step mechanism for DNA transport. Mol. Microbiol. *45*, 1–8.

Loharikar, A., Newton, A.E., Stroika, S., Freeman, M., Greene, K.D., Parsons, M.B., Bopp, C., Talkington, D., Mintz, E.D., and Mahon, B.E. (2015). Cholera in the United States, 2001-2011: a reflection of patterns of global epidemiology and travel.

Epidemiol. Infect. *143*, 695–703.

López-Pérez, M., Gonzaga, A., and Rodriguez-Valera, F. (2013). Genomic diversity of "deep ecotype" *Alteromonas macleodii* isolates: evidence for Pan-Mediterranean clonal frames. Genome Biol. Evol. *5*, 1220–1232.

Lorenz, M.G., and Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev. *58*, 563–602.

Lwoff, A. (1953). Lysogeny. Bacteriol. Rev. *17*, 269–337.

Mao, C., Carlson, N.G., and Little, J.W. (1994). Cooperative DNA-protein interactions. Effects of changing the spacing between adjacent binding sites. J. Mol. Biol. *235*, 532–544.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., *et al*. (2015). CDD: NCBI's conserved domain database. Nucleic Acids Res. *43*, D222–D226.

Marrero, J., and Waldor, M.K. (2005). Interactions between inner membrane proteins in donor and recipient cells limit conjugal DNA transfer. Dev. Cell *8*, 963–970.

Marrero, J., and Waldor, M.K. (2007a). The SXT/R391 family of integrative conjugative elements is composed of two exclusion groups. J. Bacteriol. *189*, 3302–3305.

Marrero, J., and Waldor, M.K. (2007b). Determinants of entry exclusion within Eex and TraG are cytoplasmic. J. Bacteriol. *189*, 6469–6473.

Matthew, M., Hedges, R.W., and Smith, J.T. (1979). Types of β-lactamase determined by plasmids in gram-negative bacteria. J. Bacteriol. *138*, 657–662.

Maurer, R., Meyer, B., and Ptashne, M. (1980). Gene regulation at the right operator (OR) bacteriophage λ. I. OR3 and autogenous negative control by repressor. J. Mol. Biol. *139*, 147–161.

Mauro, S.A., Pawlowski, D., and Koudelka, G.B. (2003). The role of the minor groove substituents in indirect readout of DNA sequence by 434 Repressor. J. Biol. Chem. *278*, 12955–12960.

McCool, J.D., Long, E., Petrosino, J.F., Sandler, H.A., Rosenberg, S.M., and Sandler, S.J. (2004). Measurement of SOS expression in individual *Escherichia coli* K-12 cells

using fluorescence microscopy. Mol. Microbiol. *53*, 1343–1357.

McGary, K., and Nudler, E. (2013). RNA polymerase and the ribosome: the close relationship. Curr. Opin. Microbiol. *16*, 112–117.

van der Meer, J.R., and Sentchilo, V. (2003). Genomic islands and the evolution of catabolic pathways in bacteria. Curr. Opin. Biotechnol. *14*, 248–254.

Meibom, K.L. (2005). Chitin induces natural competence in *Vibrio cholerae*. Science *310*, 1824–1827.

Van Melderen, L., and De Bast, M.S. (2009). Bacterial toxin–antitoxin systems: more than selfish entities? PLoS Genet. *5*, e1000437.

de Mello Varani, A., Souza, R.C., Nakaya, H.I., de Lima, W.C., Paula de Almeida, L.G., Kitajima, E.W., Chen, J., Civerolo, E., Vasconcelos, A.T.R., and Van Sluys, M.-A. (2008). Origins of the *Xylella fastidiosa* prophage-like regions and their impact in genome differentiation. PLoS ONE *3*.

Meyer, B.J., and Ptashne, M. (1980). Gene regulation at the right operator (OR) of bacteriophage λ. III. λ repressor directly activates gene transcription. J. Mol. Biol. *139*, 195–205.

Meyer, B.J., Maurer, R., and Ptashne, M. (1980). Gene regulation at the right operator (OR) of bacteriophage λ. II. OR1, OR2, and OR3: their roles in mediating the effects of repressor and Cro. J. Mol. Biol. *139*, 163–194.

Ministère de la santé publique et de la population d'Haïti (2015). Rapport de cas de choléra quotidien par commune.

Minoia, M., Gaillard, M., Reinhard, F., Stojanov, M., Sentchilo, V., and van der Meer, J.R. (2008). Stochasticity and bistability in horizontal transfer control of a genomic island in *Pseudomonas*. Proc. Natl. Acad. Sci. *105*, 20792–20797.

Miyazaki, R., Minoia, M., Pradervand, N., Sulser, S., Reinhard, F., and van der Meer, J.R. (2012). Cellular variability of RpoS expression underlies subpopulation activation of an integrative and conjugative element. PLoS Genet *8*, e1002818.

Morelli, M.J., Rein ten Wolde, Pieter, and Allen, Rosalind J. (2009). DNA looping provides stability and robustness to the bacteriophage λ switch. Proc. Natl. Acad. Sci. *106*, 8101–8106.

Murray, A.E., Lies, D., Li, G., Nealson, K., Zhou, J., and Tiedje, J.M. (2001). DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. Proc. Natl. Acad. Sci. *98*, 9853–9858.

Mutreja, A., Kim, D.W., Thomson, N., Connor, T.R., Lee, J.H., Kariuki, S., Croucher, N.J., Choi, S.Y., Harris, S.R., Lebens, M., *et al*. (2011). Evidence for multiple waves of global transmission within the seventh cholera pandemic. Nature *477*, 462–465.

Nair, G.B., Shimada, T., Kurazono, H., Okuda, J., Pal, A., Karasawa, T., Mihara, T., Uesaka, Y., Shirai, H., and Garg, S. (1994). Characterization of phenotypic, serological, and toxigenic traits of *Vibrio cholerae* O139 bengal. J. Clin. Microbiol. *32*, 2775–2779.

Newlove, T., Konieczka, J.H., and Cordes, M.H.. (2004). Secondary structure switching in Cro protein evolution. Structure *12*, 569–581.

Nishi, A., Tominaga, K., and Furukawa, K. (2000). A 90-kilobase conjugative chromosomal element coding for biphenyl and salicylate catabolism in *Pseudomonas putida* KF715. J. Bacteriol. *182*, 1949–1955.

Norman, A., Hansen, L.H., and Sørensen, S.J. (2009). Conjugative plasmids: vessels of the communal gene pool. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *364*, 2275–2289.

Oppenheim, A.B., Kobiler, O., Stavans, J., Court, D.L., and Adhya, S. (2005). Switches in bacteriophage λ development. Annu. Rev. Genet. *39*, 409–429.

Organisation mondiale de la Santé (2014). Choléra. Aide-mémoire #107. OMS.

Osorio, C.R., Marrero, J., Wozniak, R.A.F., Lemos, M.L., Burrus, V., and Waldor, M.K. (2008). Genomic and functional analysis of ICE*Pda*Spa1, a fish-pathogen-derived SXT-related integrating conjugative element that can mobilize a virulence plasmid. J. Bacteriol. *190*, 3353–3361.

Pacini, F. (1854). Osservazioni microscopiche e deduzioni patologiche sul cholera asiatico. Gazz. Med. Ital. *4*, 397–401.

Pearson, M.M., Sebaihia, M., Churcher, C., Quail, M.A., Seshasayee, A.S., Luscombe, N.M., Abdellah, Z., Arrosmith, C., Atkin, B., Chillingworth, T., *et al*. (2008). Complete genome sequence of uropathogenic *Proteus mirabilis*, a master of both adherence and motility. J. Bacteriol. *190*, 4027–4037.

Peters, J.M., Mooney, R.A., Grass, J.A., Jessen, E.D., Tran, F., and Landick, R. (2012). Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. Genes Dev. *26*, 2621–2633.

Peters, S.E., Hobman, J.L., Strike, P., and Ritchie, D.A. (1991). Novel mercury resistance determinants carried by IncJ plasmids pMERPH and R391. Mol. Gen. Genet. MGG *228*, 294–299.

Pinhassi, J., and Berman, T. (2003). Differential growth response of colony-forming α- and γ-proteobacteria in dilution culture and nutrient addition experiments from Lake Kinneret (Israel), the eastern Mediterranean Sea, and the Gulf of Eilat. Appl. Environ. Microbiol. *69*, 199–211.

Pirrotta, V. (1979). Operators and promoters in the OR region of phage 434. Nucleic Acids Res. *6*, 1495–1508.

Poele, E.M. te, Bolhuis, H., and Dijkhuizen, L. (2008). Actinomycete integrative and conjugative elements. Antonie Van Leeuwenhoek *94*, 127–143.

Poteete, A.R., and Ptashne, M. (1982). Control of transcription by the bacteriophage P22 repressor. J. Mol. Biol. *157*, 21–48.

Poteete, A.R., Hehir, K., and Sauer, R.T. (1986). Bacteriophage P22 Cro protein: sequence, purification, and properties. Biochemistry (Mosc.) *25*, 251–256.

Poulin-Laprade, D., and Burrus, V. (2015). A λCro-like repressor is essential for the induction of conjugative transfer of SXT/R391 elements in response to DNA damage. J. Bacteriol.

Poulin-Laprade, D., Matteau, D., Jacques, P.-É., Rodrigue, S., and Burrus, V. (2015a). Transfer activation of SXT/R391 integrative and conjugative elements: unraveling the SetCD regulon. Nucleic Acids Res. *43*, 2045–2056.

Poulin-Laprade, D., Carraro, N., and Burrus, V. (2015b). The extended regulatory networks of SXT/R391 integrative and conjugative elements and IncA/C conjugative plasmids SXT/R391 and IncA/C conjugation regulation. Front. Microbiol. *6*, 837.

Priest, D.G., Cui, L., Kumar, S., Dunlap, D.D., Dodd, I.B., and Shearwin, K.E. (2014). Quantitation of the DNA tethering effect in long-range DNA looping *in vivo* and *in vitro* using the Lac and λ repressors. Proc. Natl. Acad. Sci. U. S. A. *111*, 349–354.

Ptashne, M. (1967). Specific binding of the λ phage repressor to λ DNA. Nature *214*, 232–234.

Ptashne, M. (2004). A genetic switch: phage λ revisited (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Ptashne, M., Backman, K., Humayun, M.Z., Jeffrey, A., Maurer, R., Meyer, B., and Sauer, R.T. (1976). Autoregulation and function of a repressor in bacteriophage λ. Science *194*, 156–161.

Ramamurthy, T., Garg, S., Sharma, R., Bhattacharya, S.K., Nair, G.B., Shimada, T., Takeda, T., Karasawa, T., Kurazano, H., and Pal, A. (1993). Emergence of novel strain of *Vibrio cholerae* with epidemic potential in southern and eastern India. Lancet *341*, 703–704.

Ravatn, R., Studer, S., Springael, D., Zehnder, A.J., and van der Meer, J.R. (1998). Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. Strain B13. J. Bacteriol. *180*, 4360–4369.

Reinhard, F., and van der Meer, J.R. (2014). Life history analysis of integrative and conjugative element activation in growing microcolonies of *Pseudomonas*. J. Bacteriol. *196*, 1425–1434.

Reinhard, F., Miyazaki, R., Pradervand, N., and van der Meer, J.R. (2013). Cell differentiation to “mating bodies” induced by an integrating and conjugative element in free-living bacteria. Curr. Biol. CB *23*, 255–259.

Roberts, J.W. (1969). Termination factor for RNA synthesis. Nature *224*, 1168–1174.

Roberts, A.P., and Mullany, P. (2009). A modular master on the move: the Tn*916* family of mobile genetic elements. Trends Microbiol. *17*, 251–258.

Roberts, A.P., Chandler, M., Courvalin, P., Guédon, G., Mullany, P., Pembroke, T., Rood, J.I., Jeffery Smith, C., Summers, A.O., Tsuda, M., *et al*. (2008). Revised nomenclature for transposable genetic elements. Plasmid *60*, 167–173.

Roberts, J.W., Roberts, C.W., and Craig, N.L. (1978). *Escherichia coli recA* gene product inactivates phage λ repressor. Proc. Natl. Acad. Sci. U. S. A. *75*, 4714–4718.

Rodríguez-Blanco, A., Lemos, M.L., and Osorio, C.R. (2012). Integrating conjugative

elements as vectors of antibiotic, mercury, and quaternary ammonium compound resistance in marine aquaculture environments. Antimicrob. Agents Chemother. *56*, 2619–2626.

Roessler, C.G., Hall, B.M., Anderson, W.J., Ingram, W.M., Roberts, S.A., Montfort, W.R., and Cordes, M.H. (2008). Transitive homology-guided structural studies lead to discovery of Cro proteins with 40% sequence identity but different folds. Proc. Natl. Acad. Sci. *105*, 2343–2348.

Rosenberg, M., and Court, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. *13*, 319–353.

Roszak, D.B., and Colwell, R.R. (1987). Survival strategies of bacteria in the natural environment. Microbiol. Rev. *51*, 365–379.

Salje, J., Gayathri, P., and Löwe, J. (2010). The ParMRC system: molecular mechanisms of plasmid segregation by actin-like filaments. Nat. Rev. Microbiol. *8*, 683–692.

Salyers, A.A., Shoemaker, N.B., Stevens, A.M., and Li, L.Y. (1995). Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. Microbiol. Rev. *59*, 579–590.

Sánchez-Romero, M.A., Cota, I., and Casadesús, J. (2015). DNA methylation in bacteria: from the methyl group to the methylome. Curr. Opin. Microbiol. *25*, 9–16.

Sass, A.M., Sass, H., Coolen, M.J., Cypionka, H., and Overmann, J. (2001). Microbial communities in the chemocline of a hypersaline deep-sea basin (Urania basin, Mediterranean Sea). Appl. Environ. Microbiol. *67*, 5392–5402.

Sauret, C., Böttjer, D., Talarmin, A., Guigue, C., Conan, P., Pujo-Pay, M., and Ghiglione, J.-F. (2015). Top-down control of diesel-degrading prokaryotic communities. Microb. Ecol.

Schubert, R.A., Dodd, I.B., Egan, J.B., and Shearwin, K.E. (2007). Cro's role in the CI Cro bistable switch is critical for λ's transition from lysogeny to lytic development. Genes Amp Dev. *21*, 2461–2472.

Schumacher, M.A., Piro, K.M., Xu, W., Hansen, S., Lewis, K., and Brennan, R.G. (2009). Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB. Science *323*, 396–401.

Seed, K.D., Yen, M., Shapiro, B.J., Hilaire, I.J., Charles, R.C., Teng, J.E., Ivers, L.C., Boncy, J., Harris, J.B., and Camilli, A. (2014). Evolutionary consequences of intra-patient phage predation on microbial populations. eLife *3*, e03497.

Sendzik, W. (1997). The 1832 Montreal cholera epidemic: A study in state formation.

Sentchilo, V., Ravatn, R., Werlen, C., Zehnder, A.J.B., and van der Meer, J.R. (2003a). Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. strain B13. J. Bacteriol. *185*, 4530–4538.

Sentchilo, V., Zehnder, A.J.B., and van der Meer, J.R. (2003b). Characterization of two alternative promoters for integrase expression in the *clc* genomic island of *Pseudomonas* sp. strain B13. Mol. Microbiol. *49*, 93–104.

Singleton, F.L., Attwell, R., Jangi, S., and Colwell, R.R. (1982a). Effects of temperature and salinity on *Vibrio cholerae* growth. Appl. Environ. Microbiol. *44*, 1047–1058.

Singleton, F.L., Attwell, R.W., Jangi, M.S., and Colwell, R.R. (1982b). Influence of salinity and organic nutrient concentration on survival and growth of *Vibrio cholerae* in aquatic microcosms. Appl. Environ. Microbiol. *43*, 1080–1085.

Sneppen, K., Semsey, S., Seshasayee, A.S.N., and Krishna, S. (2015). Restriction modification systems as engines of diversity. Evol. Genomic Microbiol. *6*, 528.

Snow, J. (1855). On the mode of communication of cholera (London).

Snyder, L., Peters, J.E., Henkin, T.M., and Champness, W. (2013). Molecular genetics of bacteria (Washington, DC: ASM press).

Solovyev, V., and Salamov, A. (2011). Automatic annotation of microbial genomes and metagenomic sequences. In metagenomics and its applications in agriculture, biomedicine and environmental Studies, (Nova Science Publishers), pp. 61–78.

Soutourina, O.A., and Bertin, P.N. (2003). Regulation cascade of flagellar expression in Gram-negative bacteria. FEMS Microbiol. Rev. *27*, 505–523.

Spagnoletti, M., Ceccarelli, D., Rieux, A., Fondi, M., Taviani, E., Fani, R., Colombo, M.M., Colwell, R.R., and Balloux, F. (2014). Acquisition and evolution of SXT-R391 integrative conjugative elements in the seventh-pandemic *Vibrio cholerae* lineage. mBio *5*, e01356–14.

Springael, D., Peys, K., Ryngaert, A., Roy, S.V., Hooyberghs, L., Ravatn, R., Heyndrickx, M., van der Meer, J.-R., Vandecasteele, C., Mergeay, M., *et al*. (2002). Community shifts in a seeded 3-chlorobenzoate degrading membrane biofilm reactor: indications for involvement of *in situ* horizontal transfer of the *clc*-element from inoculum to contaminant bacteria. Environ. Microbiol. *4*, 70–80.

Stafford, G.P., Ogi, T., and Hughes, C. (2005). Binding and transcriptional activation of non-flagellar genes by the *Escherichia coli* flagellar master regulator $FlhD_2C_2$. Microbiology *151*, 1779–1788.

Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A., and Lewis, M. (2008). Crystal structure of the λ repressor and a model for pairwise cooperative operator binding. Nature *452*, 1022–1025.

Sterk, A., Schets, F.M., de Roda Husman, A.M., de Nijs, T., and Schijven, J.F. (2015). Effect of climate change on the concentration and associated risks of *Vibrio spp*. in dutch recreational waters. Risk Anal. Off. Publ. Soc. Risk Anal.

Strainic, M.G., Sullivan, J.J., Collado-Vides, J., and deHaseth, P.L. (2000). Promoter interference in a bacteriophage λ control region: Effects of a range of interpromoter distances. J. Bacteriol. *182*, 216–220.

Su, Y.A., He, P., and Clewell, D.B. (1992). Characterization of the *tet*(M) determinant of Tn*916*: evidence for regulation by transcription attenuation. Antimicrob. Agents Chemother. *36*, 769–778.

Sullivan, J.T., and Ronson, C.W. (1998). Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc. Natl. Acad. Sci. U. S. A. *95*, 5145–5149.

Summer, E.J., Enderle, C.J., Ahern, S.J., Gill, J.J., Torres, C.P., Appel, D.N., Black, M.C., Young, R., and Gonzalez, C.F. (2010). Genomic and biological analysis of phage Xfas53 and related prophages of *Xylella fastidiosa*. J. Bacteriol. *192*, 179–190.

Susskind, M.M., and Botstein, D. (1978). Molecular genetics of bacteriophage P22. Microbiol. Rev. *42*, 385–413.

Svenningsen, S.L., Costantino, N., Court, D.L., and Adhya, S. (2005). On the role of Cro in λ prophage induction. Proc. Natl. Acad. Sci. U. S. A. *102*, 4465–4469.

Szybalski, W., and Iyer, V.N. (1964). Crosslinking of DNA by enzymatically or

chemically activated mitomycins and profiromycins, bifunctionally “alkylating” antibiotics. Fed. Proc. *23*, 946–957.

Takeda, Y. (1979). Specific repression of *in vitro* transcription by the Cro repressor of bacteriophage λ. J. Mol. Biol. *127*, 177–189.

Taviani, E., Ceccarelli, D., Lazaro, N., Bani, S., Cappuccinelli, P., Colwell, R.R., and Colombo, M.M. (2008). Environmental *Vibrio spp*., isolated in Mozambique, contain a polymorphic group of integrative conjugative elements and class 1 integrons. FEMS Microbiol. Ecol. *64*, 45–54.

Thomas, C.M., and Nielsen, K.M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat. Rev. Microbiol. *3*, 711–721.

Toleman, M.A., and Walsh, T.R. (2010). ISCR elements are key players in IncA/C plasmid evolution. Antimicrob. Agents Chemother. *54*, 3534.

Toussaint, A., and Merlin, C. (2002). Mobile elements as a combination of functional modules. Plasmid *47*, 26–35.

Trotochaud, A.E., and Wassarman, K.M. (2004). 6S RNA function enhances long-term cell survival. J. Bacteriol. *186*, 4978–4985.

Vasu, K., and Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol. Mol. Biol. Rev. *77*, 53–72.

Vezzulli, L., Brettar, I., Pezzati, E., Reid, P.C., Colwell, R.R., Höfle, M.G., and Pruzzo, C. (2012). Long-term effects of ocean warming on the prokaryotic community: evidence from the vibrios. ISME J. *6*, 21–30.

Vezzulli, L., Colwell, R.R., and Pruzzo, C. (2013). Ocean warming and spread of pathogenic vibrios in the aquatic environment. Microb. Ecol. *65*, 817–825.

Vila-Sanjurjo, A., Schuwirth, B.-S., Hau, C.W., and Cate, J.H.D. (2004). Structural basis for the control of translation initiation during stress. Nat. Struct. Mol. Biol. *11*, 1054–1059.

Waldor, M.K., and Mekalanos, J.J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. Science *272*, 1910–1914.

Waldor, M.K., Colwell, R., and Mekalanos, J.J. (1994). The *Vibrio cholerae* O139 serogroup antigen includes an O-antigen capsule and lipopolysaccharide virulence

determinants. Proc. Natl. Acad. Sci. U. S. A. *91*, 11388–11392.

Waldor, M.K., Tschäpe, H., and Mekalanos, J.J. (1996). A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. J. Bacteriol. *178*, 4157–4165.

Wang, S., Fleming, R.T., Westbrook, E.M., Matsumura, P., and McKay, D.B. (2006). Structure of the *Escherichia coli* FlhDC complex, a prokaryotic heteromeric regulator of transcription. J. Mol. Biol. *355*, 798–808.

Wassarman, K.M., and Saecker, R.M. (2006). Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. Science *314*, 1601–1603.

Waters, J.L., and Salyers, A.A. (2013). Regulation of CTnDOT conjugative transfer is a complex and highly coordinated series of events. mBio *4*, e00569–13.

Wilson, G.G., and Murray, N.E. (1991). Restriction and modification systems. Annu. Rev. Genet. *25*, 585–627.

Wozniak, C.E., and Hughes, K.T. (2008). Genetic dissection of the consensus sequence for the class 2 and class 3 flagellar promoters. J. Mol. Biol. *379*, 936–952.

Wozniak, R.A.F., and Waldor, M.K. (2009). A toxin–antitoxin system promotes the maintenance of an integrative conjugative element. PLoS Genet. *5*, e1000439.

Wozniak, R.A.F., and Waldor, M.K. (2010). Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat. Rev. Microbiol. *8*, 552–563.

Wozniak, R.A.F., Fouts, D.E., Spagnoletti, M., Colombo, M.M., Ceccarelli, D., Garriss, G., Déry, C., Burrus, V., and Waldor, M.K. (2009). Comparative ICE genomics: Insights into the evolution of the SXT/R391 family of ICEs. PLoS Genet. *5*, e1000786.

Xu, J., and Koudelka, G.B. (2001). Repression of transcription initiation at 434 P_R by 434 repressor: Effects on transition of a closed to an open promoter complex. J. Mol. Biol. *309*, 573–587.

Yakhnin, A.V., Baker, C.S., Vakulskas, C.A., Yakhnin, H., Berezin, I., Romeo, T., and Babitzke, P. (2013). CsrA activates *flhDC* expression by protecting *flhDC* mRNA from RNase E-mediated cleavage: Mechanism of CsrA-mediated activation of *flhDC*. Mol. Microbiol. *87*, 851–866.

Yoh, M., Matsuyama, J., Ohnishi, M., Takagi, K., Miyagi, H., Mori, K., Park, K.-S., Ono, T., and Honda, T. (2005). Importance of *Providencia* species as a major cause of travellers' diarrhoea. J. Med. Microbiol. *54*, 1077–1082.

Zinder, N.D., and Lederberg, J. (1952). Genetic exchange in *Salmonella*. J. Bacteriol. *64*, 679–699.

Zuckerman, J.N., Rombo, L., and Fisch, A. (2007). The true burden and risk of cholera: implications for prevention and control. Lancet Infect. Dis. *7*, 521–530.