

**RÔLE ET RÉGULATION
DU SYSTÈME DE RECOMBINAISON
DES ICEs DE LA FAMILLE SXT/R391**

par

Geneviève Garriss

**Thèse présentée au Département de Biologie en vue
de l'obtention du grade docteur ès sciences (PhD)**

**FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE**

Sherbrooke, Qc, Canada

Décembre 2012



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-499-00423-9

Our file Notre référence

ISBN: 978-0-499-00423-9

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Le 19 décembre 2012

*le jury a accepté la thèse de Madame Geneviève Gariss
dans sa version finale.*

Membres du jury

Professeur Vincent Burrus
Directeur de recherche
Département de biologie

Professeur Ryszard Brzezinski
Membre
Département de biologie

Professeur George Szatmari
Membre externe
Université de Montréal

Professeur Daniel Lafontaine
Président rapporteur
Département de biologie

Side effects may include evolution.

(Cirz et al., 2006)

REMERCIEMENTS

Un merci immense à Vincent. Ces travaux n'auraient pu être possibles sans le temps, les conseils et la confiance d'un excellent mentor. Les connaissances que j'ai acquises au cours de ces six années dépassent grandement ce qui est présenté dans cette thèse.

Merci aux membres présents et passés du laboratoire Burrus, pour les discussions scientifiques (et moins scientifiques!), pour tous les beaux souvenirs et les belles amitiés. Votre participation à ce travail est plus importante que vous le croyez. Merci aux Pr Daniel Lafontaine et Ryszard Brzezinski pour les conseils et les discussions tout au long de mon parcours. Merci aux membres du jury d'évaluation pour le temps qu'ils ont dévoué à la lecture et à l'évaluation de cette thèse.

Un merci spécial à mes parents, Nicole et Matthew, pour le goût du travail bien accompli, la persévérance nécessaire à la réalisation de ces travaux et pour l'appui inconditionnel dont vous m'avez toujours témoigné. Merci à mes sœurs, Julie et Jackie, et à mon frère, Tom, pour les encouragements et le bonheur de vous avoir auprès de moi. Thanks to Lin, for believing in me all the way.

Merci aux amis, en particulier David, Daniela, Sophie, Éoui et Aurélie, pour votre fidélité et pour tous les beaux moments partagés.

Enfin, merci à Roland. Ta présence est inestimable.

TABLE DES MATIÈRES

SOMMAIRE.....	IV
REMERCIEMENTS	VII
LISTE DES TABLEAUX	XI
LISTE DES FIGURES	XII
PRÉAMBULE.....	XIV
CHAPITRE 1 - INTRODUCTION	15
1.1 Transfert horizontal.....	15
1.1.1. Transformation naturelle	16
1.1.2. Transduction.....	18
1.1.3. Transfert conjugatif	18
1.2. Recombinaison homologue chez <i>E. coli</i>	20
1.2.1. Dommages causés à l'ADN.....	21
1.2.2. RecBCD.....	23
1.2.3. RecFOR	24
1.2.4. Barrières à la recombinaison	25
1.3. La recombinaison homologue chez le bactériophage λ	25
1.3.1 λ Exo	26
1.3.2. λ Bet	27
1.3.3. λ Gam	27
1.3.4. Mécanismes de recombinaison homologue du bactériophage λ	28
1.3.4.1. Appariement de molécules simple-brin.....	29
1.3.4.2. Invasion d'une molécule double-brin.....	29
1.3.4.3. Invasion d'une fourche de réplication	31
1.3.5. Homologues du système Red	33
1.3.6. Rôle biologique du système Red	34

1.3.7. Le système Red comme outil de biologie moléculaire.....	34
1.4. RÉPONSE SOS	35
1.5 ELEMENTS INTEGRATIFS ET CONJUGATIFS	36
1.5.1. Le choléra et la découverte de SXT	39
1.5.2. La famille SXT/R391	40
1.5.3. Squelette minimal fonctionnel.....	41
1.5.3.1 Intégration site-spécifique	41
1.5.3.2. Transfert conjugatif	43
1.5.3.3. Régulation du transfert	44
1.5.4. Les ICE SXT/R391 sont apparentés aux plasmides de la famille IncA/C	45
1.5.5. Régions variables et diversité des fonctions conférées	47
1.5.6. Système toxine/antitoxine	49
1.5.7. Mobilisation de plasmides, d'îlots génomiques mobilisables et d'ADN chromosomique	51
1.5.7.1. Mobilisation de plasmides.....	51
1.5.7.2. Mobilisation d'îlots génomiques mobilisables.....	52
1.5.7.3. Mobilization d'ADN chromosomique.....	53
1.5.8. Exclusion d'entrée	54
1.5.9. Formation de tandems et d'hybrides SXT/R391	55
1.6. Hypothèse générale du projet de doctorat.....	59
1.6.1. Objectifs du projet de doctorat	59
CHAPITRE 2 – RÉSULTATS.....	60
2.1. Rôle du système de recombinaison homologue des ICEs SXT/R391 dans la formation d'hybrides	61
2.1.1 Présentation de l'article	61
2.1.2. Contribution à l'article	62
2.2. Régulation du système de recombinaison homologue des ICEs SXT/R391 et étude de la distribution phylogénétique de systèmes apparentés	101
2.2.1. Présentation de l'article	101
2.2.2. Contribution à l'article	102

CHAPITRE 3 - DISCUSSION.....	148
3.1 Formation d'hybrides.....	148
3.2 Régulation du système de recombinaison.....	152
3.3. Rôles de Ssb et OrfZ.....	157
3.4. Distribution phylogénétique et évolution	159
3.5. Aspects mécanistiques et rôle biologique.....	162
3.6. Impact évolutif.....	164
3.7. Perspectives	165
CONCLUSION	169
BIBLIOGRAPHIE	172
ANNEXE 1.....	189
ANNEXE 2.....	193

LISTE DES TABLEAUX

Tableau 1. Exemples d'éléments intégratifs et conjugatifs.	38
Table 1. Strains of <i>E. coli</i> and plasmids used in this study.	76
Table 2. Percentage of colonies containing hybrid ICEs or single elements recovered over time from a strain initially harboring a wild-type R391-SXT, or a non-transmissible $\Delta mobI$ R391-SXT tandem array.....	82
Table 1. List of <i>E. coli</i> strains and plasmids used in this study.....	114
Table 2. Mitomycin C induces the translation of <i>bet</i> and <i>exo</i>	123
Table S1. List of primers used in this study.	147
Tableau 2. Principaux membres de la famille SXT/R391.....	190

LISTE DES FIGURES

Figure 1. Transfert conjugatif de plasmides et d'éléments intégratifs et conjugatifs.....	19
Figure 2. Bris induits par la déstabilisation d'une fourche de réplication.....	22
Figure 3. Système de recombinaison Red du bactériophage λ	26
Figure 4. Modèles de mécanismes de recombinaison du système λ Red.....	30
Figure 5. Invasion d'une fourche de réplication.....	32
Figure 6. Squelette de gènes conservés des ICEs de la famille SXT/R391.....	41
Figure 7. Intégration et excision site-spécifique de SXT.....	42
Figure 8. Régulation de SXT.....	45
Figure 9. Comparaison du squelette de gènes conservés de SXT et du plasmide conjugatifs pIP1202 du groupe d'incompatibilité IncA/C.....	46
Figure 10. Régions variables des ICEs de la famille SXT/R391.....	48
Figure 11. Mobilisation des MGIs par les ICEs SXT/R391.....	53
Figure 12. Formation de tandems SXT-R391 dans le chromosome.....	56
Figure 13. Formation d'hybrides SXT/R391.....	57
Figure 1. Evidence suggesting that recombination occurs between SXT/R391 ICEs.....	71
Figure 2. Schematic of colony color-based semi-quantitative assay for the detection of hybrid ICE-containing colonies.....	73
Figure 3. Involvement of <i>recA</i> , <i>s065</i> and <i>s066</i> in the formation of hybrid ICEs.....	78
Figure 1. Schematic comparison of the recombination loci of SXT/R391 ICEs, IncA/C plasmids and bacteriophage λ	107
Figure 2. Expression of <i>bet</i> and <i>exo</i> is induced by mitomycin C.....	116
Figure 3. Operon structure of the SXT/R391 homologous recombination system.....	118
Figure 4. Promoter region of the SXT/R391 homologous recombination system.....	120
Figure 5. Impact of <i>s089</i> and <i>ssb</i> on hybrid ICE formation.....	122
Figure 6. Relative quantification of the transcription and translation of <i>bet</i> and <i>exo</i> in presence of MMC.....	124

Figure 7. The recombination locus contains two predicted translational attenuators.	126
Figure 8. Phylogenetic trees of single-stranded DNA annealing proteins and cognate exonucleases encoded by gene pairs found in sequenced microbial, plasmid and bacteriophage genomes.....	129
Figure S1. Comparison of transcription levels of <i>bet</i> and <i>exo</i> in a WT context and <i>lacZ</i> in the <i>bet::lacZ</i> and <i>exo::lacZ</i> fusions, in control (CTL) and MMC-induced conditions (MMC)....	145
Figure S2. Aligement of the intergenic regions located between <i>bet</i> and <i>orfZ</i> (Ig <i>bet</i> - <i>orfZ</i>) and between <i>orfZ</i> and <i>exo</i> (Ig <i>orfZ</i> - <i>exo</i>) in SXT.....	145
Figure S3. Aligement of the predicted translational attenuators found upstream <i>bet</i> (TAbet) and upstream <i>exo</i> (TAexo) in sequenced SXT/R391 ICEs.....	146
Figure S4. Translational attenuators predicted by RibEx upstream <i>bet</i> and <i>exo</i> in IncA/C plasmid pIP1202.....	146
Figure 14. Induction de la recombinaison homologue par Bet/Exo et par RecA.....	154

PRÉAMBULE

Les éléments intégratifs et conjugatifs de la famille SXT/R391 sont des éléments génétiques mobiles, retrouvés chez une diversité d'hôtes bactériens, qui participent grandement au transfert horizontal de gènes, notamment de gènes codant pour des résistances aux antibiotiques. Le projet ici présenté a visé à étudier le rôle d'un nouveau système de recombinaison homologue, identifié chez les ICEs de la famille SXT/R391, entre autres en déterminant sa participation à la formation d'ICEs hybrides issus de la recombinaison entre deux ICEs SXT/R391 semblables mais comportant des combinaisons différentes de gènes accessoires.

Afin de mettre en contexte le sujet et de faire un survol de la littérature appropriée permettant une meilleure compréhension des éléments génétiques mobiles et des mécanismes de recombinaison connus, l'introduction (Chapitre 1) touchera aux principaux thèmes suivants : le transfert horizontal, la recombinaison homologue chez *Escherichia coli* et le bactériophage λ et les ICEs de la famille SXT/R391. Le chapitre 2 (Résultats) est divisé en deux sections, chacune présentant un article scientifique issu des résultats obtenus au courant du projet de doctorat. Un troisième article, situé en Annexe 2, n'a pas été intégré à proprement dit dans la thèse puisque ma participation au projet qui a mené à cette publication a été moins significative. Néanmoins, cet article amène une avancée significative dans la compréhension de l'organisation génétique des ICEs de la famille SXT/R391. Il m'a donc semblé plus pertinent de l'utiliser dans l'introduction, afin de présenter un survol complet des connaissances acquises à ce jour sur cette famille d'éléments génétiques mobiles. Finalement, le chapitre 3 est une discussion générale, qui se veut une intégration des résultats présentés dans les deux articles du chapitre 2. Comme cette discussion n'est pas une répétition des discussions des publications, pour une analyse précise des résultats de chaque article, je vous prie de vous référer à leurs discussions respectives.

CHAPITRE 1 - INTRODUCTION

1.1 Transfert horizontal

Les bactéries ont la capacité de répondre rapidement à un environnement en constant changement grâce à la nature hautement plastique de leur génome (Skippington et Ragan, 2011 ; Stokes et Gillings, 2011 ; Wiedenbeck et Cohan, 2011). La plasticité des génomes bactériens est en grande partie due à leur abilité à mobiliser du matériel génétique grâce à un phénomène appelé transfert horizontal. L'impact du transfert horizontal est substantiel et il est estimé que plus de 25% de certains génomes bactériens peuvent être dérivés d'évènements de transfert horizontal (Ochman *et al.*, 2000). L'émergence de souches bactériennes multirésistantes aux antibiotiques est l'exemple le plus frappant et certainement le plus médiatisé de l'évolution bactérienne due au transfert horizontal. Les bactéries à Gram négatif, notamment les *Enterobacteriaceae*, sont particulièrement bien adaptées à échanger du matériel génétique et ainsi acquérir des gènes conférant la résistance aux antibiotiques (Partridge, 2011).

Il est aussi reconnu que l'impact du transfert horizontal sur le métabolisme bactérien dépasse grandement la résistance aux antibiotiques, et contribue de façon majeure, entre autres, à l'adaptation à de nouvelles niches écologiques et à la virulence des souches pathogènes (Ochman *et al.*, 2000). Plus la pression de sélection exercée par un environnement donné est forte, plus les gènes favorisant l'adaptation à cet environnement auront de chances d'être conservés, disséminés et de devenir prédominants dans une population donnée. Au fur et à mesure que du nouveau matériel génétique est acquis et que les pressions de sélection changent, d'autres gènes pourront devenir prédominants (Partridge, 2011). (Pour des revues

exhaustives sur le transfert horizontal voir Ochman *et al* (2000), Partridge (2011), Skippington et Ragan (2011), Stokes et Gillings (2011), ainsi que Wiedenbeck et Cohan (2011)).

Le transfert horizontal peut se diviser en deux étapes : i) la translocation physique d'un fragment d'ADN d'une cellule donneuse vers une cellule réceptrice et ii) l'intégration de ce fragment dans le génome de la cellule réceptrice, ce qui en assure la stabilité. La translocation de l'ADN entre deux cellules bactériennes s'effectue principalement selon trois mécanismes : la transformation naturelle, la transduction et la conjugaison. Le premier de ces mécanismes est entièrement médié par le génome de l'hôte, tandis que les deux derniers sont catalysés par l'action d'éléments génétiques mobiles tels les plasmides conjugatifs, les bactériophages, et les éléments intégratifs et conjugatifs. L'intégration stable des fragments acquis par la cellule réceptrice peut être entraînée par différents mécanismes : recombinaison homologue (voir section 1.2), intégration site-spécifique (voir section 1.5.3.1), transposition, recombinaison illégitime ou par reconstitution d'un réplicon autonome, tel un plasmide (Brigulla et Wackernagel, 2010 ; Stokes et Gillings, 2011). Pour des considérations d'espace et de pertinence pour le projet de doctorat, seuls la recombinaison homologue et l'intégration site-spécifique seront abordés dans ce document (voir sections indiquées). Le lecteur désireux d'obtenir plus d'information est dirigé vers les revues publiées par Stokes et Gillings (2001), Brigulla et Wackernagel (2010) et Skippington et Ragan (2011).

1.1.1. Transformation naturelle

La transformation naturelle est l'acquisition d'ADN extracellulaire par une cellule bactérienne dite compétente. Au contraire des autres modes de transfert horizontal, la transformation naturelle ne dépend pas de la présence d'éléments génétiques mobiles : c'est une caractéristique inhérente d'une espèce bactérienne, dont les fonctions sont codées par des gènes chromosomiques (Claverys *et al.*, 2009). Certaines espèces sont naturellement

compétentes, telles *Neisseria gonorrhoeae* et *Haemophilus influenzae*, mais chez la majorité des espèces bactériennes la compétence est induite par un état physiologique transitoire dépendant de processus spécifiques à chaque organisme, comme le quorum sensing, des signaux nutritifs (Chen et Dubnau, 2004) ou même par la présence de certaines molécules dans le milieu extracellulaire de la bactérie telles la chitine (*Vibrio cholerae* et espèces apparentées) (Lo Scrudato et Blokesch, 2012). La transformation naturelle se divise en trois étapes : i) acquisition d'ADN extracellulaire, ii) recombinaison de séquences homologues avec le chromosome et iii) expression phénotypique du matériel génétique acquis (Kruger et Stingl, 2011).

L'acquisition d'ADN par transformation chez les bactéries à Gram positif et à Gram négatif se fait par des mécanismes similaires et invoque la participation de certaines protéines homologues à celles qui composent les systèmes de sécrétion de type IV et de type II (Chen et Dubnau, 2004). La transformation a été particulièrement étudiée chez *Bacillus subtilis* et servira donc ici de modèle. L'acquisition d'ADN se fait en deux étapes : i) la liaison de l'ADN extracellulaire à la surface d'une cellule compétente, et ii) la translocation de l'ADN à travers la membrane cytoplasmique. La liaison de l'ADN à la surface cellulaire se fait par l'intermédiaire du pseudopilus, qui permet l'accès de l'ADN double-brin extracellulaire à un récepteur, et sa subséquente translocation à travers un canal membranaire. L'ADN transformant entre dans la cellule sous forme simple-brin, le brin complémentaire du duplex étant dégradé durant la translocation (Chen et Dubnau, 2004). L'interaction de la molécule d'ADN simple-brin entrante avec les protéines cytoplasmiques est essentielle pour la protéger de la dégradation. Par la suite, l'ADN simple-brin peut être intégré dans le chromosome par recombinaison homologue, ou bien s'il s'agit d'un plasmide, être reconstitué et maintenu stablement sous forme extrachromosomique (Chen et Dubnau, 2004).

1.1.2. Transduction

Les bactériophages sont les organismes majoritaires de notre planète et ils jouent un rôle majeur tant dans l'écosystème de la planète que dans la diversification génétique de leur hôtes bactériens (Krupovic *et al.*, 2011). Les bactériophages catalysent la mobilisation d'ADN d'une cellule à une autre par transduction, qui peut être de deux types, généralisée ou spécialisée. La transduction généralisée se produit lorsque de l'ADN bactérien est empaqueté par erreur à la place de l'ADN du phage dans la capsid virale. La particule résultante est fonctionnelle : l'ADN bactérien qu'elle contient pourra être injecté dans une nouvelle cellule et être intégré par recombinaison homologe, si une homologie suffisante y existe (Brussow *et al.*, 2004). La transduction spécialisée est catalysée par les phages tempérés qui s'intègrent dans le chromosome de leur hôte. L'excision imprécise d'un prophage mène à l'excision conjointe des gènes situés à proximité du site d'intégration. Ceux-ci seront alors répliqués, empaquetés et transportés avec le phage vers une nouvelle cellule hôte (Brigulla et Wackernagel, 2010).

La participation des phages tempérés dans la dissémination de gènes de résistance aux antibiotiques est de plus en plus reconnue (Skippington et Ragan, 2011) et leur rôle dans l'émergence de souches bactériennes virulentes est bien documentée. Notamment, plusieurs gènes codant des toxines chez des pathogènes d'importance (i.e. *Corynebacterium diphtheriae*, *Staphylococcus aureus*, *Clostridium botulinum*, *Vibrio cholerae*, *Escherichia coli*) sont contenues dans des génomes de phages (Brussow *et al.*, 2004).

1.1.3. Transfert conjugatif

Le transfert conjugatif, ou conjugaison, est le mode de transfert utilisé par les plasmides conjugatifs et les éléments intégratifs et conjugatifs (ICEs) (Figure 1) (Skippington et Ragan,

2011). Le transfert conjugatif se divise en deux étapes clés : l'initiation du transfert conjugatif par la coupure d'un brin de la molécule double-brin circulaire de l'élément, à un site nommé l'origine de transfert (*oriT*), et la formation de l'appareil conjugatif qui permettra la translocation de l'ADN entre une cellule donneuse et une cellule réceptrice.

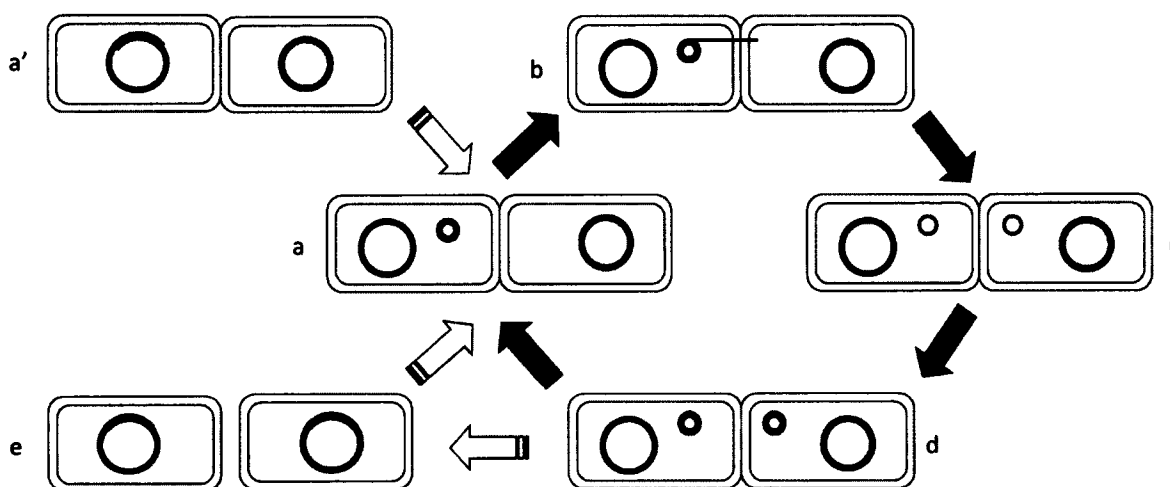


Figure 1. Transfert conjugatif de plasmides et d'éléments intégratifs et conjugatifs.

a) une cellule donneuse contenant un plasmide ou un ICE sous forme extrachromosomique entre en contact avec une cellule réceptrice. (Dans le cas d'un ICE, celui-ci s'excise préalablement du chromosome (a')). b) la molécule simple-brin clivée par la relaxase est transférée vers la cellule réceptrice à travers le pore de conjugaison, c) la molécule est recircularisée par la relaxase, d) le brin complémentaire est synthétisé, e) si l'élément conjugatif est un ICE il se réintègre dans le chromosome par recombinaison site-spécifique. Cercle orange, chromosome de la cellule donneuse; cercle vert, chromosome de la cellule réceptrice; cercle violet, plasmide ou ICE excisé; trait violet, ICE intégré.

La relaxase reconnaît et clive un brin de la molécule double-brin à l'origine de transfert afin d'initier le transfert conjugatif. La relaxase forme généralement un complexe multiprotéique avec des protéines auxiliaires de mobilisation nommé relaxosome. Une fois la molécule clivée, un seul des deux brins est transféré de la cellule donneuse vers la cellule réceptrice à travers un canal multiprotéique codé par l'élément lui-même (Figure 1) (Chen *et al.*, 2005 ; Smillie *et*

al., 2010). Dans le cas des éléments intégratifs et conjugatifs, l'excision préalable de l'élément du chromosome sous forme d'une molécule double-brin circulaire est nécessaire (Wozniak et Waldor, 2010). Suite au transfert dans la cellule réceptrice la relaxase, qui est restée liée à l'extrémité 5' de la molécule durant le transfert, permet de religuer les deux extrémités de la molécule simple-brin transloquée. Cette molécule sera ensuite dupliquée et maintenue sous forme extrachromosomique répliquative, dans le cas d'un plasmide. Dans le cas d'un élément intégratif et conjugatif, suite à la duplication il sera intégré de manière site-spécifique dans le chromosome de la cellule réceptrice (Figure 1) (Wozniak et Waldor, 2010). Dans les deux cas, la cellule donneuse et la cellule réceptrice pourront agir comme donneuses pour un autre transfert conjugatif.

Certains plasmides dits «mobilisables» possèdent une *oriT* et les gènes codant pour les protéines composant un relaxosome. Ils peuvent être mobilisés *in trans* en utilisant l'appareil conjugatif codé par un élément conjugatif (Smillie *et al.*, 2010). Plus récemment, il a été démontré que certains îlots génomiques peuvent être mobilisés par des éléments intégratifs et conjugatifs (ICEs), due à la reconnaissance de leur *oriT* similaire par le relaxosome d'un ICE présent dans la même cellule, et par leur transfert subséquent à travers le pore de conjugaison de l'ICE (Daccord *et al.*, 2010).

1.2. Recombinaison homologue chez *E. coli*

La recombinaison homologue est un processus inhérent et vital à toute cellule, car elle permet le maintien de l'intégrité génomique. La recombinaison homologue joue un rôle central dans l'intégration des gènes acquis par transfert horizontal et constitue ainsi une étape clé de la diversification génétique. Bien que le rôle de la recombinaison homologue dans la plasticité génomique soit indéniable, son rôle premier est de réparer les dommages causés à l'ADN, afin de permettre la répliquaison du matériel génétique et la division cellulaire. La réparation des

cassures double-brin dans le chromosome bactérien procède par recombinaison homologue, en utilisant une molécule d'ADN double-brin intacte afin de réparer une molécule brisée. Ce phénomène est assuré par une protéine centrale retrouvée chez pratiquement toutes les espèces bactériennes, RecA, qui catalyse les deux étapes essentielles à la réparation par recombinaison : i) la recherche d'un partenaire homologue à une molécule brisée, et ii) l'échange des brins complémentaires de ces deux molécules (Kuzminov, 1999). Outre son rôle dans la réparation par recombinaison, RecA est également responsable de l'induction de la réponse SOS (voir section 1.4. Réponse SOS).

1.2.1. Dommages causés à l'ADN

Au cours de sa vie, une cellule bactérienne doit faire face à de nombreuses conditions environnementales et agents chimiques qui endommagent son génome. Les bris causés sont de différentes natures (double-brin ou simple-brin) et la réparation des différents types de bris ou d'altérations est prise en charge par différents mécanismes cellulaires de réparation. La réparation de altérations simple-brin s'effectue soit par réversion des dommages (ex. : photoréactivation, dans le cas de dimères CC ou TT) soit par l'enlèvement des bases entourant la lésion, permettant la resynthèse en utilisant le brin complémentaire comme matrice. Ce dernier cas fait appel à plusieurs systèmes : le «base excision repair» ou BER, le «nucleotide excision repair» ou NER et le «méthyl-directed mismatch repair» (Kuzminov, 1999 ; Moat, 2002). Quoique moins fréquent, l'endommagement direct de deux sites opposés sur une même molécule par l'action directe d'agents toxiques (rayons X, les rayons gamma, les rayons UV, les radicaux oxygénés, la mitomycine C) crée des cassures double-brin. Ce genre de bris peut-être converti en une paire de cassures simple-brin par simple échange de brin avec une molécule double-brin intacte homologue. Ces lésions simple-brins pourront être prises en charge par la suite par les mécanismes nommés ci-haut (Kuzminov, 1999).

Il arrive cependant qu'une lésion simple-brin cause la déstabilisation d'une fourche de réplication passant. Dans un premier exemple, si la machinerie répliquative rencontre un site abasique elle sera déstabilisée et la fourche de réplication se désintègrera, provoquant la séparation d'une molécule double-brin du duplex en cours de réplication (Figure 2A) (Kowalczykowski, 2000 ; Kuzminov, 1999). Dans un deuxième exemple, si la fourche de réplication rencontre une lésion non-codante (ex. un dimère de thymine), elle sera momentanément bloquée et redémarrera en aval de la lésion non-codante, laissant derrière elle une région simple-brin non répliquée (Figure 2B) (Kuzminov, 1999).

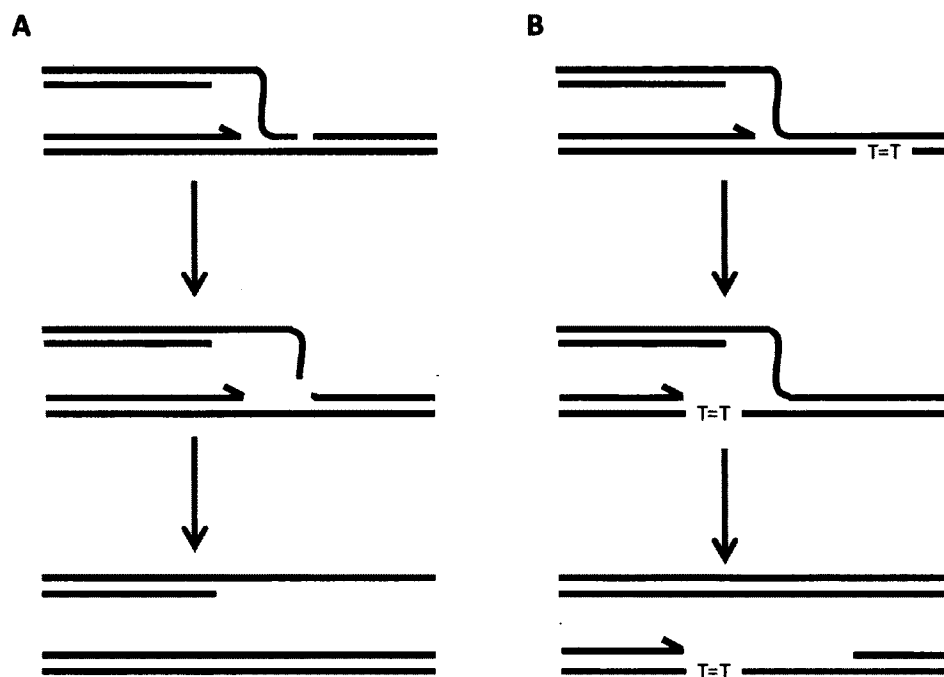


Figure 2. Bris induits par la déstabilisation d'une fourche de réplication.

A) Un site abasique cause la désintégration et le détachement de la fourche de réplication, générant une cassure double-brin. B) Une lésion simple-brin non-codante bloque l'avancement de la fourche de réplication. La réplication redémarre en aval de la lésion, laissant un segment simple-brin dans la molécule nouvellement répliquée. Inspiré de Kuzminov (1999).

La réparation des cassures double-brin peut se diviser en trois phases : i) la présynapse : préparation de l'ADN pour la recombinaison, ii) la synapse : appariement et échange de brins homologues, et iii) la postsynapse : reprise de la réplication et résolution des intermédiaires de recombinaison. Chaque étape nécessite la participation de plusieurs protéines (Kuzminov, 1999 ; Nowosielska, 2007). L'étape présynaptique de la réparation des deux types de bris induits par la réplication procède par deux mécanismes distincts qui portent le nom des complexes protéiques assistant RecA : RecBCD et RecFOR, et seront brièvement décrits dans les sections suivantes. L'étape synaptique est caractérisée par un intermédiaire de recombinaison à trois brins appelé boucle de déplacement ou «D-loop», qui est causé par l'invasion d'une molécule double-brin par le filament simple-brin auquel est lié RecA. Dans l'étape postsynaptique, la résolution des jonctions de Holliday formées par la recombinaison est réalisée par des facteurs de l'hôte, tels RuvABC et RecG (Kuzminov, 1999).

1.2.2. RecBCD

RecBCD est une hélicase/exonucléase ATP-dépendante extrêmement processive qui est nécessaire à la réparation par recombinaison homologe des fourches de réplication affaissées (Kuzminov, 1999). RecBCD lie et dégrade agressivement l'ADN double-brin linéaire, jusqu'à la rencontre du site Chi («crossover hotspot instigator » ou χ) (5' -GCTGGTGG- 3'). Cette dégradation est asymétrique, l'extrémité 3' étant dégradée beaucoup plus rapidement que l'extrémité 5'. La rencontre de Chi cause un changement de préférence de brin dégradé par RecBCD : après Chi RecBCD ne dégrade que l'extrémité 5', ce qui produit une extrémité 3' simple-brin libre qui sera ultérieurement utilisée comme substrat de recombinaison par RecA (Anderson et Kowalczykowski, 1997 ; Kuzminov, 1999 ; Smith, 2012). L'interaction entre RecBCD et Chi est donc essentielle à la recombinaison. L'extrémité d'ADN simple-brin produite est aussitôt liée par Ssb, qui inhibe initialement la liaison de RecA. Bien que compétitrice, la liaison de Ssb à l'ADN simple-brin est essentielle, car elle permet de déstabiliser les structures secondaires qui empêcheraient la polymérisation de RecA sur

l'ADN simple-brin. (Schlacher *et al.*, 2006). RecBCD catalyse ensuite le chargement de RecA sur l'ADN simple-brin lié par Ssb (Anderson et Kowalczykowski, 1997). Une fois polymérisée sur l'ADN simple-brin, RecA catalyse l'invasion d'une molécule double-brin ou l'appariement avec une molécule homologue double-brin comportant elle aussi une extrémité 3' simple-brin. Par la suite, il y aura reprise de la réplication et résolution des jonctions de Holliday produites dans les intermédiaires de recombinaison est réalisée par RuvABC ou RecG (Kuzminov, 1999). (Les détails de cette étape ne seront pas abordés ici. Pour une revue exhaustive, veuillez voir les publications de Kuzminov (1999) et Nowosielska (2007)).

1.2.3. RecFOR

La voie RecFOR est impliquée dans la réparation des lésions simple-brin produites lorsque la machinerie répliquative est déstabilisée par une lésion non-codante et laisse derrière elle un segment simple-brin. Cette voie fait appel à plusieurs protéines, RecF, RecO, RecR, RecJ, RecQ et RecN. Dans l'étape présynaptique, l'extrémité 3' simple-brin utilisée par RecA est préparée par l'hélicase RecQ et l'exonucléase 5'-3' RecJ. Par la suite, RecFOR catalyse l'assemblage de RecA sur la molécule d'ADN simple-brin complexée à Ssb. Une fois polymérisé sur l'ADN simple-brin, RecA catalyse l'invasion d'un partenaire double-brin homologue. L'étape synaptique est facilitée par le déenroulement des superenroulements causés dans le duplex intact par l'invasion (gyrase) et dans le nouveau duplex, entre le brin envahissant et le brin d'origine (topoisomérase I). Le segment simple-brin est alors dupliqué par l'ADN polymérase I et les jonctions de Holliday sont résolues par RuvABC ou RecG (Kuzminov, 1999).

1.2.4. Barrières à la recombinaison

La fidélité de la recombinaison, et conséquemment l'intégration de séquences divergentes permettant la diversification génétique, est dépendante de l'étendue et du degré de l'homologie et est contrôlée par les enzymes de recombinaison et par le système MMR ou «mismatch repair». Le système MMR, composé chez *E. coli* des protéines MutS, MutL et MutH, fonctionne de manière à reconnaître et réparer les bases mésappariées durant la réplication (Yang *et al.*, 2000). Cependant, les protéines MutS et MutL du système MMR bloquent aussi l'échange de brins par RecA entre deux molécules trop divergentes (Worth *et al.*, 1994). Entre autres, la fréquence de recombinaison après conjugaison ou transduction entre *E. coli* et *Salmonella typhimurium* augmente de 1 000 fois dans une réceptrice *mutSL* (Matic *et al.*, 1995).

1.3. La recombinaison homologue chez le bactériophage λ

À l'instar des cellules bactériennes, plusieurs bactériophages possèdent des systèmes de recombinaison. Cependant, au contraire de leurs équivalents bactériens, ces systèmes sont beaucoup plus simples sur les points de vue enzymatiques et mécanistiques. Le système de recombinaison le plus étudié chez les bactériophages est celui du phage λ . Ce système, nommé système Red, comporte trois gènes, *exo*, *bet* et *gam* qui lui sont essentiels pour catalyser la recombinaison homologue en l'absence de RecA (Figure 3) (Court *et al.*, 2002). Situés un à la suite de l'autre dans l'opéron *pL* du phage λ , les gènes du système Red se retrouvent sous le contrôle du répresseur CI et sont exprimés tôt après l'infection par le phage ou après l'induction du prophage (Court *et al.*, 2002).



Figure 3. Système de recombinaison Red du bactériophage λ .

Représentation partielle de l'opéron pL du bactériophage λ . En noir les gènes du système de recombinaison Red : *gam* code pour un inhibiteur de RecBCD, *bet* code pour une recombinase et *exo* pour une exonuclease. En gris, les autres gènes de l'opéron pL , situés entre N et le site *attP* : *n* code pour l'antitermineur N (Oppenheim *et al.*, 2005), *ral* code pour une protéine impliquée dans la résistance aux systèmes de restriction-modification (Zabeau *et al.*, 1980), *cIII* code pour une protéase qui inhibe la dégradation de *cII* (Kobiler *et al.*, 2007), *kil* code pour une protéine induisant une lyse cellulaire prématurée (Reisinger *et al.*, 1993), *ea22* et *ea8.5* sont impliqués dans le blocage de l'initiation de la réplication (Sergueev *et al.*, 2002), *int* code pour l'intégrase qui catalyse l'intégration site-spécifique du phage dans le chromosome et *xis* code pour le facteur de directionnalité de la recombinaison (excisionase) qui assiste Int à catalyser l'excision du prophage du chromosome (Oppenheim *et al.*, 2005).

1.3.1 λ Exo

Le gène λ exo code pour une exonucléase qui lie les extrémités d'ADN double-brin et en dégrade l'extrémité 5', laissant après son passage de longues extrémités simple-brin en 3'. Les extrémités 3' simple-brin générées par λ Exo seront utilisées par la recombinase codée par le gène λ bet. Exo nécessite la présence d'une extrémité double-brin pour initier le clivage; elle est incapable d'utiliser des sites où il n'y a qu'une coupure simple-brin (nick) dans l'ADN (Court *et al.*, 2002). En solution, λ Exo est une protéine trimérique en forme de tunnel dont l'entrée peut accommoder de l'ADN double-brin, mais dont la sortie ne laisse la place que pour une molécule simple-brin (Kovall et Matthews, 1997), suggérant que λ Exo glisse le long du brin 3' en clivant les mononucléotides du brin 5' et laissant l'extrémité 3' intacte (Court *et al.*, 2002). λ Exo est une enzyme très processive, capable de cliver environ 3 000 nucléotides dans

un seul évènement de liaison (Kuzminov, 1999), à une fréquence d'environ 12 nucléotides par seconde (Kovall et Matthews, 1997).

1.3.2. λ Bet

λ Bet est une protéine de liaison à l'ADN simple-brin, ou recombinaise, qui a la capacité de promouvoir l'appariement de molécules complémentaires d'ADN simple-brin et catalyser l'échange de brins (Kmiec et Holloman, 1981). Comme plusieurs autres recombinaises, λ Bet peut réaliser ces activités en présence ou absence de RecA (Iyer *et al.*, 2002 ; Kuzminov, 1999). Plusieurs protéines dérivées de différents prophages telles RecT (*Escherichia coli*), OrfC (*Listeria pneumophila*), Orf48 (*Listeria monocytogenes*), ou GP35 (*Bacillus subtilis*) sont connues pour être fonctionnellement apparentées à λ Bet : elles sont capables de catalyser la recombinaison d'ADN simple-brin avec le chromosome, et en présence de l'exonucléase qui leur est associée, elles peuvent catalyser la recombinaison d'ADN double-brin avec le chromosome (Datta *et al.*, 2008). Les modèles proposés de mécanismes par lesquels λ Bet catalyse la recombinaison d'ADN double-brin sont décrits dans la section 1.3.4.

1.3.3. λ Gam

La participation de Gam est essentielle pour conférer le potentiel maximal de recombinaison au système Red et aussi à la réplication du phage. Le rôle principal de Gam est d'inhiber l'activité de la protéine RecBCD de la cellule hôte (Karu *et al.*, 1975). Malgré le fait qu'elle soit présente en seulement quelques copies par cellule, RecBCD dégrade très agressivement la majorité de l'ADN double-brin linéaire qui s'y retrouve (Court *et al.*, 2002) et interfère donc avec la réplication tardive de λ , durant laquelle les concatémères d'ADN sont formés (Marsic

et al., 1993). Ces concatémères seront ensuite clivés et empaquetés dans la capside virale. Gam inhibe toutes les fonctions connues de RecBCD, incluant ses activités de nucléase, d'hélicase, d'ATPase et sa capacité à lier χ (Karu *et al.*, 1975). Gam inhibe l'activité nucléase de RecBCD en se liant à la sous-unité RecB du complexe (Marsic *et al.*, 1993) et inhibe sa liaison aux extrémités d'ADN double-brin (Murphy, 2007). Il est à noter qu'il n'y a pas de sites χ dans le type sauvage de λ , mais que ceux-ci peuvent apparaître grâce à des mutations. Ainsi, dans un mutant $\lambda bet/exo^-$ la recombinaison se fait par la voie RecBCD de l'hôte, mais est très faible. Le taux de recombinaison est grandement augmenté par les mutations introduisant un site χ (Stahl, 1998).

Quelques autres inhibiteurs de RecBCD existent chez d'autres bactériophages, bien que leurs mécanismes d'actions soient différents de celui de Gam. Le produit du gène 2 du phage T4 lie les extrémités d'ADN double-brin afin de les protéger de l'action de RecBCD (Lipinska *et al.*, 1989 ; Silverstein et Goldberg, 1976), tandis que la protéine Abc2 du bactériophage P22 modifie RecBCD afin de l'incorporer dans son propre système de recombinaison (Murphy et Lewis, 1993).

1.3.4. Mécanismes de recombinaison homologue du bactériophage λ

Le mécanisme par le lequel le système Red catalyse la recombinaison est le sujet d'un débat dans la littérature scientifique qui n'a pas encore atteint un consensus (Szczepanska, 2009). À ce jour, trois modèles principaux distincts de recombinaison ont été suggérés pour le système Red : l'appariement d'ADN simple-brin, l'invasion d'ADN double-brin et l'invasion d'une fourche de réplication.

1.3.4.1. Appariement de molécules simple-brin

Le premier mécanisme de recombinaison à avoir été proposé pour le système Red est l'appariement de deux molécules simple-brin homologues. Lorsque deux partenaires d'ADN double-brin homologues se retrouvent brisés, les extrémités 5' seraient dégradées par Exo, produisant les extrémités 3' simple-brin qui sont les substrats utilisés par Bet pour l'appariement (Figure 4A). Ce mécanisme est totalement indépendant de la présence de RecA. La nature non conservatrice de ce type de recombinaison la rend inutilisable pour la réparation d'un génome comme celui *E. coli*, mais plus acceptable pour les génomes phagiques, qui se trouvent généralement en plusieurs copies dans une cellule (Kuzminov, 1999). Cependant, il explique mal le haut taux de recombinaison élevé catalysé par le système Red, étant entièrement dépendant de la présence fortuite de cassures double-brin dans deux molécules homologues.

1.3.4.2. Invasion d'une molécule double-brin

Le modèle de recombinaison par invasion d'une molécule double-brin est le modèle généralement associé avec la recombinaison catalysée par RecA, qui procède par l'invasion d'une molécule double-brin homologue à la molécule simple-brin à laquelle elle est liée, en formant une boucle de déplacement (Figure 4B) (Kuzminov, 1999).

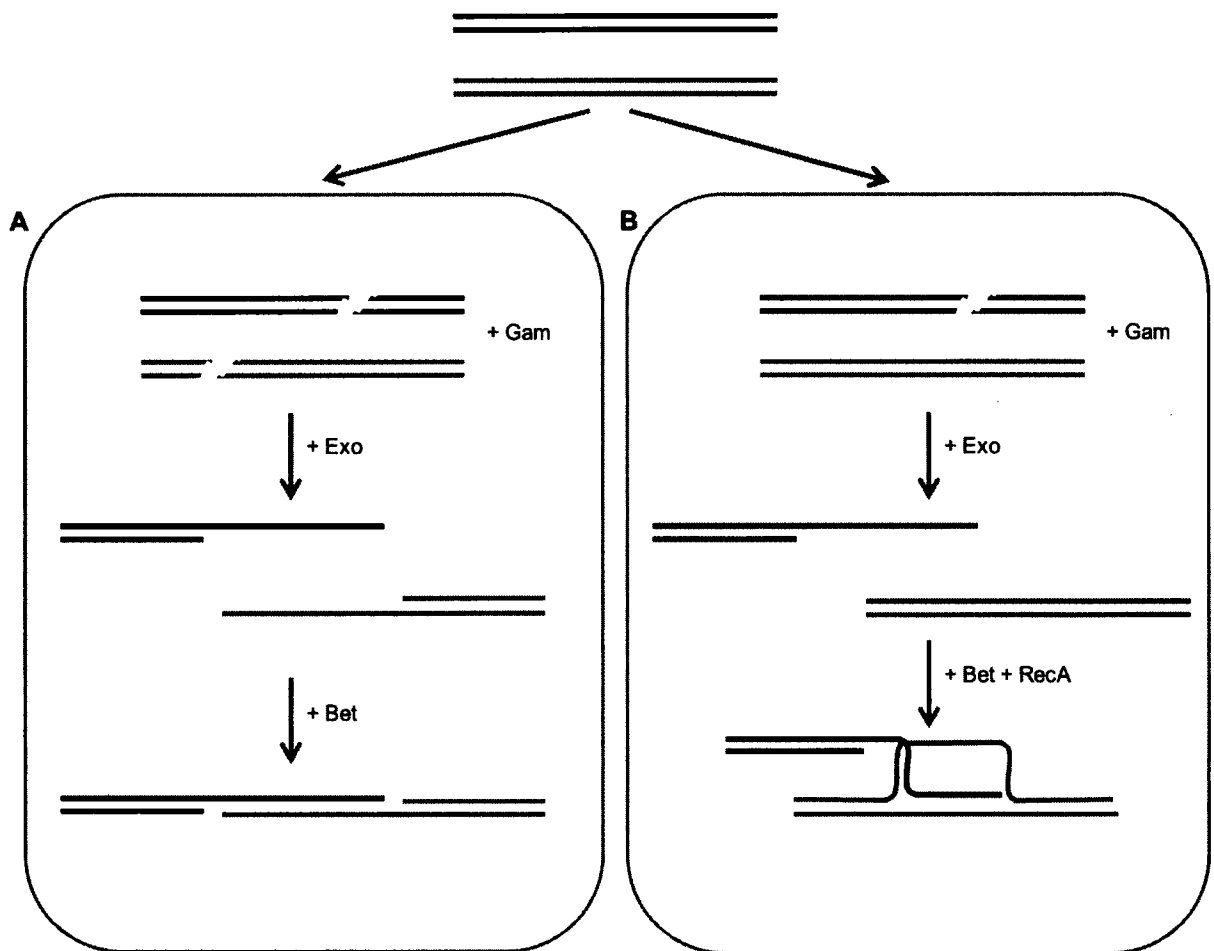


Figure 4. Modèles de mécanismes de recombinaison du système λ Red.

A) Modèle d'appariement simple-brin. Une cassure double-brin dans deux molécules homologues (traits noirs et gris) génère des extrémités double-brin qui sont dégradées par Exo, laissant des extrémités simple-brin 3'. Ces extrémités sont utilisées par Bet qui catalyse l'appariement des molécules homologues. B) Modèle d'invasion d'une molécule double-brin. Une seule des deux molécules homologues subit une cassure double-brin, dont l'extrémité 5' sera dégradée. Bet, aidée de RecA, catalyse l'invasion de la molécule double intacte et cause la formation d'une boucle de déplacement. Dans les deux modèles, Gam protège l'ADN double-brin linéaire de la dégradation par RecBCD. Inspiré de (Kuzminov, 1999) et (Poteete, 2001).

Des résultats publiés par Rybalchenko *et al* (2004) suggèrent que Bet est capable de procéder à l'invasion et à l'échange de brins d'ADN *in vitro* quand certains critères sont rencontrés. Notamment, ils ont observé que l'invasion catalysée par Bet se fait préférentiellement par

l'échange de bases A·T et qu'ainsi, un contenu élevé en G+C désavantage fortement la formation de boucles de déplacement. Cette préférence pour l'échange de bases A·T est également connue pour le mécanisme d'invasion d'autres recombinaisons, comme RecA et RecT chez *E. coli*, Rad51, Dmc1 et Rad52 chez l'humain et Rad52 chez la levure (Rybalchenko *et al.*, 2004). Cependant, il n'a pas été démontré que Bet catalyse des réactions d'invasion *in vivo* et il est pour l'instant accepté dans la littérature que Bet catalyse l'invasion d'une molécule d'ADN double-brin d'une manière strictement RecA-dépendante (Figure 4B). En contradiction complète avec cette idée, certains résultats suggèrent que le système Red puisse catalyser l'invasion dans un contexte *recA*⁻ (Muyers *et al.*, 1999).

1.3.4.3. Invasion d'une fourche de réplication

Des résultats plus récents viennent néanmoins éclaircir cette problématique. Un troisième mécanisme, l'invasion d'une fourche de réplication, a été proposé par plusieurs auteurs (Maresca *et al.*, 2010 ; Mosberg *et al.*, 2010 ; Poteete, 2008). Poteete (2008) a été le premier à suggérer que la recombinaison chez le bactériophage λ impliquerait la réplication. Il a proposé que l'extrémité 3' simple-brin générée par Exo s'insère dans la fourche de réplication et soit appariée à sa séquence homologue sur le brin tardif («lagging strand»). La réplication serait alors poursuivie en utilisant le nouveau brin comme matrice.

Des résultats plus récents ont renforcé la théorie que la réplication soit nécessaire à la recombinaison par Red (Maresca *et al.*, 2010 ; Mosberg *et al.*, 2010). Cependant, ces résultats apportent une nuance importante au mécanisme proposé par Poteete (2008). En effet, Mosberg (2010) et Maresca *et al.* (2010) ont présenté indépendamment des résultats suggérant que Exo dégrade entièrement un des deux brins de la molécule à recombiner, et non pas seulement les extrémités 5' comme proposé antérieurement. Cette molécule simple-brin envahirait alors

une fourche de réplication en s'appariant au brin tardif, à la place d'un fragment d'Okazaki. La réplication poursuivrait ensuite son cours normal (Figure 5).

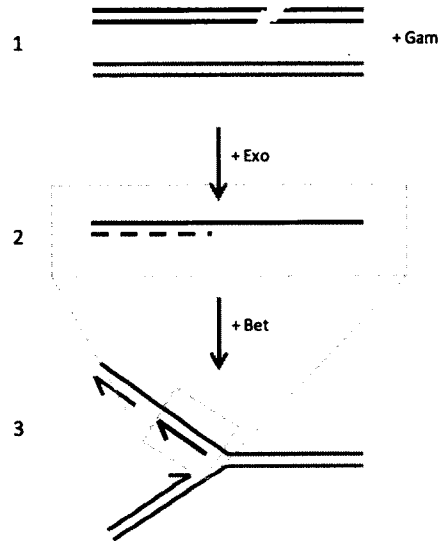


Figure 5. Invasion d'une fourche de réplication

1) Une molécule d'ADN double-brin est brisée et λ Gam protège les extrémités de la dégradation par RecBCD.
 2) λ Exo dégrade entièrement un des deux brins. 3) λ Bet catalyse l'appariement du brin non-digéré à une portion homologue située dans une fourche de réplication, à la place d'un fragment d'Okazaki.

La capacité du système Red à catalyser l'invasion d'une fourche de réplication fournit une explication cohérente à l'observation de Muyrers *et al* (1999) que la recombinaison double-brin peut être catalysée par Red en absence de RecA. La diversité des mécanismes de recombinaison proposés, appuyés de données expérimentales, suggère donc que le système Red soit capable de catalyser la recombinaison par plusieurs mécanismes, dont le choix dépend probablement des conditions cellulaires et de la disponibilité des différents partenaires : disponibilité de molécules homologues ayant des cassures double-brin, présence de RecA, présence de fourches de réplifications, taux de réplication, etc.

1.3.5. Homologues du système Red

Des analogues fonctionnels du système Red existent chez d'autres bactériophages. Bien que généralement différentes au niveau de leur séquences, les recombinases et exonucléases codées par ces systèmes sont fonctionnellement semblables à λ Bet et λ Exo (Datta *et al.*, 2008 ; Iyer *et al.*, 2002 ; Passy *et al.*, 1999 ; Poteete et Volkert, 1988). Le bactériophage P22 possède un système de recombinaison composé de quatre gènes, *arf* (accessory recombination function), *erf* (essential recombination function), *abc1* et *abc2* (anti-RecBCD) (Poteete et Fenton, 1993). La recombinase Erf est entre autres requise pour la circularisation du chromosome de P22 à son entrée dans la cellule (Iyer *et al.*, 2002). Bien que Bet et Erf soient évolutivement distantes, la délétion de Erf est facilement complétée par la présence de Bet (Iyer *et al.*, 2002). Les fonctions de recombinaisons codées par P22 peuvent également substituer celles de λ (Poteete et Fenton, 1993).

Le prophage Rac retrouvé dans le chromosome d'*E. coli* possède également des gènes de recombinaison (*recE/recT*). Normalement silencieux, ces gènes sont activés par les mutations dans le gène *sbcA* (Kuzminov, 1999). Le gène *recE* du prophage Rac code pour une exonucléase fonctionnellement apparentée à λ Exo, bien qu'elle n'ait aucune homologie de séquence au niveau génétique ou protéique. *In vitro* RecE est hautement processive et dégrade l'ADN double-brin à partir de cassures double-brin, dans la direction 5' → 3', laissant des extrémités 3' libres. *recT* code pour une recombinase capable de compléter l'activité de λ Bet. Comme Bet, RecT est capable d'apparier des molécules d'ADN simple-brin complémentaires (Kolodner *et al.*, 1994).

1.3.6. Rôle biologique du système Red

Bien que l'absence des gènes *bet* et *exo* n'affecte pas significativement le cycle lytique ou la lysogénie de λ , le système Red joue un rôle essentiel dans la biologie du phage λ , tant en ce qui a trait à la réparation efficace des cassures double-brin qu'à la stimulation de la réplication. Puisque le prophage λ est activé en présence d'agents causant des dommages à l'ADN, il est fort probable que la réplication non-passive de λ se produise dans un contexte où les dommages causés à l'ADN sont très nombreux, ce qui justifierait la présence d'un système de réparation de l'ADN (Poteete, 2001). De plus, son déplacement de cellule en cellule soumet le génome du phage à l'action des systèmes de restriction/modification. Par ailleurs, la recombinaison catalysée par Red stimule la réplication de l'ADN de λ (Poteete, 2001 ; Wu *et al.*, 2006). La recombinaison pourrait favoriser la formation de longs concatémères, le substrat préférentiel pour l'empaquetage du génome du phage dans sa capsid, notamment en permettant le changement du mode de réplication θ au mode de réplication σ (Kuzminov, 1999 ; Szczepanska, 2009 ; Weigel et Seitz, 2006).

1.3.7. Le système Red comme outil de biologie moléculaire

La recombinaison homologue *in vivo* est un moyen précis et versatile de modifier le chromosome bactérien. La découverte que le système Red est capable de recombiner de l'ADN linéaire avec le chromosome suite à une transformation a mené à l'élaboration de plusieurs techniques de biologie moléculaire et l'optimisation de ces techniques a produit la majorité des connaissances que nous possédons aujourd'hui sur les mécanismes de recombinaison du système Red. L'avantage principal d'utiliser le système de recombinaison de λ pour le génie génétique est qu'il recombine efficacement des séquences homologues dont la longueur est relativement petite (35 à 50 paires de bases) (Court *et al.*, 2002). Les propriétés intéressantes du système Red permettent de l'utiliser pour plusieurs applications, autant pour

l'étude de λ en soi (Oppenheim *et al.*, 2004), la modification de plasmides (Thomason *et al.*, 2007), la construction de fusions transcriptionnelles et traductionnelles (Ellermeier *et al.*, 2002 ; Gerlach *et al.*, 2007) et la délétion de gènes chromosomiques (Datsenko et Wanner, 2000). Ces techniques sont principalement utilisées chez *E. coli* (Datsenko et Wanner, 2000) et d'autres bactéries à Gram négatif comme *Salmonella* (Husseiny et Hensel, 2005) et *Yersinia* (Derbise *et al.*, 2003) et plus récemment, chez *Vibrio cholerae* (Yamamoto *et al.*, 2009).

1.4. RÉPONSE SOS

Outre son rôle dans la recombinaison, RecA est aussi responsable de l'induction de la réponse SOS en cas de dommages causés à l'ADN. La réponse SOS est une série de réactions concertées et hautement organisées qui permettent à une cellule de survivre à l'inhibition de la synthèse de son ADN (Kuzminov, 1999), par l'expression de plus d'une quarantaine de gènes impliqués dans l'inhibition de la division cellulaire, la réplication «error-prone», la synthèse translésion, la recombinaison et la réparation des bris simple-brin par excision (Matic *et al.*, 2004). Deux protéines sont au centre de la régulation de la réponse SOS : RecA et LexA. LexA est le répresseur principal du régulon SOS et contrôle l'expression des gènes du régulon par sa liaison à des boîtes spécifiques (boîtes SOS) situées en amont de leurs promoteurs (Butala *et al.*, 2009 ; Erill *et al.*, 2007 ; Matic *et al.*, 2004). Le site consensus des boîtes SOS est un palindrome (5'CTGT-N₈-ACAG'3) conservé chez la majorité des bactéries à Gram négatif. RecA agit comme senseur des dommages causés à l'ADN. La liaison de RecA à l'ADN simple-brin induit un état «activé» chez RecA, qui devient alors apte à promouvoir l'activité autoprotéolytique du répresseur LexA (Erill *et al.*, 2007).

En temps normal, le répresseur est actif et les gènes qu'il contrôle ne sont pas exprimés ou exprimés à un faible taux basal. L'ordre dans lequel les gènes du régulon SOS sont exprimés dépend de la force des boîtes SOS, de leur position relative au promoteur et de la force des

différents promoteurs (Butala *et al.*, 2009). La durée de l'expression des gènes dépend de la persistance d'ADN simple-brin dans la cellule (Matic *et al.*, 2004). Ainsi, certains gènes qui sont moins fortement réprimés seront activés même en réponse à des dommages mineurs (Butala *et al.*, 2009). Les premiers gènes à être exprimés sont impliqués dans la réparation des cassures simple-brin par excision (*uvrABC* : NER), dans la réparation par recombinaison (*ruvAB*), dans la réplication «error-prone» (*polB* et *dinB* qui codent respectivement pour PolIII et PolIV) et dans l'inhibition du clivage de *umuD* en *umuD'* (*dinI*) (Janion, 2008). *lexA* est également exprimé au même moment. Si l'expression des gènes de réparation des cassures simple-brin ne suffit pas à diminuer la quantité d'ADN simple-brin dans la cellule, les gènes impliqués dans la recombinaison (*recA* et *recN*) sont induits. Si les dommages sont si importants que même la recombinaison ne suffit pas, *umuCD* et *sulA* sont exprimés, ce qui permet la synthèse d'ADN translésion (*umuCD*) et l'arrêt de la division cellulaire (*sulA*). Finalement, en tout dernier recours, il y a induction de prophages et de plasmides codant pour la production de colicine, qui causent la lyse de la cellule, permettant de préserver les nutriments pour les cellules viables (Kuzminov, 1999).

Outre LexA, l'autoprotéolyse de plusieurs autres répresseurs est catalysée par RecA activée, notamment le répresseur CI du bactériophage λ (Oppenheim *et al.*, 2005) et le répresseur SetR des éléments intégratifs et conjugatifs de la famille SXT/R391 (Voir section 1.5.3.3.) (Beaber *et al.*, 2004).

1.5 ELEMENTS INTEGRATIFS ET CONJUGATIFS

Le transfert horizontal de matériel génétique entre les bactéries participe grandement à la plasticité des génomes bactériens. En plus de participer à l'évolution des génomes, les éléments génétiques mobiles bactériens sont d'importants vecteurs de dissémination de gènes

de résistance aux antibiotiques. Les intégrons, les plasmides, les transposons et les éléments intégratifs et conjugatifs (ICEs) n'en sont que quelques exemples. Les ICEs se transfèrent par conjugaison selon un mécanisme similaire à celui emprunté par les plasmides conjugatifs (voir section 1.1.3.), mais, au contraire de ces derniers, ils s'intègrent dans le chromosome de leur hôte au lieu de se maintenir sous la forme d'une molécule extrachromosomique circulaire et répliquative. Une conséquence directe de leur intégration dans le chromosome est leur transfert vertical lors de la division cellulaire. Lorsque certaines conditions environnementales sont respectées, ces éléments s'excisent du chromosome, généralement par recombinaison site-spécifique, afin de former un intermédiaire circulaire non-répliquatif qui servira de substrat pour leur transfert conjugal. À leur arrivée dans la cellule réceptrice ils reprennent une forme circulaire qui servira à leur intégration dans le chromosome (Burrus *et al.*, 2002a ; Wozniak et Waldor, 2010).

Certains ICEs se transfèrent sous forme d'ADN simple-brin, d'autre sous forme d'ADN double-brin, possèdent une haute ou faible spécificité d'intégration et des recombinases de type sérine ou tyrosine (Burrus *et al.*, 2002a). Cette diversité de caractéristiques a mené à une nomenclature très hétérogène de ce type d'éléments, incluant les plasmides intégratifs, les transposons conjugatifs, les îlots génomiques et plusieurs éléments non classifiés. Le terme ICE qui a été introduit par Burrus *et al.* (2002) pour tenter de résoudre ce problème de nomenclature inclut les éléments génétiques bactériens qui s'excisent du chromosome par recombinaison site-spécifique, transfèrent leur forme circulaire excisée par conjugaison et s'intègrent par recombinaison entre un site spécifique de l'intermédiaire circulaire et le chromosome de l'hôte. Le terme s'est par la suite élargi pour inclure tous les éléments intégratifs transmissibles par conjugaison, sans accorder d'importance au mécanisme d'intégration ou de conjugaison. Ainsi, le terme ICE inclut également des éléments tels CTnDOT ou Tn916 (Tableau 1) qui s'intègrent avec un minimum de spécificité de séquence et sont donc capables de transposition autant que de transfert intercellulaire. Bien que les facteurs qui déterminent la diversité des hôtes vers lesquels les ICEs peuvent se transférer n'ont pas été

étudiés, il est connu que la majorité des ICEs étudiés à ce jour possèdent un très large spectre d'hôtes (Wozniak et Waldor, 2010).

Tableau 1. Exemples d'éléments intégratifs et conjugatifs.

ICE	Hôte	Taille (kb)	Site d'intégration	Fonctions accessoires ^a
SXT	<i>Vibrio cholerae</i>	99.5	<i>prfC</i>	Su ^R , Tm ^R , Cm ^R , Sm ^R
R391	<i>Providencia rettgeri</i>	89	<i>prfC</i>	Kn ^R , Hg ^R
ICEBs1	<i>Bacillus subtilis</i>	20	gène tRNA ^{Leu}	NI
PAPI-1	<i>Pseudomonas aeruginosa</i>	108	gène tRNA ^{Lys}	Facteurs de virulence et formation de biofilm
ICEclc	<i>Pseudomonas knackmussii</i>	105	gène tRNA ^{Gly}	Dégradation d'acide 3- chlorobenzoïque
ICEHin1056	<i>Haemophilus influenzae</i>	49.4	gène tRNA ^{Leu}	Tet ^R , Cm ^R , Amp ^R
Tn916	<i>Enterococcus faecalis</i>	18	régions A-T riches	Tet ^R
pSAM2	<i>Streptomyces ambofaciens</i>	10.9	gène tRNA ^{Pro}	NI
CTnDOT	<i>Bacteroides</i> spp.	65	GTANNTTTTGC	Tet ^R , Erm ^R
TnGBS2	<i>Streptococcus agalactiae</i>	33.5	Régions intergéniques en amont des promoteurs σ^A	NI
ICEMISym	<i>Mesorhizobium loti</i>	502	gène tRNA ^{Phe}	Symbiose (nodulation et fixation d'azote)

^a Su^R, résistance au sulfaméthoxazole; Tm^R, résistance à la triméthoprime; Cm^R, résistance au chloramphénicol; Sm^R, résistance à la streptomycine; Kn^R, résistance à la kanamycine; Hg^R, résistance au mercure; Tet^R, résistance à la tétracycline; Amp^R, résistance à l'ampicilline; NI, non indentifié. Inspiré de Wozniak et Waldor (2010).

Les ICEs ont une structure typiquement modulaire, comprenant généralement trois modules distincts codants pour les différentes fonctions d'intégration/excision, transfert conjugatif et régulation. La majorité de ces modules sont également associés avec les fonctions codées par d'autres éléments mobiles, tels les plasmides conjugatifs, les phages tempérés et les transposons. Bien que tous les ICEs partagent une structure similaire, les mécanismes spécifiques codés par les différents modules peuvent grandement varier d'un ICE à l'autre. En

plus de ces fonctions principales, la majorité des ICEs codent pour des fonctions accessoires conférant un avantage évolutif à leur hôte (i.e. résistance aux antibiotiques, résistance aux métaux lourds, voies de dégradation de composés toxiques, facteurs de virulence, formation de biofilm) (Wozniak et Waldor, 2010).

1.5.1. Le choléra et la découverte de SXT

La majorité des ICEs étudiés à ce jour ont été identifiés dans des souches de *Vibrio cholerae*, l'agent causal du choléra, ainsi que dans des espèces bactériennes qui lui sont apparentées. Le choléra est une maladie infectieuse sévère qui est contractée par l'ingestion d'eau et d'aliments contaminés par *V. cholerae*. Cette maladie est épidémique dans les régions du globe où l'accès à l'eau potable est limité et où les conditions sanitaires inappropriées favorisent la dissémination du pathogène dans les sources d'eau de consommation. Le choléra est caractérisé par une diarrhée profuse qui provoque une perte de liquide fulgurante, causant la déshydratation sévère des patients et pouvant mener à la mort en moins de 24 heures. Bien que la simple réhydratation des patients soit le traitement le plus efficace, l'utilisation massive et abusive d'antibiotiques afin de traiter et prévenir les infections à *V. cholerae* a permis l'émergence de souches multirésistantes (Kaper *et al.*, 1995).

Les nombreuses épidémies de cholera qui ont ravagé la majeure partie des deux siècles derniers ont été causées par le sérotype O1 de *V. cholerae*, jusqu'à l'émergence au début des années 1990 du sérotype O139 (Kaper *et al.*, 1995). L'analyse du patron particulier de résistances aux antibiotiques de ce nouveau sérotype a mené à la découverte de l'élément intégratif et conjugatif SXT, qui confère la résistance aux antibiotiques triméthoprim, sulfaméthoxazole, streptomycine et chloramphénicol (Waldor et Mekalanos, 1996).

1.5.2. La famille SXT/R391

La découverte de l'ICE SXT a marqué le début de deux décennies d'étude extensive de ce nouveau genre d'élément génétique mobile, menant à la découverte de nombreux autres ICEs apparentés ou non à SXT, et également à une nouvelle analyse de nombreux éléments qui avaient à tort été classifiés comme étant des plasmides ou des transposons conjugatifs. L'une des conséquences majeures de cette réanalyse a été la reclassification de tous les plasmides appelés «facteurs R», pour «facteurs de résistance», appartenant au groupe d'incompatibilité J (Coetzee *et al.*, 1972 ; Hedges, 1974 ; Hedges, 1975), tels que R391 de *Providencia rettgeri* (avant connu sous le nom de *Proteus rettgeri*), comme étant des ICEs apparentés à SXT, et la création d'une nouvelle famille d'éléments génétiques mobiles, la famille SXT/R391 (Tableau 2, Annexe 1) (Beaber *et al.*, 2002a ; Burrus *et al.*, 2006a ; Hochhut *et al.*, 2001a). Ces éléments sont composés d'un squelette conservé de gènes codant pour leurs fonctions essentielles d'intégration, d'excision, de transfert conjugatif et de régulation, et sont tous retrouvés intégrés de manière site-spécifique dans l'extrémité 5' du gène *prfC* de leur hôte, codant pour le facteur 3 de relargage des chaînes peptidiques (RF3) (Hochhut et Waldor, 1999).

La famille SXT/R391 contient à ce jour plus de 50 éléments, découverts dans des isolats d'origines clinique et environnementale d'une grande diversité d'espèces bactériennes provenant de divers endroits du globe et isolés au cours des 40 dernières années (Tableau 2, Annexe 1) (Burrus *et al.*, 2006a ; Wozniak *et al.*, 2009). Les ICEs nouvellement identifiés sont classifiés comme membres de cette famille par la présence du squelette de gènes conservés, la présence d'une intégrase conservée et par leur intégration site-spécifique dans l'extrémité 5' de *prfC* (Burrus *et al.*, 2006a). Outre leur rôle majeur dans la dissémination de gènes de résistance aux antibiotiques, les ICEs de la famille SXT/R391 sont maintenant reconnus pour leur participation à la plasticité du génome de leurs hôtes et dans leur capacité à s'adapter à leur environnement.

1.5.3. Squelette minimal fonctionnel

La comparaison de la séquence complète de 13 éléments appartenant à la famille SXT/R391 a permis de mettre en évidence que ces éléments partagent un squelette de gènes conservés codant pour leurs fonctions principales d'intégration, d'excision, de transfert conjugatif et de régulation (Figure 6) (Wozniak *et al.*, 2009). Ce squelette comprend 52 gènes dont seulement 25, totalisant 29.7 kpb, sont nécessaires à ces fonctions. Les gènes codant pour ces fonctions sont organisés en modules fonctionnels qui gouvernent l'intégration/excision, le transfert conjugatif (mobilisation et formation de l'appareil conjugatif) et la régulation.

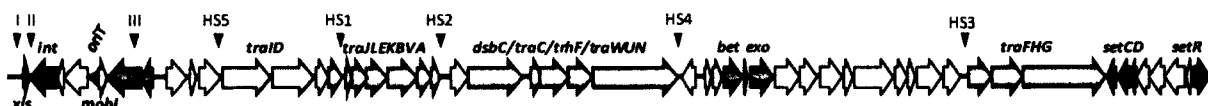


Figure 6. Squelette de gènes conservés des ICEs de la famille SXT/R391.

Les différents modules fonctionnels sont représentés avec un code de couleur. Gris, intégration/excision; jaune, mobilisation; bleu, recombinaison et réparation; orange, formation du pore de conjugaison; aqua, exclusion; vert, activation transcriptionnelle; rouge, répression transcriptionnelle; blanc, fonction inconnue. La position de régions variables présentes à différents endroits du squelette est indiquée par des pointes noires descendantes.

1.5.3.1 Intégration site-spécifique

L'intégration et l'excision sont catalysées par une recombinase site-spécifique de type P4, Int, et un facteur de directionnalité de la recombinaison, Xis, dont les gènes sont situés à l'extrémité 5' des éléments (Figure 6). L'intégration dans le chromosome est le résultat de la recombinaison site-spécifique catalysée par l'intégrase Int, entre deux séquences identiques de 17 paires de bases (pb), l'une retrouvée sur la molécule circulaire extrachromosomique de l'ICE (*attP*) et l'autre retrouvée au niveau du chromosome bactérien (*attB*) (Figure 7). Bien

que l'intégration de SXT interrompe la séquence 5' de *prfC*, elle fournit une nouvelle séquence promotrice et une nouvelle séquence 5' codante, permettant l'expression d'une protéine RF3 fonctionnelle (Hochhut et Waldor, 1999).

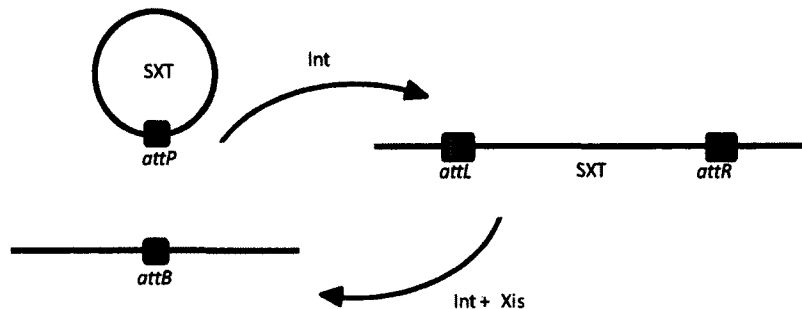


Figure 7. Intégration et excision site-spécifique de SXT.

La recombinase site-spécifique Int permet l'intégration de SXT en recombinant le site *attP* de la molécule circulaire de l'ICE avec le site *attB* retrouvé sur le chromosome. Int, aidée de Xis, catalyse la recombinaison entre *attL* et *attR*, ce qui permet l'excision de SXT du chromosome.

La réaction inverse, menant à l'excision de l'ICE du chromosome, résulte de la recombinaison site-spécifique entre les sites *attL* et *attR* retrouvés de part et d'autre de l'élément intégré et est catalysée par Int, assistée de Xis (Figure 7) (Burrus et Waldor, 2003). Par analogie avec les autres systèmes conjugatifs, il est supposé que la molécule circulaire excisée résultante sera clivée à son origine de transfert (*oriT*) et transférée sous forme d'ADN simple-brin vers une cellule réceptrice par le pore de conjugaison codé par l'ICE. Les molécules simple-brin générées dans les cellules donneuses et réceptrices seraient dupliquées et ensuite intégrées par recombinaison site-spécifique dans chacun des chromosomes.

1.5.3.2. Transfert conjugatif

En aval du module d'intégration/excision se trouve le module de mobilisation, qui est séparé en deux loci par *rumBA* (impliqués dans la réparation des dommages par les UV) et par les cadres de lectures ouverts (ORF) *s024*, *s025* et *s026* de fonction inconnue (Figure 6). Le premier locus du module de mobilisation contient la séquence reconnue et clivée par la relaxase putative TraI, l'origine de transfert (*oriT*) (Ceccarelli *et al.*, 2008). Cette séquence de 299 pb est située dans la région intergénique entre *s003* et *mobI* et est commune chez tous les membres de la famille. À l'instar des *oriT* des plasmides conjugatifs, l'*oriT* des ICEs SXT/R391 contient cinq répétitions imparfaites directes et inversées qui sont présumées être impliquées dans la reconnaissance spécifique de ce locus par la relaxase, permettant le clivage d'un des deux brins de la forme excisée de l'ICE. Le gène *mobI* retrouvé immédiatement en aval de l'origine de transfert est essentiel pour le transfert conjugatif de SXT. Bien que le rôle précis de *mobI* dans le transfert ne soit pas connu, il ne semble pas impliqué dans la formation du pore de conjugaison puisqu'il n'est pas requis pour la mobilisation en *trans* par SXT du plasmide pCLODF13 (Ceccarelli *et al.*, 2008). Le deuxième locus du module de mobilisation contient la relaxase putative *traI*, ainsi que *traD* (un facteur de couplage putatif) et *traJ*, un gène codant pour une protéine ayant une faible homologie avec la protéine TraJ du plasmide F (Wozniak *et al.*, 2009).

Le système de sécrétion de type IV qui permet la translocation de l'ADN de l'ICE dans une cellule réceptrice est codé par trois opérons : (i) *traLEKBVA*, (ii) *dsbC/traC/trhF/traWUN* et (iii) *traFHG*, qui contiennent tous les gènes nécessaires à la formation du pilus de conjugaison et de la machinerie conjugative, ainsi qu'à la formation et à la stabilisation du couple cellule-cellule nécessaire au transfert (Figure 6) (Wozniak *et al.*, 2009). Bien qu'à ce jour aucune fonction précise n'ait été attribuée au gène *s063*, situé entre *traN* et *traF*, celui-ci est inclus dans le squelette minimal fonctionnel car sa deletion résulte en une diminution d'environ 100 fois de la fréquence de transfert de SXT (Wozniak *et al.*, 2009).

1.5.3.3. Régulation du transfert

L'expression concertée des gènes requis pour les processus d'intégration/excision et transfert conjugatif est assurée par trois régulateurs codés par l'ICE : le répresseur principal *setR*, ainsi que les activateurs de transcription *setC* et *setD* (Beaber *et al.*, 2004 ; Beaber et Waldor, 2004). Le locus codant pour ces fonctions de régulation est situé à l'extrémité 3' du squelette conservé (Figure 6). SetR, à l'instar de CI du bactériophage λ , contient à la fois un domaine de liaison à l'ADN hélice-tour-hélice et un motif protéase (Beaber *et al.*, 2004). SetR réprime l'expression de *setCD* à partir d'un promoteur situé en amont de *s086* (Figure 8) (Beaber et Waldor, 2004). Comme c'est le cas pour λ , la répression maintenue par SetR est levée par l'induction du système SOS de l'hôte, ce qui favorise le transfert des ICEs SXT/R391 en réponse à des agents endommageant l'ADN, tels certains antibiotiques induisant la réponse SOS (Beaber *et al.*, 2004). Suite à la stimulation de l'activité coprotéase de RecA par les dommages causés à l'ADN, la protéine RecA activée semble catalyser l'autoprotéolyse et l'inactivation de SetR (Beaber *et al.*, 2004). La diminution de la concentration intracellulaire de SetR lève la répression maintenue sur *setC* et *setD*, causant l'induction de l'expression de *int*, de *xis*, des gènes impliqués dans la mobilisation et des trois opérons *tra* décrits plus haut, vraisemblablement menant à l'expression simultanée de tous les gènes nécessaires au transfert efficace de SXT vers une nouvelle cellule hôte (Figure 8) (Beaber *et al.*, 2002b ; Burrus et Waldor, 2003).

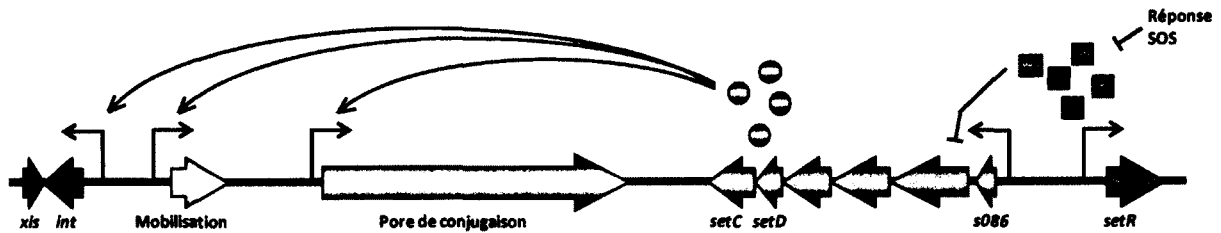


Figure 8. Régulation de SXT.

Le répresseur SetR réprime l'expression des activateurs de transcription SetC et SetD en liant un promoteur situé en amont de *s086*. L'induction de la réponse SOS lève la répression maintenue par SetR, permettant l'expression de SetCD et l'activation des gènes d'intégration/excision (gris), de mobilisation (jaune) et de formation du pore de conjugaison (orange). Inspiré de Wozniak et Waldor (2010).

1.5.4. Les ICE SXT/R391 sont apparentés aux plasmides de la famille IncA/C

Récemment, Wozniak *et al.* (2009) ont démontré que les gènes de transfert des ICE de la famille SXT/R391 sont également retrouvés chez les plasmides de la famille IncA/C. Ces plasmides conjugatifs codent pour de multiples résistances aux antibiotiques sont largement distribués chez *Salmonella* (Lindsey *et al.*, 2009) et ont été identifiés chez *Yersinia pestis* (Galimand *et al.*, 1997), l'agent causal de la peste bubonique, ainsi que chez *Vibrio cholerae* et dans des γ -protéobactéries apparentées (Fricke *et al.*, 2009 ; Pan *et al.*, 2008 ; Welch *et al.*, 2007). Outre le fait qu'ils soient leurs plus proches parents, les gènes de transfert retrouvés chez les plasmides IncA/C sont en parfaite synténie avec les 4 loci codant la machinerie conjugative des ICE SXT/R391 (Figure 9).

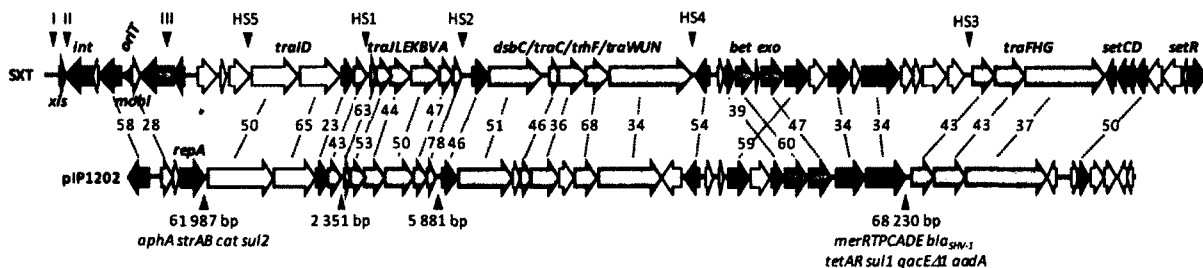


Figure 9. Comparaison du squelette de gènes conservés de SXT et du plasmide conjugatifs pIP1202 du groupe d'incompatibilité IncA/C.

Les différents modules fonctionnels sont représentés avec un code de couleur. Gris, intégration/excision; jaune, mobilisation; bleu, recombinaison et réparation; orange, formation du pore de conjugaison; aqua, exclusion; vert, activation transcriptionnelle; rouge, répression transcriptionnelle; violet, réplication; noir, gènes de fonction inconnue homologues, blanc, gènes de fonction inconnue sans homologues. Les lignes pointillées indiquent les gènes homologues entre SXT et pIP1202 et les nombres indiquent le % d'identité entre les protéines codées par les gènes homologues. La position des régions variables et des hotspots (I à III et HS1 à HS5) chez SXT sont indiquées par des pointes noires descendantes. La taille et le contenu des régions variables de pIP1202 sont indiqués et positionnés dans le squelette par des flèches noires montantes. Les résistances codées par pIP1202 sont indiquées : *aphA*, kanamycine; *strAB*, streptomycine; *cat*, chloramphénicol; *sul1* et *sul2*, sulfonamides; *merRTPCADE*, mercure; *bla_{SHV-1}*, ampicilline; *tetAR*, tetracycline; *qacEΔ1*, ammonium quaternaire; *aadA*, spectinomycine.

Cependant, l'absence du module de régulation comprenant les gènes *setR* et *setCD* des plasmides IncA/C suggère que leur transfert soit régulé différemment. De plus, au lieu de contenir les gènes nécessaires à l'intégration et l'excision, ces plasmides contiennent un réplicon de type RepA. En plus des gènes de transfert connus, dix gènes additionnels de fonction inconnue sont retrouvés dans les squelettes des ICE SXT/R391 et des plasmides IncA/C, suggérant qu'ils ont un rôle important à jouer dans la biologie de ces deux familles distinctes d'éléments mobiles (Wozniak *et al.*, 2009).

Comme c'est le cas pour les ICE SXT/R391, les plasmides IncA/C contiennent des régions variables dispersées dans leur squelette de gènes conservés. Ces régions variables sont

d'ailleurs retrouvées dans les mêmes positions relatives au squelette dans les deux familles d'éléments : en amont de *traI* et de *traL*, en aval de *traA* et en amont de *traF* (Wozniak *et al.*, 2009). Cependant, le contenu des régions variables des IncA/C diffère complètement de celui des régions variables des ICE SXT/R391, suggérant que l'ancêtre commun de ces deux familles aurait divergé avant l'acquisition des séquences variables.

1.5.5. Régions variables et diversité des fonctions conférées

Tous les membres de la famille SXT/R391 qui ont été séquencés à ce jour possèdent la même structure, composée du squelette de gènes conservés dans lequel se trouvent 5 loci contenant des segments d'ADN variable de différentes tailles (Figure 10). Ces 5 loci, nommés HS1 à HS5 pour «hotspot», sont retrouvés dans des régions intergéniques du squelette où leur acquisition a laissé intactes toutes les fonctions essentielles. Certains éléments contiennent de l'ADN variable additionnel situé dans des régions variables (VRI à VRIV). Tel est le cas pour SXT et d'autres membres de cette famille qui contiennent un élément de type ISCR2 inséré dans *rumB* (Figure 10). Les éléments ISCR2 sont apparentés aux éléments transposables de type IS91 qui sont fréquemment associés avec la résistance aux sulfonamides (Toleman *et al.*, 2006).

Les séquences retrouvées dans un «hotspot» donné sont rarement restreintes à un seul élément : elles sont généralement retrouvées dans au moins un autre ICE de la famille, dont le contenu des autres «hotspot» sera généralement complètement différent. À titre d'exemple, ICE*VchMex1* et ICE*VfInd1* contiennent le même ADN variable dans HS3 (Figure 10), mais le contenu de leurs autres «hotspot» est complètement différent. Une observation similaire peut être faite pour la majorité des ICE de cette famille.

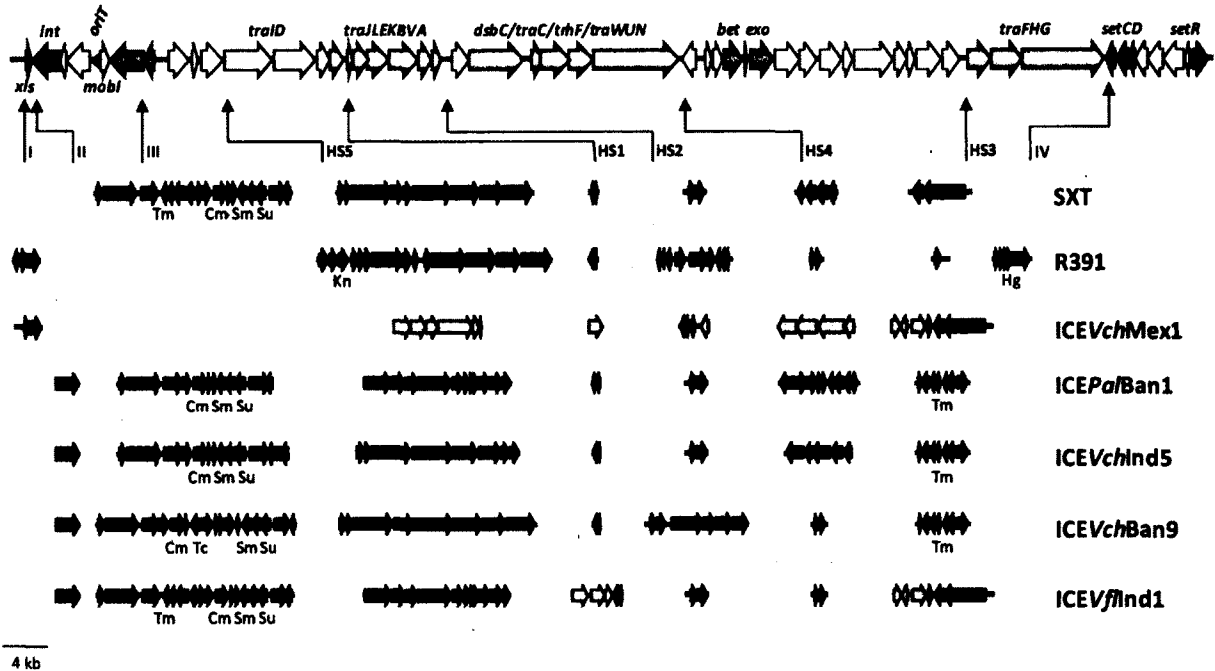


Figure 10. Régions variables des ICEs de la famille SXT/R391.

Les différents modules fonctionnels du squelette sont représentés suivant le code de couleurs utilisé dans la figure 6. Les régions variables de certains ICEs de la famille SXT/R391 sont représentées sous le squelette de gènes conservés. Les couleurs indiquent les régions variables identiques. Les résistances sont indiquées : Tm, triméthoprime; Cm, chloramphénicol; Sm, streptomycine; Su, sulfaméthoxazole; Kn, kanamycine; Hg, mercure; Tc, tétracycline.

Bien que la majorité des gènes retrouvés dans les régions variables n'aient pas d'homologue connu, certains d'entre eux codent pour des fonctions qui profitent certainement à leur hôte bactérien. La plupart des éléments appartenant à la famille SXT/R391 confèrent la résistance à de multiples antibiotiques, principalement par le biais de gènes retrouvés dans d'autres éléments mobiles associés à l'ICE, tels que les éléments ISCR2 (e.g. SXT, ICEPaBan1, et ICEVchBan9) ou des intégrons de classe IV (e.g. ICEVchInd5 and ICEVchBan9). Cependant, un certain nombre d'ICEs SXT/R391 contiennent des gènes variables offrant d'autres avantages possibles à leur hôte, tels ICEVchMex1, ICEVflInd1 et ICEVchMoz3 qui contiennent des gènes impliqués dans la synthèse du c-di-GMP, un second messager important dans la biologie de *V. cholerae* (Bordeleau *et al.*, 2009). Certains éléments tels que SXT,

ICE*Pal*Ban1, ICE*Vch*Ind5 et ICE*VfI*Ind1 contiennent les gènes *mosA* et *mosT* dans le HS2, qui forment un système toxine/antitoxine qui semble être impliqué dans leur maintien intracellulaire (Wozniak et Waldor, 2009). (Pour une revue exhaustive du contenu des régions variables, veuillez vous référer à Wozniak *et al.*, 2009).

1.5.6. Système toxine/antitoxine

Les systèmes toxines/antitoxines (TA) sont formés d'une toxine stable et d'une antitoxine labile et fonctionnent de façon à inhiber la perte d'un élément extrachromosomique (pour une revue complète voir Hayes, 2003). Si un élément transportant la paire de gènes TA est perdu dû à une erreur dans sa réplication ou à un défaut dans son maintien la toxine instable sera rapidement éliminée du complexe TA et ne pourra être resynthétisée dû à l'absence de l'élément. La toxine stable sera libérée du complexe et interagira avec sa cible chez l'hôte, inhibant sa croissance ou induisant sa mort cellulaire prématurée.

Lorsqu'ils sont présents dans leur hôte sous leur forme intégrée, les ICEs SXT/R391 sont maintenus de façon stable et sont répliqués avec le chromosome de l'hôte. Cependant, sous les conditions qui induisent leur transfert, ils s'excisent et existent de façon transitoire sous la forme d'une molécule double-brin extrachromosomique non-répliquative (Wozniak et Waldor, 2010). Par contre, tout événement d'excision ne mène pas nécessairement à un événement de transfert, comme indiqué par la fréquence d'excision qui est de 100 fois supérieure à la fréquence de formation d'exconjugants (Burrus et Waldor, 2003 ; Wozniak *et al.*, 2009), créant ainsi un équilibre entre les formes intégrées et excisées des ICEs SXT/R391 dans une cellule donnée. Ce phénomène est une menace constante à la stabilité de l'élément, puisqu'il pourra être perdu si sa cellule hôte se divise au moment où l'ICE non-répliquatif est sous sa forme extrachromosomique. Pourtant, bien qu'il y ait 3.7 événements d'excision détectables par 100 cellules contenant SXT, seulement 3.2 événements de pertes par 10⁶ cellules sont

détectés, démontrant que la majorité des évènements d'excision ne mènent pas à la perte de SXT (Burrus et Waldor, 2003 ; Wozniak et Waldor, 2009). Ce phénomène pourrait être expliqué par la présence d'un système TA.

Les gènes *mosA* et *mosT* (pour «maintenance of SXT, Antitoxin and Toxin») qui sont codés par SXT et plusieurs autres membres de la famille augmentent la stabilité de ces éléments dans leur cellule hôte (Wozniak et Waldor, 2009). Ces deux gènes sont retrouvés dans le HS2 de SXT et agissent afin de prévenir la perte de SXT. *mosA* et *mosT* semblent former un opéron qui est maintenu réprimé par l'autorépression causée par MosA lorsque SXT est sous forme intégrée. Cependant, les facteurs responsables de l'excision de SXT, soit l'excisionase Xis et les activateurs transcriptionnels SetCD, lèvent la répression sur *mosAT* et permettent leur expression lorsque SXT est excisé du chromosome. *mosT* code pour une toxine qui inhibe sévèrement la croissance de *E. coli* lors de sa surexpression, mais son activité est neutralisée par la présence de *mosA*. Par analogie avec d'autres systèmes TA, MosA serait moins stable que MosT, ce qui mènerait à la mort des cellules n'ayant pas hérité SXT si la division cellulaire se produisait alors que l'élément est sous forme extrachromosomique, ou à l'inhibition de leur croissance jusqu'à ce que l'élément se réintègre.

Curieusement, plusieurs membres de la famille SXT/R391, tels R391 (Boltner *et al.*, 2002 ; Coetzee *et al.*, 1972), ICEPdaSpa1 (Osorio *et al.*, 2008) et ICEVchMex1, (Burrus *et al.*, 2006b) ne possèdent pas les gènes *mosAT*. Des études préliminaires suggèrent néanmoins que R391 n'est que légèrement moins stable que SXT, suggérant que les membres de cette famille qui ne contiennent pas *mosAT* possèdent peut-être d'autres gènes assurant leur maintien. Entre autres, R391 et ICEVchMex1 contiennent dans HS2 des gènes apparentés de façon lointaine au système TA HipAB d'*E. coli*. De plus, une autre étude a démontré que les gènes *s044* et *s045* situés dans le HS1 de SXT et ICEPalBan1 (Tableau 2, Annexe 1) codent pour un système TA fonctionnel. Lorsque surexprimée dans *E. coli*, la toxine codée par *s045* entraîne une diminution de 100 fois de la viabilité cellulaire et cet effet est annulé par la co-expression

de *s044* (Dziewit *et al.*, 2007). De plus, lorsque cloné dans un vecteur instable, le système *s044/s045* augmente la rétention du plasmide d'environ 30% après 30 générations en absence de sélection (Dziewit *et al.*, 2007). Cependant, la deletion de *s044/s045* de SXT n'a aucun impact négatif sur sa stabilité, contrairement à ce qui est observé pour la délétion de *mosA/mosT*, posant la question si ce système est réellement fonctionnel (Wozniak et Waldor, 2009).

1.5.7. Mobilisation de plasmides, d'îlots génomiques mobilisables et d'ADN chromosomique

1.5.7.1. Mobilisation de plasmides

Les ICEs de la famille SXT/R391 ont la capacité de promouvoir la mobilisation de plasmides mobilisables mais non conjugatifs, tels pCLODF13, en leur fournissant un pore de conjugaison (Hochhut *et al.*, 2000). Plus récemment, il a été démontré que ICE*PdaSpa1*, un ICE SXT/R391 dérivé du pathogène de poisson *Photobacterium damsela* subsp. *piscicidae*, peut promouvoir le transfert du plasmide de virulence pPHDP10 qui est retrouvé dans la même souche. Cependant, le mécanisme par lequel pPDHP10 est mobilisé par ICE*PdaSpa1* est entièrement différent du transfert de pCLODF13 par SXT, puisque pPDHP10 ne contient ni origine de transfert, ni gènes de mobilisation (Osorio *et al.*, 2008). Il a été démontré que la mobilisation de pPHDP10 passe par la formation RecA-indépendante d'un co-intégrat entre le plasmide et l'ICE. Par ailleurs, ce phénomène n'est pas site-spécifique puisque pPDHP10 a été retrouvé intégré dans plusieurs loci de ICE*PdaSpa1*. Bien que le mécanisme par lequel est formé le co-intégrat ne soit pas déterminé, cette structure est transférable par conjugaison vers un nouvel hôte, dans lequel le co-intégrat sera intégré de manière site-spécifique. Le plasmide demeure par la suite intégré de manière stable dans l'ICE et le co-intégrat est par la suite transféré à une fréquence comparable à celle de ICE*PdaSpa1* (Osorio *et al.*, 2008).

1.5.7.2. Mobilisation d'îlots génomiques mobilisables

Il a été démontré très récemment que les ICEs SXT/R391 catalysent le transfert d'une classe spécifique d'îlots génomiques mobilisables (MGI) retrouvés dans les génomes de souches de *Vibrio*, par le biais de la reconnaissance d'une origine de transfert (*oriT*) apparentée (Daccord *et al.*, 2010). Cette *oriT* de 299 bp partage plus de 63% d'identité avec *oriT*_{SXT} et contient les mêmes cinq répétitions imparfaites directes et inversées qui sont présumées participer dans la reconnaissance et le clivage de l'*oriT* par la relaxase de SXT.

Ces MGIs sont intégrés de manière site-spécifique dans l'extrémité 3' du gène *yicC*, codant pour une protéine non caractérisée qui est prédite pour être induite en conditions de stress, et sont bordés par des sites *attL* et *attR* (Figure 11A). En aval de *attL* se trouve une intégrase site-spécifique de type P4 qui catalyse l'intégration du MGI dans le chromosome de son hôte (Daccord *et al.*, 2010). Ces MGIs contiennent également un facteur de directionnalité de la recombinaison, appelé *rdfM*, qui est nécessaire, en plus de l'intégrase, à leur excision (Daccord *et al.*, 2012). La mobilisation des MGIs est catalysée par la présence d'un ICE de type SXT/R391 dans la même cellule, tant par la reconnaissance de *oriT*_{MGI} par la machinerie conjugative de l'ICE que par l'induction de l'expression de *int*_{MGI} et *rdfM* par les activateurs transcriptionnels SetCD de l'ICE (Figure 11A) (Daccord *et al.*, 2010 ; Daccord *et al.*, 2012). Au contraire de *rdfM*, *int*_{MGI} est exprimée de façon constitutive à un faible taux, permettant l'intégration du MGI dans une réceptrice sans nécessiter la présence des activateurs transcriptionnels SetCD, ce qui permet l'intégration d'un MGI dans une cellule réceptrice ne contenant pas un ICE SXT/R391 (Figure 11B) (Daccord *et al.*, 2012).

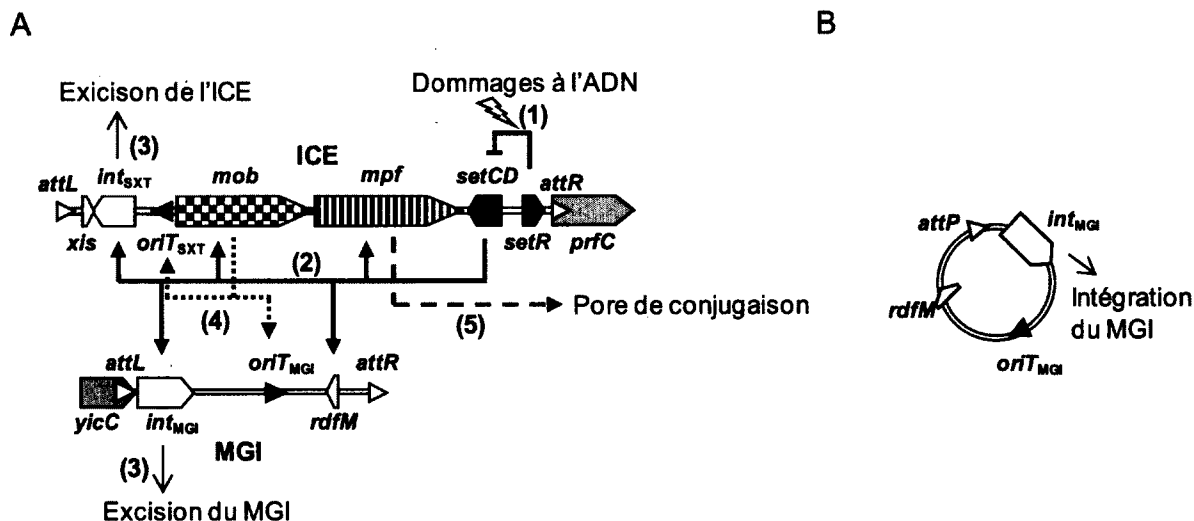


Figure 11. Mobilisation des MGIs par les ICEs SXT/R391.

A) Régulation du transfert des MGI par les ICEs de la famille SXT/R391. (1) Les agents endommageant l'ADN induisent la réponse SOS, levant la répression maintenue par SetR sur *setCD*. (2) SetCD active l'expression de *int_{SXT}*, de *int_{MGI}* et des fonctions de mobilisation (*mob*) et de conjugaison (*mpf*). (3) *int_{SXT}* et *int_{MGI}* catalysent l'excision de SXT et des MGIs, respectivement. (4) Les protéines produites par les gènes *mob* reconnaissent et clivent l'*oriT*. (5) L'expression des gènes *mpf* permet la formation de l'appareil conjugal au travers duquel le brin clivé de l'ICE ou du MGI sera transféré vers la cellule réceptrice. B) L'expression basale de l'intégrase des MGI permet leur intégration dans le chromosome de la cellule réceptrice sans l'aide d'un ICE SXT/R931.

1.5.7.3. Mobilisation d'ADN chromosomique

Les ICEs de la famille SXT/R391 sont connus pour permettre le transfert d'ADN chromosomique d'une façon similaire au transfert HFr catalysé par le plasmide F (Hochhut *et al.*, 2000). Puisque l'excision de SXT du chromosome n'est pas un prérequis à l'expression des fonctions de transfert, le transfert conjugal peut être initié à l'*oriT* d'un élément intégré et ainsi mener à la mobilisation de l'ADN localisé en 3' de *prfC* (Hochhut *et al.*, 2000). Suite au transfert, l'ADN chromosomique transféré pourra être intégré dans le chromosome de la cellule réceptrice par recombinaison homologue. Récemment, il a été démontré que les MGIs

mobilisés par les ICEs de la famille SXT/R391 peuvent catalyser le même type de transfert et permettre la mobilisation de jusqu'à 1 Mb d'ADN chromosomique situé en 5' de *yicC*, chez *E. coli* (Daccord *et al.*, 2010). Cette étude a également démontré que ICE*Vf*/IndI est capable de mobiliser jusqu'à 1.5 Mb d'ADN chromosomique, ce qui représente le triple de ce qui avait été précédemment publié pour SXT, suggérant que la capacité des ICEs à transférer de l'ADN chromosomique est beaucoup plus importante qu'initialement déterminée (Daccord *et al.*, 2010 ; Hochhut *et al.*, 2000).

1.5.8. Exclusion d'entrée

La majorité des éléments conjugatifs possèdent des systèmes d'exclusion qui fonctionnent de telle manière à éviter les transferts redondants entre cellules donneuses contenant le même élément et permet d'éviter les interactions compétitives entre des systèmes de répllication ou de partition trop similaires (Novick, 1987). Bien que SXT et R391 soient composés du même squelette de gènes conservés ils ne s'excluent pas ; ils excluent cependant l'entrée d'un deuxième élément qui leur est identique. Ainsi, SXT exclu l'entrée d'une deuxième copie de SXT, mais pas d'une copie de R391 (Marrero et Waldor, 2005). Le même phénomène est observé lorsqu'une cellule contient R391. Cette exclusion très spécifique est gouvernée par l'interaction entre des variants des protéines codées par *eex* et *traG*. La caractérisation génétique et fonctionnelle des gènes *eex* et *traG* appartenant à 21 ICEs SXT/R391 a mis en évidence qu'ils appartiennent à deux groupes : S et R (Tableau 2, Annexe 1) (Marrero et Waldor, 2007a). Ainsi, comme observé pour SXT et R391 (Marrero et Waldor, 2005), les ICEs appartenant au groupe d'exclusion S excluent l'entrée d'un autre ICE appartenant à ce groupe, mais pas d'un ICE appartenant au groupe R, et vice versa (Marrero et Waldor, 2007a).

Les protéines Eex des deux groupes d'exclusion partagent un haut degré de similarité (87% d'identité) dans les 86 premiers résidus de leur portion N-terminale. Cependant, leur extrémité

C-terminale est plus divergente (41% d'identité) et dicte leur spécificité d'exclusion (Marrero et Waldor, 2005). Des différences additionnelles subtiles dans la portion C-terminale des protéines EexS et EexR permettent de les sous-classifier en variants (Tableau 2, Annexe 1). Dans chaque ICE, le gène *traG* appartient au même groupe d'exclusion que le gène *eex*. Au contraire des protéines Eex, les protéines TraG sont très bien conservées (98% d'identité entre TraG_S et TraG_R) et seulement trois acides aminés permettent cette haute spécificité de reconnaissance (Marrero et Waldor, 2005). Les régions de TraG et Eex qui permettent l'exclusion semblent être cytoplasmique, suggérant que leur interaction se déroule dans le cytoplasme (Marrero et Waldor, 2007b). Cependant, il n'est pas clair comment des protéines cytoplasmiques de deux cellules en conjugaison puissent venir en contact. Il est possible qu'un fragment de TraG ou Eex soit transporté dans le partenaire de conjugaison, mais cela reste à déterminer. L'exclusion déterminée par les protéines TraG et Eex n'est pas parfaite puisqu'elle diminue le transfert de SXT vers une cellule SXT⁺ d'environ 100 fois (Marrero et Waldor, 2007a), un phénomène qui pourrait expliquer la formation de tandems de SXT par de multiples événements d'acquisition de l'élément dans une même cellule réceptrice (Burrus et Waldor, 2004).

1.5.9. Formation de tandems et d'hybrides SXT/R391

Lorsque les séquences de SXT (Beaber *et al.*, 2002b) et de R391 (Boltner *et al.*, 2002) furent initialement reportées, il apparaissait que les régions variables retrouvées dans différents endroits du squelette conservé étaient spécifiques à chaque élément. Cependant, au fur et à mesure que les séquences d'autres ICEs ont été obtenues il est devenu évident que ces régions sont en fait partagées par au moins deux éléments (Figure 10). Tel que discuté plus haut, certains ICEs de cette famille ne s'excluent pas, un phénomène qui permet la coexistence de deux éléments similaires dans la même cellule bactérienne. Cette caractéristique particulière, combinée avec l'intégration site-spécifique de tous les membres de cette famille dans le même

site chromosomique, permet la formation de tandems chromosomiques d'éléments semblables mais non identiques (Burrus et Waldor, 2004 ; Hochhut *et al.*, 2001a).

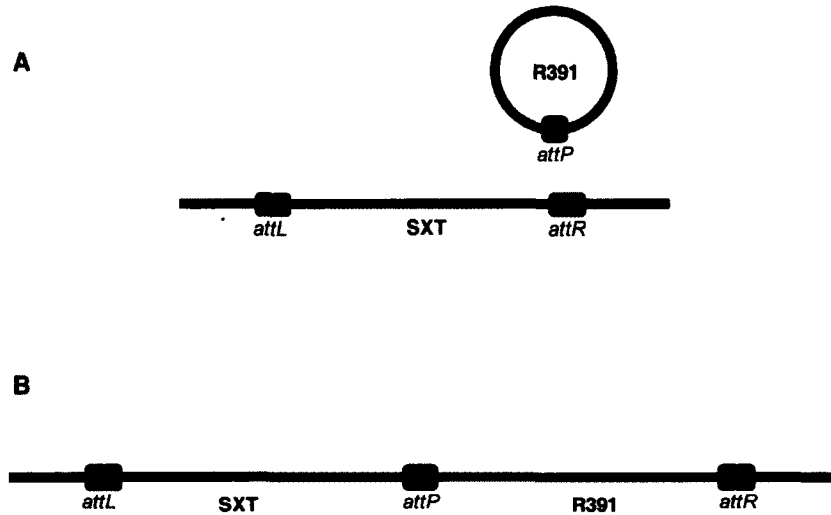


Figure 12. Formation de tandems SXT-R391 dans le chromosome.

A) R391 est transféré vers une cellule qui contient SXT intégré dans le site *attB* (*prfC*). L'intégrase catalyse la recombinaison entre *attP* de la molécule circulaire R391 et le site *attR* de SXT intégré (la recombinaison entre *attP* et *attL* est aussi possible). B) R391 et SXT se retrouvent intégrés en tandem dans le chromosome de leur cellule hôte. Inspiré de Hochhut *et al* (2001).

Ainsi, lorsque R391 entre dans une cellule contenant déjà SXT, le site *attP* de la molécule circulaire R391 recombine avec soit le site *attL* ou le site *attR* formés par l'intégration de SXT dans *prfC* (Figure 12) (Hochhut *et al.*, 2001a). Le même mécanisme se produit lorsque SXT entre dans une cellule contenant déjà R391. R391 s'intègre préférentiellement dans le site *attL* d'un élément SXT intégré, tandis que SXT semble n'avoir aucune préférence d'insertion, et s'intègre aussi fréquemment dans le site *attL* ou le site *attR* d'un élément R391 intégré. De plus, en absence de sélection, SXT et R391 sont préférentiellement perdus de la position droite du tandem (Hochhut *et al.*, 2001a). Ainsi, R391 et SXT peuvent coexister dans la même

cellule, fournissant un substrat plus que suffisant pour la recombinaison inter-ICE dû à leurs 47 kb d'ADN hautement conservé.

Dans certains cas, le transfert conjugatif entre une cellule donneuse contenant un tandem SXT/R391 et une cellule réceptrice ne contenant pas d'ICE mène à la formation d'exconjugants contenant un ICE hybride issu de la recombinaison entre les éléments parentaux de la cellule donneuse (Figure 13) (Burrus et Waldor, 2004).

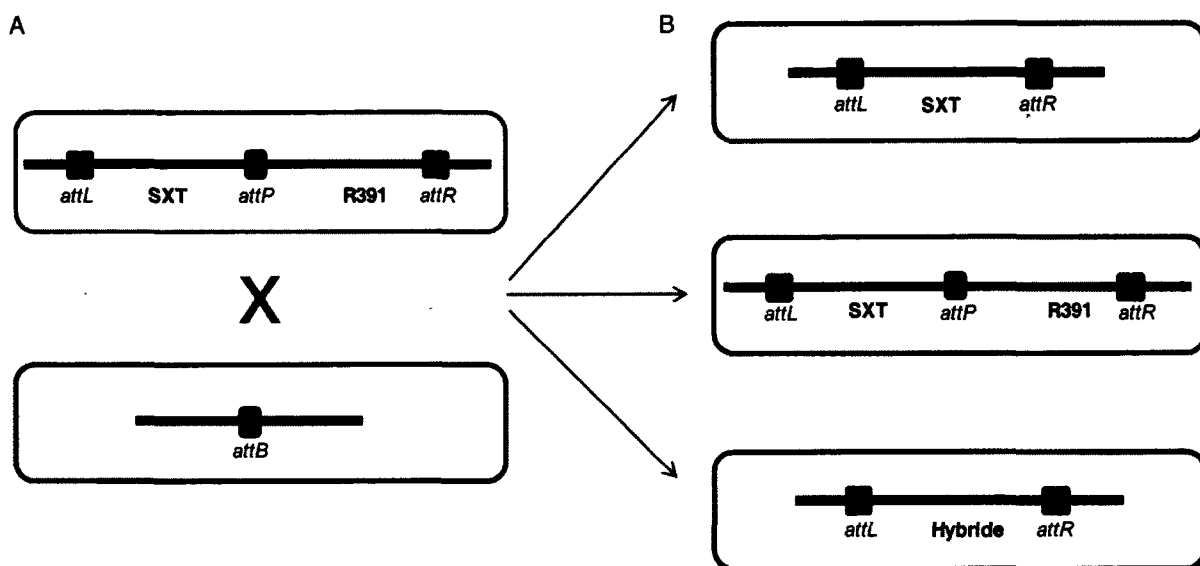


Figure 13. Formation d'hybrides SXT/R391.

A) Une cellule contenant un tandem SXT/R391 est une donneuse (orange) pour un transfert conjugatif vers une cellule réceptrice (verte) ne contenant pas d'ICE (site *attB* libre). B) Les exconjugants résultants de ce transfert peuvent contenir un ICE seul (SXT), un tandem SXT-R391 ou un ICE hybride issu de la recombinaison entre SXT et R391. SXT (bleu), R391 (rouge), hybride (violet). Inspiré de Burrus et Waldor (2004).

Différents mécanismes pourraient expliquer la formation d'hybrides SXT/R391. Dans un premier temps, un transfert conjugatif initié à l'*oriT* de l'élément situé en 5' du tandem et se

terminant à l'*oriT* de l'élément en 3' du tandem pourrait générer un ICE hybride. Cependant, pour une structure donnée de tandem (SXT-R391 ou R391-SXT), seulement un type d'hybride pourrait être formé. Les travaux de Burrus et Waldor (2004) ont démontré que plusieurs types d'hybrides peuvent être formés à partir d'une seule structure de tandem, suggérant qu'un autre mécanisme que le transfert *oriT-oriT* catalyse la formation d'hybrides. Dans un deuxième temps, la recombinaison homologe catalysée par l'enzyme RecA de l'hôte pourrait être à la source de la formation d'hybride, en permettant la recombinaison entre les deux ICEs présents dans la même cellule. Toutefois, des ICEs hybrides sont détectables dans un contexte $\Delta recA$ (Burrus et Waldor, 2004), démontrant qu'un troisième mécanisme doit permettre la formation d'hybride. La présence d'homologues du système de recombinaison du bactériophage λ dans le squelette de gènes conservés des ICEs SXT/R391 pourrait catalyser la recombinaison entre deux membres de cette famille présents dans la même cellule hôte (Burrus et Waldor, 2004).

1.6. Hypothèse générale du projet de doctorat

L'hypothèse générale du projet de doctorat a été posée comme suit : les homologues des gènes de recombinaison du bactériophage λ (*bet/exo*) retrouvés dans les ICEs de la famille SXT/R391 génèrent des ICEs hybrides en catalysant la recombinaison entre deux ICEs intégrés en tandem dans le chromosome de l'hôte.

1.6.1. Objectifs du projet de doctorat

Afin de vérifier l'hypothèse générale, les objectifs généraux et spécifiques suivants ont été définis :

- 1) Identifier les déterminants génétiques impliqués dans la formation d'hybrides SXT/R391
 - a) Mettre au point une technique permettant d'analyser à grande échelle la formation d'hybrides
 - b) Évaluer le rôle des gènes de recombinaison putatifs des ICEs SXT/R391 dans la formation d'hybrides
 - c) Réévaluer le rôle de RecA dans la formation d'hybrides
- 2) Identifier les conditions favorisant la formation d'hybrides
 - a) Déterminer les conditions induisant l'expression des gènes de recombinaison des ICEs SXT/R391
 - b) Déterminer l'organisation génétique du locus de recombinaison
- 3) Étudier la distribution de systèmes de recombinaison apparentés
 - a) Identifier les systèmes de recombinaison apparentés dans les génomes séquencés de bactéries, plasmides et bactériophages
 - b) Évaluer les relations phylogénétiques existant entre les protéines composant les systèmes identifiés

CHAPITRE 2 – RÉSULTATS

Les résultats obtenus au cours du projet de doctorat sont contenus dans deux articles scientifiques, l'un publié et l'autre soumis dans des journaux avec comité de lecture. Le présent chapitre présente ces deux articles, qui répondent aux objectifs principaux du projet de doctorat présentés dans la section précédente.

2.1. Rôle du système de recombinaison homologue des ICEs SXT/R391 dans la formation d'hybrides

2.1.1 Présentation de l'article

La première partie du projet de doctorat a visé à déterminer les gènes impliqués dans la formation d'hybrides SXT/R391. J'ai développé une méthode de détection permettant d'explorer les déterminants génétiques impliqués dans la formation d'hybrides SXT/R391. Cette méthode, qui permet la détection d'évènements de recombinaison entre deux ICE intégrés en tandem dans le chromosome, est basée sur les marqueurs de résistance aux antibiotiques naturellement portés par SXT et R391 et sur des marqueurs phénotypiques artificiellement insérés dans ces éléments. La formation d'éléments recombinants s'est avérée dépendre de deux systèmes indépendants: la protéine de l'hôte RecA et un système de recombinaison homologue codé par les ICE (*bet/exo* ou *s065/s066*, selon la nomenclature utilisée dans cette section). Ces deux gènes, fortement conservés entre tous les membres de la famille, forment un système de recombinaison fonctionnel apparenté au système de recombinaison Red du bactériophage λ et constitue le premier exemple de ce genre chez des éléments conjugatifs. J'ai démontré que ces gènes catalysent la recombinaison entre deux ICE non-identiques mais similaires (SXT et R391) intégrés en tandem dans le chromosome. La recombinaison catalysée par *bet-exo* est indépendante de la présence de RecA, mais que les deux systèmes peuvent coopérer, augmentant ainsi le potentiel de formation d'hybrides. J'ai montré que l'excision des ICE du chromosome ainsi que leur transfert conjugatif ne sont pas nécessaires à la formation d'hybrides. La conjugaison semble plutôt faciliter la ségrégation des hybrides et pourrait fournir un moyen de sélectionner des éléments recombinants fonctionnels comportant de nouvelles combinaisons de gènes de résistance aux antibiotiques. Finalement, mes résultats démontrent que les ICEs de la famille SXT/R391 peuvent promouvoir leur propre diversité génétique en générant de nouveaux éléments capables de disséminer de nouvelles combinaisons de résistances aux antibiotiques.

L'ensemble de ces résultats est présenté dans l'article qui suit, publié en 2009 dans la revue PLoS Genetics, dont la référence complète est la suivante :

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

2.1.2. Contribution à l'article

J'ai réalisé la totalité des manipulations et expériences qui ont mené à l'obtention des résultats présentés dans l'article, à l'exception de quelques constructions génétiques réalisées par V. Burrus. J'ai rédigé le manuscrit et préparé les figures, qui par la suite ont été révisés par les co-auteurs. L'approche expérimentale a été élaborée conjointement avec V. Burrus.

Title: Mobile Antibiotic Resistance Encoding Elements Promote their Own Diversity

Running head: Hybrid ICE Formation

Authors: Geneviève Garriss,¹ Matthew K. Waldor,^{2,3*} Vincent Burrus^{1*}

¹Centre d'Étude et de Valorisation de la Diversité Microbienne (CEVDM), Département de biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada

²Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

³Howard Hughes Medical Institute, Boston, MA, USA

***Corresponding authors:**

Mailing addresses:

V.B.: Département de biologie, Université de Sherbrooke, 2500 boulevard de l'Université, Sherbrooke, Québec, Canada J1K 2R1

Phone: (819) 821-8000 ext 65223

Fax: (819) 821-8049

E-mail: Vincent.Burrus@USherbrooke.ca

M.K.W.: Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115

E-mail: mwaldor@rics.bwh.harvard.edu

ABSTRACT

Integrating conjugative elements (ICEs) are a class of bacterial mobile genetic elements that disseminate via conjugation and then integrate into the host cell genome. The SXT/R391 family of ICEs consists of more than 30 different elements that all share the same integration site in the host chromosome but often encode distinct properties. These elements contribute to the spread of antibiotic resistance genes in several gram-negative bacteria including *Vibrio cholerae*, the agent of cholera. Here, using comparative analyses of the genomes of several SXT/R391 ICEs, we found evidence that the genomes of these elements have been shaped by inter-ICE recombination. We developed a high throughput semi-quantitative method to explore the genetic determinants involved in hybrid ICE formation. Recombinant ICE formation proved to be relatively frequent, and to depend on host (*recA*) and ICE (*s065* and *s066*) loci, which can independently and potentially cooperatively mediate hybrid ICE formation. *s065* and *s066*, which are found in all SXT/R391 ICEs, are orthologues of the bacteriophage λ Red recombination genes *bet* and *exo* and the *s065/s066* recombination system is the first Red-like recombination pathway to be described in a conjugative element. Neither ICE excision nor conjugative transfer proved to be essential for generation of hybrid ICEs. Instead conjugation facilitates the segregation of hybrids and could provide a means to select for functional recombinant ICEs containing novel combinations of genes conferring resistance to antibiotics. Thus, ICEs promote their own diversity and can yield novel mobile elements capable of disseminating new combinations of antibiotic resistance genes.

AUTHOR SUMMARY

Integrating and conjugative elements (ICEs) are a class of mobile elements found in diverse bacteria. ICEs of the SXT/R391 family have enabled the dissemination of genes conferring resistance to antibiotics among several important pathogens, including *Vibrio cholerae*, the agent of cholera. Here, using comparative analyses of the genomes of several SXT/R391 ICEs, we found that these elements are mosaics that have been shaped by inter-ICE recombination. We developed a plate-based method for semi-quantitative analyses of the genetic requirements for hybrid ICE formation. We discovered that hybrids form at relatively high frequencies and that both host and ICE genes can function independently and potentially cooperatively to mediate hybrid formation. The ICE-encoded recombination genes, which are found in all SXT/R391 ICEs, are related to genes that mediate recombination in bacteriophages, but have not been described previously in conjugative elements. Conjugative ICE transfer was not required for hybrid ICE formation but facilitates the segregation of hybrids. Thus, ICEs promote their own diversity and the generation of recombinant ICEs can yield novel mobile elements capable of disseminating new combinations of antibiotic resistance genes.

INTRODUCTION

Mobile genetic elements, including bacteriophages, conjugative plasmids and integrating conjugative elements (ICEs), are key mediators of bacterial genome evolution [1]. These elements can rapidly spread in bacterial populations and often confer to host bacteria selectable traits that are advantageous in particular environments or enable adaptation to new ecological niches. Transfer of ICEs and plasmids from donor to recipient bacteria occurs via conjugation, a process that requires direct cell-to-cell contact [2,3]. Conjugative transmission of ICEs and plasmids has limited the clinical usefulness of many antibiotics, since these mobile elements are potent vectors for dissemination of antibiotic resistance genes in bacterial populations [2,4,5,6,7].

ICEs integrate into and replicate along with the host cell chromosome, whereas plasmids exist as extra-chromosomal (usually circular) autonomously replicating DNA molecules. ICEs can excise from the donor cell chromosome and form circular molecules that are thought to be the substrates for the conjugative machinery. Similar to most conjugative plasmids [8], ICE conjugative DNA transfer is thought to be initiated at a specific *cis*-acting site (*oriT*) required for efficient translocation of the DNA to the recipient cell through the mating bridge. Within the recipient cell, host enzymes are thought to convert the translocated single-stranded DNA into double-stranded DNA that is circularized. An element-encoded recombinase (integrase) enables the integration of the ICE into the chromosome of the new host [2,9,10,11].

ICEs are widespread among diverse taxonomic groups of bacterial species and are able to transfer between genetically unrelated bacteria [5,10,11,12]. The SXT/R391 family of ICEs, which is one of the largest and most diverse set of ICEs studied, includes elements that have been detected in clinical and environmental isolates of several species of γ -proteobacteria from four continents over the past 40 years [13,14,15,16,17,18,19,20]. In Asia and Africa, this

family of ICEs has played an important role in the spread of genes conferring resistance to multiple antibiotics in *Vibrio cholerae*, the causative agent of cholera [17,19,21,22,23]. Currently, nearly all isolates of *V. cholerae* from cholera patients from these two continents harbor SXT, a prototypical member of the SXT/R391 family originally isolated from a 1992 Indian *V. cholerae* O139 isolate, or a closely related ICE [17,18,19,24,25,26].

The ICEs of the SXT/R391 family are grouped together because they all encode a highly conserved integrase (Int) that mediates the elements' site-specific integration into the host genome in the 5' end of *prfC*, a conserved gene encoding the peptide chain release factor RF3 [27]. Based on knowledge of the ~100-kb genomes of several SXT/R391 ICEs [15,28,29,30,31], in addition to the conserved integrase gene (*int*), these elements all contain a conserved set of ~24 genes that mediate their common functions that include: excision/integration, conjugative transfer and regulation [5]. Distinct variable regions that confer element-specific phenotypes, such as synthesis of the second messenger c-di-GMP or resistance to antibiotics or heavy metals are interspersed within this conserved and syntenous SXT/R391 backbone (see Figure 1A) [5,15,22,30,32].

In some cases, SXT/R391 ICEs do not exclude one another and can be present in the same host [33,34,35], providing the opportunity for the generation of recombinant ICEs. For example, R391, the other prototypical member of the SXT/R391 family, which was originally derived from a 1967 South African *Providencia rettgeri* isolate, and SXT can reside together in the same host [33]. A cell that contains one of these two ICEs can acquire a copy of the other ICE, yielding tandem arrangements of SXT and R391 in the host chromosome [33]. Tandem repeat structures are often excellent substrates for recombination [36] and exconjugants derived from donor strains containing such tandem arrays sometimes contain hybrid ICEs with genes from both R391 and SXT [37].

The molecular mechanisms that enable the formation of hybrid ICEs, which may contain novel combinations of genes conferring resistance to antibiotics, have not been addressed. However, two genes, *s065* and *s066*, which are highly conserved ($\geq 96\%$ identity) among all known SXT/R391 ICEs could contribute to the formation of hybrid ICEs. These genes encode proteins that are similar to the recombinase Bet (71% similarity and 55% identity) and the double-strand specific 5' to 3' exonuclease Exo (38% similarity and 26% identity) that are encoded by the temperate bacteriophage λ and several other phages [38] (Figure 1C). In λ , Bet and Exo, along with the Gam protein constitute an efficient *recA*-independent recombination system known as λ Red. Classic studies by Stahl and colleagues revealed many of the key features of the λ Red recombination system. They showed that efficient Red-mediated homologous recombination between λ chromosomes was almost entirely dependent on DNA replication [39], which generates a significant population of λ DNA with double-stranded breaks that serve as substrates for Red. Using replication-blocked crosses of phage λ chromosomes containing a single double-stranded cut, Stahl et al proposed that λ Red mediates recombination by a strand annealing mechanism [40]. Red Exo degrades 5' ends of linear double-stranded DNA, creating 3' single-stranded overhangs that can serve as templates for Red Bet to pair with complementary single-stranded DNA targets [41]. Red Gam (for which there is no SXT-encoded homologue) inactivates the *E. coli* exonuclease V (RecBCD), thereby protecting the ends of linear double-stranded DNA from degradation [41,42]. Besides providing significant amounts of double-stranded breaks, replication also provides a single-stranded DNA target for strand annealing on the lagging strand that is exposed by a passing replication fork [43]. Ordinarily, λ recombination is RecA-independent; however, when DNA replication is blocked, λ Red can also mediate efficient recombination via a strand invasion mechanism that is dependent upon RecA function [40,44]. Poteete et al suggested that the strand invasion pathway is a RecA-dependent salvage pathway for aborted Red-mediated recombination [45]. In recent years, the λ Red system has proven to be extremely useful for genetic engineering of *Escherichia coli* and closely related species [46,47,48,49]; however, investigation of the function of the Red pathway in its natural context, cells undergoing the λ

lytic cycle, has several technical challenges [48]. To our knowledge, λ Red-like recombination systems have not been described previously in conjugative elements.

Here, we found that the genomes of SXT/R391 ICEs appear to be routinely shaped by inter-ICE recombination. We explored the role of the SXT and R391 *bet* and *exo* homologues (*s065* and *s066*) and that of *recA*, a key host recombination gene, in the formation of hybrid ICEs. To accomplish this, we created a high throughput semi-quantitative screening assay that enabled the visual identification of exconjugant colonies containing hybrid ICEs. We found that *recA* mediated the formation of the majority of hybrid ICEs. Both *s065* and *s066* also contribute to the formation of hybrid ICEs and in the absence of *recA*, *s065* and *s066* appear to mediate the formation of nearly all hybrid ICEs. Conjugation was not essential for the formation of hybrid ICEs, suggesting that conjugative transfer acts as a means to segregate hybrid elements into new host cells. Thus, both host- and element-encoded recombination systems promote the formation of the mosaic genomes of SXT/R391 ICEs.

RESULTS

Evidence for recombination between SXT/R391 ICEs

When the genomes of SXT [28] and R391 [29] were originally reported, it appeared that the variable regions in this family of ICEs (shown as colored bars underneath the set of shared genes within the gray rectangle in figure 1A) were element-specific [50]. However, examination of the growing number of sequenced SXT/R391 ICE genomes suggests that even though some variable regions may be element-specific, others are shared by two or more ICEs (e.g. see ICE*PdaSpa1* and ICE*SpuPO1* in figure 1A), suggesting that this family of ICEs undergoes recombination. Closer analysis of conserved regions of these elements also suggested that recombination between SXT/R391 ICEs has shaped their genomes. Pairwise alignments of the genome sequence of ICE*PdaSpa1*, an ICE derived from the fish pathogen *Photobacterium damsela* subsp. *piscicida*, with that of SXT or R391 revealed that the majority of conserved sequences are only 95-97% identical, but that the 11.5-kb *attL-tnpB* and 0.6-kb *s021-rumB'* regions of ICE*PdaSpa1* and SXT are nearly 100% identical (Figure 1B). These comparisons suggest that a relatively recent recombination event within the 5' end of the truncated copy of *rumb'* occurred between precursors of ICE*PdaSpa1* and SXT, and support the idea that SXT/R391 ICE genomes are mosaics that have been sculpted by inter-ICE recombination. Exchange of DNA segments between these ICEs occurs when these elements are present in the same host cell. The tandem arrays that these ICEs can form in the host chromosome likely provide a suitable substrate for such recombination events to occur.

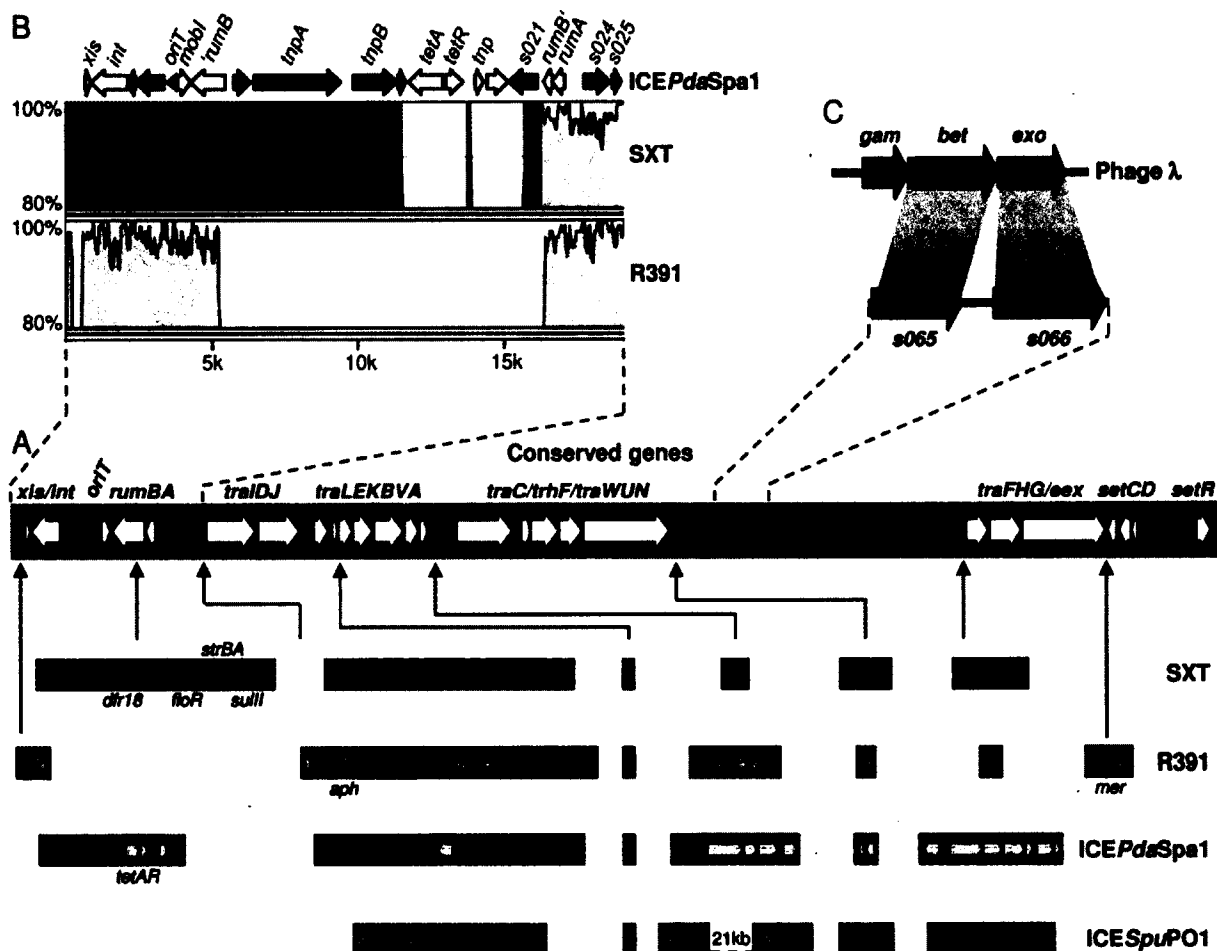


Figure 1. Evidence suggesting that recombination occurs between SXT/R391 ICEs.

(A) The middle gray box represents the set of genes (arrows) conserved in the 4 SXT/R391 genomes shown. Gray ORFs represent genes of unknown function, white ORFs represent genes of known function [28,51,52], and black ORFs correspond to *s065* and *s066*. Below, variable ICE regions are shown with colors according to the elements in which they were originally described: SXT [28] (blue), R391 [29] (red), ICEPdaSpa1 [15] (green) and ICESpuPO1 [30] (purple). (B) A close-up of the *attL*-*s025* region of ICEPdaSpa1 (accession number AJ870986) is shown in the upper left. The variation of percentage of identity was plotted using a Multi-LAGAN pairwise comparison [53] of this ICEPdaSpa1 region with the corresponding regions of SXT (accession number AY055428) and R391 (accession number AY090559) and the mVista visualization module [54] with a sliding window of 100 bp. The minimum width and the minimum percent conservation identity that must be maintained over that width for a region to be considered conserved were set at 100 bp and 70% respectively. The dark gray area highlights the large nearly identical region conserved between SXT and ICEPdaSpa1. (C) A comparison of *s065* and *s066*, which are present in all SXT/R391 ICEs, to the bacteriophage λ Red genes (numbers represent % similarity between S065 and Bet, and S066 and Exo, respectively) is shown in the upper right. *drf18* encodes

trimethoprim resistance; floR encodes chloramphenicol resistance; strAB encodes streptomycin resistance; sulII encodes sulfamethoxazole resistance; tetAR encodes tetracycline resistance; aph encodes kanamycin resistance; and mer encodes mercury resistance.

Detection of hybrid ICE formation

We developed a high throughput conjugation-based semi-quantitative screen to assess the genetic requirements for the formation of hybrid ICEs. The assay employs donor cells bearing tandem copies of modified SXT and R391 and was designed to distinguish between exconjugant colonies containing SXT-R391 tandem arrays, hybrid elements or single parental elements (Figure 2). The phenotypic markers *lacZ* and *galk* were inserted between *traG* and *eex* in SXT and between *traG* and *merR* in R391, respectively (Figure 1A and 2). The position of these two loci, near the right ends of the elements, is remote from the antibiotic resistance markers that are found near the left ends of SXT (*sulII dfr18*) and R391 (*aph*) (Figure 1A and 2), thereby maximizing the opportunity to detect recombination events occurring within tandem arrays. Both *lacZ* and *galk* were placed under control of the P_{lac} promoter to enable high-level β -galactosidase and galactokinase activities in a *lacI* background. *Escherichia coli* strains containing tandem arrays of these labeled ICEs were used as donors in mating assays using $\Delta galk lacZU118 lacI42::Tn10$ derivatives of *E. coli* MG1655 as recipient strains. Exconjugants were isolated on MacConkey indicator agar plates supplemented with galactose and X-Gal (MCGX) along with the antibiotics sulfamethoxazole (Su) and trimethoprim (Tm) to select for SXT or kanamycin (Kn) to select for R391. Using this medium, we expected to infer the ICE content of each exconjugant colony from its color and resistance phenotypes (e.g., Figure 2), and to thereby determine the percentage of exconjugants containing hybrid elements. PCR assays confirmed our expectations regarding the presence of hybrid ICEs in red colonies on Su-Tm medium and blue colonies on Kn medium (Figure 2 and data not shown). However, PCR analyses also revealed that a subset of purple exconjugant colonies contained tandem arrays consisting of a hybrid ICE coupled to a

parental ICE. Thus, our method for enumeration of recombinant ICEs formed in these assays (e.g. as red colonies in Figure 2) understates the true frequency of recombination events.

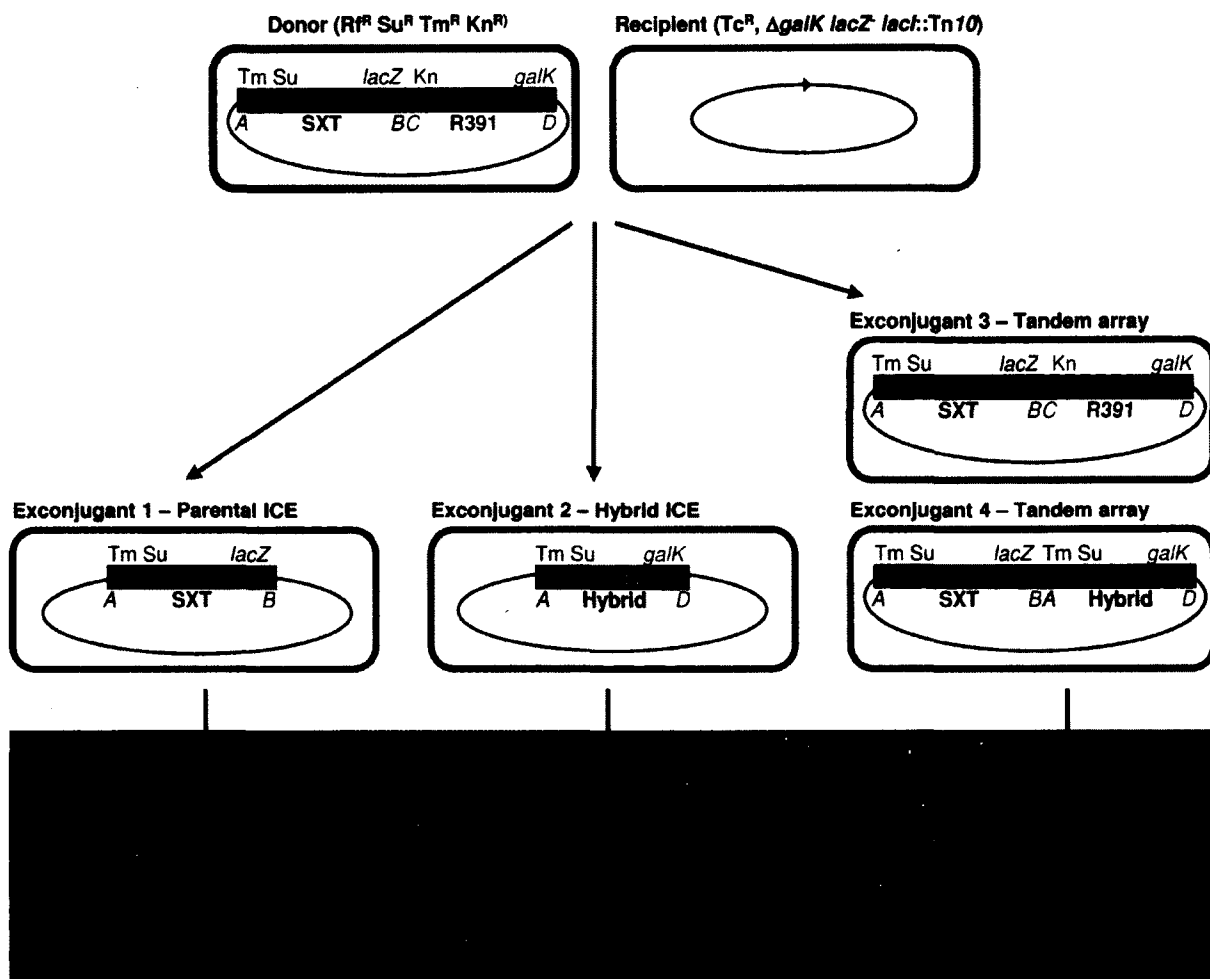


Figure 2. Schematic of colony color-based semi-quantitative assay for the detection of hybrid ICE-containing colonies.

Relative positions of resistance markers (trimethoprim (Tm), sulfamethoxazole (Su), kanamycin (Kn)) and phenotypic markers (*lacZ* and *galk*) in SXT and R391 are indicated. DNA originating from SXT is shown in blue and DNA originating from R391 is shown in red. The use of a $\Delta galK$ $lacZ^-$ $lacI::Tn10$ recipient strain allows constitutive expression of the inserted *lacZ* and *galk* from the ICEs in the exconjugant colonies. Mating between a donor cell (green) containing an SXT-R391 tandem array and a recipient cell (orange) yields exconjugants that may contain a single element, a hybrid element or a tandem array. MacConkey X-gal D-galactose indicator agar containing trimethoprim, sulfamethoxazole and tetracycline (bottom panel) reveals colonies harboring single

parental ICEs (blue colonies), hybrid ICEs (red colonies), and SXT-R391 tandem arrays (purple colonies). Purple colonies may also consist of cells containing an array composed of SXT and a hybrid element on this media (e.g. exconjugant 4). Red and purple colonies are larger on this medium because they can use D-galactose as a carbon source. A and B, SXT left and right extremities; C and D, R391 left and right extremities, respectively amplified by primer pairs VISLF/VISLR3, VISRF/VISRR, VISLF/VISLR2, VISRF/VISRR2 [37].

In pilot experiments, we found that the percentage of hybrid ICEs detected was influenced by which ICE's antibiotic resistance markers were selected. A higher percentage of exconjugants harboring a hybrid ICE was isolated on Su-Tm (6.75%) than on Kn (2.70%). This is probably a consequence of the fact that the R391 transfer frequency is about 10-fold higher than that of SXT, and hence a high frequency of colonies containing hybrids are likely to contain R391 as well, and thus cannot be distinguished from strains containing tandem arrays (39.2% tandem arrays on Su-Tm vs 10.4% on Kn). Consequently, in most subsequent studies of the genetic requirements for hybrid ICE formation, we used donors harboring SXT-R391 arrays and Su Tm to select for hybrid-harboring exconjugants; however, in some experiments we were unable to obtain SXT-R391 arrays with the desired deletions and in these cases we used donors containing R391-SXT arrays.

***recA* enables the formation of most but not all hybrid ICEs**

We suspected that the host *recA* gene might play a key role in the generation of hybrid ICEs since the SXT and R391 genomes have more than 95% identity over nearly 64 kb of DNA distributed in 11 segments ranging from 247 bp to 12,085 bp. Hybrid ICEs could form by RecA-mediated homologous recombination either in the donor cells prior to transfer or in recipient cells after transfer of both SXT and R391 from donor cells. We carried out conjugation experiments using *recA*⁺ (GG61) or *recA*⁻ (GG66) donor cells containing a tandem array of SXT and R391, and *recA*⁺ (VB38) or *recA*⁻ (VB47) recipient cells (Table 1) to

distinguish between these possibilities. However, since RecA is required in donor cells for SXT and R391 transfer, probably to alleviate the repression of expression of genes encoding the conjugative transfer machinery (*tra* genes), it was necessary to exogenously express SetC and SetD, the activators of the *tra* genes, in all *recA* donors [55]. Such exogenous activation of transfer genes generally induces a 10- to 100-fold increase in the frequency of ICE transfer ([55] and data not shown); however, since we compare the percentage of hybrids in different backgrounds, rather than the absolute frequency of hybrid formation, the increase in transfer frequency should not distort our results.

Table 1. Strains of *E. coli* and plasmids used in this study.

Strain or plasmid	Relevant genotype or phenotype ^a	Reference or source
Strains		
CAG18420	MG1655 <i>lacZU118 lacI42::Tn10kan</i> (K ^{nR})	[56]
CAG18439	MG1655 <i>lacZU118 lacI42::Tn10</i> (Tc ^R)	[56]
VB112	MG1655 R ^{fR}	[52]
VB38	CAG18439 Δ <i>galK</i> (Tc ^R)	This study
VB47	CAG18439 Δ <i>galK</i> Δ <i>recA</i> (Tc ^R)	This study
GG55	VB112 Δ <i>recA</i> (R ^{fR})	This study
GG47	GG55 pVI67 (R ^{fR} Ap ^R)	This study
HW220	CAG18439 <i>prfC::SXT</i> (Tc ^R Su ^R Tm ^R)	[27]
JO99	CAG18439 <i>prfC::R391</i> (Tc ^R K ^{nR})	[33]
VB40	CAG18439 Δ <i>lacZ prfC::SXT::lacZ</i>	This study
GG13	CAG18439 Δ <i>galK prfC::R391::galK</i>	This study
GG61	VB112 <i>prfC::[R391::galK]-[SXT::lacZ]</i> (R ^{fR} Su ^R Tm ^R K ^{nR})	This study
GG64	VB112 <i>prfC::[R391::galK Δorf68]-[SXT::lacZ Δs065]</i>	This study
GG65	VB112 <i>prfC::[R391::galK Δorf69]-[SXT::lacZ Δs066]</i>	This study
GG93	VB112 <i>prfC::[R391::galK Δ(orf68-orf69)]-[SXT::lacZ Δ(s065-s066)]</i>	This study
GG66	GG47 <i>prfC::[SXT::lacZ]-[R391::galK]</i> (R ^{fR} Ap ^R Su ^R Tm ^R K ^{nR})	This study
GG69	GG47 <i>prfC::[SXT::lacZ Δs065]-[R391::galK Δorf68]</i>	This study
GG70	GG47 <i>prfC::[SXT::lacZ Δs066]-[R391::galK Δorf69]</i>	This study
GG102	GG47 <i>prfC::[SXT::lacZ Δ(s065-s066)]-[R391::galK Δ(orf68-orf69)]</i>	This study
GG125	VB38 <i>prfC::[R391::galK ΔmobI]-[SXT::lacZ ΔmobI]</i> (Tc ^R Su ^R Tm ^R K ^{nR})	This study
GG171	VB38 <i>prfC::[SXT::lacZ Δint]-[R391::galK Δint]</i> (Tc ^R Su ^R Tm ^R K ^{nR})	This study
GG185	VB38 <i>prfC::[R391::galK]-[SXT::lacZ]</i> (Tc ^R Su ^R Tm ^R K ^{nR})	This study
GG186	VB38 <i>prfC::[SXT::lacZ]-[R391::galK]</i> (Tc ^R Su ^R Tm ^R K ^{nR})	This study
Plasmids		
pKD3	Cm ^R template for one-step chromosomal gene inactivation	[47]
pKD4	K ^{nR} template for one-step chromosomal gene inactivation	[47]
pVI36	Sp ^R template for one-step chromosomal gene inactivation	[52]
pVI40A	pVI36 <i>BamHI::P_{lac}-galK</i>	This study
pVI42B	pVI36 <i>BamHI::P_{lac}-lacZ</i>	This study
pAH57	<i>oriR101 repA101^{Ts} cl857^{ts} λP_R-xis_{λ}-int_{λ}</i> (Ts Ap ^R)	[57]

pVI67	pAH57 $\Delta(xis_{\lambda}-int_{\lambda})::setDC$ (Ts)	This study
pVI68	pAH57 $\Delta(xis_{\lambda}-int_{\lambda})::int_{SXT}$ (Ts)	This study
pMobl-B	pBAD-TOPO <i>mobil</i>	[52]

^a *s065* and *s066* of SXT correspond to *orf68* and *orf69* of R391, respectively, according to the annotation of both elements [28,29]. Ap^R, ampicillin resistant; Cm^R, chloramphenicol resistant; Kn^R, kanamycin resistant; Rf^R, rifampicin resistant; Su^R, sulfamethoxazole resistant; Sm^R, streptomycin resistant; Sp^R, spectinomycin resistant; Tc^R, tetracycline resistant; Tm^R, trimethoprim resistant; Ts, thermosensitive.

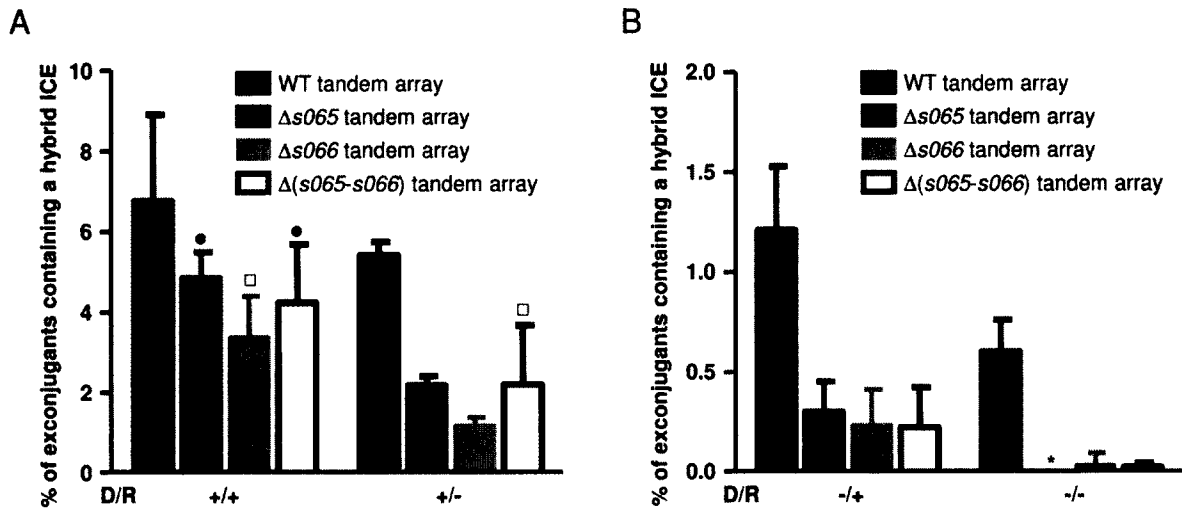


Figure 3. Involvement of *recA*, *s065* and *s066* in the formation of hybrid ICEs.

recA⁺ (panel A) or *recA*⁻ (panel B) donor strains, which contained either wild-type (WT), $\Delta s065$, $\Delta s066$ or $\Delta(s065-s065)$ SXT-R391 tandem arrays, were used as donors in these assays. The recipient strains were either *E. coli* VB38 (*recA*⁺) or *E. coli* VB47 (*recA*⁻). D/R + and – indicate the *recA* genotype of the donor and recipient strains, respectively. SetDC was expressed from a plasmid when *recA* donors were used. Bars represent the percentage of exconjugants containing hybrid ICEs and were calculated by dividing the number of exconjugants containing hybrid ICEs (red Tc^R Su^R Tm^R CFU) by the total number of exconjugants (Tc^R Su^R Tm^R CFU). The means and standard deviations obtained from at least three independent assays are shown and the number of colonies containing a hybrid ICE counted for each assay is presented in Table S2. Note the differences in the scale of the y-axis in panels A and B. One-way ANOVA with a Tukey-Kramer post-test was used to compare the means of hybrid ICE-containing exconjugant colonies. The confidence interval for the comparisons of mutant tandem arrays relatively to WT tandem arrays was P<0.001, except □ which indicates P<0.05 and ● which indicates that the difference was statistically not significant. * indicates that the percentage of exconjugants bearing a hybrid ICE was below the limit of detection (<0.01%).

Deletion of *recA* from donor cells had a significant effect on the percentage of exconjugants found to contain hybrid ICEs. Conjugation assays with *recA* donors reduced the percentage of hybrids at least 5.6-fold relative to assays with WT donors, both when WT and *recA* recipients were used (p<0.001) (Figure 3A and 3B). In contrast, deletion of *recA* from the recipient cells did not have a significant effect on the percentage of exconjugant colonies containing a hybrid

ICE when WT donor cells were used (Figure 3A). When *recA* donor cells were used, there was an ~2-fold reduction in the percent of exconjugants with hybrid elements in *recA* recipients compared to WT recipients, which was not statistically significant (Figure 3B). Finally, deletion of *recA* from both donors and recipients reduced the percentage of hybrid ICEs detected by more than 11-fold as compared to when *recA*⁺ was present in both donor and recipient. Taken together, these observations suggest that *recA*-mediated homologous recombination generates the majority of hybrid ICEs and that these recombination events happen both in donor and recipient strains. RecA's role is more readily discerned in donors; however, this may reflect a limitation of our assay in that conjugation facilitates detection of hybrids as discussed below. Notably, 0.60% of exconjugants contained hybrid ICEs even when both donor and recipient strains lacked *recA* indicating that some hybrid ICEs are generated via a *recA*-independent recombination pathway (Figure 3B).

***s065* and *s066* promote hybrid ICE formation**

We explored whether *s065* and *s066*, which encode a single-strand DNA recombinase (unpublished results and [38]) and a putative exonuclease respectively (Figure 1C), also influence the formation of hybrid ICEs, and whether they might account for *recA*-independent generation of these elements. Donor strains harboring tandem arrays of $\Delta s065$, $\Delta s066$, or $\Delta(s065-s066)$ deletion mutants of SXT and R391 were constructed in *recA*⁺ and *recA* donor strains, and these strains were used in conjugation assays with *recA*⁺ (VB38) and *recA* (VB47) recipient strains as described above. Compared to WT donors, when *recA*⁺ donors lacking *s065*, *s066* or both genes were tested, there was a consistent reduction in the frequency of hybrid formation (Figure 3A). This decrease was generally not statistically significant when *recA*⁺ donors and recipients were used; however, when *recA* was absent from either donor or recipient cells, the effect of $\Delta s065$ and/or $\Delta s066$ deletions became more pronounced. For example, when *recA* donors and WT recipients were used, the percentage of exconjugants containing hybrid ICEs was reduced ~5 fold by deletion of *s065* and/or *s066*, and when both

donors and recipients lacked *recA*, the additional mutations reduced hybrid frequency more than 20-fold (Figure 3B). Presumably, the absence of *recA*, which we have shown prevents formation of a majority of hybrid ICEs, allows the subtler effects of *s065* and/or *s066* deletions to become more apparent. Our data suggest that both *s065* and *s066* contribute to hybrid ICE formation, and that they act in a non-redundant fashion with each other. Additionally, our finding that deletion of both *s065* and *s066* has an effect comparable to that of a single gene deletion indicates that their roles may be interdependent. Since formation of hybrid ICEs was scarcely detectable when both *recA* and *s065/s066* were disrupted, it appears that *s065* and *s066* are required for the majority of *recA*-independent hybrid ICE formation.

Comparisons of the percentages of hybrid formation shown in figure 3 suggest that *recA* and *s065/s066* may cooperate in generating hybrid ICEs. Approximately 37% of hybrid formation in donor cells is attributable to *recA* as shown by the frequency of hybrid-bearing exconjugants (~2%) observed in the absence of *s065* and/or *s066* with *recA*- recipients (Figure 3A +/- all but black bar). When donor cells lack *recA* and rely on the *s065/s066*-pathway for hybrid formation we found that 0.6% of exconjugants contained hybrids, i.e. 11% of total hybrid formation (Figure 3B -/- black bar). Taken together, these frequencies cannot account for the frequency of exconjugants harboring hybrids observed in the presence of both pathways (5.4%, Figure 3A +/- black bar). Thus, these two pathways, which can function independently, may also act synergistically to promote hybrid ICE formation. However, given the variability in our data, particularly using *recA*⁺ recipients, definitive evidence for interactions between these pathways is lacking.

Conjugation is not required for the formation of hybrid ICEs

In the experiments described above, we relied on conjugative transfer to identify hybrid ICEs in exconjugant colonies. However, our observation that some hybrids appear to form in

recipient cells, after elements have transferred (as indicated by differences in hybrid formation in *recA*⁺ and *recA* recipients) suggested that the conjugative process was not necessarily a component of hybrid formation. We took advantage of our previous observations that there is little, if any, conjugative transfer of SXT in broth culture [23], to begin to explore whether conjugation was required for hybrid ICE formation. We tested whether we could detect hybrid formation in a *recA*⁺ Δ *galK lacZ* Tc^R strain (GG185) bearing a wild-type R391-SXT array (the opposite array orientation as used above) in the absence of a recipient strain. GG185 was passaged with two subcultures in LB broth for 72 h (>100 generations) and then the culture was plated on MCGX indicator medium supplemented with Tc Su Tm, to identify Su^R Tm^R hybrid ICEs (red colonies), or with Tc Kn, to identify Kn^R hybrid ICEs (blue colonies). Kn^R hybrid ICEs were detected (0.16 ± 0.05 % of colonies) at this point but Su^R Tm^R hybrids were barely detectable (Table 2). Detection of hybrid ICE formation using this experimental system requires marker loss. We observed greater loss of SXT (20.1 ± 4.9 %) than R391 (<0.02 %) in this experiment, in accord with a previous report that the ICE located at the right end of the array is more frequently lost and that in this position, R391 is more stable than SXT [33]. Thus, the few detectable Su^R Tm^R hybrids in this experiment likely reflect the lack of loss of R391 from the tandem array in GG185.

The detection of Kn^R hybrids during passage of GG185 in LB broth provides support for the idea that conjugation is not essential for hybrid ICE formation. However, it is possible that there is a low frequency of conjugative ICE transfer in broth cultures. To formally exclude a role for conjugation in hybrid ICE formation, we constructed a strain harboring an R391-SXT array where the ICEs were unable to transfer due to the deletion of *mobI*. MobI is part of the SXT/R391 DNA processing machinery and is thought to recognize and act on *oriT*; deletion of *mobI* renders SXT and R391 non-transmissible but does not impair their excision (data not shown) or the formation of a functional conjugation apparatus [52]. We constructed a Δ *mobI* R391-SXT tandem array in a *recA*⁺ Δ *galK lacZ* Tc^R strain (VB38) (Table 1). The resulting strain (GG125) was cultivated for 6 days with two daily subcultures (>250 generations) in LB broth with tetracycline as the sole antibiotic. Serial passage allowed for the loss of unselected

markers [37], thereby helping to reveal formation of possible hybrid ICEs. The culture was plated at 24, 72, and 144 h post-inoculation on the indicator medium supplemented with the same antibiotics used above to identify hybrid ICEs. As noted above with GG185, loss of SXT from the R391-SXT array in GG125, yielding a single R391 (Kn^R) ICE, occurred much more frequently than the loss of R391 from this strain (Table 2). Kn^R hybrid ICEs were detectable at 24 h, when 0.11% of colonies contained a hybrid ICE, and by 144 h this percentage increased to 0.38 % (Table 2). Su^R Tm^R hybrid ICEs were only isolated after 144 h of culture and only 0.004% of colonies contained hybrids. Potential explanations for the different frequencies with which hybrids were observed are discussed below. However, the results from both selections clearly demonstrate that conjugation is not required for the formation of hybrid ICEs. Furthermore, using a variety of PCR assays (see [37]), three distinct ICE structures were identified among 19 of the Kn^R hybrids (data not shown). Thus, the hybrids identified in these experiments cannot be explained by clonal amplification of a single cell containing a hybrid ICE.

Table 2. Percentage of colonies containing hybrid ICEs or single elements recovered over time from a strain initially harboring a wild-type R391-SXT, or a non-transmissible $\Delta mobI$ R391-SXT tandem array.

Strains	Time (h)	% colonies selected on Kn		% colonies selected on Su Tm	
		Hybrids ^a	Single	Hybrids ^a	Single
GG185	24	0.04±0.04	11.7±3.4	0.02±0.03	0.04±0.07
	72	0.16±0.05	20.1±4.9	0.03±0.05	<0.02 ^b
GG125 ($\Delta mobI$)	24	0.11±0.02	11.8±0.6	<0.01 ^b	0.02±0.03
	72	0.21±0.03	16.1±3.7	<0.01 ^b	0.02±0.03
	144	0.38±0.07	22.6±1.2	0.004±0.007	0.09±0.07

^a The numbers of colonies containing a hybrid ICE that were counted for each assay are presented in Table S3.

^b Detection limit of the assay.

Excision is not required for hybrid ICE formation

Despite existing predominantly as chromosomal-encoded elements, the position of ICEs with respect to host chromosomes is highly dynamic. ICE-encoded *int* and *xis* genes allow them to excise from host chromosomes [51], and this event is thought to be an early step in conjugation. To assess whether extrachromosomal ICE DNA is a required substrate for hybrid ICE formation, we constructed a strain carrying a Δint SXT-R391 array (GG171) (Table 1). GG171 was used in assays similar to those described above for the $\Delta mobI$ array. After only 24 h of culture, 0.9% of colonies contained a $Su^R Tm^R galk^+$ hybrid ICE, demonstrating that formation of recombinant ICEs does not depend on ICE excision. Thus, chromosomal tandem ICE arrays can serve as a recombination substrate.

Discussion

Comparative analyses of the genomes of several SXT/R391 ICEs revealed that these elements are mosaics that have been shaped by inter-ICE recombination (Figure 1A). The large set of core genes that are conserved among all SXT/R391 ICEs provides an ample substrate for inter-ICE recombination. Furthermore, the inherent ability of these elements to form tandem array structures [33,37] increases the opportunities for ICE recombination. Given the high degree of homology between SXT and R391, our finding that *recA* accounts for the generation of the majority of hybrid ICEs is understandable. However, *s065* and *s066*, which are present in all SXT/R391 ICEs, also contribute to formation of recombinant ICEs. The contribution of these ICE λ *bet* and *exo* homologues was easiest to discern in the absence of *recA*; in this context, *s065* and *s066* accounted for the formation of nearly all of the hybrids we detected. These two genes appear to function in the same recombination pathway, since deletion of *s065*, *s066*, or both genes resulted in similar reductions in hybrid formation. Neither ICE excision nor conjugative transfer proved to be essential for generation of hybrid ICEs; instead conjugation appears to facilitate the segregation of hybrids and may provide a means to select for functional recombinant ICEs.

In previous work, we used multiple PCR analyses to show that exconjugants derived from conjugations with donors bearing SXT-R391 arrays occasionally contained a hybrid ICE [37]. This technique was too cumbersome to enable either quantitative or genetic analysis of hybrid ICE formation. The high-throughput semi-quantitative detection method reported here enabled more sensitive analyses of the genetic determinants involved in hybrid ICE formation. Hybrid formation was relatively frequent, as we found that almost 7% of exconjugants selected on Su and Tm contained a recombinant ICE. Since some exconjugants scored as containing a parental ICE array (purple colonies in figure 2) actually contained a hybrid ICE and a parental ICE, 7% is an underestimation of the true frequency of hybrid formation. Thus,

formation of hybrid ICEs, which may have novel combinations of genes conferring resistance to antibiotics, may be fairly common.

While hybrid ICEs were readily detectable in exconjugants using our plate-based screening method, we found that they also form in cells containing tandem arrays of non-transmissible ICEs. Detection of non-transmissible hybrid ICEs seems to depend upon the rate of post-recombinational loss of one or the other ICE, as shown by the coincident increase over time of colonies harboring hybrids (Table 2). Different frequencies of Kn^{R} vs Su^{R} Tm^{R} hybrids formed from the non-transmissible R391-SXT array (Table 2). These differences are probably a consequence of the structure of the array used here. The relatively low frequency of hybrids in donors compared to exconjugants suggests that conjugation facilitated detection of hybrids by allowing for segregation of hybrid ICEs from parental ICEs. In nature, it is possible that conjugation serves to select for functional hybrids that are capable of transmission.

Our data indicate that both *recA* and *s065/s066* can mediate hybrid formation independently, and potentially co-operatively as well. *RecA*'s role in homologous recombination has been the subject of extensive study; we assume its mechanism of action parallels that described in previous work. Our models for how *s065* and *s066* mediate hybrid ICE formation are largely based on prior studies of phage-borne *s065* and *s066* homologues. However, there is evidence that S065, like λ Bet, can mediate single-stranded DNA recombination ([38] and our unpublished observations) and that S066 has double-stranded DNA exonuclease activity (Rory Watt, unpublished observations). Thus, it is reasonable to assume that S065 and S066 function in a similar fashion as Bet and Exo to promote ICE recombination. Double-stranded DNA ends are thought to be the principle substrate for the Red pathway in its natural context [40,58]; Exo is thought to digest the 5' end of such double-stranded DNA breaks leaving a suitable single-stranded substrate for Bet recombination [59]. Double-strand breaks in ICE DNA could occur in the chromosomal ICE, the excised circular double-stranded ICE or the

extrachromosomal circular double-stranded ICE after transfer but prior to re-integration. The latter molecule may be subject to host restriction endonucleases, generating suitable substrates for S066 and S065. Furthermore, DNA damaging agents (UV, antibiotics), which are known to trigger the conjugative transfer of SXT/R391 ICEs, also provide suitable substrates for recombination in the form of double-stranded DNA breaks. It is also possible that single-stranded ICE DNA generated in donor cells and transferred to the recipient during conjugation can be a substrate for formation of hybrid elements.

There are particularities of the lifecycles of ICEs and lambdoid phages that suggest that their respective recombination systems may function differently. Unlike λ , which can replicate autonomously as double-stranded DNA (theta replication) during its lytic cycle, SXT/R391 ICEs do not seem to replicate autonomously. This difference likely decreases the opportunities for generating double-stranded breaks that have been shown to be a major substrate for λ Red functions [39,40,44]. In addition, the absence of a *gam* ortholog in SXT/R391 ICEs suggests that either RecBCD's exonuclease activity has little impact on recombination catalyzed by S065/S066, i.e. double-stranded DNA extremities are not a significant substrate, or that ICEs encode an unrelated inhibitor of exonuclease V that remains to be identified.

To our knowledge, the *s065/s066* recombination system is the first Red-like recombination pathway to be described in a conjugative element. To date, Red-like recombination genes/systems have been exclusively identified in prophages of both gram-positive and gram-negative bacteria [38]. Interestingly, *s065* and *s066* are part of the core genome found in all SXT/R391 ICEs. Their ubiquity in this family of mobile elements suggests that the generation of diversity via inter-ICE recombination is a key feature of this family of ICEs. The routine formation of tandem ICE arrays in fresh exconjugants [37] and the lack of exclusion between certain SXT/R391 ICEs [34,35] also suggests that the modus operandi of these elements includes recombination. Recombination is also a central feature of lambdoid phages (for

review, see [60,61]) and Martinsohn et al recently proposed that the λ Red recombination pathway contributes to the mosaic genomes that characterize this family of bacteriophages [62]. Another striking parallel between SXT/R391 ICEs and lambdoid phages is that their transfer (by conjugation or transduction respectively) is greatly increased by damage to host DNA. Expression of *s065* and *s066*, like that of *exo* and *bet*, increases with UV damage to the host (Mariam Quinones, unpublished results). Thus, like the λ Red recombination pathway [41], the *s065/s066* recombination system may serve as a recombinational repair system to promote the formation of functional ICEs capable of exiting from a damaged host and re-establishing themselves in a new host.

While numerous questions regarding the action of S065 and S066 remain to be explored, collectively our findings suggest that these genes promote the plasticity of SXT/R391 ICE genomes. Besides enhancing inter-ICE recombination, it is also possible that *s065* and *s066* enable the incorporation of exogenous genetic material into ICE genomes, such as the DNA shown in colors in figure 1A. Lastly, we identified orthologs of *s065* and *s066* in IncA/C plasmids such as pIP1202 from *Yersinia pestis* biovar *Orientalis*, the causative agent of plague. These conjugative plasmids have recently been found to be broadly disseminated among multiply drug resistant zoonotic pathogens [6]. It will be interesting to explore whether these *s065/s066* orthologs contribute to the plasticity of this family of conjugative plasmids.

MATERIALS AND METHODS

Bacterial strains, plasmids and media

The bacterial strains and plasmids used in this study are described in Table 1. Bacterial strains were routinely grown in Luria-Bertani (LB) broth at 37°C in an orbital shaker and maintained at -80°C in LB broth containing 15% (v/v) glycerol. Colonies harboring hybrid ICEs were screened by plating on MacConkey agar base (Difco) plates supplemented with 0.6% galactose, 80 mg/l X-Gal (5-bromo-4-chloro-3-indoyl- β -D-galactopyranoside) (indicator medium MCGX) and the suitable antibiotics. Antibiotics were used at following concentrations: ampicillin (Ap), 100 mg/l; kanamycin (Kn), 50 mg/l; rifampicin (Rf), 100 mg/l; spectinomycin (Sp), 50 mg/l; sulfamethoxazole (Su), 160 mg/l; trimethoprim (Tm), 32 mg/l; tetracycline (Tc), 12 mg/l.

Plasmid construction

The oligonucleotides used for construction of plasmids are described in Table S1. Plasmids pVI67 and pVI68, designed to allow conditional expression of SetDC or Int_{SXT}, were constructed by replacing the 1,383-bp *EcoRI/NcoI* fragment of pAH57 [57] with either a 942-bp *EcoRI/NcoI* fragment containing the *setDC* operon of SXT or a 1,367-bp *EcoRI/NcoI* fragment containing *int_{SXT}*, respectively. *setDC* and *int_{SXT}* were amplified by PCR using primer pairs setDF/setCR and intSF/intSR, respectively, and the DNA of *E. coli* HW220 as a template. Both plasmids are temperature sensitive for replication and allow the expression of the cloned genes from λp_R under control of the thermosensitive repressor *cI857*.

Plasmids pVI40A and pVI42B were templates used in the creation of PCR products for the insertion of *lacZ* and *galK* markers into SXT and R391 with the Datsenko and Wanner protocol [47]. These templates contain *galK* or *lacZ*, both under control of P_{lac} , introduced into the *Bam*HI site of pVI36 [52]. The P_{lac} -*galK* fragment was made by amplifying by PCR *galK* and the P_{lac} promoter sequence using primer pairs galK1F/galK1R and Plac3F/Plac3R, respectively, and the DNA of *E. coli* VB112 as a template. The resulting two fragments were fused using the Splicing by Overlap Extension protocol [63]. The P_{lac} -*lacZ* of pVI42B was amplified using DNA of *E. coli* VB112 as a template and primer pair lacZ1R/Plac3F. The inserts of all plasmids constructed for this study were sequenced by DNA LandMarks Inc (St-Jean-sur-Richelieu, QC).

Construction of chromosomal deletions and insertions

The oligonucleotides used for chromosomal deletions and insertions are described in Table S1. Deletion and insertion mutants were constructed by using the one-step chromosomal gene inactivation technique of Datsenko and Wanner [47]. All deletions were designed to be non-polar. The $\Delta galK$ and $\Delta lacZ$ mutations were introduced in *E. coli* CAG18439 using primer pairs galKWF/galKWR and lacZW-B/lacZW-F, and plasmids pVI36 and pKD4 as templates. The $\Delta recA$ mutation was introduced in *E. coli* VB38 and VB112 using primer pair recAWF/recAWR and pVI36 as a template. The $\Delta s065$, $\Delta s066$, and $\Delta(s065-s066)$ mutations were introduced in SXT (in strain HW220) using primer pairs 65WF/65WR, 66WF/66WR, and 65WF/66WR, respectively, and template plasmid pVI36. The corresponding mutations $\Delta orf68$, $\Delta orf69$ and $\Delta(orf68-orf69)$ were introduced in R391 (in strain JO99) using primer pairs betWF/betWR, exoWF/exoWR, and betWF/exoWR, respectively, and pVI36 as a template. $\Delta mobI$ and Δint mutations were created in R391 using primer pairs orfXRWF/orfXRWR and intRWF/intRWR, respectively, and pKD3 as a template. SXT deletion mutants of *mobI* (VB119) and *int* (BI554) were already available [52,64].

lacZ-tagged SXT was constructed by inserting P_{lac} -*lacZ* between *traG* and *eex* using primer pair IlacWF/IlacWR and pVI42B as the template, yielding strain VB40. Similarly, *galK*-tagged R391 was created by inserting P_{lac} -*galK* between *traG* and *merR* using primer pair IgalWF/IgalWR and pVI40A as the template, yielding strain GG13. P_{lac} -*lacZ* and P_{lac} -*galK* were also introduced into strains containing SXT and R391 deletion mutants, using P1*vir* generalized transduction and *E. coli* VB40 and GG13 as donor strains. All deletion and insertion mutations were verified by PCR amplification using primers flanking the deletion, cloning and sequencing.

Construction of strains containing tandem arrays of SXT and R391

Strains containing tandem arrays were constructed by successively transferring SXT::*lacZ* or and R391::*galK* (or their corresponding deletion derivatives) into VB112, yielding strains GG61 to GG65 and GG93. The *recA* null strains GG66 to GG70 and GG102 were created in a similar fashion except that pVI67 was introduced into GG55 prior to the transfer of the ICEs. We used the *mobI* expression vector pMobI-B [52] to mobilize Δ *mobI* ICEs in the construction of strain GG125. We verified that the deletion of *mobI* did not impair SXT or R391 excision using a real-time PCR quantification assay designed to determine the relative proportion of *attP* and *attB* sites per 100 chromosomes as described previously [51]. The *int* expression vector pVI68 was used to mobilize Δ *int* ICEs in the construction of GG171. All strains harboring tandem arrays were tested to determine the relative positions of SXT and R391 in the tandem array by PCR amplification of the leftmost and rightmost ICE-chromosome junctions with primers pairs primer 6/primer 4 and primer 8/primer 9 described by Hochhut et al. [33].

Conjugation assays and detection of hybrid ICEs

Conjugation assays were performed by mixing equal volumes of overnight cultures of donor and recipient strains grown overnight at 37°C. The cells were harvested by centrifugation, washed in 1 volume of LB broth and resuspended in 1/20 volume of LB broth. The mixtures were then deposited on LB agar plates and incubated at 37°C for 6 hours. The cells were recovered from the plates in 1 ml of LB broth and serial dilutions were prepared. Donors, recipients and exconjugants were selected on LB agar plates containing appropriate antibiotics.

The *setDC* expression vector pVI67 was used in mating assays involving *recA* donor strains. In these experiments, donor strains were grown overnight at 30°C and then cultures were shifted to 42°C for 15 minutes prior to contact with the recipient strain, to induce expression of SetC and SetD.

MCGX indicator agar medium plates supplemented with appropriate antibiotics were used to determine whether SXT, R391, SXT-R391 tandem arrays, or hybrid elements were present in exconjugant colonies or in donor colonies in experiments assessing the necessity of conjugative transfer or excision in hybrid ICE formation. The hybrid ICE detection technique was validated by PCR screening of exconjugant colonies using the primer pairs VISLF/VISLR3, 10SF13/SXT1-13, YND2/ORF16, VISRF/VISRR, VISLF/VISLR2, MER104A/MER103B and VISRF/VISRR2 as described by Burrus and Waldor [37].

Molecular biology methods

Plasmid DNA was prepared using a QIAprep Spin Miniprep kit (Qiagen) according to manufacturer's instructions. All the enzymes used in this study were purchased from New England BioLabs. PCR assays were performed with the primers described in Table S1 in 20 μ l reactions with 1 U of *Taq* DNA polymerase; 1 μ l of a mixture of one colony resuspended in 10 μ l of HyPure Molecular Biology Grade Water (HyClone) was used as a template in PCR reactions. 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 appropriated annealing temperature, and 1 minute/kb at 72°C; and (iii) 5 min at 72°C. When necessary, PCR products were purified using a QIAquick PCR Purification kit (Qiagen) according to manufacturer's instructions. *E. coli* was transformed by electroporation as described by Dower et al [65] in a BioRad GenePulser Xcell apparatus set at 25 μ F, 200 Ω and 1.8 kV using 0.1 cm gap electroporation cuvettes.

Acknowledgments

We thank Davina Cloutier for technical assistance. We are grateful to Brigid Davis for helpful comments on the manuscript.

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. Burrus V, Pavlovic G, Decaris B, Guedon G (2002) Conjugative transposons: the tip of the iceberg. *Mol Microbiol* 46: 601-610.
3. Lawley TD, Klimke WA, Gubbins MJ, Frost LS (2003) F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett* 224: 1-15.
4. Bennett PM (2008) Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *Br J Pharmacol* 153 Suppl 1: S347-357.
5. Burrus V, Marrero J, Waldor MK (2006) The current ICE age: biology and evolution of SXT-related integrating conjugative elements. *Plasmid* 55: 173-183.
6. Welch TJ, Fricke WF, McDermott PF, White DG, Rosso ML, et al. (2007) Multiple antimicrobial resistance in plague: an emerging public health risk. *PLoS ONE* 2: e309.
7. Whittle G, Shoemaker NB, Salyers AA (2002) The role of *Bacteroides* conjugative transposons in the dissemination of antibiotic resistance genes. *Cell Mol Life Sci* 59: 2044-2054.
8. Gomis-Ruth FX, Coll M (2006) Cut and move: protein machinery for DNA processing in bacterial conjugation. *Curr Opin Struct Biol* 16: 744-752.
9. Salyers AA, Shoemaker NB, Li LY (1995) In the driver's seat: the *Bacteroides* conjugative transposons and the elements they mobilize. *J Bacteriol* 177: 5727-5731.
10. 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.
11. Scott JR, Churchward GG (1995) Conjugative transposition. *Annu Rev Microbiol* 49: 367-397.
12. Burrus V, Waldor MK (2004) Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol* 155: 376-386.

13. 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.
14. Burrus V, Quezada-Calvillo R, Marrero J, Waldor MK (2006) SXT-related integrating conjugative element in New World *Vibrio cholerae*. Appl Environ Microbiol 72: 3054-3057.
15. Osorio CR, Marrero J, Wozniak RA, Lemos ML, Burrus V, et al. (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.
16. Juiz-Rio S, Osorio CR, de Lorenzo V, Lemos ML (2005) Subtractive hybridization reveals a high genetic diversity in the fish pathogen *Photobacterium damsela* subsp. *piscicida*: evidence of a SXT-like element. Microbiology 151: 2659-2669.
17. Iwanaga M, Toma C, Miyazato T, Insisiengmay S, Nakasone N, et al. (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.
18. Ehara M, Nguyen BM, Nguyen DT, Toma C, Higa N, et al. (2004) Drug susceptibility and its genetic basis in epidemic *Vibrio cholerae* O1 in Vietnam. Epidemiol Infect 132: 595-600.
19. Dalsgaard A, Forslund A, Sandvang D, Arntzen L, Keddy K (2001) *Vibrio cholerae* O1 outbreak isolates in Mozambique and South Africa in 1998 are multiple-drug resistant, contain the SXT element and the *aadA2* gene located on class 1 integrons. J Antimicrob Chemother 48: 827-838.
20. Coetzee JN, Datta N, Hedges RW (1972) R factors from *Proteus rettgeri*. J Gen Microbiol 72: 543-552.
21. Hochhut B, Lotfi Y, Mazel D, Faruque SM, Woodgate R, et al. (2001) Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins. Antimicrob Agents Chemother 45: 2991-3000.

22. Taviani E, Ceccarelli D, Lazaro N, Bani S, Cappuccinelli P, et al. (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.
23. 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.
24. Opintan JA, Newman MJ, Nsiah-Poodoh OA, Okeke IN (2008) *Vibrio cholerae* O1 from Accra, Ghana carrying a class 2 integron and the SXT element. *J Antimicrob Chemother* 62: 929-933.
25. Mwansa JC, Mwaba J, Lukwesa C, Bhuiyan NA, Ansaruzzaman M, et al. (2007) Multiply antibiotic-resistant *Vibrio cholerae* O1 biotype El Tor strains emerge during cholera outbreaks in Zambia. *Epidemiol Infect* 135: 847-853.
26. Pugliese N, Maimone F, Scrascia M, Materu SF, Pazzani C (2009) SXT-related integrating conjugative element and IncC plasmids in *Vibrio cholerae* O1 strains in Eastern Africa. *J Antimicrob Chemother* 63: 438-442.
27. Hochhut B, Waldor MK (1999) Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. *Mol Microbiol* 32: 99-110.
28. 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.
29. Boltner D, MacMahon C, Pembroke JT, Strike P, Osborn AM (2002) R391: a conjugative integrating mosaic comprised of phage, plasmid, and transposon elements. *J Bacteriol* 184: 5158-5169.
30. 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.
31. Wozniak RAF, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, et al. (2009) Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet*: in press.

32. Bordeleau E, Brouillette E, Robichaud N, Burrus V (2009) 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: In press.
33. 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.
34. Marrero J, Waldor MK (2005) Interactions between inner membrane proteins in donor and recipient cells limit conjugal DNA transfer. Dev Cell 8: 963-970.
35. Marrero J, Waldor MK (2007b) Determinants of entry exclusion within Eex and TraG are cytoplasmic. J Bacteriol 189: 6469-6473.
36. Davis BM, Waldor MK (2000) CTXphi contains a hybrid genome derived from tandemly integrated elements. Proc Natl Acad Sci U S A 97: 8572-8577.
37. Burrus V, Waldor MK (2004) Formation of SXT tandem arrays and SXT-R391 hybrids. J Bacteriol 186: 2636-2645.
38. 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.
39. Stahl FW, McMilin KD, Stahl MM, Crasemann JM, Lam S (1974) The distribution of crossovers along unreplicated lambda bacteriophage chromosomes. Genetics 77: 395-408.
40. Stahl MM, Thomason L, Poteete AR, Tarkowski T, Kuzminov A, et al. (1997) Annealing vs. invasion in phage lambda recombination. Genetics 147: 961-977.
41. Kuzminov A (1999) Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. Microbiol Mol Biol Rev 63: 751-813.
42. Unger RC, Clark AJ (1972) Interaction of the recombination pathways of bacteriophage lambda and its host *Escherichia coli* K12: effects on exonuclease V activity. J Mol Biol 70: 539-548.

43. Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* 98: 6742-6746.
44. Poteete AR, Fenton AC (1993) Efficient double-strand break-stimulated recombination promoted by the general recombination systems of phages lambda and P22. *Genetics* 134: 1013-1021.
45. Poteete AR, Fenton AC, Nadkarni A (2004) Chromosomal duplications and cointegrates generated by the bacteriophage lambda Red system in *Escherichia coli* K-12. *BMC Mol Biol* 5: 22.
46. Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, et al. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A* 97: 5978-5983.
47. 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.
48. 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.
49. Sawitzke JA, Thomason LC, Costantino N, Bubunenko M, Datta S, et al. (2007) Recombineering: in vivo genetic engineering in *E. coli*, *S. enterica*, and beyond. *Methods Enzymol* 421: 171-199.
50. 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.
51. Burrus V, Waldor MK (2003) Control of SXT integration and excision. *J Bacteriol* 185: 5045-5054.
52. 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.

53. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, et al. (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 13: 721-731.
54. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32: W273-279.
55. Beaber JW, Hochhut B, Waldor MK (2004) SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 427: 72-74.
56. Singer M, Baker TA, Schnitzler G, Deischel SM, Goel 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.
57. Haldimann A, Wanner BL (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol* 183: 6384-6393.
58. Poteete AR (2008) Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol Microbiol* 68: 66-74.
59. Little JW (1967) An exonuclease induced by bacteriophage lambda. II. Nature of the enzymatic reaction. *J Biol Chem* 242: 679-686.
60. Hatfull GF (2008) Bacteriophage genomics. *Curr Opin Microbiol* 11: 447-453.
61. Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, et al. (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol* 299: 27-51.
62. Martinsohn JT, Radman M, Petit MA (2008) The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. *PLoS Genet* 4: e1000065.
63. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* 77: 61-68.

64. 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.
65. Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res 16: 6127-6145.

2.2. Régulation du système de recombinaison homologue des ICEs SXT/R391 et étude de la distribution phylogénétique de systèmes apparentés

2.2.1. Présentation de l'article

Le deuxième objectif du projet de doctorat a visé à déterminer les facteurs et les conditions impliqués dans la régulation de l'expression du système de recombinaison homologue des ICEs SXT/R391. Par des analyses transcriptomiques, j'ai démontré que les gènes *bet* et *exo* sont induits lors de la réponse SOS de l'hôte, par l'intermédiaire des activateurs de transcription principaux de l'ICE, SetC et SetD. L'analyse de l'organisation génétique du locus comprenant les gènes de recombinaison a permis de mettre en évidence qu'ils font partie d'un opéron polycistronique de plus de 11 kb qui comprend 12 autres gènes du squelette conservé des ICEs. Ces gènes, dont la fonction précise n'est pas déterminée, ne sont pas impliqués dans les fonctions d'intégration/excision ou de transfert conjugatif (Wozniak *et al.*, 2009). Cet opéron est exprimé à partir d'un promoteur inductible par les dommages causés à l'ADN situé dans la région intergénique entre les gènes *s063* (mobilisation) et *s089* du squelette conservé. Mes résultats démontrent que bien que ce transcrit soit produit à haut niveau en réponse aux dommages à l'ADN, une forte régulation traductionnelle agit sur l'expression de *bet* et *exo*, probablement par l'intermédiaire d'atténuateurs de traduction localisés en 5' de chaque gène. Mes résultats montrent que la délétion du gène *ssb*, situé dans la même unité transcriptionnelle que *bet* et *exo*, augmente de façon significative la formation d'hybrides.

Afin de répondre au troisième objectif, j'ai procédé à une analyse *in silico* visant à identifier tous les systèmes de recombinaison composés de protéines ayant les mêmes domaines fonctionnels que Bet (RecT) et Exo (YqaJ) présents dans les génomes séquences de microbes, plasmides et bactériophages. Cela a permis de démontrer que des systèmes orthologues sont

grandement distribués dans des souches bactériennes génétiques très éloignées et qu'ils sont également transportés par divers éléments génétiques mobiles. Finalement, les résultats de l'analyse des relations phylogénétiques entre les protéines composant ces systèmes suggèrent que les recombinaisons et exonucléases qui composent un même système ont parfois des origines évolutives différentes, soulignant la pression sélective importante qui agit pour maintenir la fonctionnalité de systèmes composés de protéines coopérantes mais d'origine différentes.

Ces résultats font l'objet d'un article soumis à la revue *Journal of Bacteriology*, sous le titre *DNA Damaging Agents Induce the RecA-Independent Homologous Recombination Functions of SXT/R391 ICEs* (Numéro de soumission : JB02090-12).

2.2.2. Contribution à l'article

J'ai réalisé tout le travail menant à l'obtention et à la présentation des résultats présentés dans cet article : approche expérimentale, manipulations et expériences, analyses génétiques et informatiques, rédaction du manuscrit et préparation des figures. Le manuscrit a été révisé et corrigé par V. Burrus.

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

Running Title: Activation of SXT/R391 homologous recombination

Key words: Integrating conjugative elements, horizontal gene transfer, recombination, exonuclease, SOS response, regulation, evolution, IncA/C plasmids

Authors: Geneviève Garriss, Vincent Burrus#.

Affiliation: Département de Biologie, Université de Sherbrooke, 2500 boulevard de l'Université, Sherbrooke, Québec, Canada, J1K 2R1.

vincent.burrus@usherbrooke.ca

Abstract

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 they encode. Here, we report that expression of this recombination system, which consists of the single-strand DNA annealing protein/recombinase Bet (RecT homologue) and the exonuclease Exo (YqaJ homologue), 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 ICE-conserved 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 subjected to additional strong regulatory mechanisms at the translational level. We also show that an ICE-encoded putative single-strand 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 distant 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.

Introduction

Bacteria have the capacity of rapidly adapting 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. The highly plastic nature of bacterial genomes relies mainly on two key factors: horizontal gene transfer (HGT) and homologous recombination. Homologous recombination generates genetic diversity by recombining similar yet non identical sequences and by stably integrating features, acquired by HGT, that will then be vertically transmitted (22). HGT occurs through transformation, transduction or conjugation and allows for individual bacterial cells to share genes that provide them with selective advantages. HGT events are mediated by an array of mobile genetic elements (MGEs) such as conjugative plasmids, bacteriophages and integrating conjugative elements (ICEs), also known as conjugative transposons (25, 34, 45, 54).

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 γ -proteobacterial species (55, 56). These elements are found site-specifically integrated in the chromosome of their host. Under certain conditions ICEs of the SXT/R391 family excise from their host's chromosome as a circular, covalently closed molecule and transfer via conjugation to a new recipient cell in the form of a single-stranded DNA substrate (13). 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 (55). 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 in the 5' end of *prfC*, which encodes the peptide chain release factor RF3 (11). 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-signalling pathways (2, 5-7, 10, 28, 32, 35, 47). 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 in conditions that trigger the host's SOS response. In non-inducing conditions, the master repressor SetR encoded by these elements prevents the expression of the transcriptional activators SetCD. DNA-damaging agents such as mitomycin C, UV light or the antibiotic ciprofloxacin alleviate this repression, allowing for SetCD-mediated activation of the conserved genes involved in excision and conjugative transfer (8).

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 (26). The recombination system carried by SXT/R391 ICEs comprises two genes: *s065*, encoding a single-strand DNA annealing protein (SSAP) (21), and *s066*, encoding a 5'-3' exonuclease (17). These proteins are respectively related to λ Bet and λ exo, and we will thus subsequently 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 enterobacterial species from agricultural sources and it has been recently demonstrated that most of the conserved genes of SXT/R391 ICEs are also present in this family of conjugative plasmids (55).

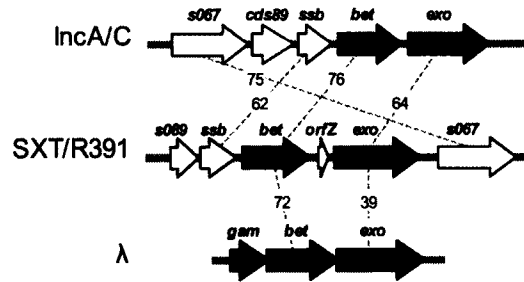


Figure 1. Schematic comparison of the recombination loci of SXT/R391 ICEs, IncA/C plasmids and bacteriophage λ .

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 grey. Dotted lines and associated numbers indicate the percent of similarity between orthologous genes.

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 (18, 48). λ *gam*, an inhibitor of the host's exonuclease RecBCD found directly upstream of λ *bet*, is transcribed with the recombination genes (Fig. 1). The product of λ *gam* protects the phage's linear dsDNA by inhibiting the ability of RecBCD to bind dsDNA ends (43). 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 DNA damaging agents via the main transcriptional activators SetCD. Our results also indicate that the recombination functions are further regulated at the

translational level, most likely by two conserved translational attenuators located upstream of *bet* and *exo*. Furthermore, deletion of *ssb* significantly increased Bet/Exo mediated hybrid ICE formation, suggesting that it could act as a modulator of the recombination activity. Finally, *in silico* analyses 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 on 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 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 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% w/v) or mitomycin C (100 or 200 ng/ml).

Plasmid and strain construction. The oligonucleotides used are described in Table S1. Deletion and fusion mutants were constructed by using the one-step chromosomal gene inactivation technique of Datsenko and Wanner (20). All deletions were designed to be non-polar. The *ΔrecA* mutation was introduced in *E. coli* BW25113 using primer pair *recAWF/recAWR* (26) and plasmid pVI36 as the template. The *Δs089* and *Δssb* mutations were introduced in SXT (in strain HW220) using primer pairs *s089WF/s089WR* and *ssb1WF/ssb1WR*, respectively, and template plasmid pVI36. The corresponding mutations were introduced in R391 (in strain JO99) using 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 P1_{vir} generalized 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 primers pairs *65lacZ42BF/65lacZ42BR* and *66lacZ42BF/66lacZ42BR*, respectively and plasmid pVI42B as the template. 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 primer pairs setDF/setC2R, subcloning into pCR2.1TOPO (Invitrogen) and recloning into pBAD-TOPO (Invitrogen) following manufacturer's instructions. Plasmid pGG7 was similarly constructed using primer pairs orfZTOPOF/corfZR. Plasmid pGG32 was constructed by amplifying the intergenic region comprised between *s063* and *s089* in SXT from *E. coli* HW220 using primer pair 6389F/6389R and cloning into pCR2.1TOPO (Invitrogen). All plasmid inserts were verified by sequencing by Centre d'Innovation Génome Québec (McGill University, Montreal, Quebec, Canada). *E. coli* was transformed by electroporation in 1mm-gap cuvettes in a GenePulser Xcell apparatus (BioRad) set at 2.5 μ F, 200 Ω and 1.8 kV.

RNA extraction and cDNA synthesis. Briefly, RNA extractions were performed as follows. Cultures were grown for 14-16 hrs in LB broth containing appropriated antibiotics, diluted 1:100 in fresh media in absence of antibiotics (except when ampicillin was added for the maintenance of the pGG2B vector) and grown to an OD₆₀₀ of 0.2. They were then diluted a second time 1:100 and grown to an OD₆₀₀ of 0.2. Each culture was then split in 2 flasks and one flask for each culture was induced with 100 ng/ml mitomycin C (MMC) or 0.02% l-arabinose (when using pGG2B), for 2 hours at 37°C. A 1 mL sample of each culture was used for total RNA extraction using RNeasy Mini Kit 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 manufacturer's instructions. Complementary DNA was synthesized using 1 μ g of RNA, 50 ng of random hexamers or 2 pmoles of gene-specific primers exoRT or s073RT (Integrated DNA Technologies) (Table S1) and the reverse transcriptase SuperScript II (Invitrogen), following manufacturer's instructions. Control reactions in the absence of reverse transcriptase (NoRT) 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*) (19) or of each gene of interest (GOI) (Table S1). The expression ratios were calculated using the ($2^{-\Delta\Delta C_t}$) method and the difference in the transcript levels of each GOI was calculated compared to the transcript level of *rpoZ* and *gyrA*, using the ($2^{-\Delta C_t}$) 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 absence (CTL) or presence of mitomycin C (MMC) at 100 ng/ml, with radio-labeled TSS89up4 primer ($[\gamma\text{-}^{32}\text{P}]\text{-ATP}$, 3000 ci/mmol, 10 mCi/ml (Perkin Elmer)) and the Primer Extension System-AMV-Reverse Transcriptase kit (Promega) following manufacturer's instructions. A control reaction in which water replaced the RNA was performed. 3 μl 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 (BioRad) alongside the sequencing reaction of the intergenic region comprised 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, $[\alpha\text{-}^{32}\text{P}]\text{-dATP}$ (800 Ci/mmol, 10 mCi/ml) (Perkin Elmer), primer TSS89up4 and the Sequenase Version 2.0 kit (Affymetrix). Signal strength was detected using a Storm 860 Molecular Imager (GMI).

Semi-quantitative PCR. PCR reactions aimed at amplifying *s063*, *s989* and *bet* were carried out using cDNA prepared with genes-specific primers exoRT or s073RT (Table S1). Each reaction contained 1 μM each primer, 100 μM each dNTPs and 1 U *Taq* DNA polymerase (New England Biolabs); 1 μl of a 1:10 dilution of the cDNA synthesis reactions, 1 μl of a 1:10

dilution of the NoRT reactions (negative control), or 500 pg BW25113 SXT gDNA (positive control) were used as template for PCR amplification. PCR conditions were as follows: i) 3 min at 95°C, ii) 28 cycles of 30 sec at 95°C, 30 sec at the appropriate annealing temperature and 30 sec at 72°C, iii) 5 min at 72°C. After PCR, the samples were mixed with loading dye (orange G in 30% glycerol v/v) at a final concentration of 1X and 15 μ l of each reaction were migrated in a 2% agarose gel in 1X TAE buffer, alongside 1 μ g 2-Log DNA Ladder (New England Biolabs). Migrated gels were colored in an ethidium bromide bath (3 μ g/ml) for 30 min and then soaked for 30 min in distilled water. Gels were visualized under UV light in a GelDocXR (BioRad), and analyzed using the Quantity One v6.3 software (BioRad).

β -galactosidase assays. β -galactosidase assays were carried out using translational fusions of *lacZ* to *bet* or *exo* in BW25113 SXT, under induced or control conditions. Briefly, strains were grown for 14-16 hrs in LB with the appropriated antibiotics then refreshed 1:100 and grown to an OD₆₀₀ of 0.2 in fresh LB broth in absence of antibiotics, except when using ampicillin to maintain pGG2B. Cultures were split in 4 ml samples and grown in absence or presence of MMC (100 ng/ml or 200 ng/ml) for 2 or 16hrs. β -galactosidase activity was assessed using 500 μ l samples as described elsewhere (42).

Bioinformatics analyses. The predicted proteomes of 2,714 plasmids, 4,093 viruses and 1,706 complete microbial genomes available in March 2012 were extracted from the RefSeq database (49) and analyzed to identify putative recombination systems similar to the SXT/R391 system. We screened this dataset using profile Hidden Markov models of the RecT, YqaJ and Gam families with HMMsearch from the HMMER v3.0 software package (23). The HMM profiles of Bet, Exo and λ Gam were recovered from the Pfam 25.0 database: Bet: RecT, PF03837; Exo: YqaJ, PF09588; λ Gam: Gam, PF06064. The sequences corresponding to each identified protein were downloaded and aligned using MUSCLE multiple sequence alignment software (24). Maximum-likelihood phylogenies of RecT and

YqaJ proteins were generated using the PhyML v3.0 program (27) 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 non-parametric bootstrap (100 replicates). Phylogenetic analyses were computed using amino acid alignment generated by MUSCLE and the poorly aligned regions were removed with the trimAl v1.2 software using the automated heuristic approach (15) prior to phylogenetic analyses. Phylogenetic trees were viewed in iTOL v2 (38) and are available as a shared project on the iTOL website (<http://itol.embl.de/shared/GGarriss>). Sequence alignments were performed using Clustal W (37).

Table 1. List of *E. coli* strains and plasmids used in this study.

Strain or Plasmid	Relevant genotype or phenotype	Source or reference
<i>E. coli</i>		
BW25113	F ⁻ , Δ (<i>araD-araB</i>)567, Δ <i>lacZ</i> 4787(:: <i>rrnB</i> -3), λ , <i>rph</i> -1, Δ (<i>rhaD-rhaB</i>)568, <i>hsdR</i> 514	(20)
CAG18439	MG1655 <i>lacZ</i> U118 <i>lacI</i> 42::Tn10 (Tc ^R)	(50)
HW220	CAG18439 <i>prfC</i> ::SXT (Tc ^R Su ^R Tm ^R)	(31)
JO99	CAG18439 <i>prfC</i> ::R391 (Tc ^R Kn ^R)	(30)
VB40	CAG18439 Δ <i>lacZ</i> <i>prfC</i> ::SXT:: <i>lacZ</i>	(26)
GG13	CAG18439 Δ <i>galK</i> <i>prfC</i> ::R391:: <i>galK</i>	(26)
VB17	BW25113 <i>prfC</i> ::SXT	VB
GG247	BW25113 Δ <i>recA</i> <i>prfC</i> ::SXT	This study
DPL247	BW25113 <i>prfC</i> ::SXT Δ <i>setDC</i>	D. Poulin-Laprade
GG55	MG1655 Rf ^R Δ <i>recA</i>	(26)
GG66	VB112 <i>prfC</i> ::[SXT:: <i>lacZ</i>]-[R391:: <i>galK</i>]	(26)
GG67	VB112 <i>prfC</i> ::[SXT:: <i>lacZ</i> Δ <i>s089</i>]-[R391:: <i>galK</i> Δ <i>s089</i>]	This study
GG68	VB112 <i>prfC</i> ::[SXT:: <i>lacZ</i> Δ <i>ssb</i>]-[R391:: <i>galK</i> Δ <i>ssb</i>]	This study
VB47	CAG18439 Δ <i>galK</i> Δ <i>recA</i>	(26)
GG209	BW25113 <i>prfC</i> ::SXT Δ <i>bet</i> :: <i>lacZ</i>	This study
GG215	BW25113 <i>prfC</i> ::SXT Δ <i>exo</i> :: <i>lacZ</i>	This study
Plasmids		
pGG2B	pBAD30:: <i>setDC</i>	This study
pGG7	pBAD:: <i>orfZ</i>	This study
pGG32	pCR2.1::SXT IgR <i>s063-s089</i>	This study
pVI67	pAH57 Δ (<i>xis₁-int₁</i>):: <i>setDC</i> (Ts)	(26)
pVI36	Sp ^R template for one-step chromosomal gene inactivation	(16)
pVI42B	pVI36 BamHI::P _{<i>lac</i>} - <i>lacZ</i>	(26)

Ap^R, ampicillin resistant; Cm^R, chloramphenicol resistant; Kn^R, kanamycin resistant; Rf^R, rifampicin resistant; Su^R, sulfamethoxazole resistant; Sm^R, streptomycin resistant; Sp^R, spectinomycin resistant; Tc^R, tetracycline resistant; Tm^R, trimethoprim resistant; Ts, thermosensitive.

Results

DNA damaging agents induce the expression of ICE recombination genes

In order to gain 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 in conditions that trigger their host's SOS response, we decided to verify whether DNA damaging agents would also promote the transcription of the recombination genes. We measured by RT-qPCR the effect of the exposure to mitomycin C (MMC) on the transcript levels of *bet* and *exo*, using *E. coli* BW25113 harboring SXT (VB17). Upon exposure to MMC, the relative expressions ($2^{-\Delta\Delta C_t}$) of *bet* and *exo* were ~50 times higher than in the control condition (Fig. 2A). However, when using an isogenic *recA* null mutant (GG247), 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 $\Delta setCD$ mutant (DPL247). Our results showed that in 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 and resulted in a more than 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*.

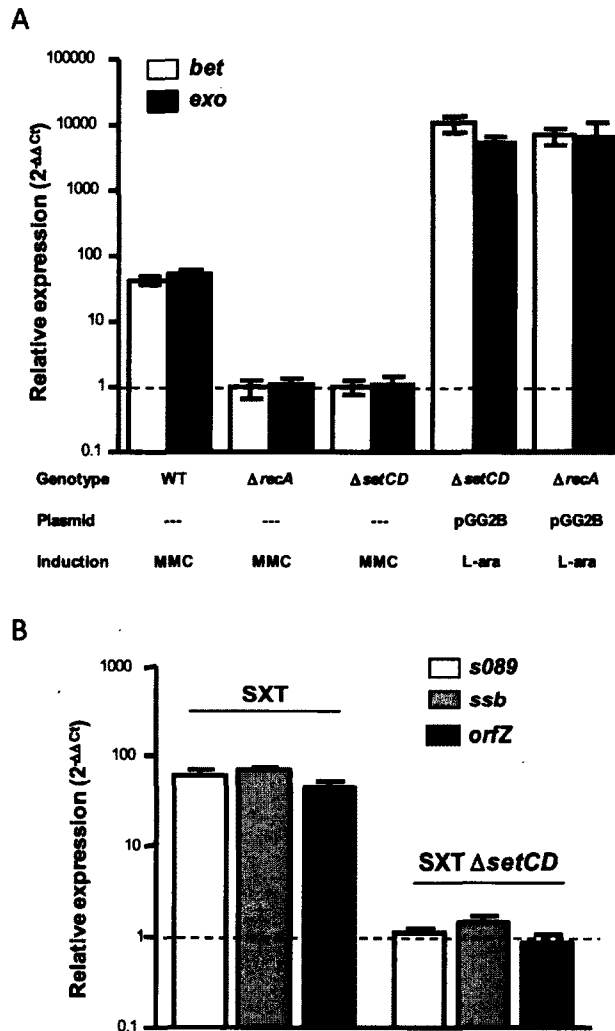


Figure 2. Expression of *bet* and *exo* is induced by mitomycin C.

(A) Relative expression of *bet* and *exo*. Genotype of ICE or strain, presence of *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 SXT or SXT $\Delta setCD$ backgrounds upon induction with MMC. Bars represent the ratio ($2^{-\Delta\Delta C_t}$) between an induced and the non-induced conditions, using *rpoZ* as an internal reference as determined by RT-qPCR, and represent mean values of 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 condition).

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 induction 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 ICE conserved DNA and is driven from a MMC-induced promoter located upstream *s089*

The similar response of *bet* and *exo* to the exposure to MMC, combined with the fact that in λ the recombination genes are present in a single transcript (18), led us to investigate if a similar operon structure is found in SXT/R391 ICEs. To determine whether *bet* and *exo* are part of the same transcript we carried out reverse transcription using a primer complementary to the 3' end of the non-template strand of *exo* (*exoRT*), on RNA extracted from VB17 under non-inducing and MMC-induced conditions (Fig. 3A). We then performed a PCR assay aiming at amplifying a 171-bp fragment at the 5' end of *bet* (Fig. 3A and 3B). Consistent with the observations presented above, our results show that the non-induced sample does not yield any amplicon for *bet*, whereas the MMC-induced one does yield an amplicon of the expected size (Fig. 3B). These results show that *bet* and *exo* are part of the same MMC-inducible transcript.

The recombination genes are located within a ~11-kb conserved region of the SXT/R391 ICEs positioned between two inversely transcribed genes, *s063* (conserved DNA) and *s074* (non-conserved gene present in SXT) (Fig. 3A) (55). Eight nearly identical and perfectly syntenous genes are located downstream of *exo* in this region. These genes play no role in conjugative

transfer and their respective functions remains obscure (55). The overall genetic organization of this locus led us to further investigate whether all the genes comprised between *s063* and *s074* are co-transcribed with *bet-exo*. We carried out the same reverse transcription experiment as above using this time a primer located at the 3' end of *s073* (*s073RT*) and RNA samples obtained from MMC-treated VB17 cells (Fig. 3A). We then amplified by PCR fragments located at the 5' end 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. 3C) and that expression of all of the genes from *s089* to *s073* is induced by MMC. Moreover, the impossibility to amplify *s063* from the same cDNA sample indicates that this transcript originates within the intergenic region lying between divergently transcribed *s063* and *s089* (Fig. 3A and 3C).

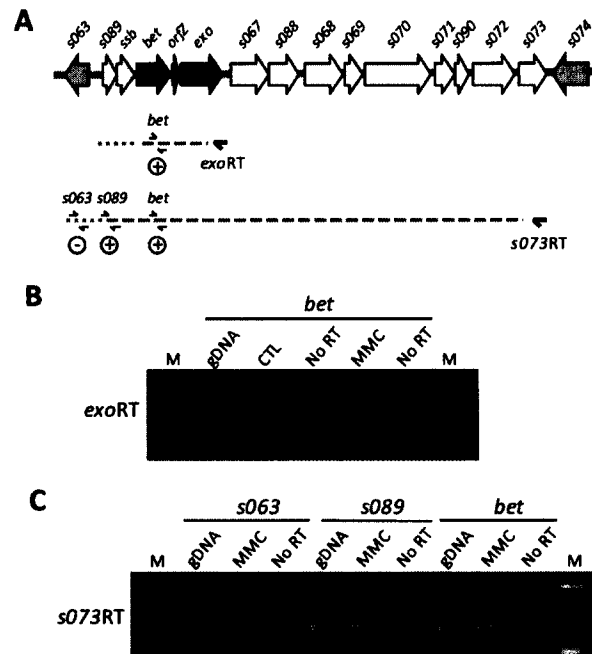


Figure 3. Operon structure of the SXT/R391 homologous recombination system.

(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; grey, immediate borders of *s089-s073* locus. Relative position of reverse transcription primers *exoRT* and *s073RT* as well as PCR primers

used to amplify *bet*, *s063* and *s089* are indicated. Dotted lines shows reverse transcription products and results of PCR are indicated as positive, (circled +) or negative (circled -). B) 2% agarose gel of assay aiming at amplifying *bet* on exoRT product. C) 2% agarose gel of assay aiming at amplifying *s063*, *s089* and *bet* on s073RT product. VB17 genomic DNA (gDNA) and reverse transcription samples in the absence of reverse transcriptase (NoRT) were used respectively as positive and negative PCR controls. non-induced control sample, CTL; MMC-induced sample, MMC; molecular weight marker, M.

In order to determine the precise location of the promoter driving the transcription of the recombination genes, we analysed the 367-bp intergenic region located between *s063* and *s089* by primer extension, using total RNA from *E. coli* VB17 prepared in non-induced 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 non-transcribed DNA strand (Fig. 4A and 4B). Despite the presence of a large 5'UTR we found no evidence of another ORF laying in the region located between the transcriptional and translation start sites of the recombination system. As expected, the intensity of the band is dependent upon 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 might be due to imprecise transcription initiation (Fig. 4A). Further analysis of the sequence upstream of C⁺¹ revealed near-consensus -35 (TTCTAA) and -10 (TACACT) promoter elements (Fig. 4B). No other DNA elements that contribute to promoter recognition by *E. coli* σ^{70} (29) could be identified in this region.

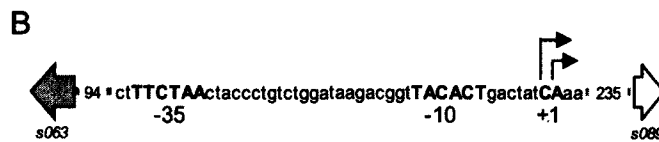
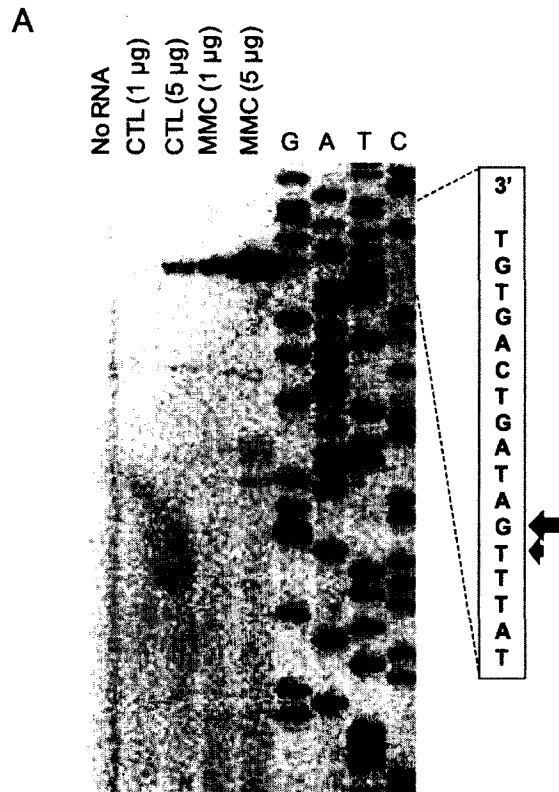


Figure 4. Promoter region of the SXT/R391 homologous recombination system.

(A) Primer extension analysis of the transcribed strand of the *s063-s089* intergenic region. NoRNA, negative control; CTL, non-induced 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. Partial read of the sequencing reaction is located on the right hand side 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 grey and white arrows, respectively. Lengths in base pairs (bp) of the spacers between represented regions are indicated.

A gene encoding a single-stranded 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 $\Delta s089$ and Δssb 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 (26). These experiments were carried out in a $\Delta recA$ background to prevent RecA-mediated recombination and ensure that all hybrid ICEs are formed by Bet/Exo-mediated recombination (26). We bypassed the requirement for RecA for the activation of expression of the conjugative transfer genes (8, 26) 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 for Bet/Exo to more efficiently catalyze inter-ICE recombination. As an ssDNA-binding protein, Ssb might compete with Bet for binding to ssDNA substrates, reducing annealing.

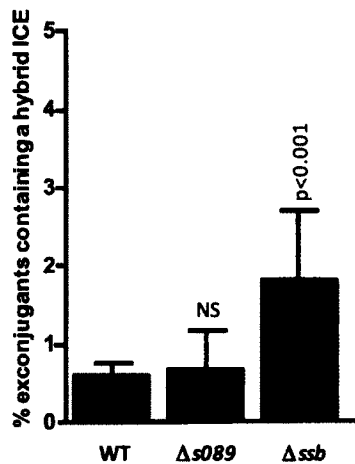


Figure 5. Impact of *s089* and *ssb* on hybrid ICE formation.

recA donor strains containing wild-type SXT-R391 tandem arrays or tandem arrays of $\Delta s089$ and Δssb mutants of SXT and R391 were used in mating experiments with a *recA* recipient strain. Bars represent the mean of at least three independent biological replicates. Standard deviations are indicated. One-way ANOVA with Dunnett's Multiple Comparison Test was used to compare the means of hybrid-ICE containing exconjugants colonies of each deletion mutant with the WT. The confidence intervals (p) for the comparisons are indicated. NS indicates the difference was not statistically different from the WT.

Expression of *bet* and *exo* is subjected to strong translational regulation

To determine whether the translation of *bet* and *exo* is consistent with the strong induction of their transcription, 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 these genes only occurred 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 in conditions similar to those used for our RT-qPCR analysis (2hrs with 100 ng/ml MMC). We confirmed by RT-qPCR with primers targeting *lacZ* that the transcript level of *lacZ* for each fusion was comparable to that obtained

for the corresponding wild-type genes (Fig. S1). 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* is increased by MMC in a WT strain (Table 2).

Table 2. Mitomycin C induces the translation of *bet* and *exo*.

Fusion	β -galactosidase activity ^a	
	CTL	MMC
<i>Δbet::lacZ</i>	0.00 ± 0.01	3.56 ± 0.93
<i>Δexo::lacZ</i>	0.00 ± 0.01	0.52 ± 0.23

^aResults are expressed in Miller Units and represent the mean and standard deviations of at least three independent biological replicates.

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 transcript amounts relatively to the reference gene *rpoZ* ($2^{-\Delta Ct}$) (Fig. 6A). However, β -galactosidase assay results demonstrate that *exo* is significantly less translated (near 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 that the difference observed between transcriptional and translational data was due to possible polar effects resulting from the *lacZ* fusions, our results suggest that expression of *bet* and *exo* is subjected to an additional translational regulatory mechanism.

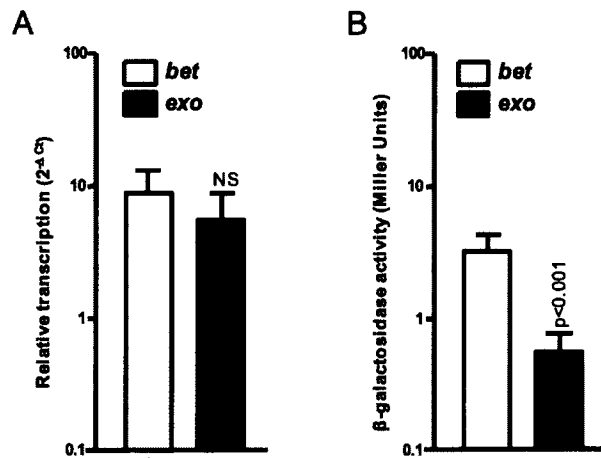


Figure 6. Relative quantification of the transcription and translation of *bet* and *exo* in presence of MMC. (A) RT-qPCR transcription results are expressed as the relative quantification ($2^{-\Delta C_t}$) of each gene relatively to *rpoZ*, in a WT background upon induction with MMC. (B) Translation results are expressed as Miller Units obtained from β -galactosidase assays with *lacZ* translational fusions of *bet* and *exo* in a WT background upon induction with MMC. Bars represent the mean value of at least three independent biological replicates. Standard deviations are indicated. One-way ANOVA with Tukey's Multiple Comparison Test was used to compare the means between *bet* and *exo* transcription and translation in a WT background. The mean of Miller Units obtained for *exo* was statistically significant from *bet* with a confidence interval of $p < 0.001$ and is indicated (A). Transcription of *exo* is not statistically different from *bet* (NS) (B).

The SXT/R391 recombination systems 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 transcription of *bet* and *exo*. Relative expression ($2^{-\Delta\Delta C_t}$) of *bet* and *exo* remained unaffected by *orfZ* overexpression (expression ratio of ≈ 1). We also tested its impact on the translation of *bet* and *exo* using single-copy chromosomal *lacZ* translational fusions and β -galactosidase assays in 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 (Fig. S2). 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 translational attenuators are found upstream *bet* and *exo*

We used the RibEx server (1) 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* (TAbet) and *exo* (TAexo) (Fig. 7). Moreover, the attenuator located upstream of *exo* has a higher predicted stability ($\Delta G = -25.50 \text{ kcal.mol}^{-1}$) than the one found upstream of *bet* ($\Delta G = -21.60 \text{ kcal.mol}^{-1}$), which could also explain the stronger translational attenuation observed for the production of Exo. These attenuators are highly conserved in all SXT/R391 ICEs (Fig. S3) and RibEx analyses place similar structures upstream the *bet* and *exo* genes found in IncA/C plasmids (Fig. S4). The presence of these putative translational attenuators is consistent with our experimental finding that the expression of *bet* and *exo* is subjected to a strong translational regulation.

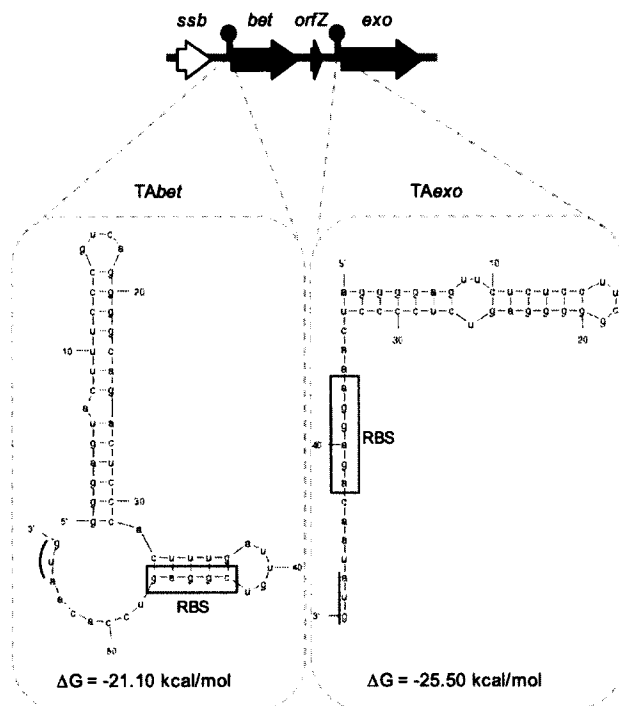


Figure 7. The recombination locus contains two predicted translational attenuators.

Presence of attenuators was determined with the RibEx server (1) and are represented by black lollipops upstream of *bet* and *exo*. Dotted rectangles contain mfold v4.6 (57) folding predictions for each attenuator, along with the sequence located between the attenuator and the translation start sites of *bet* or *exo*. Positions of the ribosome binding sites (RBS) (boxes), translation start sites (underlining) and ΔG values are indicated for each attenuator.

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 the HMMER3 software (23). Based on this analysis, we found 412 proteins containing a RecT domain and 632 proteins containing a YqaJ domain. 163 recombination systems were identified based on the presence of a RecT-like recombinase and an YqaJ-like exonuclease in the same locus (Table S2)¹.

The vast majority (125 out of 163) of the recombination systems we identified were found in sequenced microbial genomes. Each system was further analyzed to determine if it is part of 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, tail tape measure protein) in the vicinity of the recombination system. Four of the 125 systems identified in microbial genomes belonged to SXT/R391 ICEs: ICES_{puPO1} (CP000503), ICE_{PmiUSA1} (AM942759), ICE_{VchBan9} (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 out 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, pDK1), one in a plasmid of the IncT group (Rts1), as well as two in non-conjugative plasmids pMAQU02 and pALVIN02.

We analyzed the distribution of the identified recombination systems according to the taxonomic order of their host. As expected, most (69/163) recombination systems were identified in strains belonging to the *Enterobacteriales* order, the predominant hosts of

¹ Table S2 is available upon request.

lambdoid phages. 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 species belonging to the *Rhizobiales* and *Thermonoanaerobacterales* orders (Table S2 and Fig. 8).

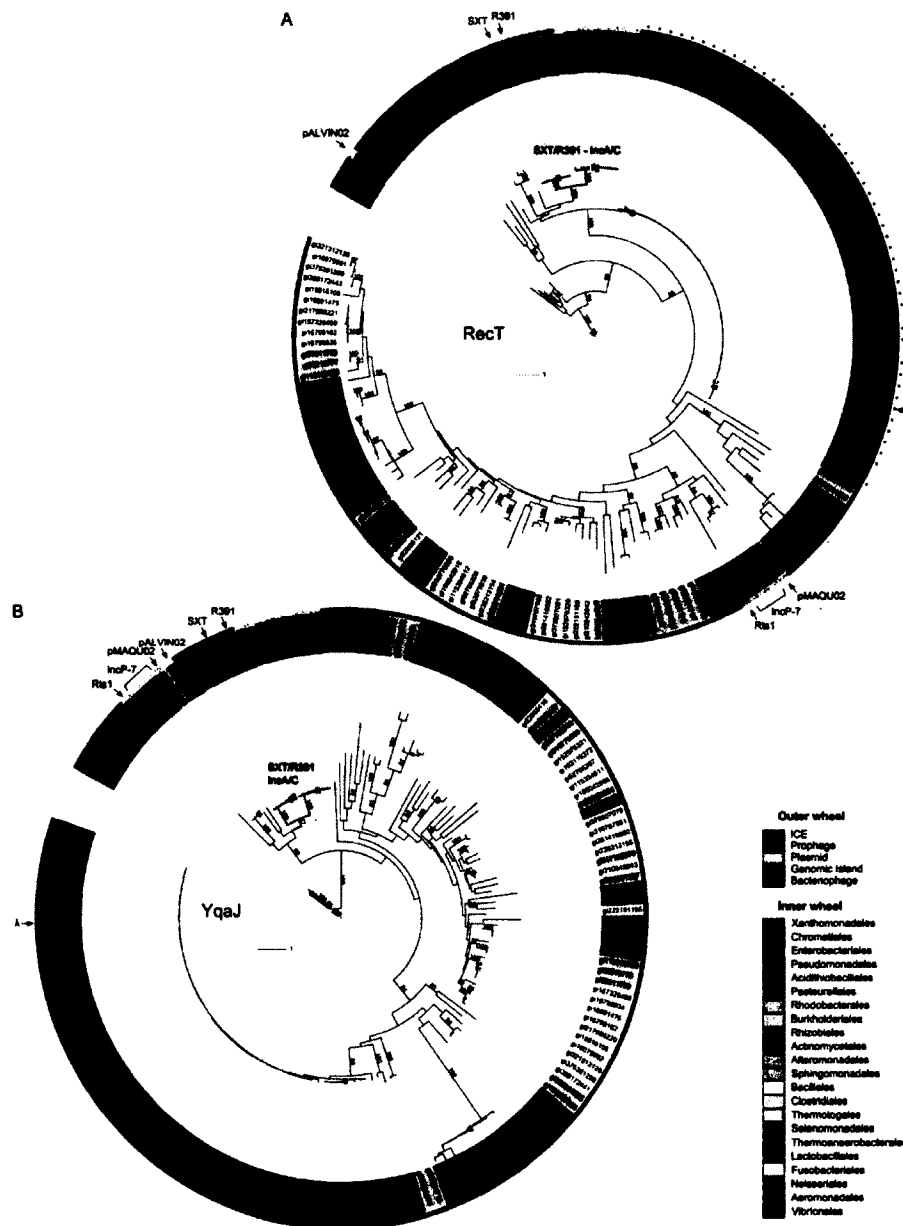


Figure 8. Phylogenetic trees of single-stranded DNA annealing proteins and cognate exonucleases encoded by gene pairs found in sequenced microbial, plasmid and bacteriophage genomes. (A) Recombination proteins containing a RecT domain. (B) Exonucleases containing a YqaJ domain. Label colors indicate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

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 and YqaJ proteins identified (Fig. 8). As SXT and R391 are found in genomes that are not fully assembled, thus absent from the RefSeq database, we manually added their respective Bet and Exo sequences to our dataset. Our analysis that RecT and YqaJ proteins identified in phages or GIs seemed to cluster relatively to their host's taxonomic order. However, proteins encoded by ICEs and plasmids tend to cluster relatively to the family of mobile genetic element to which they belong (Fig. 8). For instance, RecT and YqaJ proteins 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).

Surprisingly, our analysis shows that RecT and YqaJ partners 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 YqaJ proteins are very distantly related to those encoded by lambdoid phages (Fig. 8B). Moreover, while the YqaJ proteins encoded by all the plasmid and ICE-borne recombination systems derive from the same common ancestor (Fig. 8B), their cognate RecT proteins have different phylogenetic relationships: RecT from SXT/R391 ICEs and IncA/C plasmids form a monophyletic group, while RecT from IncP-7 plasmids (pCAR1, pCAR1.2, pDK1), 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 RecT gene (Table S2). Interestingly, pALVIN02 is the only plasmid that harbours a RecT orthologue unrelated to other plasmid-borne RecT proteins. Although most bacteriophages from the *Enterobacteriales* order contain a homologue of

λgam, none was found in the recombination systems carried by ICEs or plasmids (Fig. 8A). Finally, the only 6 occurrences of an ORF related to *orfZ* located between the *RecT* and *YqaJ* 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 (14, 26). 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 (26). 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 were 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 (7-9). 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 (8), 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 co-induction during SOS-response highlights the extent to which these elements take advantage of their host's cellular function to enhance their own diversity (26). Additionally, in the assay used to study hybrid ICE formation (26), 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. In 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 presence of DNA-

damaging agents. Baharoglu and colleagues (3, 4) 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. 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, accumulation of SetR is likely slowed down by its simultaneous auto-cleavage which is promoted by the co-protease activity of RecA bound to the incoming single-stranded DNA. Together with our results, this observation sheds a new light on the major impact that these elements can have on their host's genome plasticity.

In bacteriophage λ , λbet and λexo are part of the *pL* operon, which is also induced by DNA-damaging agents (18, 46). We, as well as others, have discussed in the past that these homologous recombination systems could act as a repair pathway for salvaging ICE or phage DNA broken by exposure to DNA-damaging agents (26, 48). Our finding that the SXT/R391 *bet* and *exo* genes are specifically expressed in 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 syntenous 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 (55). The conservation of these genes as well as their co-expression 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 slightly differs (55). To the best of our knowledge, the stimuli that trigger the conjugative transfer of IncA/C plasmids have not been determined. 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 co-expression 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-strand binding protein located immediately upstream *bet*, caused a significant near three-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 ssDNA (36, 51). 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* co-expression of SXT's Ssb-Bet-Exo diminishes by more than two-fold the recombination frequency of a linear dsDNA substrate with the *E. coli* chromosome, compared with the co-expression of only Bet-Exo (17). Additionally, like λ Bet, SXT Bet is able to recombine ssDNA sequences that possess very short homologous sequences (36 bp) (21) and it has been demonstrated that λ Bet is able to promote strand exchange and displacement using short oligonucleotides containing several mismatches as a substrate (39). 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 non-functional elements unable of transferring to a new host.

Our *in silico* analysis of the structure of several recombination systems show that while a homologue of λ Gam is encoded by numerous phages, none is found in the plasmid or ICE-

borne systems. λ Gam protects the phage linear dsDNA from degradation by inhibiting the host's RecBCD endonuclease (40, 43). Although we haven't assessed the presence of other RecBCD-inhibitors (51), the absence of λ Gam 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 non-conjugative plasmid pALVIN02, contain a gene encoding an Ssb upstream the gene encoding the RecT-like protein. 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 over-activation of the host's SOS response. Although Ssb proteins are typically recognized for their positive effect on phage homologous recombination, their specific role in the recombination catalyzed by systems found in conjugative elements has not been investigated yet.

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 (Fig. S2) suggests that it results from an insertion event and supports the idea that it is not relevant to the recombination mechanism of RecT-YqaJ systems.

The expression profiles and genetic organization of *bet* and *exo* revealed that although these genes are transcribed at a high level in presence of MMC, their translation is weak. The presence of two translational attenuators, one located upstream of *bet* (TAbet) and the other upstream of *exo* (TAexo), likely account for the weak translation of the Bet- and Exo-LacZ fusion proteins. The higher predicted stability of TAexo could account for the significantly

lower expression of Exo compared to Bet. 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 (44). The process by which these attenuators regulate the expression of *bet* and *exo* is currently under investigation, however, their presence seems to limit the production of a certain level of the Bet and Exo proteins, regardless of the amount of transcript produced. Interestingly, while IncA/C plasmids lack an *orfZ* gene, attenuators similar to those predicted in SXT/R391 are found in the same relative positions (upstream the *bet* and *exo* IncA/C genes) providing additional evidence for a common mode of translational regulation between these two recombination systems. Combined with our previously published results (26), our observation indicate that only very low levels of both Bet and Exo are sufficient to catalyze the formation of hybrid ICEs at high frequency.

RecT-YqaJ recombination systems 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 RecT and YqaJ proteins from different origins suggests that those derived from phages cluster relatively to the taxonomic order of their host, while those encoded by plasmids and ICEs cluster relatively to the incompatibility group or family. This trend probably results from the narrow host range of phages, which would lead to fewer opportunities of interspecies gene exchanges compared to plasmids and ICEs. Most mobile genetic elements are usually considered as combinations of interchangeable functional modules that were originally defined as belonging to bacteriophages, conjugative plasmids or transposons (12, 52, 53, 56). The widespread presence of RecT/YqaJ recombination systems in these mobile genetic elements might provide a suitable explanation for the exchange of such modules between different types of mobile elements, with requirement for only very short homology regions on each side the different modules. The differences in the evolutionary relationships between RecT and YqaJ proteins suggests 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 RecT-YqaJ pair. While it is surprising that functionally cooperating proteins have been acquired from different origins, it is known that RecT proteins are sometimes found associated with exonucleases belonging to the RecE family, which has led to the assumption that the similar operon structure of different systems could arise independently due the selective pressure for the co-expression of cooperating proteins (33).

The prevalence of similar recombination genes in many mobile genetic elements found in nearly all bacterial orders suggest 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 the mean of homologous recombination systems requiring much shorter identity region than RecA-dependent pathways.

Aknowledgments

We thank Dominic Poulin-Laprade for the gift of strain DPL247. 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. GG was the recipient a doctoral fellowship (Fonds de Recherche du Québec, Nature et Technologies, Qc, Canada) and VB holds a Canada Research Chair in bacterial genetics and is a member of the FRSQ-funded Centre de Recherche Clinique Étienne-Le Bel.

References

1. **Abreu-Goodger, C., and E. Merino.** 2005. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res* **33**:W690-2.
2. **Ahmed, A. M., S. Shinoda, and T. Shimamoto.** 2005. A variant type of *Vibrio cholerae* SXT element in a multidrug-resistant strain of *Vibrio fluvialis*. *FEMS Microbiol Lett* **242**:241-7.
3. **Baharoglu, Z., D. Bikard, and D. Mazel.** 2010. Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLoS Genet* **6**:e1001165.
4. **Baharoglu, Z., and D. Mazel.** 2011. *Vibrio cholerae* triggers SOS and mutagenesis in response to a wide range of antibiotics, a route towards multi-resistance. *Antimicrob Agents Chemother* doi:10.1128/AAC.01549-10.
5. **Bani, S., P. N. Mastromarino, D. Ceccarelli, A. Le Van, A. M. Salvia, Q. T. Ngo Viet, D. H. Hai, D. Bacciu, P. Cappuccinelli, and M. M. Colombo.** 2007. Molecular characterization of ICE V_{ch} Vie0 and its disappearance in *Vibrio cholerae* O1 strains isolated in 2003 in Vietnam. *FEMS Microbiol Lett* **266**:42-8.
6. **Beaber, J. W., V. Burrus, B. Hochhut, and M. K. Waldor.** 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**:2065-70.
7. **Beaber, J. W., B. Hochhut, and M. K. Waldor.** 2002. Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *J Bacteriol* **184**:4259-69.
8. **Beaber, J. W., B. Hochhut, and M. K. Waldor.** 2004. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* **427**:72-4.

9. **Beaber, J. W., and M. K. Waldor.** 2004. Identification of operators and promoters that control SXT conjugative transfer. *J Bacteriol* **186**:5945-9.
10. **Bordeleau, E., E. Brouillette, N. Robichaud, and V. Burrus.** 2009. 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**:510-23.
11. **Burrus, V., J. Marrero, and M. K. Waldor.** 2006. The current ICE age: biology and evolution of SXT-related integrating conjugative elements. *Plasmid* **55**:173-83.
12. **Burrus, V., G. Pavlovic, B. Decaris, and G. Guedon.** 2002. The ICES₁ 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.
13. **Burrus, V., and M. K. Waldor.** 2003. Control of SXT integration and excision. *J Bacteriol* **185**:5045-54.
14. **Burrus, V., and M. K. Waldor.** 2004. Formation of SXT tandem arrays and SXT-R391 hybrids. *J Bacteriol* **186**:2636-45.
15. **Capella-Gutierrez, S., J. M. Silla-Martinez, and T. Gabaldon.** 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972-3.
16. **Ceccarelli, D., A. Daccord, M. Rene, and V. Burrus.** 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.
17. **Chen, W. Y., J. W. Ho, J. D. Huang, and R. M. Watt.** 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.
18. **Court, D. L., A. B. Oppenheim, and S. L. Adhya.** 2007. A new look at bacteriophage lambda genetic networks. *J Bacteriol* **189**:298-304.

19. **Daccord, A., M. Mursell, D. Poulin-Laprade, and V. Burrus.** 2012. Dynamics of the SetCD-regulated integration and excision of Genomic Islands mobilized by Integrating Conjugative Elements of the SXT/R391 family. *J Bacteriol.*
20. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**:6640-5.
21. **Datta, S., N. Costantino, X. Zhou, and D. L. Court.** 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**:1626-31.
22. **Didelot, X., and M. C. Maiden.** 2010. Impact of recombination on bacterial evolution. *Trends Microbiol* **18**:315-22.
23. **Eddy, S. R.** 2009. A new generation of homology search tools based on probabilistic inference. *Genome Inform* **23**:205-11.
24. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792-7.
25. **Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint.** 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**:722-32.
26. **Garriss, G., M. K. Waldor, and V. Burrus.** 2009. Mobile antibiotic resistance encoding elements promote their own diversity. *PLoS Genet* **5**:e1000775.
27. **Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**:307-21.
28. **Harada, S., Y. Ishii, T. Saga, K. Tateda, and K. Yamaguchi.** 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**:3545-50.
29. **Haugen, S. P., W. Ross, and R. L. Gourse.** 2008. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nature Reviews. Microbiology* **6**:507-19.

30. **Hochhut, B., J. W. Beaber, R. Woodgate, and M. K. Waldor.** 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**:1124-32.
31. **Hochhut, B., and M. K. Waldor.** 1999. Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. *Mol Microbiol* **32**:99-110.
32. **Iwanaga, M., C. Toma, T. Miyazato, S. Insiengmay, N. Nakasone, and M. Ehara.** 2004. Antibiotic resistance conferred by a class I integron and SXT constin in *Vibrio cholerae* O1 strains isolated in Laos. *Antimicrob Agents Chemother* **48**:2364-9.
33. **Iyer, L. M., E. V. Koonin, and L. Aravind.** 2002. Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* **3**:8.
34. **Juhas, M., J. R. van der Meer, M. Gaillard, R. M. Harding, D. W. Hood, and D. W. Crook.** 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* **33**:376-93.
35. **Juiz-Rio, S., C. R. Osorio, V. de Lorenzo, and M. L. Lemos.** 2005. Subtractive hybridization reveals a high genetic diversity in the fish pathogen *Photobacterium damsela* subsp. *piscicida*: evidence of a SXT-like element. *Microbiology* **151**:2659-69.
36. **Kuzminov, A.** 1999. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol Mol Biol Rev* **63**:751-813.
37. **Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins.** 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947-8.
38. **Letunic, I., and P. Bork.** 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**:W475-8.

39. **Li, Z., G. Karakousis, S. K. Chiu, G. Reddy, and C. M. Radding.** 1998. The beta protein of phage lambda promotes strand exchange. *J Mol Biol* **276**:733-44.
40. **Marsic, N., S. Roje, I. Stojiljkovic, E. Salaj-Smic, and Z. Trgovcevic.** 1993. *In vivo* studies on the interaction of RecBCD enzyme and lambda Gam protein. *J Bacteriol* **175**:4738-43.
41. **McCool, J. D., E. Long, J. F. Petrosino, H. A. Sandler, S. M. Rosenberg, and S. J. Sandler.** 2004. Measurement of SOS expression in individual *Escherichia coli* K-12 cells using fluorescence microscopy. *Mol Microbiol* **53**:1343-57.
42. **Miller, J. F.** 1992. A short course in bacterial genetics. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
43. **Murphy, K. C.** 2007. The lambda Gam protein inhibits RecBCD binding to dsDNA ends. *J Mol Biol* **371**:19-24.
44. **Muyrers, J. P., Y. Zhang, F. Buchholz, and A. F. Stewart.** 2000. RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev* **14**:1971-82.
45. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**:299-304.
46. **Oppenheim, A. B., O. Kobiler, J. Stavans, D. L. Court, and S. Adhya.** 2005. Switches in bacteriophage lambda development. *Annu Rev Genet* **39**:409-29.
47. **Osorio, C. R., J. Marrero, R. A. Wozniak, M. L. Lemos, V. Burrus, and M. K. Waldor.** 2008. Genomic and functional analysis of ICE*PdaSpa1*, a fish-pathogen-derived SXT-related integrating conjugative element that can mobilize a virulence plasmid. *J Bacteriol* **190**:3353-61.
48. **Poteete, A. R.** 2001. What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiol Lett* **201**:9-14.
49. **Pruitt, K. D., T. Tatusova, W. Klimke, and D. R. Maglott.** 2009. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res* **37**:D32-6.

50. **Singer, M., T. A. Baker, G. Schnitzler, S. M. Deischel, M. Goel, W. Dove, K. J. Jaacks, A. D. Grossman, J. W. Erickson, and C. A. Gross.** 1989. A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*. *Microbiol Rev* **53**:1-24.
51. **Szczepanska, A. K.** 2009. Bacteriophage-encoded functions engaged in initiation of homologous recombination events. *Crit Rev Microbiol* **35**:197-220.
52. **Toleman, M. A., and T. R. Walsh.** 2011. Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. *FEMS Microbiol Rev* **35**:912-35.
53. **Toussaint, A., and C. Merlin.** 2002. Mobile elements as a combination of functional modules. *Plasmid* **47**:26-35.
54. **Wiedenbeck, J., and F. M. Cohan.** 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* **35**:957-76.
55. **Wozniak, R. A., D. E. Fouts, M. Spagnoletti, M. M. Colombo, D. Ceccarelli, G. Garriss, C. Dery, V. Burrus, and M. K. Waldor.** 2009. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet* **5**:e1000786.
56. **Wozniak, R. A., and M. K. Waldor.** 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* **8**:552-63.
57. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**:3406-15.

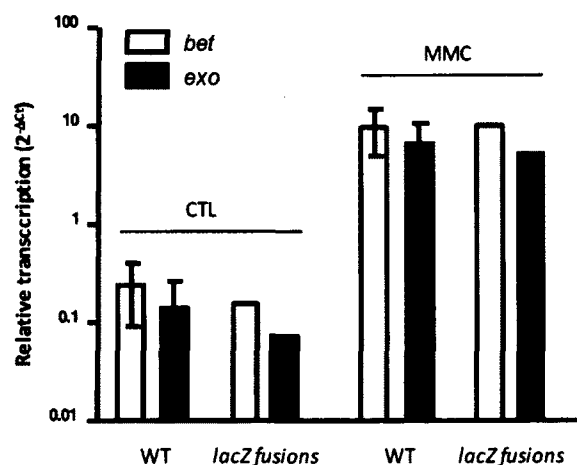


Figure S1. Comparison of transcription levels of *bet* and *exo* in a WT context and *lacZ* in the *bet::lacZ* and *exo::lacZ* fusions, in control (CTL) and MMC-induced conditions (MMC).

Results are expressed as the relative transcription of each gene compared to *rpoZ* ($2^{-\Delta C_t}$).

```

(1) 1      10      20      30      40      50      60      70      80
SXT Ig bet-orfZ (1) CGCCACGCATTTCATTTACT - AACCCGTGACGGGATTATTCTCCCGTCAGGGGGAAGGTCTCGTCTTTTTTTGGAGATCTT
SXT Ig orfZ-exo (1) -----CCACTTAACCCTGAAGGGGAG - TTCTCTCCTTCGGGGGA -- GTCTCCCTCAA -- GGAGACAAT
Consensus (1)          ACT AACCCGTGA GGG TTCTC C T GGGGGA GTCTC CT GGAGA T
Section 2
(81) 882
SXT Ig bet-orfZ (80) CC
SXT Ig orfZ-exo (61) --
Consensus (81)

```

Figure S2. Alignment of the intergenic regions located between *bet* and *orfZ* (Ig bet-orfZ) and between *orfZ* and *exo* (Ig orfZ-exo) in SXT.

	(1)	1	10	20	30	40	56
SXT_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
R391_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
ICEPalBan1_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
ICEPdaSpa1_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
ICEPmiUSA1_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
ICEVchBan10_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
ICEVchBan5_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
ICEVchBan8_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
ICEVchInd4_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
ICEVchInd5_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
ICEVchMex1_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
ICEVchMoz10_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
ICEVfiInd1_TAbet	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			
Consensus	(1)	GGGAGTACTTTCCCGTCAGGG	CAGACTCCCACTTTGATTGT	CGGAGTCCCAAAATG			

	(1)	1	10	20	30	40	51
SXT_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
R391_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
ICEPalBan1_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
ICEPdaSpa1_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
ICEPmiUSA1_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
ICEVchBan10_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
ICEVchBan5_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
ICEVchBan8_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
ICEVchInd4_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
ICEVchInd5_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
ICEVchMex1_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
ICEVchMoz10_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
ICEVfiInd1_TAexo	(1)	AGGGGAGT	CTCTCCTTCGGGGG	-AGTCTCCC	TCAAAGGAGACAATATG		
Consensus	(1)	AGGGGAGT	CTCTCCTTCGGGGG	AGTCTCCC	TCAAAGGAGACAATATG		

Figure S3. Alignment of the predicted translational attenuators found upstream *bet* (TA_{bet}) and upstream *exo* (TA_{exo}) in sequenced SXT/R391 ICEs.

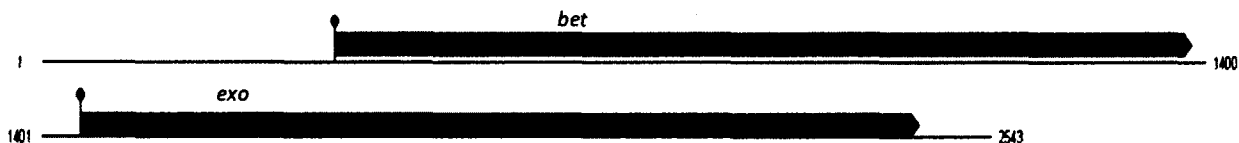


Figure S4. Translational attenuators predicted by RibEx upstream *bet* and *exo* in IncA/C plasmid pIP1202.

Attenuators are indicated by the grey lollipop upstream *bet* and *exo*.

Table S1. List of primers used in this study.

Primer	Sequence 5'-3'	Purpose
s089WF	GGGTGACTCCTTCAAAACCAACATAAGGAGTCACCAGTGTAGGC TGGAGCTGCTTCG	Deletion of <i>s089</i> from SXT
s089WR	GAGTGTTACTTGGTTTTTCATGATGTTCTCCTCGTCATTCCGGGG ATCCGTCGACC	Deletion of <i>s089</i> from SXT and R391
ssb1WF	GTGTGCACTGAGTTCAACTAAGACGAGGAGAACATCGTGTAGG CTGGAGCTGCTTCG	Deletion of <i>ssb</i> from SXT
ssb1WR	AAGTACTCCCGGCTGGGTAAAGCGTTTTTGTCTGGATTCCGGG GATCCGTCGACC	Deletion of <i>ssb</i> from SXT
R66WF	GGGTGACTCCTTCAAAACCAAAATTATGGAGTCACCAGTGTAGGC TGGAGCTGCTTCG	Deletion of <i>s089</i> from R391
ssb2WF	GTGTGCACAGAGTTCAACTAAGACGAGGAGAACATCGTGTAGG CTGGAGCTGCTTCG	Deletion of <i>ssb</i> from R391
ssb2WR	AAGTACTCCCGGCTGGGTAAAGCGTTTTCTGCTCGGATTCCGGG GATCCGTCGACC	Deletion of <i>ssb</i> from R391
65lacZ42BF	ATTGTCTGGAGTCCACAATGGAAAAACCAAAGCTAATCCAAGCC GTCGTTTTACAACGTCGT	LacZ fusion of <i>bet</i> in SXT
65lacZ42BR	AGAATAATCCCGTCAGGGTTAGTAAAATGAATGCGTGGCGGTGT AGGCTGGAGCTGCTTCG	LacZ fusion of <i>bet</i> in SXT
66lacZ42BF	CCTCAAAGGAGACAATATGAAGGTTATCGACCTATCACAAGCCG TCGTTTTACAACGTCGT	LacZ fusion of <i>exo</i> in SXT
66lacZ42BR	AGCGCATGAGAAAGGAGCCGAAATCGGCTCCAAGTGAACGGTG TAGGCTGGAGCTGCTTCG	LacZ fusion of <i>exo</i> in SXT
setDF	CTGGAATTCGCAAAGAGTGCTTTCTATCTA	Construction of pGG2B
setC2R	TTGCTTAAGTAGTACTCTCAAAGTGCACG	Construction of pGG2B
orfZTOPOF	TAATTTGGAGATCTTCCATGACTAA	Construction of pGG7
corfZR	TTACTTTGCAGCTTGTGCTAAC	Construction of pGG7
6389F	CGATTCCGTAAAGCGCACTGACC	Construction of pGG32
6389R	CGAGCAAGCAATCGCAAAGCG	Construction of pGG32
s089-C	TGCAAAACTGAGCTGGTGC	RT-qPCR <i>s089</i>
s089-D	GTCGTA CTGATGTGGCTCTTTG	RT-qPCR <i>s089</i>
ssb-A	TTCGAGATCGTGGTGGATTAA	RT-qPCR <i>ssb</i>
ssb-B	CAGATGCTTTGTTACATTATCAAG	RT-qPCR <i>ssb</i>
bet-A	AAACAAAGTCTCCCTGGATGG	RT-qPCR <i>bet</i>
bet-B	ATGGACCATCTACACGGTAAGG	RT-qPCR <i>bet</i>
exo-A	ACTTTCTTCTGCCGTTATGTGC	RT-qPCR <i>exo</i>
exo-B	GATGCTGTACTTGCACCCAATA	RT-qPCR <i>exo</i>
orfZ-A	GACTAAATCAGCCTCACTTTTTTCG	RT-qPCR <i>orfZ</i>
orfZ-B	CTTTGCAGCTTGTGCTAACGC	RT-qPCR <i>orfZ</i>
RTlacZF	GTGACGTCTCGTTGCTGCAT	RT-qPCR <i>lacZ</i>
RTlacZR	CACCCTGCCATAAAGAAACTG	RT-qPCR <i>lacZ</i>
exoRT	CTCGCGGAGTCACCTGGCAG	Reverse transcription <i>exo</i>
s073RT	GCATGACGAGTCACCACGACAG	Reverse transcription <i>s073</i>
TSS89up4	GTTTAAGGAGGCTAGCTACCTC	Primer extension analysis

CHAPITRE 3 - DISCUSSION

3.1 Formation d'hybrides

Les avancées récentes dans l'étude des ICEs de la famille SXT/R391 démontrent que ces éléments participent de façon majeure à la plasticité de leurs hôtes bactériens, tant par l'apport de gènes leur conférant un avantage sélectif (résistances aux antibiotiques, résistances aux métaux lourds, formation de biofilm, etc) que par leur mobilisation *en trans* d'îlots génomiques, d'ADN chromosomique et de plasmides de virulence (Bordeleau *et al.*, 2009 ; Daccord *et al.*, 2010 ; Hochhut *et al.*, 2000 ; Osorio *et al.*, 2008). L'analyse comparative des séquences des différents membres de la famille SXT/R391 a mis en évidence que ces éléments ont une structure mosaïque vraisemblablement générée par la recombinaison entre deux ICEs (Garriss *et al.*, 2009 ; Wozniak *et al.*, 2009). Trois caractéristiques essentielles de ces éléments, soit l'exclusion partielle d'entrée (Marrero et Waldor, 2005), l'intégration site-spécifique de tous les membres de cette famille dans le même site chromosomique (Hochhut *et al.*, 2001a) et la présence d'un squelette de gènes hautement conservé (Wozniak *et al.*, 2009), fournissent le prérequis essentiel à la recombinaison entre deux éléments : la coexistence dans une même cellule de deux éléments hautement similaires. Dans une étude précédente, Burrus et Waldor (2004) ont démontré que le transfert conjugatif entre une cellule donneuse contenant un tandem SXT-R391 et une cellule réceptrice ne contenant pas d'ICE mène parfois à la détection d'un ICE hybride, issu de la recombinaison entre les deux éléments parentaux, dans les exconjugants. Leurs résultats ont également démontré que la formation d'hybrides était observable en absence de RecA, suggérant que d'autres déterminants génétiques favorisent la recombinaison inter-ICE. L'identification de gènes codant pour des protéines homologues à la recombinase Bet et à l'exonucléase Exo du bactériophage λ dans le squelette de gènes conservés des ICEs SXT/R391 a soulevé l'hypothèse principale du projet qui

proposait que ces gènes forment un système de recombinaison fonctionnel responsable de la formation d'hybrides².

Une méthode de détection semi-quantitative basée sur la conjugaison a été mise au point afin d'étudier les déterminants génétiques impliqués dans la formation d'hybrides SXT/R391. Cette technique, qui repose sur les résistances aux antibiotiques naturellement portées par SXT et R391 et sur des marqueurs phénotypiques artificiellement insérés dans chaque ICE, a permis une analyse quantitative et génétique de la formation d'hybrides, tant dans la cellule donneuse que dans la cellule réceptrice. Cette méthode est une avancée significative en comparaison avec l'analyse PCR utilisée par Burrus et Waldor (2004), qui était beaucoup trop laborieuse pour effectuer l'analyse de l'impact de différents contextes génétiques sur la recombinaison inter-ICE. La mise au point de cette méthode de détection a permis d'évaluer et mesurer le rôle de *bet* et *exo* dans la formation d'hybrides, ainsi que de réévaluer et mesurer le rôle de RecA. Le pourcentage d'exconjugants contenant un ICE hybride obtenu avec cette technique est cependant une sous-estimation du nombre réel d'hybrides fournis, puisqu'elle ne permet pas la discrimination entre un tandem SXT-R391 et un tandem formé de deux éléments hybrides. Il est donc fort probable que le nombre réel d'hybrides formés dépasse celui détecté.

Les résultats présentés par Burrus et Waldor (2004) démontrent qu'en absence de RecA des ICEs hybrides sont toujours détectables. Bien que ce résultat suggère la participation d'autres facteurs dans la formation d'hybrides, la méthode non-quantitative utilisée dans cette étude ne permet pas d'exclure que RecA participe à la formation d'hybride. De plus, l'ample homologie existant entre les squelettes conservés de SXT et R391 rend plus que probable l'implication de RecA. La réévaluation du rôle de RecA a permis de confirmer l'hypothèse que RecA participe effectivement à la formation d'hybrides. Cependant, les résultats obtenus

² N.B. : Afin d'éviter toute confusion, la nomenclature suivante sera utilisée: $\lambda bet/\lambda Bet$ et $\lambda exo/\lambda Exo$ désignent les gènes/protéines de recombinaison du phage λ , *bet*/*Bet* et *exo*/*Exo* désignent les gènes/protéines de recombinaison des ICEs SXT/R391.

en absence de RecA de la cellule donneuse et de la cellule réceptrice viennent confirmer que des facteurs supplémentaires catalysent la formation d'hybrides.

Le rôle indépendant de *bet* et *exo* dans la formation d'hybrides en présence et en absence de *recA* de la donneuse et de la réceptrice a été évalué et les résultats obtenus démontrent que ces deux gènes catalysent la formation d'hybrides d'une manière RecA-indépendante. En absence de *bet/exo* et de *recA*, la formation d'hybride était à peine détectable, confirmant que Bet et Exo sont responsables de la presque totalité des hybrides formés en absence de RecA. Les quelques exconjugants contenant un hybride dans le contexte Δexo peuvent être attribuables à l'utilisation par Bet d'extrémités simple-brin générées par un autre mécanisme que par l'activité exonucléotidique d'Exo, notamment par RecBCD. La présence de quelques sites χ dans la séquence de SXT rend cette hypothèse plausible (résultats non publiés). De plus, au contraire du bactériophage λ , les ICEs SXT/R391 ne codent pas pour un inhibiteur de RecBCD. Les différences majeures entre le mode de vie des ICEs et phages lambdoïdes pourraient expliquer l'absence d'un homologue de λ Gam chez les ICEs SXT/R391, puisque ces derniers n'existent pas sous forme double-brin linéaire, le substrat de RecBCD.

La comparaison des pourcentages d'exconjugants obtenus dans différents contextes génétiques suggèrent une certaine coopération entre RecA et Bet/Exo. En effet, 37% de la formation d'hybrides est attribuable à RecA et 11% est attribuable à l'action de Bet/Exo en absence de RecA. Additionnées, ces valeurs ne peuvent expliquer le total d'hybrides formés. Par analogie avec les mécanismes de recombinaisons décrits pour le bactériophage λ , il est plausible que la présence de RecA aide Bet à catalyser la recombinaison, en permettant l'invasion d'une molécule double-brin (Poteete, 2001). De plus, il est également envisageable que les extrémités d'ADN simple-brin générées par Exo soient utilisables par la recombinaison catalysée par RecA.

La comparaison des pourcentages d'exconjugants contenant des hybrides obtenus en utilisant différentes combinaisons de donneuses et de réceptrices RecA⁺ et RecA⁻ a permis de mettre en évidence que le RecA participe principalement à la formation d'hybrides dans la cellule donneuse. Ce phénomène pourrait être simplement dû au manque de substrat dans la cellule réceptrice, puisque les deux ICEs de la cellule donneuse doivent être transférés vers la même réceptrice afin que celle-ci contienne un tandem permettant la formation d'hybrides. Bien que le mécanisme de formation d'un tandem après un transfert conjugatif ne soit pas encore déterminé, l'explication la plus plausible est que les deux ICEs de la cellule donneuse s'excisent et se transfèrent indépendamment. Ainsi, la possibilité qu'une cellule réceptrice reçoive deux ICEs est moins probable que la possibilité qu'elle n'en reçoive qu'un. Par ailleurs, mes résultats non publiés démontrent également que la fréquence de formation de tandems dans la cellule réceptrice est limitée par la fréquence de transfert de l'ICE non-sélectionné. En effet, le pourcentage d'exconjugants contenant un tandem SXT-R391 est plus élevé si la sélection est faite sur un milieu contenant du sulfaméthoxazole et de la triméthoprime (sélection pour SXT) que si elle est faite sur un milieu contenant de la kanamycine (sélection pour R391). Cela est explicable par le fait que R391 se transfère à une fréquence plus élevée que SXT. Ainsi, un exconjugant ayant reçu SXT a de plus grandes chances de recevoir R391 que l'inverse. Cette idée est également appuyée par des résultats publiés qui indiquent que le transfert conjugatif et l'excision ne sont pas nécessaires à la formation d'hybrides (Garriss *et al.*, 2009).

L'effet similaire causé par la délétion de *bet* ou *exo* suggère que ces deux gènes fonctionnent de manière complémentaire. Des résultats publiés par d'autres équipes viennent appuyer ce résultat. D'une part, il a été démontré que Bet est capable de catalyser la recombinaison entre une molécule d'ADN simple-brin et le chromosome d'une manière similaire à λ Bet (nos résultats non publiés et Datta *et al* (2008). Par ailleurs, des résultats publiés suite à notre étude ont confirmé *in vitro* qu'Exo est une exonucléase fonctionnelle qui dégrade l'ADN double-brin dans une polarité 5'-3', laissant des extrémités 3' simple-brin (Chen *et al.*, 2011).

Conjointement avec nos résultats, cela confirme que Bet/Exo est un système de recombinaison fonctionnel, capable de générer des ICEs hybrides de manière RecA-indépendante, où Exo dégrade les extrémités double-brin afin de fournir un substrat 3' simple-brin à Bet pour la recombinaison.

Les résultats obtenus au cours de ce projet permettent également de réanalyser certains résultats publiés sur l'incompatibilité entre R391 et R997 (Pembroke et Murphy, 2000). R997 est un ICE de la famille SXT/R391, initialement classifié avec R391 comme étant un plasmide du groupe d'incompatibilité IncJ. R997 confère la résistance à l'ampicilline et sa fréquence de transfert est environ 100 fois supérieure à celle de R391. Pembroke et Murphy (2000) ont observé que si R391 et R997 sont transférés vers une même cellule RecA⁺, il est possible de récupérer des exconjugants qui permettent le co-transfert des marqueurs de résistances des deux éléments à la même fréquence. Cette observation peut être expliquée par la formation d'hybrides R391/R997, permettant le transfert des déterminants de résistance des deux éléments parentaux. Leurs résultats montrent également que si la cellule dans laquelle sont transférés les deux éléments est RecA⁻, ils restent «indépendants» et leurs marqueurs de résistance sont subséquemment transférés à leurs fréquences normales respectives. Cette observation est probablement une conséquence du fait que l'échantillon d'exconjugants testés est relativement petit, et que tel qu'observé dans nos expériences, la formation d'hybrides est grandement diminuée dans un contexte RecA⁻ (Garriss *et al.*, 2009).

3.2 Régulation du système de recombinaison

Les gènes *bet/exo* sont induits par les dommages causés à l'ADN, par l'intermédiaire des activateurs de transcription SetCD, qui régulent l'expression concertées des gènes impliqués dans l'intégration/excision et le transfert conjugatif. L'absence de boîtes SOS, les sites de

liaison du répresseur LexA du régulon SOS, dans la région promotrice de *bet/exo*, ainsi que la dépendance vis-à-vis de SetCD, indiquent que l'induction de la transcription des gènes de recombinaison est sous le contrôle du répresseur principal SetR. Ainsi, les gènes favorisant la formation d'hybrides sont exprimés dans les mêmes conditions que celles permettant la sélection des éléments hybrides fonctionnels, c'est-à-dire capables de se transférer vers un nouvel hôte. De plus, l'expression du système Bet/Exo au même moment que RecA est logique en vue des résultats suggérant la coopération de ces deux systèmes dans la formation d'hybrides. Par ailleurs, des résultats récents ont démontrés que le transfert conjugatif induit la réponse SOS dans la cellule réceptrice (Baharoglu *et al.*, 2010). Conjointement avec nos résultats, cela suggère qu'il est possible que la formation d'hybrides soit induite au moment de l'entrée dans la cellule d'un deuxième ICE (ex. : R391 transféré vers une cellule contenant déjà SXT). Cela expliquerait la possibilité de former des ICEs hybrides malgré la faible stabilité des tandems *in vivo* (Burrus et Waldor, 2004). Nos résultats, en vue de ceux publiés par Baharoglu *et al* (2010) démontrent la relation intime qui existe entre les ICEs SXT/R391 et leurs hôtes bactériens, où la réponse SOS est induite par le transfert conjugatif des ICEs, ce qui permet la mise en place des conditions nécessaires, par l'intermédiaire de RecA et SetCD, à la génération de leur diversité (induction de la recombinaison) et à leur dissémination (induction des gènes de transfert).

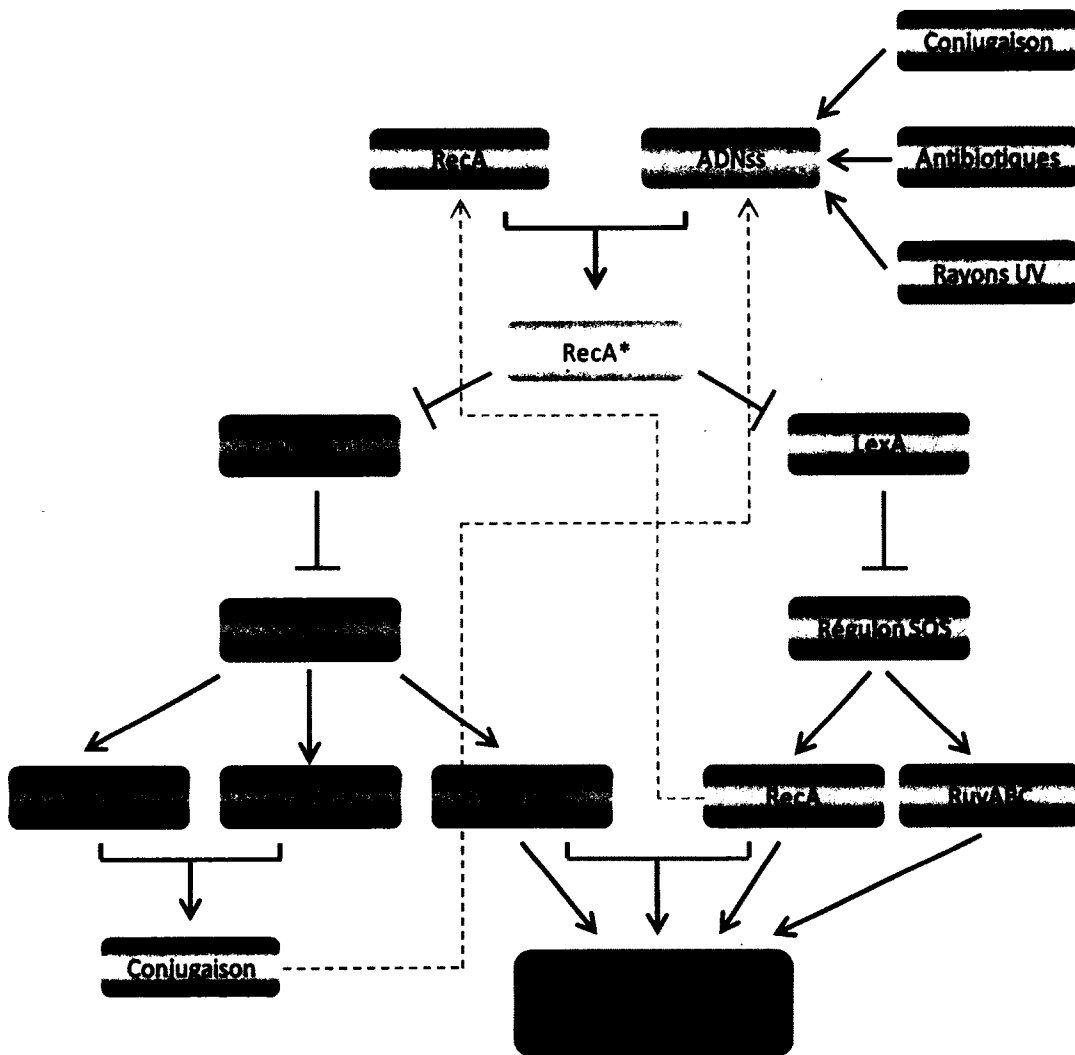


Figure 14. Induction de la recombinaison homologue par Bet/Exo et par RecA.

RecA est activé (RecA*) par la présence d'ADN simple-brin généré par différents mécanismes (antibiotiques, rayons UV, conjugaison) (encadrés oranges) et permet de lever la répression maintenue par SetR (ICEs SXT/R391) et LexA (hôte), sur leurs cibles respectives, ce qui mène ultimement à l'expression des gènes participant dans la recombinaison : Bet/Exo, RecA et certains autres gènes du régulon SOS tels UvrABC (résolution des jonctions de Holliday). Les fonctions codées par les ICEs sont encadrées en mauves, celles codées par leur hôte en vert. Les flèches vertes et les lignes rouges indiquent respectivement l'induction et la répression.

L'analyse de l'organisation transcriptionnelle des gènes de recombinaison a mis en évidence qu'ils font partie d'un opéron polycistronique contenant 12 autres gènes du squelette conservé des ICEs SXT/R391. Le rôle précis de ces gènes n'a pas été déterminé, mais il a été démontré par Wozniak *et al* (2009) qu'ils ne sont pas impliqués dans le transfert conjugatif. Les protéines codées par les gènes de ce locus ne possèdent aucun domaine conservé, à l'exception de Ssb, S067 (domaines CobS et ATPase), S070 (domaines CobT et NorD), S071 (domaine RadC) et S072 (domaine primase). De plus, 7 des 14 gènes de ce transcrit, incluant *ssb*, *bet* et *exo* sont également retrouvés chez les plasmides de la famille IncA/C. La question se pose sur la raison de la conservation de ces gènes, ainsi que leur inclusion dans un opéron codant pour des fonctions de recombinaison. L'ordre légèrement différent de ces gènes chez les IncA/C suggère une pression évolutive pour maintenir la présence de ces gènes dans la même unité transcriptionnelle, sans que leur ordre soit important. Il est possible que ces gènes apportent un certain avantage évolutif aux hôtes des ICEs SXT/R391, ou bien qu'ils jouent un rôle positif dans le maintien ou la stabilité de ces éléments.

Bien que les profils transcriptionnels de *bet/exo* soient très similaires en conditions d'induction par la mitomycine C, leurs profils traductionnels sont différents. En effet, la production de Bet et Exo semble régulée au niveau traductionnel par la présence d'atténuateurs de traduction situés directement en amont de *bet* et *exo*. La présence de ces atténuateurs viendrait expliquer la faible traduction de Bet et Exo observée expérimentalement. De plus, la plus grande stabilité prédite pour TA*exo* amènerait également une explication convenable à la plus faible traduction d'Exo par rapport à Bet. Considérant que le transcrit contenant *bet* et *exo* est d'une taille supérieure à 11 kb, il est tout à fait plausible que la régulation traductionnelle des gènes de ce transcrit permet une modulation plus fine du niveau et du moment de production des différentes protéines. Cette idée est appuyée par l'analyse RibEx (Abreu-Goodger et Merino, 2005) du transcrit, qui indique la présence d'autres atténuateurs de traduction dans le transcrit (en amont des gènes *s068*, *s069* et *s070*) (résultats non publiés). De plus, des atténuateurs de

traduction sont également présents en amont des gènes homologues de *bet* et *exo* retrouvés chez les plasmides IncA/C.

À ma connaissance, aucune étude ne s'est penchée à ce jour sur la régulation de l'expression des gènes de recombinaison du bactériophage λ . Cependant, il est connu qu'une légère surexpression des recombinases λ Bet et RecT par rapport à leur exonucléase respective (λ Exo et RecE) favorise la recombinaison (Muyrers *et al.*, 2000). La présence d'un atténuateur plus stable en amont d'Exo pourrait justement permettre un meilleur potentiel de recombinaison en assurant la surexpression de Bet par rapport à Exo.

Par ailleurs, la régulation traductionnelle de *bet* et *exo* amène une explication plausible à une observation particulière entre les profils transcriptionnels (Garriss et Burrus, 2012) et la formation d'hybrides en condition de surexpression de SetCD (Garriss *et al.*, 2009). Dans les essais de formation d'hybrides, l'absence de RecA, qui est nécessaire à l'activation des gènes de transfert, a été complétée par la surexpression de SetCD à partir du vecteur pVI67. L'effet de la surexpression de SetCD, dans un contexte $\Delta recA$, sur la transcription de *bet* et *exo* a été mesurée par RT-qPCR et a démontré une induction massive de la transcription (> 10 000 fois), ce qui dépasse largement ce qui est obtenu dans un contexte RecA⁺ où la réponse SOS est induite par la mitomycine C (Garriss et Burrus, 2012). Considérant ce résultat, on s'attendrait à ce que la surexpression de SetCD dans un contexte $\Delta recA$ cause une induction massive de la formation d'hybrides. Pourtant, ce n'est pas ce qui a été observé, bien que la surexpression de SetCD soit bien vérifiée par le transfert conjugatif qui était grandement augmentée dans ces conditions (\approx 1 000 fois plus). (Garriss *et al.*, 2009). Cette observation pourrait avoir deux explications. Dans un premier temps, bien que la surexpression de SetCD cause une augmentation massive de la transcription de *bet* et *exo*, leur traduction atteint à peine les valeurs obtenues en condition sauvage d'induction de la réponse SOS (résultats non publiés). Ainsi, il est fort possible qu'il n'y ait pas une augmentation démesurée de la

production de Bet et Exo dans le contexte des essais de formation d'hybrides, ce qui confirme que ces résultats sont d'autant plus représentatifs d'un contexte naturel. Dans un deuxième temps, il est aussi possible que le nombre d'évènements de recombinaison soit augmenté sans que le nombre d'hybrides fonctionnels formés le soit. Des hybrides non fonctionnels ne seraient pas détectés par notre essai qui, étant basé sur le transfert conjugatif, implique que l'élément recombinant puisse être transféré. Finalement, l'intégration des résultats de formation d'hybrides et des analyses traductionnelles permettent de conclure qu'une faible production de Bet et Exo est suffisante pour efficacement catalyser la formation d'hybrides.

3.3. Rôles de Ssb et OrfZ

La présence d'autres gènes dans le même opéron que *bet* et *exo* a soulevé la question de leur implication dans la formation d'hybrides. Les rôles de *s089* et *ssb*, situés en amont de *bet*, ainsi que le rôle d'*orfZ*, situé entre *bet* et *exo*, ont été évalués. L'absence de *s089* n'a pas eu d'effet, tandis que la délétion de *ssb* a causé une augmentation significative de la formation d'hybrides. Cet effet observé ne peut être attribué à l'augmentation de l'expression de *bet/exo* par rapprochement du promoteur situé en amont de *s089*, puisque la délétion de ce dernier gène, dont la taille est similaire à *ssb*, devrait causer un effet similaire.

Les protéines Ssb sont reconnues pour leur implication dans la recombinaison, la réplication et la réparation, en agissant de manière coopérative et compétitive avec un grand nombre d'autres protéines (Kuzminov, 1999 ; Szczepanska, 2009). Bien que le rôle précis de Ssb dans la recombinaison catalysée par Bet/Exo ne soit pas déterminé, il est envisageable que Ssb limite l'accès de Bet à son substrat en compétitionnant pour la liaison à l'ADN simple-brin. Cette hypothèse est d'ailleurs appuyée par de récents résultats qui démontrent que la co-expression de Ssb-Bet-Exo diminue par plus de deux fois la recombinaison *in vivo* d'une

molécule d'ADN double-brin avec le chromosome d'*E. coli* (Chen *et al.*, 2011). Par ailleurs, Bet est capable de recombiner des molécules d'ADN simple-brin possédant seulement de courtes séquences homologues (36 bp) (Datta *et al.*, 2008) et il est également connu que λ Bet est capable de promouvoir l'échange et le déplacement de brins en utilisant des oligonucléotides contenant plusieurs bases non appariées (Li *et al.*, 1998). Dans un contexte où les fonctions de recombinaison des ICEs sont induites par les dommages causés à l'ADN, donc, lorsqu'une grande quantité d'ADN simple-brin est disponible, il est possible qu'il soit favorable de contrôler le nombre d'évènements de recombinaison.

L'analyse *in silico* de la structure de nombreux systèmes de recombinaison démontre qu'au contraire de ce qui est observé chez la majorité des bactériophages, aucun système de recombinaison retrouvé sur un élément conjugatif ne possède un inhibiteur de RecBCD (λ Gam). λ Gam inhibe l'activité de l'exonucléase RecBCD qui dégrade agressivement l'ADN double-brin linéaire du phage (Marsic *et al.*, 1993 ; Murphy, 2007). Il est probable qu'un homologue de Gam soit absent des éléments conjugatifs puisque ceux-ci n'existent à aucun moment sous la forme d'une molécule double-brin linéaire au cours de leur processus normal de dissémination. Cependant, nous avons trouvé que tous les systèmes de recombinaison retrouvés sur des plasmides possèdent un gène codant pour une protéine Ssb putative, à l'exception d'un système retrouvé sur un plasmide non-conjugatif (pALVIN02 de *Allochromatium vinosum*). La présence d'une Ssb pourrait servir à protéger l'ADN simple-brin qui est le substrat pour le transfert conjugatif, ou pour limiter des évènements de recombinaison avant la duplication du brin transféré vers la cellule réceptrice. Les études portant sur les protéines Ssb des bactériophages démontrent qu'elles se lient à l'ADN simple-brin pour en empêcher la dégradation (Szczepanska, 2009). Les Ssb codées par *E. coli* et certains bactériophages sont également reconnues pour déstabiliser les structures secondaires dans l'ADN simple-brin, facilitant ainsi l'appariement lors de la recombinaison (Kuzminov, 1999 ; Szczepanska, 2009). Bien que les protéines Ssb soient généralement reconnues pour

leur impact positif sur la recombinaison chez les bactériophages, leur rôle dans la recombinaison chez les éléments conjugatifs n'a pas été étudié.

La délétion d'*orfZ*, situé entre *bet* et *exo*, a étonnamment mené à une augmentation encore plus considérable dans la formation d'hybrides que la délétion de *ssb* (résultats non publiés). La détection d'activité β -galactosidase suite à la fusion traductionnelle d'*orfZ* avec *lacZ* suggère que la protéine OrfZ soit produite. Cependant, la surexpression d'*orfZ* n'a mené à aucune modulation du profil transcriptionnel ou traductionnel de *bet* et *exo*, indiquant qu'OrfZ n'agit probablement pas en tant que régulateur de l'expression des gènes de recombinaison. L'effet de la délétion d'*orfZ* sur la formation d'hybrides serait plutôt attribuable à la délétion de TA*exo*, situé dans la région intergénique entre *orfZ* et *exo*. La délétion de cet atténuateur de traduction permettrait la production d'une plus grande quantité de molécules d'Exo et la production subséquente d'une plus grande quantité d'ADN simple-brin disponible pour la recombinaison par Bet. De plus, *orfZ* n'est présent que dans le système de recombinaison des ICEs SXT/R391. L'absence d'homologues d'*orfZ* dans tous les autres systèmes, incluant le système des plasmides IncA/C qui sont les plus proches parents des ICEs SXT/R391, suggèrent qu'il n'est pas impliqué dans la recombinaison catalysée par Bet/Exo. L'analyse des régions intergéniques situées de part et d'autre d'*orfZ* a mis en évidence qu'elles partagent une certaine homologie. Cette observation permet de proposer qu'*orfZ* a été acquis par un événement d'insertion dans l'ancêtre commun des ICEs SXT/R391 qui a mené à la duplication de la région intergénique *bet-exo*.

3.4. Distribution phylogénétique et évolution

Les systèmes de recombinaison que nous avons identifiés chez les ICEs de la famille SXT/R391 et chez les plasmides de la famille IncA/C sont les premiers exemples de systèmes

de recombinaison apparentés au système Red jamais décrits sur des éléments conjugatifs. Afin d'évaluer l'incidence de tels systèmes nous avons utilisé une méthode d'analyse bioinformatique permettant l'identification de systèmes de recombinaison apparentés à Bet/Exo dans les génomes séquencés de bactéries, plasmides et bactériophages, en utilisant les domaines conservés RecT et YqaJ respectivement associés à Bet et Exo. Les résultats de cette analyse ont mis en évidence la large distribution de systèmes RecT/YqaJ au sein d'espèces bactériennes très divergentes. Une analyse plus approfondie des séquences entourant les systèmes identifiés au sein de génomes bactériens a permis de déterminer que la majorité d'entre eux sont codés par des prophages et qu'un petit nombre est retrouvé sur des ICEs de la famille SXT/R391, déjà connus (ICE*Spu*PO1, ICE*Pmi*USA, ICE*Vch*Ban9) ou nouvellement identifiés (génome de *V. cholerae* 2010EL-1786 (NC_016445)). Les exemples restant qui n'ont pu être associés avec certitude à un prophage ou un ICE font probablement partie de vestiges de prophages. L'identification de systèmes de recombinaison composés de protéines possédant les mêmes domaines fonctionnels que Bet et Exo dans une diversité d'éléments génétiques tels des prophages, des ICEs ainsi que des plasmides appartenant à divers groupes d'incompatibilité suggère que leur rôle n'est pas restreint au mode de vie des ICEs et des phages lambdaïdes. Par ailleurs, notre étude s'étant centrée sur les génomes entièrement séquencés et assemblés de la base de données RefSeq, il est possible que les exemples identifiés ici ne représentent qu'une fraction des systèmes de recombinaison similaires. De plus, la prévalence de systèmes composés de l'association d'une exonucléase du type RecE, au lieu du type YqaJ, avec une recombinase RecT n'a pas été évaluée.

La comparaison des relations phylogénétiques entre les protéines RecT et YqaJ d'origines diverses suggère des patrons distincts d'évolution en fonction du type d'élément auquel est associé chaque protéine. En effet, les protéines RecT et YqaJ retrouvées sur des prophages ou des îlots génomiques ont tendance à se regrouper en fonction de l'ordre taxonomique de leur hôte, tandis que celles retrouvées sur des plasmides ou des ICEs se regroupent en fonction de la famille d'élément ou du groupe d'incompatibilité : les plasmides IncA/C forment un groupe

situé à proximité du groupe des ICEs SXT/R391 et les plasmides des groupes IncP-7 et IncT se regroupent. Cette observation peut être une conséquence du spectre d'hôte plus restreint des phages comparé à celui des plasmides et des ICEs, ce qui limiterait les échanges génétiques entre différents ordres.

L'observation des relations phylogénétiques des protéines RecT et YqaJ suggère que parfois les protéines composant un même système de recombinaison proviennent d'origines différentes. Ainsi, les recombinases (RecT) retrouvées chez les ICEs SXT/R391 et les plasmides IncA/C semblent avoir été acquises des phages lambdaïdes. Cependant, leurs exonucléases partenaires ne sont pas apparentées aux exonucléases retrouvées chez les phages lambdaïdes. De plus, les exonucléases (YqaJ) de tous les plasmides et ICEs se regroupent, tandis que les recombinases qui leur sont associées ont des relations phylogénétiques différentes : celles retrouvées chez les ICEs SXT/R391 et les plasmides IncA/C forment un groupe, tandis que celles identifiées chez les plasmides IncP-7 (pCAR1, pCAR1.2, pDK1) forment un groupe avec les recombinases des plasmides Rts1 (IncT) et pMAQU02. Notamment, tous les ICEs et plasmides, à l'exception du plasmide pALVIN02, possèdent à proximité du gène codant pour RecT un gène codant pour une Ssb. Le plasmide non-conjugatif pALVIN02 est le seul plasmide contenant une protéine RecT qui n'est pas apparentée aux protéines RecT codées par les autres plasmides, suggérant qu'une pression de sélection agit afin de maintenir l'association d'un gène codant pour une protéine Ssb avec le gène codant pour la recombinase.

L'apparente divergence dans les origines évolutives des protéines Rec et YqaJ composant les systèmes de recombinaison des ICEs SXT/R391 et des plasmides IncA/C suggère qu'ils ont été formés par l'acquisition successive de gènes codant pour une recombinase et une exonucléase, plutôt que par l'acquisition d'un couple RecT/YqaJ. Bien qu'il soit étonnant que des protéines ayant des fonctions qui coopèrent aient des origines distinctes, l'exemple que

nous présentons n'est pas le premier. En effet, il a été démontré dans le passé que des recombinases de type RecT sont parfois retrouvées associées avec des exonucléases appartenant à la famille RecE (Iyer *et al.*, 2002). Ces observations permettent de proposer qu'une structure similaire peut apparaître indépendamment dans différents systèmes grâce à la pression de sélection favorisant la co-expression de protéines qui coopèrent.

3.5. Aspects mécanistiques et rôle biologique

Bien que le but de ce projet n'ait pas été de caractériser le (ou les) mécanisme de recombinaison employé par le système Bet/Exo, il est tout de même possible de proposer certaines avenues intéressantes pour des études futures. Étant donnée la similitude au niveau des domaines conservés (RecT et YqaJ) entre Bet/Exo et λ Bet/ λ Exo, il est probable que leurs mécanismes de recombinaison soient similaires. Par ailleurs, les caractéristiques biochimiques d'Exo, tant au niveau de sa structure tri-dimensionnelle qu'au niveau de sa spécificité et polarité de dégradation, démontrent qu'elle est fonctionnellement apparentée à λ Exo (Chen *et al.*, 2011). De plus, bien que Bet soit moins efficace que λ Bet, sa capacité à recombiner de l'ADN simple-brin avec le chromosome et sa préférence pour l'utilisation d'un oligonucléotide ciblant le brin tardif suggèrent qu'elle catalyse la recombinaison selon les mêmes mécanismes (Datta *et al.*, 2008). De plus, des résultats obtenus par Datta *et al.* (2008) et par Chen *et al.* (2011) démontrent qu'à l'instar de RecT/RecE et λ Bet/ λ Exo (Muyrers *et al.*, 2000), le rôle d'Exo dans la recombinaison catalysée par Bet ne peut être substitué par une autre exonucléase : Bet est incapable de promouvoir la recombinaison double-brin en présence de λ Exo (Datta *et al.*, 2008), mais peut le faire en présence d'Exo (Chen *et al.*, 2011). Ces analogies fonctionnelles entre le système de recombinaison des ICEs SXT/R391 et du bactériophage λ suggèrent que des mécanismes de recombinaison similaires soient utilisés.

Notre étude démontre pour la première fois l'existence de systèmes de recombinaison homologue sur des éléments conjugatifs et la raison précise de leur présence chez ce type d'éléments est encore inconnue. De plus, le rôle réel du système Red dans la biologie du bactériophage λ demeure spéculatif. L'induction des fonctions de recombinaison homologue des ICEs SXT/R391 et du phage λ par les dommages causés à l'ADN laissent croire que ces systèmes ont un rôle dans la réparation de l'ADN. Puisque le transfert de ces deux types d'éléments est également induit dans les conditions qui induisent la réponse SOS, il est plausible que Bet/Exo et λ Bet/ λ Exo servent à réparer l'ADN de l'ICE ou du phage endommagé afin d'assurer la dissémination d'éléments fonctionnels. Les résultats de Chen *et al* (2011) démontrant la dégradation d'extrémités d'ADN double-brin par Exo et la capacité de Bet/Exo à catalyser la recombinaison d'ADN double-brin avec le chromosome permet de proposer que les cassures double-brin produites par des dommages à l'ADN peuvent être utilisées comme substrat par Bet/Exo. De plus, nos résultats démontrent *in vivo* que l'excision n'est pas nécessaire à la formation d'hybrides, donc que la molécule circulaire double-brin de l'ICE n'est pas le substrat utilisé pour la formation d'hybrides.

La recombinaison homologue semble jouer un rôle important dans la réplication de λ , notamment en permettant la transition entre les modes θ et σ de réplication (Kuzminov, 1999 ; Weigel et Seitz, 2006). L'intermédiaire circulaire extrachromosomique des ICEs de la famille SXT/R391, qui est le substrat de leur transfert conjugatif, est considéré comme étant non-répliatif. Cependant, certaines observations suggèrent qu'il en est peut-être autrement. En effet, Burrus et Waldor (2003) ont observé, dans un essai visant à mesurer l'excision de SXT, un ratio 4 : 1 *attP* : *attB* indiquant la présence de 4 molécules excisées de SXT par chromosome. Si effectivement les ICEs SXT/R391 ont la capacité de se répliquer, il est possible que la région simple-brin générée lors de leur réplication serve de cible pour la recombinaison catalysée par Bet/Exo. Cette hypothèse est d'autant plus plausible en vue des récents résultats publiés par Maresca *et al* (2010) et Mosberg *et al* (2010) démontrant que la recombinaison catalysée par λ Red se produit à la fourche de réplication. Il est également

envisageable que la réplication en cercle-roulant lors du transfert conjugatif ou que la molécule simple-brin servant au transfert soient des substrats utilisables pour la recombinaison. Les gènes impliqués dans la potentielle réplication des ICEs SXT/R391 ne sont pas identifiés, et ils ne contiennent pas d'homologue du gène de réplication *repA* des plasmides IncA/C. Néanmoins, par analogie avec les autres gènes essentiels à leur fonctionnement, il est attendu qu'ils soient localisés dans le squelette de gènes conservés. Notamment, la protéine produite par le gène *s072*, situé dans le locus *s089-s073*, contient un domaine primase putatif (résultats non publiés). Les primases jouent un rôle essentiel dans la réplication, en synthétisant les fragments d'Okazaki utilisés pour la réplication du brin tardif (Hamdan et van Oijen, 2010). Puisque la presque totalité des gènes du squelette conservé sont induits par les agents endommageant l'ADN, il est possible d'imaginer que les fonctions de réplication, si elles existent, soient également induites dans les mêmes conditions. Cela aurait comme conséquence de favoriser la recombinaison, non seulement par l'induction des fonctions de recombinaison, mais également par l'augmentation des cibles (fourches de réplication) et des substrats (molécules répliquées).

3.6. Impact évolutif

La présence d'un système de recombinaison RecA-indépendant, capable de générer des éléments hybrides, chez les ICEs SXT/R391 démontre que ces éléments participent à leur propre diversité en permettant la création de nouveaux éléments comportant, entre autres, de nouveaux patrons de résistances aux antibiotiques. La découverte de systèmes de recombinaison apparentés dans plusieurs éléments génétiques mobiles apparentés et non-apparentés aux ICEs SXT/R391 permet d'envisager que ce type de système joue un rôle important chez les éléments génétiques mobiles en général. Il est d'ailleurs reconnu dans la littérature que la majorité des éléments génétiques mobiles sont des combinaisons interchangeables de modules fonctionnels qui ont été à l'origine décrits comme appartenant à

un type précis d'élément (bactériophage, plasmide, transposon, etc) (Burrus *et al.*, 2002b ; Toleman et Walsh, 2011 ; Toussaint et Merlin, 2002 ; Wozniak et Waldor, 2010). La présence de systèmes de recombinaison au sein de ces éléments pourrait permettre notamment l'échange de ces différents modules et pourrait avoir permis l'émergence de nouveaux types d'éléments, tels les ICEs. Ces systèmes de recombinaison pourraient également avoir permis l'intégration d'une origine de transfert similaire à celle des ICEs SXT/R391 au sein d'îlots génomiques, permettant leur subséquente mobilisation par les ICEs SXT/R391 (Daccord *et al.*, 2010).

Outre la formation d'ICEs hybrides par la recombinaison entre deux éléments préexistants, le système Bet/Exo pourrait être responsable de l'acquisition de nouvelles régions variables dans le squelette conservé des ICEs, permettant la constante évolution des fonctions accessoires conférées par ces éléments. De plus, Bet/Exo pourraient augmenter directement la plasticité du génome des hôtes des ICEs SXT/R391, notamment en leur permettant d'acquérir de nouvelles fonctions ou de nouveaux allèles conférant un avantage sélectif.

3.7. Perspectives

Plusieurs aspects biologiques et mécanistiques de la formation d'hybrides SXT/R391 et de l'impact de la présence du système Bet/Exo pour les ICEs eux-mêmes et pour leurs hôtes demeurent inconnus et offrent des avenues intéressantes d'étude. Dans un premier temps, l'analyse plus approfondie du mécanisme de recombinaison de Bet/Exo pourrait permettre de mieux comprendre son rôle biologique, notamment en déterminant l'impact de la réplication potentielle de l'ICE sur la recombinaison.

Dans un deuxième temps, comprendre l'impact de la recombinaison Bet/Exo sur l'acquisition de matériel génétique exogène dans le squelette des ICEs SXT/R391 pourrait amener une explication à l'acquisition des régions variables. Jusqu'à ce jour, l'origine et le mécanisme par lequel les régions variables des ICEs SXT/R391 sont acquises n'ont pas été étudiés. Une telle étude pourrait également inclure d'autres systèmes conjugatifs, tels les plasmides du groupe d'incompatibilité IncA/C, qui possèdent eux aussi des régions variables conférant différentes fonctions accessoires (Fricke *et al.*, 2009). À titre d'exemple, l'évaluation de l'incorporation d'un gène conférant un phénotype aisément détectable (ex. : *galk*) dans le squelette de ces éléments, en utilisant différents contextes génétiques (présence ou absence de Bet, Exo, Bet/Exo, RecA) pourrait être facilement réalisable.

Dans un troisième temps, l'impact de la recombinaison Bet/Exo sur l'évolution du génome des hôtes des ICEs SXT/R391 n'a pas été évalué. Néanmoins, il est fort probable que ce système de recombinaison puisse participer à l'acquisition de nouveaux gènes dans le chromosome de leurs hôtes, par exemple en permettant la recombinaison de gènes acquis par transformation naturelle avec le chromosome. Cela pourrait permettre l'émergence de souches ayant de nouvelles propriétés. À titre d'exemple, la souche de *V.cholerae* dans laquelle a été découvert SXT a été le premier exemple de souche O139 causant une épidémie de choléra. Il a plus tard été démontré qu'un évènement de conversion de sérotype O1-O139 a mené à l'apparition de cette souche et que l'acquisition du locus codant pour le sérotype O139 par une souche de sérotype O1 par transformation naturelle pouvait expliquer ce phénomène (Blokesch et Schoolnik, 2007 ; Comstock *et al.*, 1996). Cependant, l'impact de la présence de SXT dans l'échange des locus codant pour les différents sérotypes n'a pas été évalué. La présence de SXT, dans différents contextes génétiques (présence et absence de Bet/Exo et RecA) et en condition d'induction par des agents endommageant l'ADN, pourrait favoriser la conversion de sérotype.

Par ailleurs, il serait intéressant de connaître les conditions qui induisent le transfert des plasmides IncA/C. Bien que les gènes codant pour leurs fonctions de transfert conjugatif soient apparentés à ceux des ICEs SXT/R391, ces plasmides ne contiennent pas d'homologue de SetR, le répresseur principal des ICEs. Cette observation suggère que ces plasmides ne sont pas induits en conditions d'induction de la réponse SOS. Il serait intéressant de connaître les conditions régulant l'expression des gènes de transfert des plasmides IncA/C, afin de déterminer si les mêmes conditions induisent l'expression de leurs homologues de *bet* et *exo*.

La ressemblance au niveau fonctionnel entre les protéines composant les systèmes de recombinaison des ICE SXT/R391 et du bactériophage λ suggèrent qu'ils fonctionnent selon un mécanisme semblable. Cependant, l'absence de Gam et la présence de Ssb dans le système Bet/Exo soulève l'hypothèse que des différences existent néanmoins soit dans leur mécanisme de recombinaison ou dans leur rôle dans la biologie de ces deux familles d'éléments.

Finalement, il serait possible de mettre au point un outil de biologie moléculaire permettant la délétion facile de gènes chromosomiques chez *Vibrio*. Le système Red du bactériophage λ est largement utilisé pour modifier le chromosome de *E. coli* : insertions de marqueurs, délétions de gènes, fusions transcriptionnelles et traductionnelles (Datsenko et Wanner, 2000 ; Ellermeier *et al.*, 2002 ; Gerlach *et al.*, 2007). La technique originale, publiée par Datsenko et Wanner (2000), permet de recombiner avec le chromosome des fragments d'ADN double-brin générés par PCR et comportant seulement 36 pb d'homologie avec la région à modifier en surexprimant les gènes $\lambda gam/\lambda bet/\lambda exo$ à partir d'un plasmide. Bien que cette technique soit utilisable chez d'autres bactéries à Gram négatif, elle nécessite normalement des modifications, généralement au niveau de la longueur des régions d'homologies ajoutées au fragment à recombiner. Ces modifications rendent parfois cette technique laborieuse. Notamment, l'utilisation de cette technique chez *Vibrio* nécessite des régions d'homologie de 100-1000 pb au lieu de 36 pb (Yamamoto *et al.*, 2009). Les raisons pour lesquelles le système

Red est moins efficace chez *V. cholerae* que chez *E. coli* ne sont pas connues. Il est possible que des facteurs additionnels présents chez *E. coli* soient nécessaires afin de maximiser le potentiel de recombinaison. L'utilisation du système Bet/Exo, naturellement retrouvé chez *V. cholerae*, pourrait peut-être permettre de contourner ce problème.

CONCLUSION

Le but de ce projet de doctorat était d'étudier un système de recombinaison homologue retrouvé chez les éléments intégratifs et conjugatifs de la famille SXT/R391, en déterminant son rôle dans la formation d'ICEs hybrides et en identifiant les facteurs régulant son expression. Ce système de recombinaison, composé des gènes *bet* et *exo*, est apparenté au système de recombinaison Red du bactériophage λ .

Les résultats obtenus au cours du projet ont permis de déterminer que les gènes *bet/exo* des ICEs SXT/R391 sont capables de générer des ICEs hybrides en catalysant la recombinaison entre deux ICEs intégrés en tandem dans le chromosome de leur hôte. La mise au point d'une méthode de détection semi-quantitative des colonies contenant des ICEs hybrides a permis d'évaluer le rôle de RecA et de Bet/Exo dans la formation d'hybrides. Les résultats de cette portion du projet démontrent que Bet/Exo forment un système de recombinaison fonctionnel et capable de catalyser la recombinaison en absence de RecA. Nos résultats indiquent également que RecA participe à la formation d'hybrides principalement dans la cellule donneuse.

De plus, nos résultats démontrent que la présence de RecA et de Bet/Exo permet de générer un plus grand nombre d'hybrides que la somme de leurs deux actions indépendantes, suggérant un mécanisme coopératif entre ces deux systèmes de recombinaison. Finalement, nous avons déterminé que l'excision et le transfert conjugatif ne sont pas nécessaires à la formation d'hybrides, mais servent plutôt de moyen de ségrégation permettant de sélectionner des éléments fonctionnels. Dans l'ensemble, les résultats obtenus dans cette première portion du projet démontrent que les ICEs SXT/R391 sont capables de promouvoir leur propre diversité

en formant des hybrides par l'intermédiaire du système de recombinaison RecA-indépendant qu'ils codent.

Dans le deuxième volet du projet, les facteurs régulant l'expression de Bet et Exo ont été évalués, et les résultats obtenus montrent que la transcription de *bet* et *exo* est induite par les dommages causés à l'ADN, via les activateurs transcriptionnels SetCD de l'ICE. Ainsi, la transcription des gènes de recombinaison répond aux mêmes stimuli que la transcription des gènes impliqués dans l'intégration, l'excision et le transfert conjugatif des ICEs SXT/R391. L'analyse de l'organisation génétique du locus contenant *bet* et *exo* a mis en évidence que ces gènes font partie d'un opéron polycistronique, contenant 12 autres gènes du squelette conservé, dont la transcription est initiée à partir d'un promoteur inductible par les agents endommageant l'ADN situé en amont du gène *s089*. De surcroît, 7 des 14 gènes de ce locus, dont *ssb*, *bet* et *exo*, sont également présents chez les plasmides du groupe d'incompatibilité IncA/C, les plus proches parents des ICEs SXT/R391.

Toutefois, la traduction de Bet et Exo semble être soumise à un niveau additionnel de régulation, vraisemblablement dépendant de la présence de deux atténuateurs de traduction situés directement en amont de Bet et Exo. Ces atténuateurs de traduction sont également présents en amont des homologues de *bet* et *exo* retrouvés chez les plasmides IncA/C.

D'autre part, nous avons constaté que la délétion du gène codant pour protéine putative de liaison à l'ADN simple-brin (*ssb*), situé immédiatement en amont de *bet* et dans le même transcrit, cause une augmentation significative de la formation d'hybrides. Bien que le rôle précis de Ssb dans le mécanisme de recombinaison ne soit déterminé, il est plausible qu'elle agisse en tant que compétiteur pour la liaison à l'ADN simple-brin, limitant ainsi l'accès de Bet à son substrat.

L'analyse de la distribution phylogénétique de systèmes de recombinaison apparentés à Bet/Exo démontre qu'ils sont présents chez une diversité d'espèces bactériennes appartenant à des ordres taxonomiques éloignés. De plus, elle montre pour la première fois que des systèmes de recombinaison homologues sont présents chez une diversité de plasmides conjugatifs appartenant à différents groupes d'incompatibilité. Les relations phylogénétiques des recombinases et exonucléases composant ces systèmes montrent que les protéines composant un même système ont parfois des origines évolutives distinctes, suggérant qu'une pression de sélection agit afin d'assurer la co-expression de ces protéines qui coopèrent.

En conclusion, ces travaux ont permis une meilleure compréhension des facteurs favorisant l'évolution des éléments intégratifs et conjugatifs de la famille SXT/R391 en démontrant leur participation à leur propre diversité et suggère que cette capacité puisse être étendue à un grand nombre d'éléments génétiques mobiles. Finalement, ce projet démontre la relation étroite qui lie les ICEs SXT/R391 avec la réponse de leur hôte au stress et souligne leur potentiel insoupçonné d'influencer la plasticité du génome de leurs hôtes.

BIBLIOGRAPHIE

- Abreu-Goodger, C. and Merino, E. (2005). RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res.* 33, W690-692.
- 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.
- Anderson, D.G. and Kowalczykowski, S.C. (1997). The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell* 90, 77-86.
- 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.
- Bani, S., Mastromarino, P.N., Ceccarelli, D., Le Van, A., Salvia, A.M., Ngo Viet, Q.T., Hai, D.H., Bacciu, D., Cappuccinelli, P. and Colombo, M.M. (2007). Molecular characterization of ICE V_{chVie0} and its disappearance in *Vibrio cholerae* O1 strains isolated in 2003 in Vietnam. *FEMS Microbiol. Lett.* 266, 42-48.
- 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.
- 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.

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

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

Blokesch, M. and Schoolnik, G.K. (2007). Serogroup conversion of *Vibrio cholerae* in aquatic reservoirs. *PLoS Pathog.* 3, e81.

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.

Bordeleau, E., Brouillette, E., Robichaud, N. and Burrus, V. (2009). 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.

Brigulla, M. and Wackernagel, W. (2010). Molecular aspects of gene transfer and foreign DNA acquisition in prokaryotes with regard to safety issues. *Appl. Microbiol. Biotechnol.* 86, 1027-1041.

Brussow, H., Canchaya, C. and Hardt, W.D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* 68, 560-602.

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

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

Burrus, V., Pavlovic, G., Decaris, B. and Guedon, G. (2002b). 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.

Burrus, V., Quezada-Calvillo, R., Marrero, J. and Waldor, M.K. (2006b). SXT-related integrating conjugative element in New World *Vibrio cholerae*. *Appl. Environ. Microbiol.* 72, 3054-3057.

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

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

Butala, M., Zgur-Bertok, D. and Busby, S.J. (2009). The bacterial LexA transcriptional repressor. *Cell. Mol. Life Sci.* 66, 82-93.

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 ICEs. *J. Bacteriol.* 190, 5328-5338.

Ceccarelli, D., Spagnoletti, M., Bacciu, D., Danin-Poleg, Y., Mendiratta, D.K., Kashi, Y., Cappuccinelli, P., Burrus, V. and Colombo, M.M. (2011). ICE*VchInd5* is prevalent in epidemic *Vibrio cholerae* O1 El Tor strains isolated in India. *Int. J. Med. Microbiol.* 301, 318-324.

Chen, I., Christie, P.J. and Dubnau, D. (2005). The ins and outs of DNA transfer in bacteria. *Science* 310, 1456-1460.

Chen, I. and Dubnau, D. (2004). DNA uptake during bacterial transformation. *Nat. Rev. Microbiol.* 2, 241-249.

- Chen, W.Y., Ho, J.W., Huang, J.D. and Watt, R.M. (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.
- Cirz, R.T., Gingles, N. and Romesberg, F.E. (2006). Side effects may include evolution. *Nat. Med.* 12, 890-891.
- Claverys, J.P., Martin, B. and Polard, P. (2009). The genetic transformation machinery: composition, localization, and mechanism. *FEMS Microbiol. Rev.* 33, 643-656.
- Coetzee, J.N., Datta, N. and Hedges, R.W. (1972). R factors from *Proteus rettgeri*. *J. Gen. Microbiol.* 72, 543-552.
- Comstock, L.E., Johnson, J.A., Michalski, J.M., Morris, J.G., Jr. and Kaper, J.B. (1996). Cloning and sequence of a region encoding a surface polysaccharide of *Vibrio cholerae* O139 and characterization of the insertion site in the chromosome of *Vibrio cholerae* O1. *Mol. Microbiol.* 19, 815-826.
- Court, D.L., Sawitzke, J.A. and Thomason, L.C. (2002). Genetic engineering using homologous recombination. *Annu. Rev. Genet.* 36, 361-388.
- 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.
- 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.*
- Das, B., Halder, K., Pal, P. and Bhadra, R.K. (2007). Small chromosomal integration site of classical CTX prophage in Mozambique *Vibrio cholerae* O1 biotype El Tor strain. *Arch. Microbiol.* 188, 677-683.

- 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.
- Datta, S., Costantino, N., Zhou, X. and Court, D.L. (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.
- Derbise, A., Lesic, B., Dacheux, D., Ghigo, J.M. and Carniel, E. (2003). A rapid and simple method for inactivating chromosomal genes in *Yersinia*. *FEMS Immunol. Med. Microbiol.* *38*, 113-116.
- 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.
- Ehara, M., Nguyen, B.M., Nguyen, D.T., Toma, C., Higa, N. and Iwanaga, M. (2004). Drug susceptibility and its genetic basis in epidemic *Vibrio cholerae* O1 in Vietnam. *Epidemiol. Infect.* *132*, 595-600.
- Ellermeier, C.D., Janakiraman, A. and Slauch, J.M. (2002). Construction of targeted single copy lac fusions using lambda Red and FLP-mediated site-specific recombination in bacteria. *Gene* *290*, 153-161.
- Erill, I., Campoy, S. and Barbe, J. (2007). Aeons of distress: an evolutionary perspective on the bacterial SOS response. *FEMS Microbiol. Rev.* *31*, 637-656.
- Fricke, W.F., Welch, T.J., McDermott, P.F., Mammel, M.K., LeClerc, J.E., White, D.G., Cebula, T.A. and Ravel, J. (2009). Comparative genomics of the IncA/C multidrug resistance plasmid family. *J. Bacteriol.* *191*, 4750-4757.

Galimand, M., Guiyoule, A., Gerbaud, G., Rasoamanana, B., Chanteau, S., Carniel, E. and Courvalin, P. (1997). Multidrug resistance in *Yersinia pestis* mediated by a transferable plasmid. *N. Engl. J. Med.* 337, 677-680.

Garriss, G. and Burrus, V. (2012). DNA Damaging Agents Induce the RecA-Independent Homologous Recombination Functions of Integrating Conjugative Elements of the SXT/R391 Family. *J. Bacteriol.* *Submitted*,

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

Gerlach, R.G., Holzer, S.U., Jackel, D. and Hensel, M. (2007). Rapid engineering of bacterial reporter gene fusions by using Red recombination. *Appl. Environ. Microbiol.* 73, 4234-4242.

Hamdan, S.M. and van Oijen, A.M. (2010). Timing, coordination, and rhythm: acrobatics at the DNA replication fork. *J. Biol. Chem.* 285, 18979-18983.

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.

Hedges, R.W. (1974). R factors from Providence. *J. Gen. Microbiol.* 81, 171-181.

Hedges, R.W. (1975). R factors from *Proteus mirabilis* and *P. vulgaris*. *J. Gen. Microbiol.* 87, 301-311.

Hochhut, B., Beaver, J.W., Woodgate, R. and Waldor, M.K. (2001a). 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.

Hochhut, B., Lotfi, Y., Mazel, D., Faruque, S.M., Woodgate, R. and Waldor, M.K. (2001b). Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins. *Antimicrob Agents Chemother* 45, 2991-3000.

- 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. and Waldor, M.K. (1999). Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. *Mol. Microbiol.* 32, 99-110.
- Husseiny, M.I. and Hensel, M. (2005). Rapid method for the construction of *Salmonella enterica* Serovar *typhimurium* vaccine carrier strains. *Infect. Immun.* 73, 1598-1605.
- Iwanaga, M., Toma, C., Miyazato, T., Insiengmay, S., Nakasone, N. and 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.
- Iyer, L.M., Koonin, E.V. and Aravind, L. (2002). Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* 3, 8.
- Janion, C. (2008). Inducible SOS response system of DNA repair and mutagenesis in *Escherichia coli*. *Int J Biol Sci* 4, 338-344.
- Juiz-Rio, S., Osorio, C.R., de Lorenzo, V. and Lemos, M.L. (2005). Subtractive hybridization reveals a high genetic diversity in the fish pathogen *Photobacterium damsela* subsp. *piscicida*: evidence of a SXT-like element. *Microbiology* 151, 2659-2669.
- Kaper, J.B., Morris, J.G., Jr. and Levine, M.M. (1995). Cholera. *Clin. Microbiol. Rev.* 8, 48-86.
- Karu, A.E., Sakaki, Y., Echols, H. and Linn, S. (1975). The gamma protein specified by bacteriophage gamma. Structure and inhibitory activity for the recBC enzyme of *Escherichia coli*. *J. Biol. Chem.* 250, 7377-7387.
- Kmiec, E. and Holloman, W.K. (1981). Beta protein of bacteriophage lambda promotes renaturation of DNA. *J. Biol. Chem.* 256, 12636-12639.

- Kobiler, O., Rokney, A. and Oppenheim, A.B. (2007). Phage lambda CIII: a protease inhibitor regulating the lysis-lysogeny decision. *PLoS ONE* 2, e363.
- Kolodner, R., Hall, S.D. and Luisi-DeLuca, C. (1994). Homologous pairing proteins encoded by the *Escherichia coli* *recE* and *recT* genes. *Mol. Microbiol.* 11, 23-30.
- Korichi, M.N., Belhocine, S. and Rahal, K. (1997). [Inc J plasmids identified for the first time in *Vibrio cholerae* El Tor]. *Med. Trop. (Mars.)* 57, 249-252.
- Kovall, R. and Matthews, B.W. (1997). Toroidal structure of lambda-exonuclease. *Science* 277, 1824-1827.
- Kowalczykowski, S.C. (2000). Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* 25, 156-165.
- Kruger, N.J. and Stingl, K. (2011). Two steps away from novelty--principles of bacterial DNA uptake. *Mol. Microbiol.* 80, 860-867.
- Krupovic, M., Prangishvili, D., Hendrix, R.W. and Bamford, D.H. (2011). Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol. Mol. Biol. Rev.* 75, 610-635.
- Kuzminov, A. (1999). Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol. Mol. Biol. Rev.* 63, 751-813.
- Li, Z., Karakousis, G., Chiu, S.K., Reddy, G. and Radding, C.M. (1998). The beta protein of phage lambda promotes strand exchange. *J. Mol. Biol.* 276, 733-744.
- Lindsey, R.L., Fedorka-Cray, P.J., Frye, J.G. and Meinersmann, R.J. (2009). IncA/C plasmids are prevalent in multidrug-resistant *Salmonella enterica* isolates. *Appl. Environ. Microbiol.* 75, 1908-1915.

- Lipinska, B., Rao, A.S., Bolten, B.M., Balakrishnan, R. and Goldberg, E.B. (1989). Cloning and identification of bacteriophage T4 gene 2 product *gp2* and action of *gp2* on infecting DNA *in vivo*. *J. Bacteriol.* *171*, 488-497.
- Lo Scrudato, M. and Blokesch, M. (2012). The regulatory network of natural competence and transformation of *Vibrio cholerae*. *PLoS Genet.* *8*, e1002778.
- Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y. and Stewart, A.F. (2010). Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol Biol* *11*, 54.
- 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. (2007). The SXT/R391 family of integrative conjugative elements is composed of two exclusion groups. *Journal of bacteriology* *189*, 3302-3305.
- 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.
- Marsic, N., Roje, S., Stojiljkovic, I., Salaj-Smic, E. and Trgovcevic, Z. (1993). *In vivo* studies on the interaction of RecBCD enzyme and lambda Gam protein. *J. Bacteriol.* *175*, 4738-4743.
- Matic, I., Rayssiguier, C. and Radman, M. (1995). Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. *Cell* *80*, 507-515.
- Matic, I., Taddei, F. and Radman, M. (2004). Survival versus maintenance of genetic stability: a conflict of priorities during stress. *Res. Microbiol.* *155*, 337-341.

Matthew, M., Hedges, R.W. and Smith, J.T. (1979). Types of beta-lactamase determined by plasmids in gram-negative bacteria. *J. Bacteriol.* *138*, 657-662.

Moat, A.G., Foster, J.W., Spector, M.P.

Mosberg, J.A., Lajoie, M.J. and Church, G.M. (2010). Lambda red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. *Genetics* *186*, 791-799.

Murphy, K.C. (2007). The lambda Gam protein inhibits RecBCD binding to dsDNA ends. *J. Mol. Biol.* *371*, 19-24.

Murphy, K.C. and Lewis, L.J. (1993). Properties of *Escherichia coli* expressing bacteriophage P22 Abc (anti-RecBCD) proteins, including inhibition of Chi activity. *J. Bacteriol.* *175*, 1756-1766.

Muyrers, J.P., Zhang, Y., Buchholz, F. and Stewart, A.F. (2000). RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev.* *14*, 1971-1982.

Muyrers, J.P., Zhang, Y., Testa, G. and Stewart, A.F. (1999). Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* *27*, 1555-1557.

Novick, R.P. (1987). Plasmid incompatibility. *Microbiol. Rev.* *51*, 381-395.

Nowosińska, A. (2007). Bacterial DNA repair genes and their eukaryotic homologues: 5. The role of recombination in DNA repair and genome stability. *Acta Biochim. Pol.* *54*, 483-494.

Ochman, H., Lawrence, J.G. and Groisman, E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* *405*, 299-304.

Oppenheim, A.B., Kobilier, O., Stavans, J., Court, D.L. and Adhya, S. (2005). Switches in bacteriophage lambda development. *Annu. Rev. Genet.* *39*, 409-429.

- Oppenheim, A.B., Rattray, A.J., Bubunenko, M., Thomason, L.C. and Court, D.L. (2004). *In vivo* recombineering of bacteriophage lambda by PCR fragments and single-strand oligonucleotides. *Virology* 319, 185-189.
- 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.
- Pan, J.C., Ye, R., Wang, H.Q., Xiang, H.Q., Zhang, W., Yu, X.F., Meng, D.M. and He, Z.S. (2008). *Vibrio cholerae* O139 multiple-drug resistance mediated by *Yersinia pestis* pIP1202-like conjugative plasmids. *Antimicrob. Agents Chemother.* 52, 3829-3836.
- Partridge, S.R. (2011). Analysis of antibiotic resistance regions in Gram-negative bacteria. *FEMS Microbiol. Rev.* 35, 820-855.
- Passy, S.I., Yu, X., Li, Z., Radding, C.M. and Egelman, E.H. (1999). Rings and filaments of beta protein from bacteriophage lambda suggest a superfamily of recombination proteins. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4279-4284.
- Pearson, M.M., Sebahia, M., Churcher, C., Quail, M.A., Seshasayee, A.S., Luscombe, N.M., Abdallah, Z., Arrosmith, C., Atkin, B., Chillingworth, T., Hauser, H., Jagels, K., Moule, S., Mungall, K., Norbertczak, H., Rabinowitsch, E., Walker, D., Whithead, S., Thomson, N.R., Rather, P.N., Parkhill, J. and Mobley, H.L. (2008). Complete genome sequence of uropathogenic *Proteus mirabilis*, a master of both adherence and motility. *J. Bacteriol.* 190, 4027-4037.
- 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.

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.

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.* *228*, 294-299.

Poteete, A.R. (2001). What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiol. Lett.* *201*, 9-14.

Poteete, A.R. (2008). Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol. Microbiol.* *68*, 66-74.

Poteete, A.R. and Fenton, A.C. (1993). Efficient double-strand break-stimulated recombination promoted by the general recombination systems of phages lambda and P22. *Genetics* *134*, 1013-1021.

Poteete, A.R. and Volkert, M.R. (1988). Activation of recF-dependent recombination in *Escherichia coli* by bacteriophage lambda- and P22-encoded functions. *J. Bacteriol.* *170*, 4379-4381.

Reisinger, G.R., Rietsch, A., Lubitz, W. and Blasi, U. (1993). Lambda kil-mediated lysis requires the phage context. *Virology* *193*, 1033-1036.

Rybalchenko, N., Golub, E.I., Bi, B. and Radding, C.M. (2004). Strand invasion promoted by recombination protein beta of coliphage lambda. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 17056-17060.

Schlacher, K., Pham, P., Cox, M.M. and Goodman, M.F. (2006). Roles of DNA polymerase V and RecA protein in SOS damage-induced mutation. *Chem. Rev.* *106*, 406-419.

Sergueev, K., Court, D., Reaves, L. and Austin, S. (2002). E.coli cell-cycle regulation by bacteriophage lambda. *J. Mol. Biol.* *324*, 297-307.

- Silverstein, J.L. and Goldberg, E.B. (1976). T4 DNA injection. II. Protection of entering DNA from host exonuclease V. *Virology* 72, 212-223.
- Skippington, E. and Ragan, M.A. (2011). Lateral genetic transfer and the construction of genetic exchange communities. *FEMS Microbiol. Rev.* 35, 707-735.
- 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.
- Smith, G.R. (2012). How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. *Microbiol. Mol. Biol. Rev.* 76, 217-228.
- Stahl, F.W. (1998). Recombination in phage lambda: one geneticist's historical perspective. *Gene* 223, 95-102.
- Stokes, H.W. and Gillings, M.R. (2011). Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiol. Rev.* 35, 790-819.
- Szczepanska, A.K. (2009). Bacteriophage-encoded functions engaged in initiation of homologous recombination events. *Crit. Rev. Microbiol.* 35, 197-220.
- 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.
- Taviani, E., Grim, C.J., Chun, J., Huq, A. and Colwell, R.R. (2009). Genomic analysis of a novel integrative conjugative element in *Vibrio cholerae*. *FEBS Lett.* 583, 3630-3636.
- Thomason, L.C., Costantino, N., Shaw, D.V. and Court, D.L. (2007). Multicopy plasmid modification with phage lambda Red recombineering. *Plasmid* 58, 148-158.

- Thungapathra, M., Amita, Sinha, K.K., Chaudhuri, S.R., Garg, P., Ramamurthy, T., Nair, G.B. and Ghosh, A. (2002). Occurrence of antibiotic resistance gene cassettes *aac(6)-Ib*, *dfrA5*, *dfrA12*, and *ereA2* in class I integrons in non-O1, non-O139 *Vibrio cholerae* strains in India. *Antimicrob. Agents Chemother.* *46*, 2948-2955.
- Toleman, M.A., Bennett, P.M. and Walsh, T.R. (2006). ISCR elements: novel gene-capturing systems of the 21st century? *Microbiol. Mol. Biol. Rev.* *70*, 296-316.
- Toleman, M.A. and Walsh, T.R. (2011). Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. *FEMS Microbiol. Rev.* *35*, 912-935.
- Toma, C., Nakasone, N., Song, T. and Iwanaga, M. (2005). *Vibrio cholerae* SXT element, Laos. *Emerg. Infect. Dis.* *11*, 346-347.
- Toussaint, A. and Merlin, C. (2002). Mobile elements as a combination of functional modules. *Plasmid* *47*, 26-35.
- Waldor, M.K. and Mekalanos, J.J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* *272*, 1910-1914.
- 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.
- Weigel, C. and Seitz, H. (2006). Bacteriophage replication modules. *FEMS Microbiol. Rev.* *30*, 321-381.
- Welch, T.J., Fricke, W.F., McDermott, P.F., White, D.G., Rosso, M.L., Rasko, D.A., Mammel, M.K., Eppinger, M., Rosovitz, M.J., Wagner, D., Rahalison, L., Leclerc, J.E., Hinshaw, J.M., Lindler, L.E., Cebula, T.A., Carniel, E. and Ravel, J. (2007). Multiple antimicrobial resistance in plague: an emerging public health risk. *PLoS ONE* *2*, e309.

- Wiedenbeck, J. and Cohan, F.M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* *35*, 957-976.
- Worth, L., Jr., Clark, S., Radman, M. and Modrich, P. (1994). Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proc. Natl. Acad. Sci. U. S. A.* *91*, 3238-3241.
- 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.
- Wozniak, R.A. and Waldor, M.K. (2009). A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genet.* *5*, e1000439.
- 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.
- Wu, Z., Xing, X., Bohl, C.E., Wisler, J.W., Dalton, J.T. and Bell, C.E. (2006). Domain structure and DNA binding regions of beta protein from bacteriophage lambda. *J. Biol. Chem.* *281*, 25205-25214.
- Yamamoto, S., Izumiya, H., Morita, M., Arakawa, E. and Watanabe, H. (2009). Application of lambda Red recombination system to *Vibrio cholerae* genetics: Simple methods for inactivation and modification of chromosomal genes. *Gene*
- Yang, W., Junop, M.S., Ban, C., Obmolova, G. and Hsieh, P. (2000). DNA mismatch repair: from structure to mechanism. *Cold Spring Harb. Symp. Quant. Biol.* *65*, 225-232.
- Yokota, T. and Kuwahara, S. (1977). Temperature-sensitive R plasmid obtained from naturally isolated drug-resistant *Vibrio cholerae* (biotype El Tor). *Antimicrob. Agents Chemother.* *11*, 13-20.

Zabeau, M., Friedman, S., Van Montagu, M. and Schell, J. (1980). The *ral* gene of phage lambda. I. Identification of a non-essential gene that modulates restriction and modification in *E. coli*. *Mol. Gen. Genet.* 179, 63-73.

ANNEXE 1

Tableau 2. Principaux membres de la famille SXT/R391.

ICE	Origine				Résistance ^c	Allèle <i>eex</i> ^d	Référence
	Hôte ^a	Localisation	Année	Type ^b			
ICE <i>Vch</i> Ban1	<i>Vibrio cholerae</i> O1	Bangladesh	1998	C	Su Tm Cm Sm	S ^a	(Hochhut <i>et al.</i> , 2001b)
ICE <i>Vch</i> Ban2	<i>V. cholerae</i> O1	Bangladesh	2005	C	Su Tm Tc	R1	(Marrero et Waldor, 2007a)
ICE <i>Vch</i> Ban3	<i>V. cholerae</i> O1	Bangladesh	2005	C	Su Tm Cm	R1	(Marrero et Waldor, 2007a)
ICE <i>Vch</i> Ban4	<i>V. cholerae</i> O1	Bangladesh	1998	C	Su Tm	R1	(Marrero et Waldor, 2007a)
ICE <i>Vch</i> Ban5	<i>V. cholerae</i> O1	Bangladesh	1998	C	Su Tm Cm Sm	R1	(Marrero et Waldor, 2007a ; Wozniak <i>et al.</i> , 2009)
ICE <i>Vch</i> Ind1	<i>V. cholerae</i> O1	Inde	1994	C	Su Tm Cm Sm	R1	(Hochhut <i>et al.</i> , 2001b ; Marrero et Waldor, 2007a)
ICE <i>Vch</i> Ind2	<i>V. cholerae</i> O1	Inde	1994	C	Su Tm	R1	(Marrero et Waldor, 2007a)
pJY1 ^{e,f}	<i>V. cholerae</i> O1	Philippines	1973	C	Su Cm Sm	nd	(Yokota et Kuwahara, 1977)
ICE <i>Vch</i> Lao1 ^f	<i>V. cholerae</i> O1	Lao PDR	1998	C	Su Tm Cm Sm Tc	nd	(Iwanaga <i>et al.</i> , 2004 ; Toma <i>et al.</i> , 2005)
ICE <i>Vch</i> Vie0 ^f	<i>V. cholerae</i> O1	Vietnam	1990	C	aucune	nd	(Bani <i>et al.</i> , 2007)
ICE <i>Vch</i> Vie1 ^f	<i>V. cholerae</i> O1	Vietnam	2000	C	Su Cm Sm Tc	nd	(Ehara <i>et al.</i> , 2004)
ICE <i>Vch</i> Alg1 ^f	<i>V. cholerae</i> O1	Algérie	1994	C	Su Tm Cm Sm Tc	nd	(Korichi <i>et al.</i> , 1997)
ICE <i>Vch</i> Ind5	<i>V. cholerae</i> O1	Inde	1994-2005	C	Su Tm Cm Sm	R1 ^g	(Ceccarelli <i>et al.</i> , 2011 ; Wozniak <i>et al.</i> , 2009)
ICE <i>Vch</i> Ban9	<i>V. cholerae</i> O1	Bangladesh	1994	C	Cm Tc Sm Su Tm	R1 ^g	(Taviani <i>et al.</i> , 2009 ; Wozniak <i>et al.</i> , 2009)
ICE <i>Vch</i> Moz10	<i>V. cholerae</i> O1 B33	Mozambique	2004	C	Cm Tc Sm Su Tm	R1 ^g	(Das <i>et al.</i> , 2007 ; Wozniak <i>et al.</i> , 2009)
ICE <i>Vch</i> Ban10	<i>V. cholerae</i> O1 CIRS101	Bangladesh	2010	C	Su Tm Cm Sm	R1 ^g	(Ceccarelli <i>et al.</i> , 2011)
ICE <i>Vch</i> Moz3 ^f	<i>V. cholerae</i> non-O1	Mozambique	2002	E	Pn Sp β-lactams	nd	(Taviani <i>et al.</i> , 2008)
ICE <i>Vch</i> Mex1	<i>V. cholerae</i> non-O1/O139	Méxique	2001	E	aucune	R3	(Burrus <i>et al.</i> , 2006b ; Wozniak <i>et al.</i> , 2009)
ICE <i>Vch</i> Ind6 ^f	<i>V. cholerae</i> O2	Inde	1998	E	na	nd	(Thungapathra <i>et al.</i> , 2002)
ICE <i>Vch</i> Ind7 ^f	<i>V. cholerae</i> O130	Inde	1998	E	na	nd	(Thungapathra <i>et al.</i> , 2002)
ICE <i>Vch</i> Ind8 ^f	<i>V. cholerae</i> O144	Inde	1998	E	na	nd	(Thungapathra <i>et al.</i> , 2002)
ICE <i>Vch</i> Ind9 ^f	<i>V. cholerae</i> O150	Inde	1998	E	na	nd	(Thungapathra <i>et al.</i> , 2002)
ICE <i>Vch</i> Ind10 ^f	<i>V. cholerae</i> O151	Inde	1998	E	na	nd	(Thungapathra <i>et al.</i> , 2002)
SXT ^{MO10}	<i>V. cholerae</i> O139 MO10	Inde	1992	C	Su Tm Cm Sm	S	(Hochhut <i>et al.</i> , 2001b ; Marrero et Waldor, 2007a ; Waldor <i>et al.</i> , 1996 ; Wozniak <i>et al.</i> , 2009)

ICE	Origine			Type ^b	Résistance ^c	Allèle <i>eex</i> ^d	Référence
	Hôte ^a	Localisation	Année				
ICEVchInd4	<i>V. cholerae</i> O139 AS207	Inde	1997	C	Su Cm Sm	S	(Hochhut <i>et al.</i> , 2001b ; Marrero et Waldor, 2007a ; Wozniak <i>et al.</i> , 2009)
ICEVchBan7	<i>V. cholerae</i> O139 2125/98	Bangladesh	1998	C		S	(Hochhut <i>et al.</i> , 2001b ; Marrero et Waldor, 2007a)
ICEVchHKo1 ^f	<i>V. cholerae</i> O139	Hong Kong	1993	C	aucune	nd	(Hochhut <i>et al.</i> , 2001b)
ICEVchSL1	<i>V. cholerae</i> O139 E712	Sri Lanka	1994	E	Su Tm Cm Sm	S5	(Hochhut <i>et al.</i> , 2001b ; Marrero et Waldor, 2007)
ICEVflInd1	<i>V. fluvialis</i> H-08942	Inde	2002	C	Su Tm Cm Sm	S1	(Ahmed <i>et al.</i> , 2005 ; Marrero et Waldor, 2007a ; Wozniak <i>et al.</i> , 2009)
R391 ^e	<i>Providencia rettgeri</i>	Afrique du Sud	1967	C	Kn Hg	R	(Boltner <i>et al.</i> , 2002 ; Coetzee <i>et al.</i> , 1972 ; Marrero et Waldor, 2007a)
R748 ^{e,f}	<i>Providencia</i> spp.	Afrique du Sud	1974	na	Kn Hg	nd	(Hedges, 1974)
ICEPa/Ban1	<i>P. alcalifaciens</i>	Bangladesh	1999	C	Su Tm Cm Sm	R2	(Hochhut <i>et al.</i> , 2001b ; Marrero et Waldor, 2007a ; Wozniak <i>et al.</i> , 2009)
R997 ^e	<i>Proteus mirabilis</i>	Inde	1977	C	Su Sm Ap	S2	(Marrero et Waldor, 2007a ; Matthew <i>et al.</i> , 1979)
ICEPmiJpn1	<i>P. mirabilis</i> TUM4660	Japon	2010	C	Céphalosporines	S6 ^g	(Harada <i>et al.</i> , 2010)
ICEPmiUSA1	<i>P. mirabilis</i>	États-Unis	1986	C	aucune	S7 ^g	(Pearson <i>et al.</i> , 2008 ; Wozniak <i>et al.</i> , 2009)
R705 ^{e,f}	<i>P. vulgaris</i>	Afrique du Sud	1975	na	Kn Hg	nd	(Hedges, 1975)
ICEPdaSpa1	<i>Photobacterium damsela</i> subsp. <i>piscicidae</i>	Espagne	2001	P	Tc	S3	(Juiz-Rio <i>et al.</i> , 2005 ; Marrero et Waldor, 2007a ; Osorio <i>et al.</i> , 2008 ; Wozniak <i>et al.</i> , 2009)
pMERPH ^e	<i>Shewanella putrefaciens</i>	Royaume Uni	1990	E	Hg	R4	(Marrero et Waldor, 2007a ; Peters <i>et al.</i> , 1991)
ICESpuPO1	<i>S. putrefaciens</i> W3-18-1	Océan Pacifique	na	E	Pompe efflux putative	S4	(Hochhut <i>et al.</i> , 2001b ; Marrero et Waldor, 2007a ; Pembroke et Piterina, 2006 ; Wozniak <i>et al.</i> , 2009)

^a Lorsque que pertinent la souche est indiquée.

^b Type d'isolat: C, clinique; E, environmental; P, poisson malade.

^c Ap, ampicilline; Cm, chloramphénicol; Kn, kanamycine; Pn, pénicilline; Sm, streptomycine; Su, sulfaméthoxazole; Tc, tétracycline; Tm, triméthoprime; Hg, mercure.

^d Variants des déterminants de l'exclusion d'entrée. S and R indiquent les groupes d'exclusion S et R, respectivement.

^e Membres de l'ancienne famille IncJ. Les autres membres sont R392 et R397, qui sont des clones probables de R391, R706, un clone probable de R705, et R749, clone probable de R748.

^f Membres de la famille SXT/R391 dont la séquence complète n'est pas disponible.

^g Allèle *eex* déterminé par notre analyse.

^h Allèle *eex* corrigé de Marrero and Waldor (2007a).

nd: données non disponibles.

ANNEXE 2

Comparative ICE Genomics: Insights into the Evolution of the SXT/R391 Family of ICEs

Rachel A. F. Wozniak^{1,2,3}, Derrick E. Fouts³, Matteo Spagnoletti⁴, Mauro M. Colombo⁴, Daniela Ceccarelli⁵, Geneviève Garriss⁵, Christine Déry⁵, Vincent Burrus^{5*}, Matthew K. Waldor^{1,2,6*}

1 Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **2** Department of Genetics, Tufts Medical School, Boston, Massachusetts, United States of America, **3** J. Craig Venter Institute, Rockville, Maryland, United States of America, **4** Dipartimento di Biologia Cellulare e dello Sviluppo, Università di Roma La Sapienza, Rome, Italy, **5** Centre d'Étude et de Valorisation de la Diversité Microbienne, Département de Biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada, **6** Howard Hughes Medical Institute, Chevy Chase, Maryland, United States of America

Abstract

Integrating and conjugative elements (ICEs) are one of the three principal types of self-transmissible mobile genetic elements in bacteria. ICEs, like plasmids, transfer via conjugation; but unlike plasmids and similar to many phages, these elements integrate into and replicate along with the host chromosome. Members of the SXT/R391 family of ICEs have been isolated from several species of gram-negative bacteria, including *Vibrio cholerae*, the cause of cholera, where they have been important vectors for disseminating genes conferring resistance to antibiotics. Here we developed a plasmid-based system to capture and isolate SXT/R391 ICEs for sequencing. Comparative analyses of 13 SXT/R391 ICEs derived from diverse hosts and locations revealed that they contain 52 perfectly syntenic and nearly identical core genes that serve as a scaffold capable of mobilizing an array of variable DNA. Furthermore, selection pressure to maintain ICE mobility appears to have restricted insertions of variable DNA into intergenic sites that do not interrupt core functions. The variable genes confer diverse element-specific phenotypes, such as resistance to antibiotics. Functional analysis of a set of deletion mutants revealed that less than half of the conserved core genes are required for ICE mobility; the functions of most of the dispensable core genes are unknown. Several lines of evidence suggest that there has been extensive recombination between SXT/R391 ICEs, resulting in re-assortment of their respective variable gene content. Furthermore, our analyses suggest that there may be a network of phylogenetic relationships among sequences found in all types of mobile genetic elements.

Citation: Wozniak RAF, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, et al. (2009) Comparative ICE Genomics: Insights into the Evolution of the SXT/R391 Family of ICEs. *PLoS Genet* 5(12): e1000786. doi:10.1371/journal.pgen.1000786

Editor: Ivan Matic, Université Paris Descartes, INSERM U571, France

Received: October 7, 2009; **Accepted:** November 24, 2009; **Published:** December 24, 2009

Copyright: © 2009 Wozniak et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: MKW acknowledges support from NIH (AI R37-42347) and HHMI. GG is the recipient of a PhD research scholarship from FQRNT. VB holds a Canada Research Chair in molecular biology, impact and evolution of bacterial mobile elements and is grateful for support from the Natural Sciences and Engineering Research Council of Canada (Discovery Grant Program). MS was supported by a grant from PRIN 2007 - Italy, and DC was supported by a fellowship from Cenci Bolognini-Institut Pasteur Foundation, Italy. Sequencing at the JCVI was funded by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract number N01-AI-30071. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Vincent.Burrus@USherbrooke.ca (VB); mwaldor@rics.bwh.harvard.edu (MKW)

Introduction

There are three types of self-transmissible mobile genetic elements: plasmids, bacteriophages and integrative conjugative elements (ICEs). All three classes of elements enable horizontal transmission of genetic information and all have had major impacts on bacterial evolution [1–4]. ICEs, (aka conjugation transposons), like plasmids, are transmitted via conjugation; however, unlike plasmids, ICEs integrate into and replicate along with the chromosome. Following integration, ICEs can excise from the chromosome and form circular molecules that are intermediates in ICE transfer. Plasmids and phages have been the subject of more extensive study than ICEs and while there is growing understanding of the molecular aspects of several ICEs [5–10], to date there have been few reports of comparative ICE genomics [11,12] and consequently understanding of ICE evolution is only beginning to be unraveled.

Diverse ICEs have been identified in a variety of gram-positive and gram-negative organisms [13]. These elements utilize a

variety of genes to mediate the core ICE functions of chromosome integration, excision and conjugation. In addition to a core gene set, ICEs routinely contain genes that confer specific phenotypes upon their hosts, such as resistance to antibiotics and heavy metals [14–18], aromatic compound degradation [19] or nitrogen fixation [20].

SXT is an ~100 Kb ICE that was originally discovered in *Vibrio cholerae* O139 [16], the first non-O1 serogroup to cause epidemic cholera [21]. SXT encodes resistances to several antibiotics, including sulfamethoxazole and trimethoprim (which together are often abbreviated as SXT) that had previously been useful in the treatment of cholera. Since the emergence of *V. cholerae* O139 on the Indian subcontinent in 1992, SXT or a similar ICE has been found in most clinical isolates of *V. cholerae*, including *V. cholerae* serogroup O1, from both Asia and Africa. Other vibrio species besides *V. cholerae* have also been found to harbor SXT-related ICEs [22]. Furthermore, SXT-like ICEs are not restricted to vibrio species, as such ICEs have been detected in *Photobacterium damsela*, *Shewanella putrefaciens* and *Providencia alcalifaciens* [23–25].

Author Summary

Integrative and conjugative elements (ICEs) are a class of mobile genetic elements that are key mediators of horizontal gene flow in bacteria. These elements integrate into the host chromosome, yet are able to excise and transfer via conjugation. Our understanding of ICE evolution is rudimentary. Here, we developed a method to capture ICEs on plasmids, thus facilitating their sequencing. Comparative analyses of the DNA sequences of ICEs from the same family revealed that they have an identical genetic structure consisting of syntenous, highly conserved core genes that are interrupted by clusters of diverse variable genes. Unexpectedly, many genes in the core backbone proved non-essential for ICE transfer. Comparisons of the variable gene content in the ICEs analyzed revealed that these elements are mosaics whose genomes have been shaped by inter-ICE recombination. Finally, our work suggests that ICEs contribute to a larger gene pool that connects all types of mobile elements.

Moreover, Hochhut et al [26] found that SXT is genetically and functionally related to the so-called 'Inc J' element R391, which was derived from a South African *Providencia rettgeri* strain isolated in 1967 [27]. It is now clear that Inc J elements are SXT-related ICEs that were originally misclassified as plasmids. In the laboratory, SXT has a fairly broad host range and can be transmitted between a variety of gram-negative organisms [16].

The SXT/R391 family of ICEs is now known to include more than 30 elements that have been detected in clinical and environmental isolates of several species of γ -proteobacteria from disparate locations around the globe [28]. SXT/R391 ICEs are grouped together as an ICE family because they all encode a nearly identical integrase, *Int*, a tyrosine recombinase, is considered a defining feature of these elements because it enables their site-specific integration into the 5' end of *prfC*, a conserved chromosomal gene that encodes peptide chain release factor 3 [29]. *Int* mediates recombination between nearly identical element and chromosome sequences, *attP* and *attB* respectively [29]. When an SXT/R391 ICE excises from the chromosome, *Int*, aided by *Xis*, a recombination directionality factor, mediates the reverse reaction - recombination between the extreme right and left ends (*attR* and *attL*) of the integrated element - thereby reconstituting *attP* and *attB* [6,29]. The excised circular SXT form is thought to be the principal substrate for its conjugative transfer. The genes that encode activities required for SXT transfer (*tra* genes) were originally found to be distantly related to certain plasmid *tra* genes [30–32]. The *tra* genes encode proteins important for processing DNA for transfer, mating pair formation and generating the conjugation machinery. Regulation of SXT excision and transfer is at least in part governed by a pathway that resembles the pathway governing the lytic development of the phage lambda. Agents that damage DNA and induce the bacterial SOS response are thought to stimulate the cleavage and inactivation of *SetR*, an SXT encoded λ CI-related repressor, which represses expression of *setD* and *setC*, transcription activators that promote expression of *int* and *tra* genes [5].

The complete nucleotide sequences of SXT (99.5kb) and R391 (89kb) were the first SXT/R391 ICE family genomes to be reported [14,32]. Comparative [33] and functional genomic analyses [5,32] revealed that these 2 ICEs share a set of conserved core genes that mediate their integration/excision (*int* and *xis*), conjugative transfer (various *tra* genes), and regulation (*setR*, *setCD*). In addition to the conserved genes, these 2 ICEs contain element

specific genes that confer element specific properties such as resistance to antibiotics or heavy metals. Interestingly, many of these genes were found in identical locations in SXT and R391, leading Beaber et al [33] to propose that there are 'hotspots' where SXT/R391 ICEs can acquire new DNA. The genomes of two additional SXT/R391 ICEs, ICE*PdaSpa1*, isolated from *Photobacterium damsela* [23], and ICE*SpuPO1*, derived from an environmental isolate of *Shewanella putrefaciens* [24] are now also known. These two genomes also share most of the conserved set of core genes present in SXT and R391 and contain element specific DNA.

Determination of the sequences of SXT/R391 family ICE genomes was a fairly arduous task due to their size and predominantly chromosomal localization. Here, we developed a method to capture and then sequence complete SXT/R391 ICE genomes. In addition, we identified 3 as yet unannotated SXT/R391 ICE genomes in the database of completed bacterial genomes. Comparative analyses of the 13 SXT/R391 ICE genomes now available allowed us to greatly refine our understanding of the organization and conservation of the core genes that are present in all members of this ICE family. Comparative and functional analyses also facilitated our proposal of the minimal functional SXT/R391 ICE genome. Furthermore, this work provides new knowledge of the considerable diversity of genes and potential accessory functions encoded by the variable DNA found in these mobile elements. Finally, this comparative genomics approach has allowed us to garner clues regarding the evolution of this class of mobile elements.

Results/Discussion

An ICE capture system

To date, ICE sequencing has been cumbersome because it has typically required construction of chromosome-derived cosmid libraries and screening for sequences that hybridize to ICE probes [23,32]. We constructed a vector (pIceCap) that enables capture of complete SXT/R391 ICE genomes on a low-copy plasmid to simplify the protocol for ICE sequencing. This plasmid is a derivative of the single-copy modified F plasmid pXX704 [34,35], which contains a minimal set of genes for F replication and segregation but lacks genes enabling conjugation. We modified pXX704 to include an ~400bp fragment that encompasses the SXT/R391 attachment site (*attB*) and thereby enabled *Int*-catalyzed site-specific recombination between *attB* on pIceCap and *attP* on an excised and transferred ICE to drive ICE capture (Figure 1). Conjugations between an SXT/R391 ICE-bearing donor strain and an *E. coli* recipient deleted for *prfC* (and thus chromosomal *attB*) and harboring pIceCap yielded exconjugants containing the transferred ICE integrated into pIceCap (Figure 1). We used the $\Delta prfC$ recipient to bias integration of the transferred ICE into pIceCap rather than the chromosome. In these experiments, we selected for exconjugants containing the transferred ICE integrated into pIceCap, using an antibiotic marker present on the ICE as well as a marker present in pIceCap. The low copy IceCap::ICE plasmid was then isolated and used as a substrate for shotgun sequencing. We also found that the IceCap::ICE plasmids were transmissible. Thus, in principle this technique should facilitate capture of ICEs that do not harbor genes conferring resistance to antibiotics, by mating out the IceCap::ICE plasmid into a new recipient and selecting for the marker on pIceCap.

SXT/R391 ICEs included in this analysis

A list of the 13 SXT/R391 ICEs whose genomes were analyzed and compared in this study is shown in Table 1. All of the ICEs

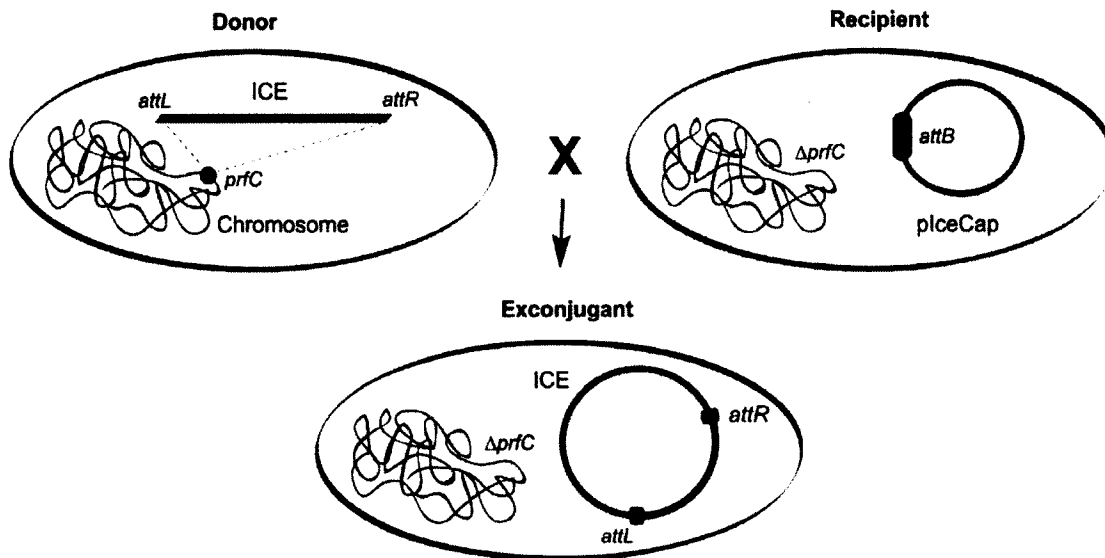


Figure 1. Schematic of the ICE capture system. Conjugation between a donor strain bearing a chromosomal ICE and a $\Delta prfC$ recipient strain harboring pIceCap, which contains *attB*, yields exconjugants that contain the transferred ICE integrated into pIceCap. Exconjugants were selected for using a marker on pIceCap and on the ICE. *attR* and *attL* represent the right and left ICE-chromosome junctions.
doi:10.1371/journal.pgen.1000786.g001

included in our analyses contain an *int* gene that was amplifiable using PCR primers for *int_{sxt}* [29]. They were isolated on 4 continents and from the Pacific Ocean during a span of more than 4 decades. They are derived from 7 different genera of γ -proteobacteria and the ICEs derived from *V. cholerae* strains are from both clinical and environmental isolates of 3 different *V. cholerae* serogroups.

Five of these ICE genome sequences were determined at the J. Craig Venter Institute (JCVI) using the ICE capture system described above (Table 1, rows 1–5). In addition, we sequenced ICE V_{Ind1} , also at the JCVI, by isolating cosmids that encompassed this *V. fluvialis* derived ICE prior to developing the ICE capture technique (Table 1, row 6). Table 1 (rows 7–10) also includes 4 previously unannotated ICE genomes that we found in BLAST searches of the NCBI database of completed but as yet unannotated genomes; 3 of these ICEs are clearly members of SXT/R391 ICE family since they are integrated into their respective host's *prfC* locus and contain *int* genes that are predicted to encode Int proteins that are 99% identical to Int_{sxt}. The fourth element, ICE V_{ChBan8} does not encode an Int_{sxt} orthologue; however, this element contains nearly identical homologues of most of the known conserved core SXT/R391 ICE family genes. ICE V_{ChBan8} will be discussed in more detail below but since it does not contain an Int_{sxt} orthologue it is not considered a member of the SXT/R391 family of ICEs and thus not included in our comparative study. Finally, Table 1 also includes the 4 SXT/R391 ICEs that were previously sequenced (Table 1, rows 11–14).

Despite the diversity of our sources for SXT/R391 ICEs, the genomes of two pairs of ICEs that we analyzed proved to be very similar. SXT^{MO10} and ICE V_{ChInd4} only differed by 13 SNPs in 7 genes and by the absence from ICE V_{ChInd4} of *dfr18*, a gene conferring trimethoprim resistance. These ICEs were derived from *V. cholerae* O139 strains isolated in India from different cities at different times: SXT^{MO10} from Chennai in 1992 and ICE V_{ChInd4} from Kolkata in 1997. The high degree of similarity of these two ICE genomes suggests that ICEs can be fairly stable over time. ICE V_{ChBan9} and ICE $V_{ChMoz10}$ were also extremely similar although ICE $V_{ChMoz10}$ lacks *dfrA1*, another allele for trimethoprim

resistance. These two ICEs were derived from *V. cholerae* O1 strains from Bangladesh (1994) and Mozambique (2004) respectively. The great similarity of these ICEs suggests that there has been spread of SXT-related ICEs between Asia and Africa in recent times. Studies of CTX prophage genomes have also suggested the spread of *V. cholerae* strains between these continents [36].

General structure and sizes of SXT/R391 genomes

The ICEs listed in Table 1 were initially compared using MAUVE [37] and LAGAN [38], programs that enable visualization of conserved and variable regions on a global scale. All of the SXT/R391 ICEs we analyzed share a common structure and have sizes ranging from 79,733 bp to 108,623 bp (Table 1 and Figure 2). They contain syntenous sets of 52 conserved core genes (Figure 2A) that total approximately 47kb and encode proteins with an average of 97% identity to those encoded by SXT. All of the individual ICEs also contain DNA that is relatively specific for individual elements (Figure 2B); the differences in the sizes of the variable regions accounts for the range in ICE sizes.

Five sites within the conserved SXT/R391 ICE structure have variable DNA present in all of the ICEs in Figure 2. Four of these sites were previously termed 'hotspots' for ICE acquisition of new DNA [33]. Due to similarities between SXT and R391, the fifth hotspot only became apparent through our comparison of the 13 ICEs examined here. Each of these hotspots (HS1 to HS5 in Figure 2B) is found in an intergenic region (see below), suggesting that the acquisition of these variable DNA regions has not interrupted core ICE gene functions. In addition, some of the ICEs have variable DNA inserted in additional intergenic locations or in *rumB* (labeled I–IV in Figure 2B). Previous analyses [32] indicated that the insertion in *rumB*, did not impair SXT transmissibility. Overall, comparison of these 13 SXT/R391 ICE genomes suggests that: 1) these elements consist of the same perfectly syntenous and nearly identical 52 core genes that serve as a scaffold (see below) capable of mobilizing a large range of variable DNA; and 2) selection pressure to maintain ICE mobility has restricted insertions of variable DNA into sites that do not interrupt core functions.

Table 1. SXT/R391 ICE family members analyzed in this study.

ICE	Host strain	Site and year of isolation	Size (bp)	% Identity to Int _{SXT}	Resistance profile	Notable Variable Genes	Genbank Accession Number	Strain or ICE References
ICEVchMex1	<i>Vibrio cholerae</i> non O1-O139	San Luis Potosi, Mexico 2001	82839	99% (410/413)	-	Fic family protein, diguanylate cyclase, restriction modification system	GQ463143	[66]
ICEVchInd4	<i>Vibrio cholerae</i> O139	Kolkata, India 1997	95491	100% (413/413)	<i>floR, strBA, sul2</i>	Toxin-antitoxin system	GQ463141	[54]
ICEVchInd5	<i>Vibrio cholerae</i> O1	Sevagram, India 1994	97847	99% (409/413)	<i>floR, strBA, sul2, dfrA1</i>	AraC family transcription regulator, glyoxylase abx resistance	GQ463142	This study
ICEVchBan5	<i>Vibrio cholerae</i> O1	Bangladesh, 1998	102131	99% (409/413)	<i>floR, strBA, sul2, dfrA1</i>	AraC family transcription regulator, glyoxylase abx resistance	GQ463140	[29]
ICEPaBan1	<i>Providencia alcalifaciens</i>	Bangladesh, 1999	96586	99% (409/413)	<i>floR, strBA, sul2, dfrA1</i>	Toxin-antitoxin system, phenazine biosynthesis protein, lysine exporter	GQ463139	[25]
ICEVflInd1	<i>Vibrio fluvialis</i>	Kolkata, India 2002	91369 ^(a)	99% (409/413)	<i>dfr18, floR, strBA, sul2</i>	Toxin-antitoxin system	GQ463144	[22]
ICEVchMoz10	<i>Vibrio cholerae</i> O1	Beira, Mozambique 2004	104495	99% (409/413)	<i>floR, strBA, sul2, tetA'</i>	AraC family transcription regulator, glyoxylase abx resistance, ATP-dependent Lon protease	ACHZ00000000	[67]
ICEPmiUsa1	<i>Proteus mirabilis</i>	Maryland, United States 1986	79733	99% (409/413)	-	ATP-dependent helicase	AM942759	[68]
ICEVchBan9	<i>Vibrio cholerae</i> O1	Matlab, Bangladesh 1994	106124	99% (409/413)	<i>floR, strBA, sul2, dfrA1, tetA'</i>	AraC family transcription regulator, glyoxylase abx resistance, ATP-dependent Lon protease	CP001485	[69]
ICEVchBan8	<i>Vibrio cholerae</i> non O1-O139	Bangladesh, 2001	105790 ^(a)	25% (76/301)	-	Toxin-antitoxin system	NZ_AAUU00000000	This study
SXT ^{MO10}	<i>Vibrio cholerae</i> O139	Chennai, India 2002	99452	100%	<i>dfr18, floR, strBA, sul2</i>	Toxin-antitoxin system	AY055428	[16]
R391	<i>Providencia rettgeri</i>	Pretoria, South Africa 1967	88532	99% (410/413)	<i>kanR, merRTPCA</i>	Sulfate transporter, universal stress protein	AY090559	[27]
ICEPdaSpa1	<i>Photobacterium damselae</i>	Galicia, Spain 2003	102985	99% (412/413)	<i>tetAR</i>	ATP-dependent Lon protease, heat-shock protein	AJ870986	[23]
ICESpuPO1	<i>Shewanella putrefaciens</i>	630m, Pacific Ocean 2000	108623	99% (409/413)	-	Zn/Co/Cd efflux system, restriction modification system	CP000503	[24]

^(a)The sequence is not complete and therefore the true size is not known.
doi:10.1371/journal.pgen.1000786.t001

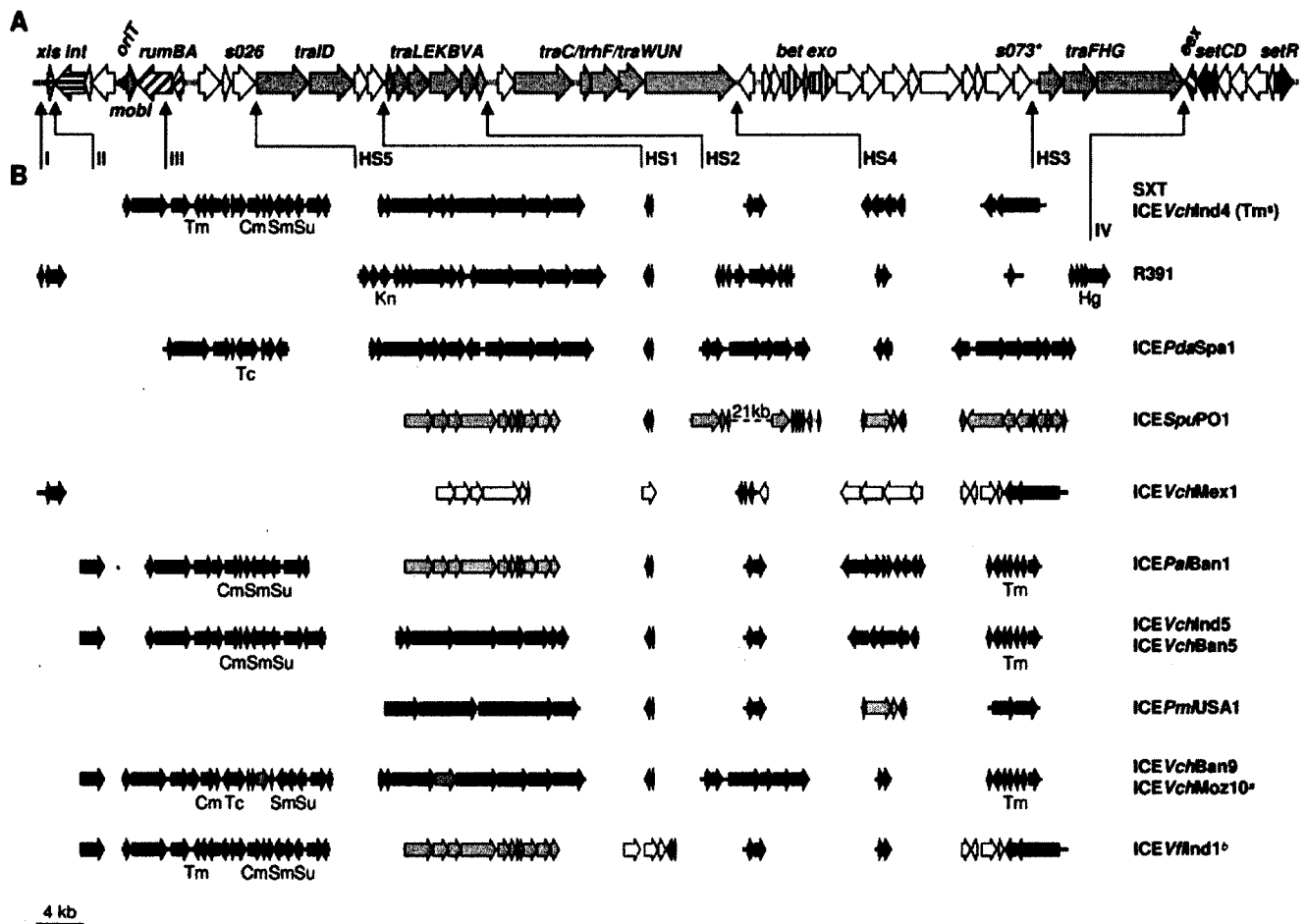


Figure 2. Structure of the genomes of 13 SXT/R391 ICEs. (A) The upper line represents the set of core genes (thick arrows) and sequences common to all 13 SXT/R391 genomes analyzed. Hatched ORFs indicate genes involved in site-specific excision and integration (*xis* and *int*), error-prone DNA repair (*rumAB*), DNA recombination (*bet* and *exo*) or entry exclusion (*ex*). Dark gray ORFs correspond to genes involved in regulation (*setCDR*). Light gray ORFs represent genes encoding the conjugative transfer machinery, and white ORFs represent genes of unknown function. (B) Variable ICE regions are shown with colors according to the elements in which they were originally described: SXT (blue), R391 (red), ICEPdaSpa1 (green), ICESpuPO1 (purple), ICEVchMex1 (yellow), ICEPaBan1 (orange), ICEVchInd5 (turquoise), ICEPmiUSA1 (olive), ICEVchBan9 (pink), ICEVflnd1 (light green). Thin arrows indicate the sites of insertion for each variable region and HS1–HS5 represent hotspots 1–5. Roman numerals indicate variable regions not considered true hotspots. Cm, chloramphenicol; Hg, mercury; Kn, kanamycin; Sm, streptomycin; Su, sulfamethoxazole; Tc, tetracycline; Tm, trimethoprim. * indicates that *s073* is absent from ICEPdaSpa1. ^a ICEVchMoz10, which lacks *dfrA1* in the integron structure, does not confer resistance to Tm. ^b The purple gene content of ICEVflnd1 was deduced from partial sequencing, PCR analysis and comparison with ICESpuPO1. doi:10.1371/journal.pgen.1000786.g002

The SXT/R391 ICE core genes

The 52 core genes present in all the SXT/R391 ICEs analyzed include sets of genes that are known to be required for the key ICE functions of integration/excision, conjugative transfer and regulation [32] as well as many genes of unknown function. Most genes of known or putative (based on homology) function (coded by gray shading or hatch marks in Figure 2A) are clustered with genes that have related functions. For example, *int* and *xis*, genes required for integration and excision, are adjacent and *setR*, and *setC/D*, the key SXT regulators are near each other at the extreme 3' end of the elements, although separated by 4 conserved genes of unknown function. Each ICE also has four gene clusters implicated in conjugative DNA processing and transfer (shown in light gray in Figure 2A). Finally, each of the ICEs has a nearly identical origin of transfer (*oriT*), a cis-acting DNA site that is thought to be nicked to initiate DNA processing events during conjugative transfer [39], in the same relative location.

The conserved core genes include approximately as many genes of unknown function as genes of known function. Some of the

genes of unknown function are found either interspersed amongst gene clusters that likely comprise functional modules (e.g. *s091* between *traD* and *s043*) while others are grouped together (e.g. most genes between *traN* and *traF*). In several cases, the interspersed genes appear to be part of operons with genes of known function (e.g. *s086-s082* maybe in an operon with *setDC*).

Variable ICE DNA

In addition to sharing 52 core genes, all of the ICE genomes analyzed contain variable DNA regions, ranging in size from 676 to 29,210 bp. Most of the variable DNA sequences are found in 5 intergenic hotspots (Figure 2B). However, some ICEs contain additional variable DNA inserts outside the 5 hotspots. For example, SXT and five other ICEs in Figure 2 have variable DNA segments, corresponding to related ISCR2 elements, disrupting *rumB* (Figure 2B, site III). ISCR2 elements are IS91-like transposable elements that tend to accumulate antibiotic resistance genes [40]. Interestingly, it is unusual for the contents of the hotspots and other variable regions to be found in only one ICE.

Instead, the variable gene content of most of the ICEs shown in Figure 2B is found in more than one ICE. For example, ICE*Spu*PO1, ICE*Pa*Ban1, and ICE*Vf*Ind1, all have identical contents in hotspot 5 (lavender genes in hotspot 5 in Figure 2B); however, the contents of the other hotspots in these 3 elements are almost entirely different. Thus, the variable gene content of the SXT/R391 ICEs reveals that these elements are mosaics. The overlapping distribution of variable DNA segments seen in the ICEs in Figure 2B suggests that recombination among this family of mobile elements may be extensive. In addition, in some instances, the variable regions appear subject to additional genetic modifications. For example, ICE*Pda*Spa1 and ICE*Vch*Ban9 contain ICE-specific DNA nested within the shared sequences inserted at hotspot 5 DNA (the green and pink genes in hotspot 5 in these elements, Figure 2B).

The variable genes encode a large array of functions and only a few will be discussed here. A complete list of the diverse genes found in the hotspots is found in Table S1. Although we cannot predict functions for many genes found in the hotspots, since they lack homology to genes of known function, at least a subset of the known genes seem likely to confer an adaptive advantage upon their hosts. Most of the ICE antibiotic resistance genes are found within transposon-like structures (e.g., the ISCR2 elements noted above) but four ICEs contain a *dhfr*A1 cassette, which confers resistance to trimethoprim [25], in a class IV integron located in hotspot 3. A disproportionate number of variable genes are likely involved in DNA modification, recombination or repair, as they are predicted to encode diverse putative restriction-modification systems, helicases and endonucleases. Such genes may provide the host with barriers to invasion by foreign DNA including phage infection and/or promote the integrity of the ICE genome during its transfer between hosts. Three ICEs contain genes that encode diguanylate cyclases [41] in hotspot 3. These enzymes catalyze the formation of cyclic-diguanosine monophosphate (c-di-GMP), a second messenger molecule that regulates biofilm formation, motility and virulence in several organisms including *V. cholerae* [42,43]. Most SXT/R391 ICEs contain *mosA* and *mosT* in hotspot 2. These two genes encode a novel toxin-antitoxin pair that promotes SXT maintenance by killing or severely inhibiting the growth of cells that have lost this element [44]. Not all ICEs in the SXT/R391 family contain *mosAT*; however, those lacking these genes may encode similar systems to prevent ICE loss. For instance, R391 and ICE*Vch*Mex1 contain two genes (*orf2* and *orf3*) encoding a predicted HipA-like toxin and a predicted transcriptional repressor distantly related to the antitoxin HipB.

Locations of the ICE variable genes

The variable regions found in the 5 hotspots are found exclusively in intergenic regions, punctuating the conserved ICE backbone (Figure 2). The boundaries between the conserved and variable sequences were mapped on the nucleotide level and compared (Figure 3A–3E). Each hotspot had a distinct boundary. Remarkably, even though the contents of the variable regions markedly differ, with few exceptions the left and right boundaries between conserved and variable DNA for each hotspot was identical among all the ICEs (Figure 3). For example, the left junctions of the inserts in hotspot 2 immediately follow the stop codon of *traA* and the right junctions are exactly 79 bp upstream of the start of *s05A* (Figure 3B), despite the fact that the DNA contents within these borders greatly differ. In hotspot 2, the right junction appears to begin with a 15 bp sequence that has two variants (Figure 3B, brown & light brown sequence). These sequences may reflect the presence of earlier insertions that have since been partially replaced. A similar pattern was found adjacent

to the left boundary of hotspot 4 in several ICEs (Figure 3D, lines 3–6). Once an insertion is acquired, the number of permissive sites for the addition of new variable DNA likely increases.

There are two exceptions to the precise boundaries between variable and conserved DNA. Hotspot 1 and hotspot 3 in ICE*Vch*Mex1 and ICE*Pda*Spa1, respectively, contain variable DNA that extends beyond the boundary exhibited by all the other ICEs in these locations (Figure 3A, line 3, and Figure 3C, line 7). The only boundary that could not be identified was the left border of hotspot 5, the region containing genes between *s026* and *traI*. As discussed below, *s026* is the least conserved core gene and its variability obscured any consensus sequence abutting the variable DNA. Perhaps this border has eroded because *s026* is not required for ICE mobility [32].

The relative precision of most boundaries between conserved and variable DNA sequences in all the ICEs analyzed suggests that a particular recombination mechanism, such as *bet/exo*-mediated recombination, may explain the acquisition of the variable regions. However, at this point, we cannot exclude the possibility that the precise locations for variable DNA insertions simply reflects selection for optimal ICE fitness; i.e., ICEs can optimally accommodate variable DNA in these locations while preserving their essential functions.

Similarity of SXT/R391 ICE and IncA/C plasmid core genes

Unexpectedly, BLAST analyses revealed that most of the conserved core SXT/R391 genes are also present in IncA/C conjugative plasmids. These multidrug resistance plasmids are widely distributed among *Salmonella* and other enterobacterial isolates from agricultural sources [45,46]. Recently, members of this family of plasmids have also been identified in *Yersinia pestis*, including from a patient with bubonic plague [47], and in aquatic γ -proteobacteria [48], including *Vibrio cholerae* [49,50]. To date, the closest known relatives of the SXT/R391 transfer proteins are found in the IncA/C plasmids. Every predicted SXT transfer protein is encoded by the IncA/C plasmid pIP1202 isolated from *Y. pestis* [50] and the identities of these predicted protein sequences vary from 34 to 78% (Figure 4A). Furthermore, there is perfect synteny between the four gene clusters encoding the respective conjugative machineries of these two mobile elements (yellow and orange genes in Figure 4A). Despite the extensive similarity of the SXT and IncA/C conjugative transfer systems, these plasmids lack homologues of *setR* and *setD/C* as well as *int/xis*, suggesting that regulation of conjugative transfer differs between these elements.

The similarity of IncA/C plasmids and SXT/R391 ICEs is not limited to genes important for conjugal DNA transfer. Ten genes of unknown function (shown in black in Figure 4A), some of which are interspersed within likely *tra* gene operons and some of which are clustered together between *traN* and *traF*, are similar in the two elements. Furthermore, most of these ten genes are in identical locations in the two elements. Both elements also contain homologs of *bet* and *exo* (shown in green in Figure 4A); these are the only known homologs of the λ Red recombination genes found outside of bacteriophages. Together, the similarity of DNA sequences and organization of SXT/R391 ICEs and IncA/C plasmids suggests that these elements have a common ancestor. The fact that the contents of the hotspots in the two classes of elements are entirely distinct suggests that their evolutionary paths diverged prior to acquisition of these variable DNA segments.

The minimal functional SXT/R391 ICE gene set

The conservation of the 52 core genes in all 13 SXT/R391 ICEs analyzed suggested that many or even all of these genes would be required for key ICE functions of excision/integration,

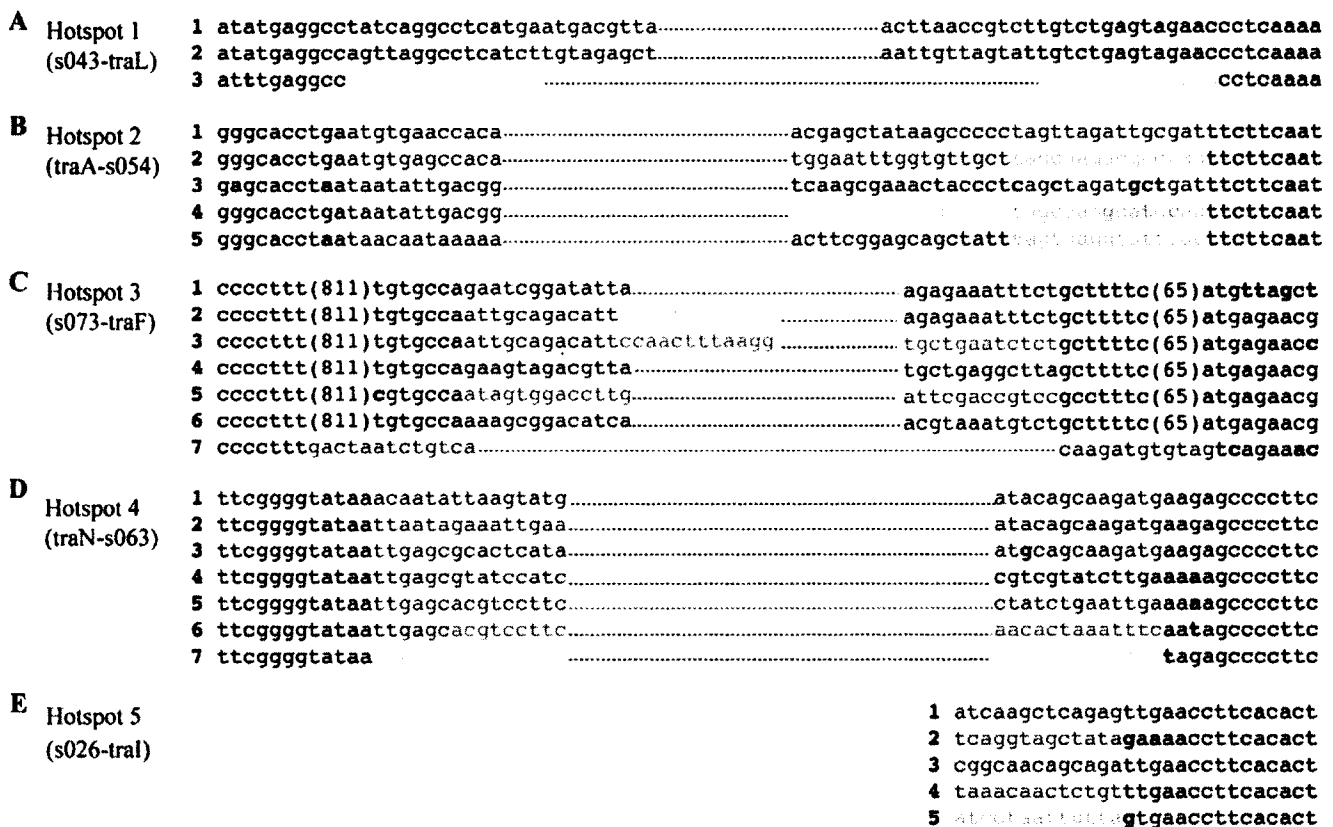


Figure 3. The boundaries of the 5 hotspots. The boundaries between conserved and hotspot variable regions are shown. Black typeface indicates conserved sequence, while color indicates variable sequence. Numbers in parentheses indicate the number of intervening nucleotides. The thin dotted lines indicate continuations of variable DNA. Bold letters indicate a non-conserved base. (A) Hotspot 1, which is present between *traJ* and *traL*. Line 1: SXT, ICEVchInd4, ICEPaBan1; Line 2: R391, ICEPdaSpa1, ICEVchBan5, ICEVchInd5, ICEPmiUSA, ICESpuPO1, ICEVflnd, ICEVchMoz10, ICEVchBan9; Line 3: ICEVchMex1. (B) Hotspot 2, which is present between *traA* and *s054*. Line 1: SXT, ICEVchInd4, ICEPmiUSA, ICEVflnd, ICEVchInd5, ICEPaBan1, ICEVchBan5; Line 2: ICEPdaSpa1, ICEVchMoz10, ICEVchBan9; Line 3: R391; Line 4: ICEVchMex1; Line 5: ICESpuPO1. (C) Hotspot 3, which is present between *s073* and *traF*. Line 1: SXT, ICEVchInd4; Line 2: ICEVchMex1, ICEVflnd; Line 3: ICEVchMoz10, ICEVchBan9, ICEVchInd5, ICEPaBan1, ICEVchBan5. Line 4: R391; Line 5: ICEPmiUSA; Line 6: ICESpuPO1; Line 7: ICEPdaSpa1. (D) Hotspot 4, which is present between *traN* and *s063*. Line 1: SXT, ICEVchInd4. Line 2: ICEVchInd5, ICEVchBan5; Line 3: ICESpuPO1, ICEPmiUSA; Line 4: R391, ICEVchMoz10, ICEVchBan9, ICEVflnd. Line 5: ICEPdaSpa1; Line 6: ICEPaBan1; Line 7: ICEVchMex1. (E) Hotspot 5, which is present between *s026* and *traI*. Line 1: SXT, ICEVchInd4, ICEPdaSpa1, R391, ICEVchMoz10, ICEVchBan9; Line 2: ICEPmiUSA; Line 3: ICESpuPO1, ICEPaBan1, ICEVflnd1; Line 4: ICEVchInd5, ICEVchBan5; Line 5: ICEVchMex1. doi:10.1371/journal.pgen.1000786.g003

conjugative transfer and regulation. The presence of ten ICE core genes of unknown function in IncA/C plasmids (black genes in Figure 4A) is also consistent with the hypothesis that these genes might be required for ICE transfer. However, our previous work demonstrated that not all genes recognized here as part of the conserved core gene set are required for SXT transfer. Beaver et al showed that deletion of *numB*–*s026* (which includes 5 core genes) from SXT had no detectable influence on SXT excision or transfer [32]. Therefore, we systematically deleted all of the core ICE genes whose contributions had not previously been assessed, in order to explore the hypothesis that these genes (especially those also present in IncA/C plasmids) would be essential for ICE transfer and to define the minimum functional SXT/R391 gene set.

Surprisingly, deletion of most of the ICE core genes of unknown function, including genes with homologues in IncA/C plasmids, did not alter SXT transfer efficiency. Deletion of *s002* or *s003*, which are located downstream of *int* in all SXT/R391 ICEs, did not alter the frequency of SXT transfer; similarly, deletion of *s082*, *s083*, and *s084*, core genes of unknown function that are found near the opposite end of SXT/R391 ICEs but not in IncA/C plasmids, also did not influence SXT transfer frequency

(Figure 4B). Furthermore, deletion of *s091*, which is found between *traD* and *s043* in ICEs and IncA/C plasmids, did not reduce SXT transfer (Figure 4B). In contrast, deletion of *s043*, which has weak homology to *traJ* in the F plasmid (a gene important in DNA processing) and is located in a transfer cluster containing *traI* and *traD*, abolished transfer (Figure 4B, Δd), suggesting that *s043*, here re-named *traJ* is required for SXT transfer. It is unlikely that the transfer defect of SXT $\Delta traJ$ can be explained by polar effects of the deletion on downstream genes, since *traJ* appears to be the last gene of an operon found immediately upstream of hotspot 1. Similarly, deletion of *s054*, which is found immediately 5' of *traC* and is homologous to a disulfide-bond isomerase *dsbC*, also abolished transfer (Figure 4B, Δe). Interestingly, disulfide bond-isomerases are present in several other conjugative systems [51]. However, it is not clear at this point if the deletion of *s054* from SXT accounts for the transfer defect of SXT $\Delta s054$, since we could not restore transfer by complementation.

Additionally, Beaver et al found that deletion of *s060* through *s073* in SXT, which includes 7 genes that are also found in IncA/C plasmids reduced SXT transfer more than 100-fold [32]. We constructed several smaller deletions in this region and found that

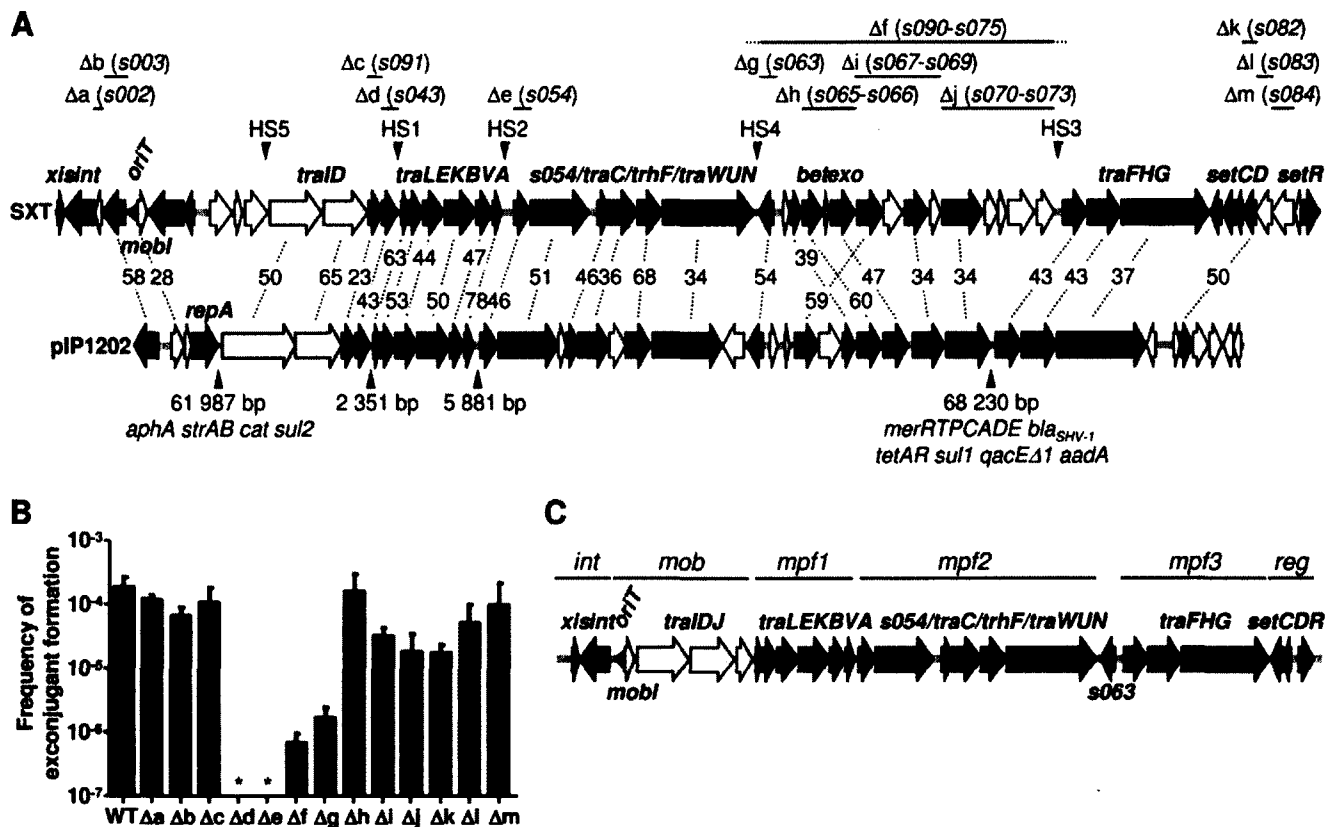


Figure 4. Comparison of the SXT/R391 core genome with the genome of pIP1202 and defining the minimal functional SXT/R391 gene set. (A) Alignment of the conserved core genes of SXT/R391 ICEs with the genome of the IncA/C conjugative plasmid pIP1202 from *Yersinia pestis*. The top line shows the same core ICE genes shown in Figure 2A. ORFs are color coded as follows: DNA processing, yellow; mating pair formation, orange; DNA recombination and repair, green; integration/excision, red; replication, purple; regulation, gray; entry exclusion, blue; homologous genes of unknown function, black; genes without corresponding counterparts in ICEs and pIP1202, white. Numbers shown in the middle represent % identity between the orthologous proteins encoded by SXT and pIP1202 [GenBank:NC_009141]. The positions of the hotspots in SXT/R391 ICEs are marked by downward pointing arrowheads. For pIP1202, the size of the sequences (which include IncA/C backbone DNA as well as variable DNA) found at these locations as well as resistance markers are indicated by upward pointing arrowheads. *aphA*, *aadA* and *strAB* confer resistance to aminoglycosides. *sul1* and *sul2* confer resistance to sulfonamides. *cat*, *bla_{SHV-1}*, *tetAR*, *qacED1* and *merRTPCADE* confer resistance to chloramphenicol, β -lactams, tetracyclines, quaternary ammonium compounds and mercury ions, respectively. Detailed descriptions of the conserved backbone of the IncA/C conjugative plasmids have been published elsewhere [48,50]. Regions that were deleted from SXT to investigate the function of genes of unknown function (see panel B) are indicated with straight lines. Dotted lines indicate that the deletion included DNA in the adjacent hotspots. (B) Influence of deletion of genes of unknown function on the frequency of SXT transfer. The mean values and standard deviations from three independent experiments are shown. * indicates that the frequency of transfer was below the detection level ($<10^{-8}$). Deletion mutants SXT Δ a, SXT Δ k and SXT Δ l, transferred at frequencies that were not significantly different from that of wild-type SXT (data not shown). (C) Proposed minimal set of genes necessary for a functional SXT/R391 ICE. *int*, integration/excision module; *mob*, DNA processing module; *mpf*, mating pair formation modules; *reg*, regulation module.

doi:10.1371/journal.pgen.1000786.g004

deletion of *s063*, which is also found in pIP1202, reduced the transfer frequency of SXT by ~ 100 -fold, nearly the same amount as deleting the entire region (Figure 4B). Complementation analyses revealed that the absence of *s063* accounted for the transfer defect of SXT Δ *s063* (data not shown). Even though SXT Δ *s063* was still capable of transfer, in our view, the drastic reduction in the transfer frequency of this mutant warrants inclusion of *s063* into the minimum functional SXT ICE genome (shown in Figure 4C). Other deletions in this region, including deletions of *bet*, *exo*, *s067*, *s068* and *s070*, which have orthologues in IncA/C plasmids, resulted in ≤ 10 -fold reductions in transfer frequency. We therefore did not include these genes in the minimal functional core SXT/R391 genome (Figure 4C).

The findings from our experiments testing the transfer frequencies of SXT derivatives harboring core gene deletions (shown in Figure 4B), coupled with our previous work demonstrating the requirements for the predicted SXT *tra* genes in the

clement's transfer [32], suggest a minimal functional SXT/R391 ICE structure as shown in Figure 4C. This minimum element is ~ 29.7 kb and consists of 25 genes. Genes with related functions, which in some cases encode proteins that likely form large functional complexes (such as the conjugation apparatus), are grouped together in the minimal genome. At the left end of the minimum ICE genomes are *xis* and *int*, the integration/excision module of SXT/R391 ICEs. In the minimal ICE genome, the ICE *oriT* and *mobI*, which encodes a protein required for SXT transfer [39], are no longer separated from the other genes (*traID*) that are also thought to play roles in the DNA processing events required for conjugative DNA transfer. The genes required for formation of the conjugation machinery, including the pilus, and mating pair formation and stabilization [32,39] are divided between three clusters (denoted mpf1-3 in Figure 4C). Finally, at the right end of the minimal functional genome are the genes that regulate ICE transfer (*setC/D* and *setR*). Thus, the minimal

functional SXT/R391 ICE is relatively small and organized into 3 discrete functional modules that mediate excision/integration, conjugation, and regulation.

Even though deletion of 27 out of 52 SXT/R391 ICE core genes proved to have little or no effect on SXT transfer frequency, and hence these genes were not included in Figure 4C, it is reasonable to presume that these genes encode functions that enhance ICE fitness given their conservation. For example, the presence of highly conserved *bet* and *exo* genes in all SXT/R391 ICEs suggests that there has been selection pressure to maintain this ICE-encoded recombination system that promotes ICE diversity by facilitating inter ICE recombination (G Garriss, MK Waldor, V Burrus, in press). A key challenge for future studies will be to determine how core genes of unknown function promote ICE fitness.

Variations in the similarity of core genes

To identify genes in the SXT/R391 core genome that may be subject to different selection pressures, we compared the percent identity of each ICE's core genes to the corresponding SXT gene (Figure 5). Most of the ICEs' core genes exhibited 94% to 98% identity on the nucleotide level to SXT's core genes. There was no discernable difference in the degree of conservation of most core genes that were or were not part of the minimal ICE, suggesting that there are equal selective pressures on essential and non-

essential genes. However, we identified 8 genes (*s026*, *traI*, *orfZ*, *s073*, *traF*, *eex*, *s086*, and *setR*) that exhibit significantly different degrees of conservation (Figure 5 and Figure S1). Three of these showed unusually high conservation, while the other 5 had below average conservation. Two of the highly conserved genes, *setR* and *s086*, are found at the extreme 3' end of the elements. The conservation of *setR* may reflect the key role of this gene in controlling SXT gene expression. *S086* may also play a role in regulating SXT transfer [52]. The other highly conserved gene, *orfZ*, is found between *bet* and *exo* and has no known function.

s026 and *s073* are the most divergent of all the genes in the backbone. *s026* encodes a hypothetical protein with homologues in many gram negative organisms. Although *S026* is predicted to contain a conserved domain, COG2378, which has a putative role in transcription regulation, this protein is not required for SXT transfer [32]. The significant divergence of *s026* along with its lack of essentiality suggests that this gene could become a pseudogene. A similar argument could be made for *s073*, which encodes a hypothetical protein that is also not required for ICE transfer. However, this argument does not hold for *traI* or *traF*, two genes which are essential for ICE transfer. Although the reasons which account for the different degrees of conservation of these 8 core genes are hard to ascertain at this point, the data in Figure 5 suggests that individual core genes are subject to different evolutionary pressures.

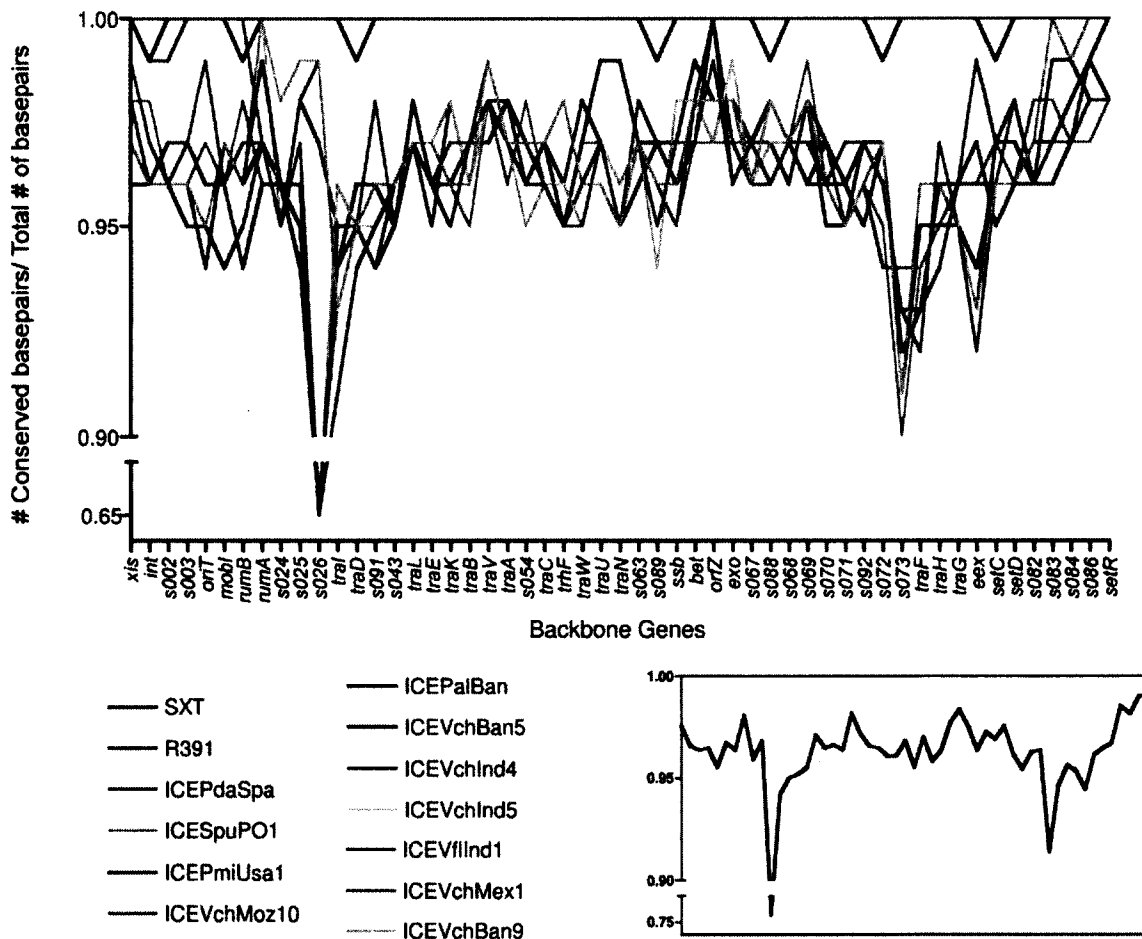


Figure 5. Variations in the nucleotide conservation of core ICE genes. The nucleotide sequence of each core gene from each ICE was compared to the corresponding sequence in SXT using pairwise BLASTn analyses to determine the percent identities. The average values for all of the ICEs, excluding SXT and ICEVchInd4, are shown in the inset. doi:10.1371/journal.pgen.1000786.g005

Comparisons of core gene phylogenies

We created phylogenetic trees for each core gene based on their respective nucleotide sequences to further explore the evolution of the conserved backbone of SXT/R391 ICEs. Since we found such a high degree of conservation for most of the core genes, the bootstrap values for most of these trees were relatively low. Thus, we concentrated on the most polymorphic genes found in Figure 5, *s026*, *s073*, *traI*, and *eex*, for phylogenetic analyses. As shown in Figure 6A, the trees for *s026*, *traI* and *s073* exhibit 3 distinct branching patterns. The lack of similarity in these phylogenetic trees suggests that either individual core genes have evolved independently or that high degrees of recombination mask their common evolutionary history. The latter hypothesis seems more likely since experimental findings have revealed that SXT/R391

ICEs can co-exist in a host chromosome in tandem [26] and recombination between tandem elements can yield novel hybrid ICEs with considerable frequency [53] (G Garriss, MK Waldor, V Burrus, in press). Also, as noted above, the distributions of variable genes among the ICEs shown in Figure 2 also supports the idea that inter-ICE recombination is commonplace.

Unlike most core genes, the trees for *traG* and *eex* were similar. In these two trees, the ICEs segregate into two evolutionarily distinct groups (Figure 6B), confirming and extending previous observations that revealed that there are two groups of *eex* and *traG* sequences in SXT/R391 ICEs [54]. These two groups correspond to the two functional SXT/R391 ICE exclusion groups. Interactions between *traG* and *eex* of the same group mediate ICE exclusion [55]. Thus, the identical 2 clusters of *traG*- and *eex*

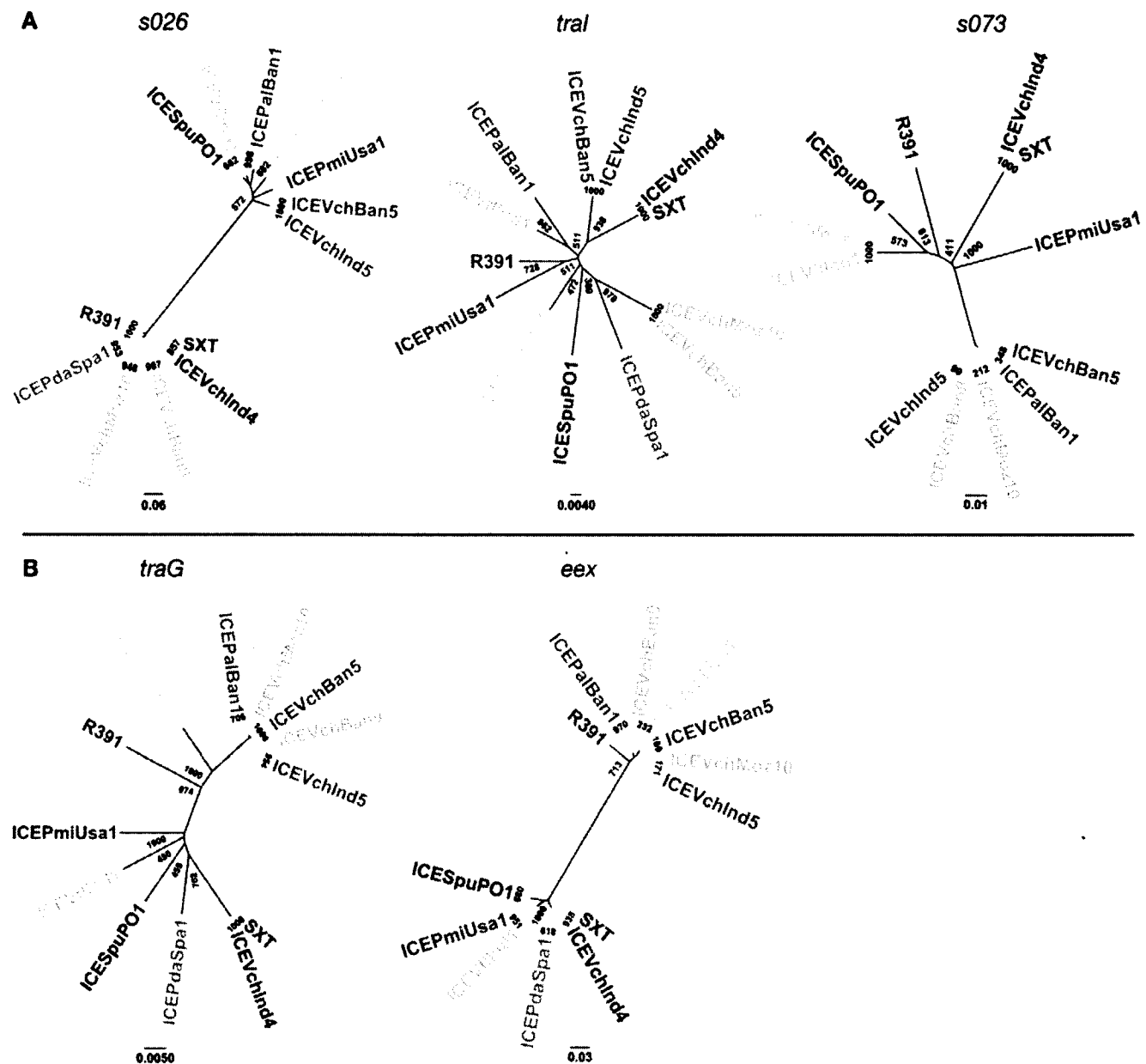


Figure 6. Phylogenetic analysis of several core ICE genes. Nucleotide sequences of the indicated core genes were used to generate the phylogenetic trees shown. Bootstrap values are indicated at branch points. The individual scale bars represent genetic distances and reflect the number of substitutions per residue. doi:10.1371/journal.pgen.1000786.g006

sequences observed in their respective trees reveals the co-evolution of the *traG/exx* functional unit. The two groups of *exx* sequences can also be observed in Figure 5 where the bifurcating pattern reveals the 2 exclusion groups. This pattern is difficult to discern for *traG*, perhaps because of the large size of this multi-functional gene.

ICEVchBan8, an SXT-like ICE that lacks Int_{SXT}

The sequence of ICEVchBan8, which was derived from a non-O1, non-O139 *V. cholerae* strain, is incomplete but it appears to contain 49 out of 52 SXT/R391 core genes. However, since this strain lacks Int_{SXT} it was not included in our comparative analyses above. It is not known if ICEVchBan8 is capable of excision or transmission; however, it contains a P4-like integrase and a putative *xis*. It is tempting to speculate that the genome of ICEVchBan8 provides an illustration of how acquisition (presumably via recombination) of a new integration/excision module may generate a novel ICE family.

Perspectives

Comparative analysis of the genomes of the 13 SXT/R391 ICEs studied here has greatly refined our understanding of this group of mobile genetic elements. These elements, which have been isolated from 4 continents and the depths of the Pacific Ocean, all have an identical genetic structure, consisting of the same syntenous set of 52 conserved core genes that are interrupted by clusters of diverse variable genes. All the elements have insertions of variable DNA segments in the same five intergenic hotspots that interrupt the conserved backbone. Furthermore, some of the elements have additional insertions outside the hotspots; however, in all cases the acquisition of variable DNA has not compromised the integrity of the core genes required for ICE mobility. Functional analyses revealed that less than half of the conserved genes are necessary for ICE transmissibility and the contributions of the 27 core genes of unknown function to ICE fitness remains an open question. Finally, several observations presented here suggest that recombination between SXT/R391 ICEs has been a major force in shaping the genomes of this widespread family of mobile elements.

Although comparisons of the 13 ICE genomes analyzed here strongly suggest that these mobile elements have undergone extensive recombination during their evolutionary histories, there is a remarkable degree of similarity among the SXT/R391 ICEs. All of these ICEs consist of the same syntenous and nearly identical 52 genes. In contrast, other families of closely related mobile elements, such as lambdaoid or T4-like phages for example, exhibit greater diversity [56,57]. Since the elements that we sequenced were isolated from several different host species and from diverse locations, the great degree of similarity of the SXT/R391 ICE family does not likely reflect bias in the elements that we sequenced. It is possible that this family of mobile elements is a relatively recent creation of evolution and has yet to undergo significant diversification.

To date, relatively few formal comparative genomic analyses of other ICE families have been reported. Mohd-Zain et al [11] identified several diverse ICEs and genomic islands that shared a largely syntenous set of core genes with ICEHin1056, an ICE originally identified in *Haemophilus influenzae*. However, even though these elements share a similar genomic organization, they exhibit far greater variability in the sites of insertion of variable DNA and in the degree of conservation in their core genes compared to SXT/R391 ICEs. Thus, although this group of elements appears to share a common ancestor, they seem to have diverged earlier in evolutionary history than the SXT/R391 ICEs.

However, when comparative genomic analyses were restricted to ICEHin1056-related ICEs found in only two *Haemophilus sp.*, Juhas et al found that, like the SXT/R391 family of ICEs, these 7 ICEHin1056-related ICEs share greater than 90% similarity at the DNA level in their nearly syntenous set of core genes [12]. It will be interesting to learn the extent of conservation of genetic structure and DNA sequence in additional ICE families to obtain a wider perspective on ICE evolution.

Comparative genomic studies of bacteriophages have led to the idea that the full range of phage sequences are part of common but extremely diverse gene pool [58,59]. The SXT/R391 ICE genomes suggest that there may be an even larger network of phylogenetic relationships linking sequences found in all types of mobile genetic elements including phages, plasmids, ICEs and transposons. The genomes of SXT/R391 ICEs appear to be amalgams of genes commonly associated with other types of mobile elements. Many of the ICE core genes are usually associated with phages, such as *int*, *bet*, *exo* and *setR*, or with plasmids, such as the *tra* genes. Additionally, the SXT/R391 ICEs and IncA/C plasmids clearly have a common ancestor, as we found that the entire set of SXT/R391 *tra* genes are also present in IncA/C plasmids. Thus, the genes present in all types of mobile genetic elements appear to contribute to a common gene pool from which novel variants of particular elements (such as ICEVchBan8) or perhaps even novel types of mobile genetic elements can arise.

Materials and Methods

ICE Sequencing

ICEPalBan1, ICEVchMex1, ICEVchInd4, ICEVchInd5 and ICEVchBan5 were isolated using the plasmid capture system described in Figure 1. The SXT chromosomal attachment sequence, *attB*, was introduced into the modified F plasmid pXX704 [34] to create pIceCap. This plasmid was then introduced into a Δ *prfC* derivative of the Tc^R *E. coli* strain CAG18439. Exconjugants derived from matings between this strain and those harboring the 5 ICEs listed above resulted in strains carrying a pIceCap::ICE plasmid. Once captured, the plasmids were isolated using the Qiagen plasmid midi kit for low-copy plasmids (Qiagen). Isolated pIceCap::ICE plasmids were then sequenced.

ICEVfInd genome was determined by sequencing several overlapping cosmids that encompassed this ICE's genome. Briefly, genomic DNA from a *Vibrio fluvialis* strain carrying ICEVfInd was prepared using the GNome DNA kit (QBIOgene). Sau3A1 restricted genomic DNA was used to create a SuperCos1 (Stratagene)-based cosmid library according to the manufacturer's instructions. The library was subsequently screened for cosmids containing ICE-specific sequences using PCR with primers to conserved core ICE sequences. Four cosmids containing overlapping ICEVfInd sequences were identified and sequenced.

The genomes of 6 ICEs were sequenced by the Sanger random shotgun method [60]. Briefly, small insert plasmid libraries (2–3 kb) were constructed by random nebulization and cloning of pIceCap::ICE DNA or of cosmid DNA for ICEVfInd. In the initial random sequencing phase, 8–12 fold sequence coverage was achieved. The sequences of either pIceCap or pSuperCos were subtracted and the remaining sequences were assembled using the Celera Assembler [61]. An initial set of open reading frames (ORFs) that likely encode proteins was identified using GLIMMER [62], and those shorter than 90 base pairs (bp) as well as some of those with overlaps eliminated.

Bioinformatics

Nucleotide and amino acid conservation were assessed with the appropriate BLAST algorithms. ICEs were aligned using clustalW with default settings [63]. MAUVE [37] and LAGAN [38] were used to identify core genes in Figure 2. To map the boundaries of the hotspots, sequence comparisons were made using MAUVE and then manually compared to find boundaries between conserved and variable DNA as shown in Figure 3.

Phylogenetic trees were generated from alignments of nucleotide sequences using the neighbor-joining method as implemented by ClustalX software, version 2.011 [64]. The reliability of each tree was subjected to a bootstrap test with 1000 replications. Trees were edited using FigTree 1.22 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Generation and testing of SXT deletion mutants

CAG81439 harboring SXT was used as the host strain to create the SXT deletion mutants shown in Figure 3; the deletions were constructed using one-step gene inactivation as previously described [44,65]. The primers used to create the deletion mutants are available upon request. Matings were conducted as previously described [16,44] using deletion mutants and a Kn^R *E. coli* recipient, CAG18420. Exconjugants were selected on LB agar plates containing chloramphenicol, 20 µg/ml (for SXT selection) and kanamycin, 50 µg/ml. The frequency of exconjugant formation was calculated by dividing the number of exconjugants by the number of donors.

References

- Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 3: 722–732.
- Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* 3: 679–687.
- Lawrence JG, Hendrickson H (2005) Genome evolution in bacteria: order beneath chaos. *Curr Opin Microbiol* 8: 572–578.
- 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.
- Beaber JW, Hochhut B, Waldor MK (2004) SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 427: 72–74.
- Burrus V, Waldor MK (2003) Control of SXT integration and excision. *J Bacteriol* 185: 5045–5054.
- Auchtung JM, Lee CA, Monson RE, Lehman AP, Grossman AD (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.
- Caparon MG, Scott JR (1989) Excision and insertion of the conjugative transposon Tn916 involves a novel recombination mechanism. *Cell* 59: 1027–1034.
- Cheng Q, Paszkiet BJ, Shoemaker NB, Gardner JF, Salyers AA (2000) Integration and excision of a *Bacteroides* conjugative transposon, CTnDOT. *J Bacteriol* 182: 4035–4043.
- Hagege J, Pernodet JL, Friedmann A, Guerin M (1993) Mode and origin of replication of pSAM2, a conjugative integrating element of *Streptomyces ambofaciens*. *Mol Microbiol* 10: 799–812.
- Mohd-Zain Z, Turner SL, Cerdeno-Tarraga AM, Lilley AK, Inzana TJ, et al. (2004) Transferable antibiotic resistance elements in *Haemophilus influenzae* share a common evolutionary origin with a diverse family of syntenic genomic islands. *J Bacteriol* 186: 8114–8122.
- Juhas M, Power PM, Harding RM, Ferguson DJ, Dimopoulou ID, et al. (2007) Sequence and functional analyses of *Haemophilus* spp. genomic islands. *Genome Biol* 8: R237.
- Burrus V, Waldor MK (2004) Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol* 155: 376–386.
- Boltner D, MacMahon C, Pembroke JT, Strike P, Osborn AM (2002) R391: a conjugative integrating mosaic comprised of phage, plasmid, and transposon elements. *J Bacteriol* 184: 5158–5169.
- Whittle G, Shoemaker NB, Salyers AA (2002) The role of *Bacteroides* conjugative transposons in the dissemination of antibiotic resistance genes. *Cell Mol Life Sci* 59: 2044–2054.

Supporting Information

Figure S1 Variations in the conservation of individual core ICE genes. The percent identity of the nucleotide sequence of each core gene and *oriT* versus the corresponding sequence in SXT was calculated for all ICEs studied. The average values for each gene (as shown in the inset of Figure 5) were then used in one-way ANOVA comparisons to determine genes that exhibit significantly more or less conservation compared to other core genes. p-values of one-way ANOVA comparisons of each core ICE gene are shown. The grid represents all pair-wise comparisons, and the color indicates the level of significance as follows: red: $p < .001$, orange: $p < .01$, and yellow: $p < .05$. Genes that exhibited a p -value $< .05$ when compared with at least 50% of all other core genes are discussed in the text.

Found at: doi:10.1371/journal.pgen.1000786.s001 (0.56 MB TIF)

Table S1 Contents of the hotspots.

Found at: doi:10.1371/journal.pgen.1000786.s002 (0.14 MB DOC)

Acknowledgments

We thank Brigid Davis and Frédérique Le Roux for helpful comments on the manuscript. We thank Robert Hall for facilitating the work described here and Yoshiharu Yamaichi for suggesting the design of *pIceCap*.

Author Contributions

Conceived and designed the experiments: RAFW VB MKW. Performed the experiments: RAFW DEF MS DC GG CD VB. Analyzed the data: RAFW MS DC VB MKW. Contributed reagents/materials/analysis tools: MMC. Wrote the paper: RAFW VB MKW.

- 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.
- Shoemaker NB, Barber RD, Salyers AA (1989) Cloning and characterization of a *Bacteroides* conjugative tetracycline-erythromycin resistance element by using a shuttle cosmid vector. *J Bacteriol* 171: 1294–1302.
- Franke AE, Clewell DB (1981) Evidence for a chromosome-borne resistance transposon (Tn916) in *Streptococcus faecalis* that is capable of “conjugal” transfer in the absence of a conjugative plasmid. *J Bacteriol* 145: 494–502.
- Ravatt R, Studer S, Springael D, Zehnder AJ, van der Meer JR (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.
- Sullivan JT, Ronson CW (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.
- Cholera Working Group (1993) Large epidemic of cholera-like disease in Bangladesh caused by *Vibrio cholerae* O139 synonym Bengal. *Lancet* 342: 387–390.
- 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.
- Osorio CR, Marrero J, Wozniak RA, Lemos ML, Burrus V, et al. (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.
- 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.
- Hochhut B, Lotfi Y, Mazel D, Faruque SM, Woodgate R, et al. (2001) Molecular analysis of antibiotic resistance gene clusters in *vibrio cholerae* O139 and O1 SXT constans. *Antimicrob Agents Chemother* 45: 2991–3000.
- 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.
- Coetzee JN, Datta N, Hedges RW (1972) R factors from *Proteus rettgeri*. *J Gen Microbiol* 72: 543–552.
- Burrus V, Marrero J, Waldor MK (2006) The current ICE age: biology and evolution of SXT-related integrating conjugative elements. *Plasmid* 55: 173–183.

29. Hochhut B, Waldor MK (1999) Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. *Mol Microbiol* 32: 99–110.
30. Maeda K, Nojiri H, Shintani M, Yoshida T, Habe H, et al. (2003) Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain CA10 indicates its mosaicity and the presence of large catabolic transposon Tn4676. *J Mol Biol* 326: 21–33.
31. Murata T, Ohnishi M, Ara T, Kaneko J, Han CG, et al. (2002) Complete nucleotide sequence of plasmid Rts1: implications for evolution of large plasmid genomes. *J Bacteriol* 184: 3194–3202.
32. 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.
33. 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.
34. Niki H, Hiraga S (1997) Subcellular distribution of actively partitioning F plasmid during the cell division cycle in *E. coli*. *Cell* 90: 951–957.
35. Yamaichi Y, Niki H (2004) *migS*, a *cis*-acting site that affects bipolar positioning of *oriC* on the *Escherichia coli* chromosome. *Embo J* 23: 221–233.
36. Faruque SM, Tam VC, Chowdhury N, Diraphat P, Dziejman M, et al. (2007) Genomic analysis of the Mozambique strain of *Vibrio cholerae* O1 reveals the origin of El Tor strains carrying classical CTX prophage. *Proc Natl Acad Sci U S A* 104: 5151–5156.
37. Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14: 1394–1403.
38. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, et al. (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 13: 721–731.
39. 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.
40. Toleman MA, Bennett PM, Walsh TR (2006) ISCR elements: novel gene-capturing systems of the 21st century? *Microbiol Mol Biol Rev* 70: 296–316.
41. Bordeleau E, Brouillette E, Robichaud N, Burrus V (2009) 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*.
42. Jenal U, Malone J (2006) Mechanisms of cyclic-di-GMP signalling in bacteria. *Annu Rev Genet* 40: 385–407.
43. Romling U, Gomelsky M, Galperin MY (2005) C-di-GMP: the dawning of a novel bacterial signalling system. *Mol Microbiol* 57: 629–639.
44. Wozniak RA, Waldor MK (2009) A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genet* 5: e1000439. doi:10.1371/journal.pgen.1000439.
45. Lindsey RL, Fedorka-Cray PJ, Frye JG, Meinersmann RJ (2009) Inc A/C plasmids are prevalent in multidrug-resistant *Salmonella enterica* isolates. *Appl Environ Microbiol* 75: 1908–1915.
46. Llanes C, Gabant P, Couturier M, Bayer L, Plesiat P (1996) Molecular analysis of the replication elements of the broad-host-range RepA/C replicon. *Plasmid* 36: 26–35.
47. Galimand M, Guiyoule A, Gerbaud G, Rasoamanana B, Chanteau S, et al. (1997) Multidrug resistance in *Yersinia pestis* mediated by a transferable plasmid. *N Engl J Med* 337: 677–680.
48. Welch TJ, Fricke WF, McDermott PF, White DG, Rosso ML, et al. (2007) Multiple antimicrobial resistance in plague: an emerging public health risk. *PLoS One* 2: e309. doi:10.1371/journal.pone.0000309.
49. Pan JC, Ye R, Wang HQ, Xiang HQ, Zhang W, et al. (2008) *Vibrio cholerae* O139 multiple-drug resistance mediated by *Yersinia pestis* pIP1202-like conjugative plasmids. *Antimicrob Agents Chemother* 52: 3829–3836.
50. Fricke WF, Welch TJ, McDermott PF, Mammel MK, LeClerc JE, et al. (2009) Comparative genomics of the IncA/C multidrug resistance plasmid family. *J Bacteriol* 191: 4750–4757.
51. Elton TC, Holland SJ, Frost LS, Hazes B (2005) F-like type IV secretion systems encode proteins with thioredoxin folds that are putative DsbC homologues. *J Bacteriol* 187: 8267–8277.
52. Beaber JW, Waldor MK (2004) Identification of operators and promoters that control SXT conjugative transfer. *J Bacteriol* 186: 5945–5949.
53. Burrus V, Waldor MK (2004) Formation of SXT tandem arrays and SXT-R391 hybrids. *J Bacteriol* 186: 2636–2645.
54. Marrero J, Waldor MK (2007) The SXT/R391 family of integrative conjugative elements is composed of two exclusion groups. *J Bacteriol* 189: 3302–3305.
55. Marrero J, Waldor MK (2005) Interactions between inner membrane proteins in donor and recipient cells limit conjugal DNA transfer. *Dev Cell* 8: 963–970.
56. Juhala RJ, Ford ME, Duda RL, Youton A, Hatfull GF, et al. (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdaoid bacteriophages. *J Mol Biol* 299: 27–51.
57. Filee J, Bapteste E, Susko E, Krisch HM (2006) A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. *Mol Biol Evol* 23: 1688–1696.
58. Hendrix RW, Hatfull GF, Smith MC (2003) Bacteriophages with tails: chasing their origins and evolution. *Res Microbiol* 154: 253–257.
59. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A* 96: 2192–2197.
60. Fouts DE, Mongodin EF, Mandrell RE, Miller WG, Rasko DA, et al. (2005) Major structural differences and novel potential virulence mechanisms from the genomes of multiple campylobacter species. *PLoS Biol* 3: e15. doi:10.1371/journal.pbio.0030015.
61. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, et al. (2000) A whole-genome assembly of *Drosophila*. *Science* 287: 2196–2204.
62. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27: 4636–4641.
63. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
64. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
65. 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.
66. Burrus V, Quezada-Calvillo R, Marrero J, Waldor MK (2006) SXT-related integrating conjugative element in New World *Vibrio cholerae*. *Appl Environ Microbiol* 72: 3054–3057.
67. Das B, Halder K, Pal P, Bhadra RK (2007) Small chromosomal integration site of classical CTX prophage in Mozambique *Vibrio cholerae* O1 biotype El Tor strain. *Arch Microbiol* 188: 677–683.
68. Pearson MM, Sebahia M, Churcher C, Quail MA, Seshasayee AS, et al. (2008) Complete genome sequence of uropathogenic *Proteus mirabilis*, a master of both adherence and motility. *J Bacteriol* 190: 4027–4037.
69. Nair GB, Faruque SM, Bhuiyan NA, Kamruzzaman M, Siddique AK, et al. (2002) New variants of *Vibrio cholerae* O1 biotype El Tor with attributes of the classical biotype from hospitalized patients with acute diarrhea in Bangladesh. *J Clin Microbiol* 40: 3296–3299.