# Consequences of a Mobile Genetic Element Integrated at Secondary Locations



By

Kayla L. Menard

B.S., Microbiology
University of Wisconsin-Madison, 2008

Submitted to the Microbiology Graduate Program in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the Massachusetts Institute of Technology
June 2013



Signature of Author: _____
Kayla L. Menard
Microbiology Graduate Program
May 23, 2013

Certified by: _____
Alan D. Grossman
Professor of Biology
Thesis Supervisor

Accepted by: _____
Michael Laub
Professor of Biology
Director, Microbiology Graduate Program

# Consequences of a Mobile Genetic Element Integrated at Secondary Locations

By

Kayla L. Menard

Submitted to the Microbiology Graduate Program on May 23, 2013 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

## ABSTRACT

Integrative and conjugative elements (ICEs) are widespread mobile genetic elements that are integrated in bacterial chromosomes, but can excise and transfer to a recipient through conjugation. ICEs are important agents of evolution, contributing to the acquisition of new traits, including antibiotic resistance. Many ICEs are site-specific in that they integrate preferentially into a primary attachment site in the bacterial chromosome. Site-specific ICEs can integrate into secondary locations, but little is known about the consequences of integration. Using ICE*Bs1*, a site-specific ICE from *Bacillus subtilis*, I found that integration into secondary attachment sites is detrimental to both ICE*Bs1* and the host cell.

**Integration at secondary locations is detrimental to ICE*Bs1***. Once integrated in the chromosome, excision of ICE*Bs1* from all secondary attachment sites analyzed was either reduced (4 sites) or undetectable (3 sites) compared to ICE*Bs1* excision from the primary site. Additionally, from two of the four secondary sites that exhibited reduced but detectable excision, the excised, circular form of ICE*Bs1* was present at lower levels than expected, indicating that circular ICE*Bs1* may be unstable. Defects in excision and stability of ICE*Bs1* severely limit its ability to spread to other cells.

**Integration at secondary locations is detrimental to the host cell**. Induction of ICE*Bs1* gene expression in secondary integration sites resulted in a defect in cell proliferation and/or viability, as well as induction of the SOS response. These effects are likely due to DNA damage resulting from plasmid-like, rolling-circle replication of the excision-defective ICE*Bs1* in the chromosome. Consistent with this model, deletion of ICE*Bs1* replication genes (*nicK* and *helP*) alleviated the proliferation and viability defects.

**Implications for the evolution of ICEs**. These previously unrecognized detrimental effects may provide selective pressure against propagation of ICE*Bs1* in secondary attachment sites. Such detrimental effects could explain the maintenance and prevalence of site-specific integration among ICEs.

Thesis Supervisor: Alan D. Grossman
Title: Professor of Biology

**Acknowledgements**

I acknowledge my advisor Alan for always taking the time to teach me how to do science. I learned a lot from him and I know I will take that with me with whatever I do in the future. I also want to thank my thesis committee members: Graham Walker, Michael Laub, and Michael Gilmore, for being helpful and available throughout my graduate studies.

During my time in the Grossman lab, I was fortunate to have great coworkers who made the lab a very positive environment. This includes C. Lee, Ana, Chris, Janet, Carla, Jacob, Tracy, Charlotte, Tyler, Laurel, Melanie, and Houra. C. Lee cannot be thanked enough for everything does she does for the lab, including for answering my numerous questions. She was always a positive, cheerful, and helpful presence in the lab. Ana was a great mentor to me beginning with my rotation in the lab; my positive experience rotating with her contributed to my joining the lab. Chris has been a great baymate; when he first joined the lab, I was pleasantly surprised to realize it was nice to have company in my bay (where I'd mostly been the only occupant). He was always helpful and friendly. Carla and Jacob were not only helpful labmates, but I also had a lot of fun hanging out with them outside of work. Tracy, my "work spouse", and I went through graduate school together, including numerous sessions where we studied for qualifying exams. He is a great friend and it is difficult to imagine that we will not be seeing each other every day going forward.

I also want to acknowledge my friends and family for listening, being supportive, and with whom I shared a lot of good times during graduate school. I really appreciate all of the people that made the effort to travel to Boston to visit me, as well as the fun times with the people I became friends with here. In particular, I will mention my boyfriend Tony who I met while at MIT and who has been a solid, positive presence in my life ever since.

**Table of Contents**

**List of Tables**

**List of Figures**

# Chapter 1

# Introduction

Outline

A. Horizontal gene transfer is mediated by integrative and conjugative elements (ICEs)

-Mechanisms of horizontal gene transfer

-Introduction to integrative and conjugative elements (ICEs)

-ICEs mediate horizontal gene transfer

B. ICE*Bs1* is an integrative and conjugative element in *Bacillus subtilis*

-Overview of ICE*Bs1*

-Regulation of ICE*Bs1*

-Replication of ICE*Bs1*

-ICE*Bs1* as a model ICE

B. Mechanisms of ICE integration

-Integration via DDE transposases

-Integration via serine recombinases

-Integration via tyrosine recombinases

C. Locations of ICE Integration

-Patterns of integration

-tRNA genes as hotspots for integration

-Secondary integration sites of site-specific ICEs

D. Effects of ICE integration

-Genes that confer a phenotype to the host

-Mutations to the host genome

-Effects on transcription of host genes

- Proliferation/viability defect of strains with integrations into secondary attachment sites

Thesis Summary

## A. Horizontal gene transfer is mediated by integrative and conjugative elements (ICEs)

Horizontal gene transfer (HGT) is a very important process in bacteria. This section discusses the mechanisms of horizontal gene transfer in bacteria, including the types of mobile genetic elements that carry out these different mechanisms. In particular, this work focuses on a type of mobile genetic element, called an integrative and conjugative element (or ICE), that is an important agent of horizontal gene transfer in bacteria.

**Mechanisms of horizontal gene transfer.** HGT is the physical transfer of DNA from one pre-existing organism to another without vertical transmission (spread of DNA from parent to offspring through reproduction) (Frost et al., 2005; Ochman et al., 2000; Redfield, 2001). HGT drives bacterial evolution and occurs through three general mechanisms: transformation, transduction, and conjugation. Transformation is the uptake, by a bacterium, of naked DNA from the environment. Transduction occurs when a phage transfers bacterial DNA to a new host. Conjugation (or mating) is when two bacteria transfer DNA through cell-cell contact.

Two of these mechanisms of horizontal gene transfer, conjugation and transduction, are carried out by mobile genetic elements (Frost et al., 2005). Mobile genetic elements are segments of DNA capable of moving either between or within organisms. Plasmids, the type of mobile genetic element often associated with conjugation, exist separate from the bacterial chromosome and are often small, circular DNA elements. Conjugative plasmids encode the genes necessary for conjugation, allowing them to spread horizontally.

Phages, another type of mobile genetic element, are responsible for transduction (Frost et al., 2005; Ochman et al., 2000). Phages, or bacterial viruses, mediate HGT in two ways: generalized transduction and specialized transduction. In generalized transduction, instead of incorporating its own DNA, the phage packages bacterial DNA that gets transferred to the new

host. In specialized transduction, lysogenic phages—which are quiescently inserted into the chromosome—excise improperly and package chromosomal genes next to the site of insertion. These additional genes then get carried to the new host along with the phage genome.

This thesis focuses on a particular type of mobile element, called an integrative and conjugative element (or ICE), that includes features of both plasmids and lysogenic phages. Despite being less well known than plasmids and phages, ICEs are widespread and important agents of horizontal gene transfer in bacteria (Burrus et al., 2002a).

**Introduction to integrative and conjugative elements (ICEs).** ICEs reside integrated into the host chromosome, similar to a lysogenic phage (or transposon) (Burrus and Waldor, 2004; Seth-Smith and Croucher, 2009; Wozniak and Waldor, 2010). Under certain conditions, they can be induced to excise from the chromosome and form a circular intermediate, similar to a plasmid (Figure 1). This circular intermediate uses self-encoded conjugation genes to mate from one cell to another. Because only a single, linear strand of ICE appears to be transferred to the recipient, this single strand is circularized and a second strand synthesized before integration into the recipient genome. A copy of ICE still remains in the donor cell and re-integrates into the chromosome of the donor strain. In this way, ICE is stably maintained in both the donor and the recipient cell after conjugation.

Even though ICEs are less well known than plasmids or phages, ICEs are extremely common elements in bacteria (Guglielmini et al., 2011; Seth-Smith and Croucher, 2009). Present in approximately 20% of sequenced bacteria, they appear to be more common than conjugative plasmids (Guglielmini et al., 2011). Ranging from Gram-positive to Gram-negative bacteria, they are present across the major divisions of bacteria (Burrus et al., 2002a; Burrus and Waldor,

2004; Guglielmini et al., 2011). Additionally, there is even at least one example of an ICE in Archaea, in the species *Aciduliprofundum boonei* (Guglielmini et al., 2011).

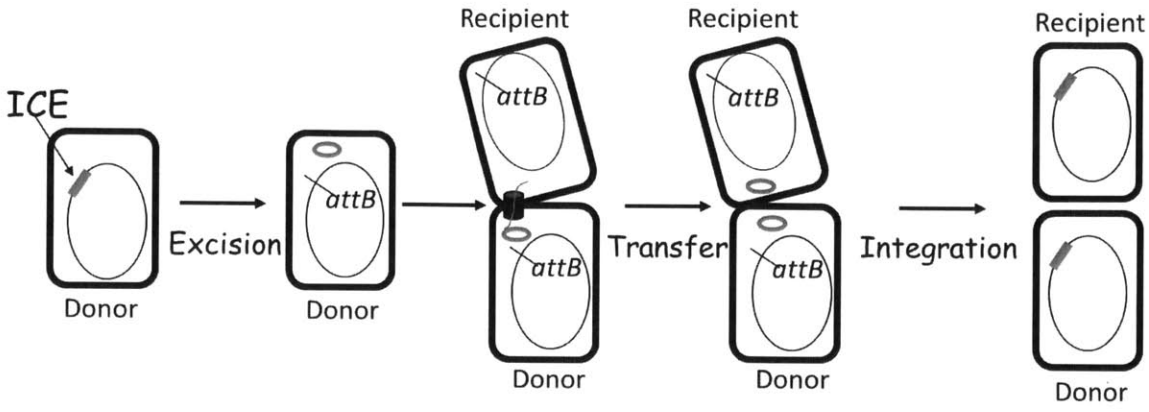



**Figure 1: Life cycle of integrative and conjugative elements (ICEs).** ICEs (dark grey) reside integrated into the chromosome (black oval) of a bacterial cell (black rectangle). Many ICEs integrate into one primary chromosomal attachment site referred to as *attB*. Under certain conditions, ICEs can be induced to excise from the chromosome and form a circular intermediate similar to a plasmid. This circular, plasmid-like intermediate is nicked and a single-strand transferred to a recipient bacterium via the element's self-encoded conjugation machinery (small, black cylinder). In the recipient, the single-strand of ICE is circularized and a second strand is synthesized before integration into the chromosome of the recipient. ICE also reintegrates into the chromosome of the donor since a copy of the element yet remains.

**ICEs mediate horizontal gene transfer.** Not only are ICEs prevalent in bacteria, they are also important agents of evolution due to their ability to mediate horizontal gene transfer. I will discuss two ways that ICEs mediate evolution; additional ways will be discussed further in Section D (Effects of ICE Integration). In this section, I will also discuss how the ability of ICEs to spread via HGT makes them useful tools for studying genetically intractable organisms.

One way ICEs mediate evolution is through their ability to carry genes that convey an advantage to the host in which they reside. This includes genes for processes such as antibiotic resistance, virulence, metabolism, or symbiosis (Seth-Smith and Croucher, 2009; Wozniak and Waldor, 2010). Historically, ICEs have been predominately studied for their role in the spread of

antibiotic resistance genes. For example, the SXT element of the pathogen Vibrio cholera carries resistance to sulfamethoxazole, trimethoprim, chloramphenicol, and streptomycin (Burrus and Waldor, 2003). Study of horizontal gene transfer and ICEs may eventually lead to ways to inhibit either the spread of antibiotic resistance genes or organisms with antibiotic resistance genes (Barlow, 2009).

Another way in which ICEs are important for the evolution of bacteria is in their ability to mobilize other non-conjugative mobile elements (Salyers et al., 1995; Wozniak and Waldor, 2010). By providing the conjugation genes necessary for mating, they can assist the horizontal dissemination of elements lacking dedicated conjugation functions. Mobilization possibly accounts for the prevalence of mobile elements (such as non-conjugative plasmids or genome islands) without conjugation functions (Guglielmini et al., 2011).

Besides their significant role in the evolution of bacteria, ICEs are also useful as tools to study genetically intractable organisms. Since ICEs can conjugate from one organism to another, this provides an attractive means to transfer DNA into the genome of organisms where direct transformation of DNA is difficult. For example, the ICE Tn*916*, has been used to perform mutagenesis studies in several bacteria, such as *Erysipelothrix rhusiopathiae* and *Desulfitobacterium dehalogens* (Shimoji et al., 1998; Smidt et al., 1999). In addition to mutagenesis studies, ICEs may be useful for adding or deleting particular genes from the chromosome (Lee et al., 2007).

The use of ICEs to study genetically intractable organisms will be aided by a thorough understanding of ICEs and how they interact with the cell. This thesis focuses on two ways that ICEs interact with the host cell: 1) specificity/location of integration and 2) effect of integration on the host cell. Relatively little is known about the interaction between ICEs and their host

(Juhas et al., 2009). In particular, I will place emphasis on interactions between a specific ICE and its host, ICE*Bs1* in *Bacillus subtilis*.

## B. ICE*Bs1* is an integrative and conjugative element in *Bacillus subtilis*

**Overview of ICE*Bs1*.** ICE*Bs1* is an approximately 20 kb (Fig. 2) integrative and conjugative element with ~24 putative open reading frames (Auchtung et al., 2005; Berkmen et al., 2011; Berkmen et al., 2010; Burrus et al., 2002b; Lee et al., 2010). Many of its genes are similar to genes in other ICEs, including similarity to ~10 genes in Tn*916*, the first conjugative transposon identified (Franke and Clewell, 1981a). ICE*Bs1* genes necessary for nicking (*nicK*), mating (*conE*, *conG*, *cwlT*), replication (*ydcS*), regulation (*immR*), and integration (*int*, but only distantly) all share homology to Tn*916*.

As is consistent with all known integrative and conjugative elements, ICE*Bs1* has a modular structure, wherein genes encoding for particular functions, such as mating, DNA processing, or regulation, are grouped together (Figure 2) (Berkmen et al., 2011). Many of the genes necessary for mating and replication are similar to plasmid genes (such as *nicK*, *ydcP*, *conE*, and *conG*). Likewise, many of the genes for excision, integration, and regulation share homology to phage genes (such as *int*, *immR*, and *immA*).

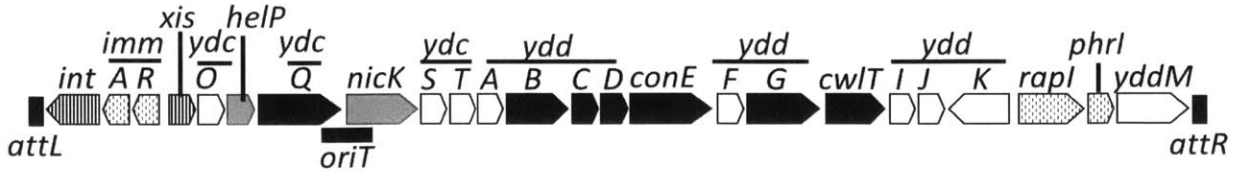



**Figure 2: Schematic of ICE*Bs1* integrated into the chromosome**. The 24 putative open reading frames of ICE*Bs1* are color coded according to their role in ICE*Bs1* processes, such as: mating (black), integration and excision (vertical black lines), regulation (dotted vertical lines), replication (grey), and unknown (white). The origin of transfer (*oriT*) and left and right attachment sites (*attL* and *attR*) are indicated. The *attL* and *attR* sites are recognized by the recombinase (Int) to mediate excision of the element.

Unlike most studied ICEs, ICE*Bs1* is not known to contain any accessory genes that provide an obvious benefit to the host. While most ICEs were identified for the benefit they confer on their host, ICE*Bs1* was experimentally identified for its ability to be regulated by cell-cell signaling (Figure 3) (Auchtung et al., 2005). Whether or not ICEs need to confer a benefit in order to be successfully maintained in the host is an open question in the field. Preliminary evidence from ICE*Bs1* suggests that it protects cells from infection by the *B.subtilis* phage SPβ (Johnson and Grossman, personal communication), possibly providing a reason for its presence in the *B.subtilis* genome.

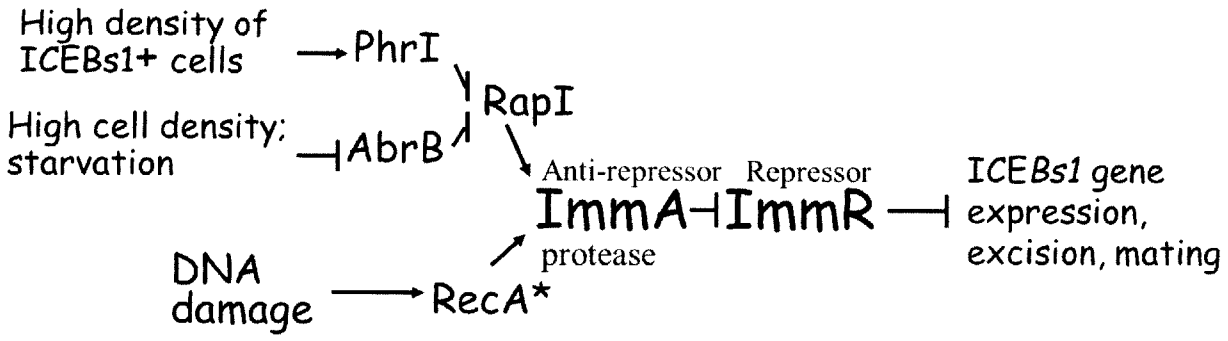



**Figure 3: Regulation of ICE*Bs1* gene expression, excision, replication, and mating.** The expression of most ICE*Bs1* genes is under the control of one promoter P*xis* (Auchtung et al., 2007; Auchtung et al., 2005). Under normal cellular conditions, P*xis* is repressed by the ICE*Bs1* encoded protein ImmR. During DNA damage or when there is a high density of *Bacillus subtilis* cells without ICE*Bs1*, ImmR is cleaved by the ICE*Bs1* encoded protease ImmA leading to induction of ICE*Bs1* gene expression, excision, nicking, replication, and initiation of mating. While the mechanism is unknown, ImmA is activiated to cleave ImmR by either RecA or RapI (Bose et al., 2008). RecA is induced to activate ImmA under conditions of DNA damage. RapI is induced to activate ImmA only when lows levels of PhrI and active AbrB are present, as these both inhibit the activity of RapI. PhrI is an extracellular protein produced by ICE*Bs1* that prevents induction of P*xis* when high numbers of other cells containing ICE*Bs1* are present. Under conditions of starvation or high cell density, AbrB no longer represses RapI.

**Regulation of ICE*Bs1*.** Most ICE*Bs1* genes appear to be under the control of one promoter P*xis* (Auchtung et al., 2007 2011; Auchtung et al., 2005; Bose et al., 2008; Lee et al., 2010). Induction from this promoter leads to ICE*Bs1* excision, nicking, rolling-circle replication, and

initiation of mating. P*xis* is controlled by the ICE*Bs1* repressor protein ImmR; after induction, the repressor ImmR is degraded by the protease ImmA, resulting in activation of ICE*Bs1* gene expression (Bose et al., 2008) (Figure 3). While the mechanism is still not clear, the proteolytic reaction is activated by RecA (after DNA damage) or by the ICE*Bs1*-encoded protein RapI (when the cell is surrounded by other *B.subtilis* cells lacking ICE*Bs1*). When ICE*Bs1* gene expression is artificially induced by overexpression of RapI, ICE*Bs1* excises from the chromosome in >90% (typically 95-99%) of cells, and leads to ICE*Bs1* nicking, replication, and initiation of mating. This high level of experimental induction is unique among studied ICEs and makes ICE*Bs1* very useful for studying basic properties of ICEs.

**Replication of ICE*Bs1*.** As mentioned previously, after activation of ICE*Bs1* gene expression from P*xis*, ICE*Bs1* excises from the chromosome and replicates autonomously (Lee et al., 2010; Thomas et al., 2013). This autonomous replication occurs via a plasmid-like, rolling-circle mechanism of replication (Figure 4). Replication begins with nicking of ICE*Bs1* within its origin of transfer (*oriT*) by the ICE*Bs1* protein NicK. Replication then proceeds unidirectionally from the 3' end created by the nick. This process creates a double-stranded circle and a single-stranded, linear copy of ICE*Bs1*. The linear single-strand is circularized and a second strand synthesized. This replication process is known to require two ICE*Bs1* genes: *helP* (helicase processivity factor) and *nicK* (nickase). Replication also requires several host genes: *polC* (DNA polymerase), *pcrA* (helicase), *ssb* (single-strand DNA binding protein), *dnaN* (β-clamp), and likely others. While it is known that ICE*Bs1* and some other ICEs replicate autonomously (Carraro et al., 2011; Ramsay et al., 2006 2010, waldor; Sitkiewicz et al., 2011) it is currently unclear how widespread this phenomenon is among ICEs.

**ICE*Bs1* as a model ICE.** Because of the conservation of many of its functions, the ease of

manipulating *B. subtilis*, and the high efficiency of experimental induction of gene expression, ICE*Bs1* is extremely useful for studying basic properties of ICEs.

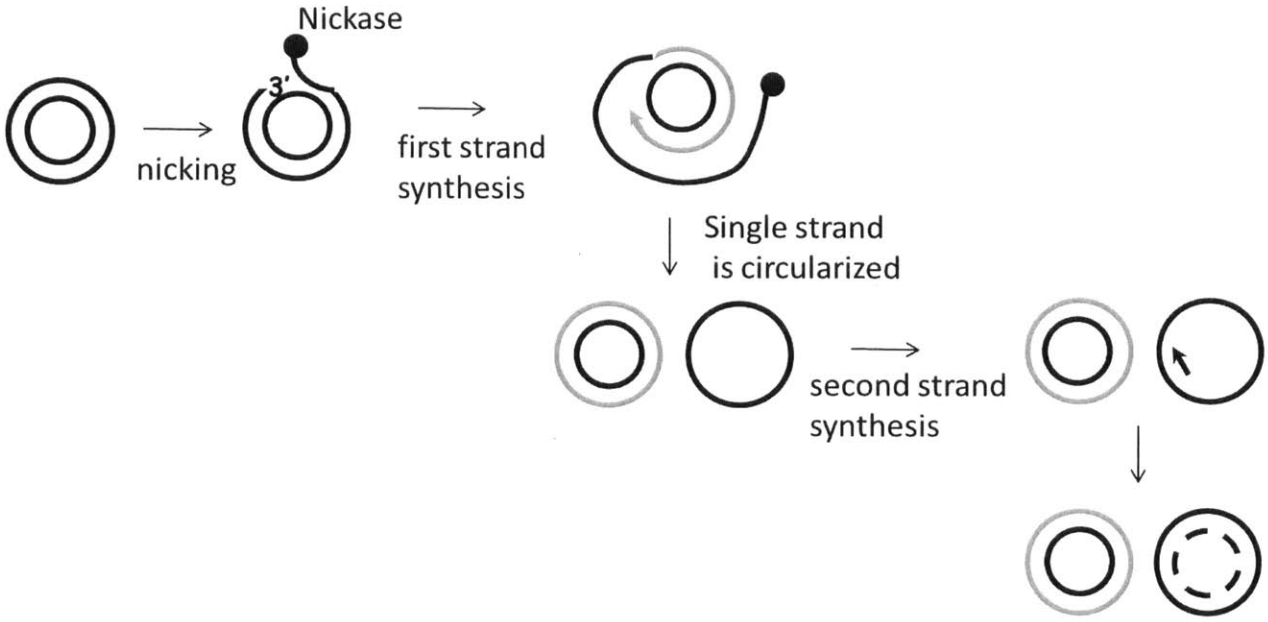



**Figure 4: Model for ICE*Bs1* rolling circle replication.** After excision from the chromosome, ICE*Bs1* replicates autonomously via a plasmid-like, rolling circle mechanism (Lee et al., 2010). Excised, circular ICE*Bs1* (black circle) is nicked by the ICE*Bs1* protein NicK (nickase), indicated by the black dot. Nicking creates a free 3' end from which replication proceeds unidirectionally. The new strand created from this replication is indicated in grey. The linear, single-strand copy of ICE*Bs1* is circularized and a second strand synthesized (dashed black line).

### C. Mechanisms of ICE integration

One great example of the modular, mosaic nature of ICEs is the various mechanisms of integration they employ. While most studied ICEs, including ICE*Bs1*, integrate into the chromosome via site-specific recombination using a tyrosine recombinase, two other mechanisms of integration are known to occur among ICEs: DDE transposition and site-specific recombination using a serine recombinase (Brochet et al., 2009; Burrus and Waldor, 2004; Wozniak and Waldor, 2010). This section will discuss integration via these three mechanisms, with a strong emphasis on integration via tyrosine recombinases, as this is the mechanism by

which ICE*Bs1* and most ICEs integrate. Since this thesis is largely focused on aspects of integration of ICEs (namely the site specificity and effects of their integration), it is useful to first discuss the mechanisms by which ICEs integrate into the chromosome.

**Integration via DDE transposases.** Recently, an ICE in *Streptococcus agalactiae* was discovered, named TnGBS2, that integrates via a DDE transposition mechanism, (Brochet et al., 2009). The DDE transposon family is extremely diverse and includes transposons of the bacteriophage Mu and some of the insertion sequences (or ISs) (Brochet et al., 2009). While many related ICEs have been identified through sequence analysis, this family of DDE transposase ICEs appears to be mostly limited to *Streptococcal* species (Guerillot et al., 2013). Mechanistically, DDE transposition is different from site-specific recombination. While it still remains to be validated, this family of ICEs appears to use a mechanism similar to the DDE transposase of IS911. IS911 does not use a cut-and-paste mechanism (like site-specific recombinases) but instead uses a copy-and-paste mechanism that leaves an integrated copy of the element in the donor while also creating a double-stranded, circular intermediate that can then integrate into the target DNA (Brochet et al., 2009; Duval-Valentin et al., 2004).

**Integration via serine recombinases.** Site-specific recombination via serine recombinases is unrelated to site-specific recombination via tyrosine recombinases and occurs through a separate mechanism (Grindley et al., 2006). As opposed to the successive single-strand breaks created by tyrosine recombinases, serine recombinases create two double-strand breaks, one within the ICE and one within the target site, before strand exchange occurs (Grindley et al., 2006). Opposite strands are brought into contact by a 180° degree shift to one-half of the DNA-enzyme complex and strands are rejoined (Grindley et al., 2006). One ICE with a serine recombinase has been identified experimentally, Tn5397 of *Clostridium difficile*, and a few others have been identified

computationally in *Entercoccus faecalis*, *Streptococcus equi*, and *Clostridium difficile* (Burrus et al., 2002b; Burrus and Waldor, 2004; Wang and Mullany, 2000). Well characterized examples of serine recombinases are the Hin invertase and the Tn3 resolvase (Smith and Thorpe, 2002).

**Integration via tyrosine recombinases.** The recombination reaction is carried out by a recombinase protein called Int, which is required for both integration and excision (Brochet et al., 2009; Grindley et al., 2006; Lee et al., 2007; Van Duyne, 2002). The most well studied member of the tyrosine recombinase family is the phage lambda integrase, and most ICEs, including ICE*Bs1*, are thought to integrate using a similar mechanism. Int, via a topoisomerase-like mechanism, mediates integration by binding to attachment sites present on both the element (referred to as *attP*) and the chromosome (referred to as *attB*) (Figure 5). Often (but not always), *attB* and *attP* are identical and contain inverted repeats (often 5-7 bp) separated by a spacer (or crossover) region (of ~6-8 bp). Bound to the inverted repeats of *attB* and *attP*, Int subunits make one single-strand nick in the crossover region of both *attB* and *attP*, covalently attaching to the 3' end of both nicks, creating two free 5' ends. The free 5' ends attack the covalently attached 3' end on the opposite attachment site and replace the covalent bond between Int and the att site. The two attachment sites are now partially joined. Then, in a similar manner, Int makes single strand nicks in the two strands that were not nicked previously and both stands are swapped resulting in an integrated product. The integrated product has two copies of the attachment site, one on the leftmost end of ICE and the other on the rightmost end (referred to as *attL* and *attR*, respectively).

Integration of ICE*Bs1* appears to follow this same model (Lee et al., 2007). Integration requires the recombinase (named Int). Circular ICE*Bs1* contains an *attP* that is identical to its

preferred *attB* site; *attP* and *attB* are 17 bp with 5 bp inverted repeats on each side of the 7 bp crossover region.

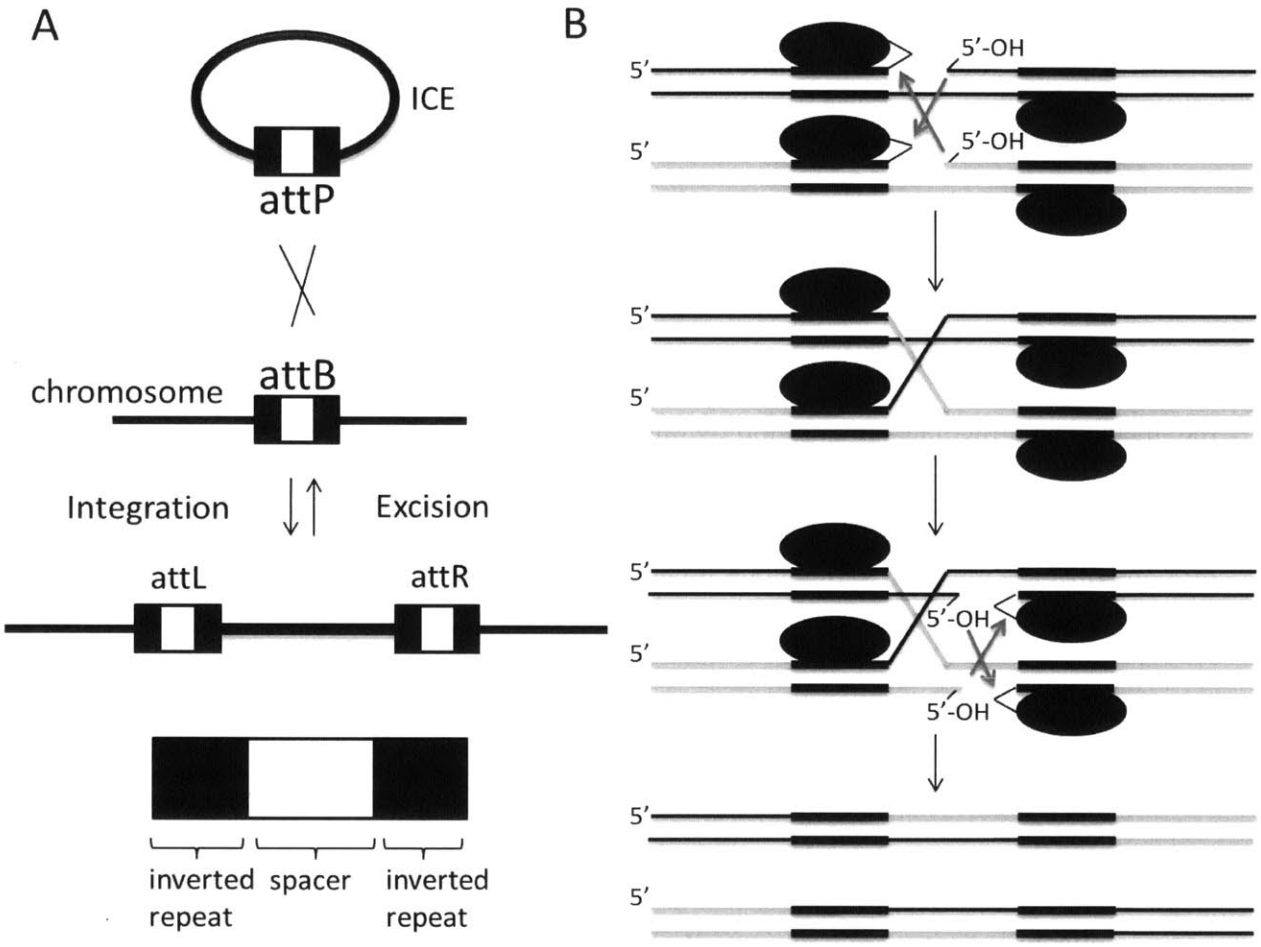



**Figure 5: Mechanism of site-specific tyrosine recombination.** A) Site-specific recombination occurs between two attachment sites, one on the ICE (referred to as *attP*) and one on the chromosome (referred to as *attB*). *attP* and *attB* are often identical; each contains an inverted repeat separated by a crossover region. B) Four subunits of the integrase (black oval) are bound to *attB* and *attP*, one at each inverted repeat (black rectangle). Initially, two of the subunits, one for each attachment site, create a single-strand nick and covalently attach to the 3'end of the DNA (indicated by thin black line). The free 5' end of the crossover region attacks the covalently attached 3' end on the opposite attachment site. The two attachment sites are now linked by a single-strand. The other two subunits of the integrase then nick and covalently attach to the DNA, allowing a second round of strand swapping and resulting in an integrated product. Adapted from (Watson et al., 2008).

The excision reaction of tyrosine recombinases occurs in much the same way, except that additional protein(s) are often needed to bend the DNA to bring to two attachment sites (*attL* and *attR*) close enough to favor the reverse reaction (Lee et al., 2007; Wozniak and Waldor, 2010). For example, a small, charged protein, called excisionase (Xis), is often necessary for excision, as in the case for ICE*Bs1* (Lee et al., 2007).

### D. Locations of ICE integration

With regard to location of integration, ICEs employ various patterns of integration, both within and between the three different mechanisms of integration. This section reviews the different patterns of integration among ICEs, including the use of tRNA genes as hotspots for tyrosine recombinase-mediated integration, as well as the use of secondary attachment sites for site-specific ICEs.

**Patterns of integration**. Some tyrosine recombinase ICEs integrate at many sites in the host chromosome, whereas others are site-specific with regards target site, integrating into only one or a few sites (Burrus and Waldor, 2004). Tn*916*, the first ICE discovered (Franke and Clewell, 1981a), integrates at many sites in the host chromosome of *Enterococcus faecalis*, having a preference for AT-rich or bent DNA (Wozniak and Waldor, 2010). In fact, because of Tn*916*'s promiscuous nature of integration, ICEs were originally referred to as conjugative transposons. Later, it was discovered that many, likely most, "conjugative transposons" are actually quite site-specific, choosing to integrate into only one or a few sites. Given the interrelatedness of elements using these two different behaviors of integration, "conjugative transposons" were renamed integrative and conjugative elements to incorporate both types of elements. (Burrus et al., 2002a, b).

Each strategy for specificity has its benefits. Spread of more promiscuous elements is not limited to organisms with a similar site, whereas the site-specific elements are much less likely to disrupt important genes or have other negative effects on the cell. The trade-offs of these different strategies and other effects of integration are discussed further in the next section.

Interestingly, ICEs with a DDE transposase appear to insert only into intergenic regions, usually 15 or 16 bp upstream of the -35 sequence of sigma A promoters (Brochet et al., 2009). Studies indicate that insertion of TnGBS2 upstream of sigma A promoters does not appear to influence transcriptional levels. While not all DDE transposases insert into intergenic regions, this pattern of integration has been observed for some ISs with a DDE transposase similar to that of TnGBS2.

Tn5397, which integrates via a serine recombinase, appears to integrate into locations with a central GA nucleotide (Wang and Mullany, 2000).

**tRNA genes as hotspots for integration.** Site-specific ICEs (as well as many other mobile genetic elements) often have preferred target sites in tRNA genes. This is true for ICE*Bs1*, which integrates into the middle of a leucine tRNA gene (*trnS-leu2*), recreating the right end of the tRNA and leaving it still functional. It is not entirely understood why tRNA genes are favored as integration targets; however, it is thought to occur, at least in part, because tRNA genes: 1) are highly conserved, and 2) often contain inverted repeats that are used as integration targets for site-specific recombinases (Williams, 2002). Even though the experiments presented in this thesis do not directly address why tRNA genes are favored, they provide insight into the selective pressures that maintain tRNA genes as integration sites.

**Secondary integration sites of site-specific ICEs.** Even though many ICEs integrate site-specifically, site-specific ICEs will integrate into secondary locations, particularly if its primary

site is deleted. A couple of studies have looked at secondary attachment sites for site-specific ICEs. Two secondary sites have been identified for the mobilizable *Salmonella* Genomic Island 1 (Doublet et al., 2008) Additionally, two secondary integration sites were identified for the SXT element of *V.cholerae* (Burrus and Waldor, 2003). Both studies found a reduced integration frequency when the primary site is absent. The study by Burrus and Waldor also looked at the excision frequency of one of the two sites and found it was reduced 3.5-fold (Burrus and Waldor, 2003). Since both papers determined the location of only a couple transconjugants for each element, extraction of features common to the secondary sites is difficult. Additionally, neither study looked at the effects of integration into secondary locations, which may have implications for the evolution of site-specificity among ICEs.

In this thesis, I present data identifying 15 secondary attachment sites for ICE*Bs1*. One of the secondary insertions, *yrkM*, is strongly preferred over the other 14 insertions and was isolated 41% of the time. Surprisingly, some of the insertions had very little similarity to the primary site, indicating that factors besides sequence specificity play a role in target site selection.

Additionally, I found that excision from all secondary insertions analyzed is impaired and that, after excision, the circular product of ICE*Bs1* may be unstable (based on the lower detectable levels of circular ICE*Bs1* from some secondary sites). Both of these effects (reduced excision and lower levels of the circular product) reduce the ability of ICE*Bs1* to spread to other organism and may provide selective pressure against integration into secondary locations, thereby maintaining site-specific integration among ICEs.

### C. Effects of ICE integration

As discussed previously, the different patterns for integration of ICEs into the chromosome

(in particular, site-specific versus non-site-specific) have different trade-offs. This sections expands on those trade-offs by providing further examples of the effects of integration of ICEs on the host cell. A more site-specific strategy makes it less likely that the element will have a detrimental effect on the cell. However, it also makes it less likely the element will alter the genome, perhaps in a way that increases its selective advantage. Thus, this section provides insight into the myriad ways ICEs affect evolution. This section focuses on a few ways that ICEs affect the host cell, including: genes that confer a phenotype to the host, mutations to the host genome, effects on transcription of host genes, and proliferation/viability defect after integration into secondary attachment sites.

**Genes that confer a phenotype to the host.** As mentioned previously, ICEs often contain genes that confer a phenotype to the host cell. This includes antibiotic resistance genes but also genes for processes such as pathogenesis, symbiosis, or metabolism (Seth-Smith and Croucher, 2009; Wozniak and Waldor, 2010). For example, ICE*Kp*1 is found in virulent strains of *Klebseilla pneumonia* (less so in non-virulent strains) and contains genes for iron acquisition and virulence (Lin et al., 2008).

**Mutations to the host genome**. ICEs can cause mutations in the host genome in several ways, including: disruption of host genes, large chromosomal deletions, and exchange of DNA from the target site after excision.

Disruption of host genes. When an ICE integrates at many sites in the chromosome, it is likely to cause disruption of the genes into which it integrates (Shimoji et al., 1998; Smidt et al., 1999). Disruption of host genes by mobile elements can inactivate genes unnecessary for survival, particularly in new environments (Moran, 2002; Parkhill et al., 2001), but it also can be detrimental or even fatal to individual cells.

Large chromosomal deletions. Two ICEs in the same genome can interact to form a large chromosomal deletion. For example, two related ICEs, Tn*916* and Tn5386, were both cut at their left end by one of the integrases (most likely the integrase of Tn*916*) and joined to form a circle of 178 kb that included the chromosomal genes located in between the two mobile elements. Thus, the ability of ICE integrases to have a broad substrate range allows them to interact with other mobile elements in a way that shapes bacterial genomes. (Rice et al., 2005).

Exchange of DNA in the target site after excision. Some ICEs with a tyrosine recombinase integrate into target sites where the sequences between *attP* and *attB* are not identical in the overlap (or crossover) region—the region between the inverted repeats where the successive single strand exchanges occur (Figure 5; also, Section C: Mechanisms of Integration). Excision of the element from these sites can ultimately result in removal of the DNA originally present at the target site *attB* and introduction of the *attP* overlap sequence. If this exchange results in a change to the amino acid sequence, this can also inactivate the gene into which the ICE originally integrated (Shimoji et al., 1998). While stable gene deletion is useful for purposes of genetic engineering, this could be detrimental to the host strain in which this occurs.

**Effects on transcription of host genes.** Integration of ICEs can alter host cell transcription in several ways (Gaillard et al., 2008; Ike et al., 1992; Moon et al., 2007). This includes generating localized effects, such as hyperexpression of a gene where ICE integrated nearby (Ike et al., 1992). It also includes larger scale effects on gene expression, including activation of other co-resident ICEs and transcriptional changes to operons (such as transport of small molecules and glycolate metabolism) during stationary phase (Gaillard et al., 2008; Moon et al., 2007).

**Proliferation/viability defect of strains with integrations into secondary attachment sites.** In this thesis, I detail the proliferation and/or viability defect associated with integration of

ICE*Bs1* into secondary attachment sites after induction of ICE*Bs1* gene expression. This proliferation/viability defect is dependent on ICE*Bs1* rolling-circle replication from the chromosome because ICE*Bs1* is unable to efficiently excise from secondary attachment sites. Such a rolling circle replication-dependent defect was previously unknown for integration of ICEs into secondary attachment sites. This detrimental effect of integration into secondary sites may maintain the site-specific nature of integration of many ICEs by providing selective pressure against integration into alternative sites.

**E. Thesis Summary**

Many ICEs integrate site specifically into one or a few locations in the host chromosome. The experiments outlined in this thesis provide insight into the forces that maintain the site-specific nature of integration observed for many ICEs. This helps us to understand the evolution of ICEs, a type of element that plays a large role in the spread of antibiotic resistances.

Chapter 2 outlines the identification and characterization of 15 secondary integration sites of ICE*Bs1*. It then examines the effect of integration into secondary integration sites on ICE*Bs1*'s 1) mobility and 2) cellular proliferation/viability, after activation of ICE*Bs1* gene expression. Last, Chapter 2 describes how the viability defect is due to ICE*Bs1* replication when unable to properly excise from the secondary attachment sites. This work is being prepared for publication.

Appendix A describes how integration into secondary attachment sites occurs at a low frequency even when the primary site is present.

Appendix B describes how ICE*Bs1* returns to the primary attachment site (*attB*) after mating from a secondary attachment site.

Appendix C establishes that when two primary attachment sites are present on the same

chromosome, ICE*Bs1* prefers the exogenous site over the endogenous site.

Appendix D describes the marked difference in mating efficiency between two similar but different Δ*attB* recipients of *B.subtilis*.

Appendix E describes how deletion of the ICE*Bs1* integrase (Int) results in a worse viability defect than expected. An additional ICE*Bs1* deletion, downstream of Int, creates the expected phenotype.

Chapter 3 discusses the work presented in Chapter 2 and Appendices A, B, C. It comments on various aspects of site-specificity among ICEs, including: 1) the possibility of host factors influencing site-specificity, 2) the frequency of use of non-primary sites, 3) the evolution of altered specificities, and 4) the selective pressures acting to limit the evolution of altered site-specificities.

# Chapter 2

# Consequences of an integrative and conjugative element at secondary integration sites

Kayla L. Menard and Alan D. Grossman

This chapter is being prepared for publication

**Abstract**

Integrative and conjugative elements (ICEs) are widespread mobile genetic elements that are usually found integrated in bacterial chromosomes. They are important agents of evolution and contribute to the acquisition of new traits, including antibiotic resistances. ICEs can excise and transfer to recipients by conjugation. Many ICEs are site-specific in that they integrate preferentially into a primary attachment site in the bacterial chromosome. ICEs can integrate into secondary locations, particularly if the primary site is absent. However, little is known about the consequences of integration of ICEs into alternative attachment sites or what drives the apparent maintenance and prevalence of the many ICEs that use a single attachment site. Using ICE*Bs1*, a site-specific ICE from *Bacillus subtilis* that integrates into a tRNA gene, we found that integration into secondary sites was detrimental to both ICE*Bs1* and the host cell. Excision of ICE*Bs1* from secondary sites was impaired either partially or completely, limiting the spread of ICE*Bs1*. Furthermore, induction of ICE*Bs1* gene expression caused a substantial defect in cell proliferation and/or viability within three hours, and this defect was dependent on rolling-circle replication of ICE*Bs1* from the chromosome. Together, these detrimental effects may provide selective pressure against the survival and dissemination of ICEs that have integrated into secondary sites. Such detrimental effects could explain the maintenance and prevalence of site-specific integration among ICEs.

**Introduction**

Integrative and conjugative elements (ICEs, also known as conjugative transposons) are mobile genetic elements that encode conjugation machinery that mediates their transfer from cell to cell. ICEs often carry additional genes that confer phenotypes to the host cell. This can

include genes involved in pathogenesis, symbiosis, and antibiotic resistances {reviewed in (Wozniak and Waldor, 2010)}. ICEs are typically found integrated in the host bacterial chromosome and can excise to form a circular product that is the substrate for conjugation. Their ability to spread to other organisms through conjugation makes ICEs important agents of horizontal gene transfer in bacteria, and they appear to be more numerous than plasmids (Guglielmini et al., 2011). ICEs can also facilitate transfer (mobilization) of other genetic elements (Lee et al., 2012; Salyers et al., 1995; Wozniak and Waldor, 2010).

Some ICEs have a specific integration (attachment or insertion) site in the host genome whereas others are more promiscuous and can integrate into many locations. For example, SXT, an ICE in *Vibrio cholera* has one primary site of integration in the 5' end of *prfC* (Hochhut and Waldor, 1999). In contrast, Tn*916* has a preference for AT-rich DNA in many different hosts and integrates into many different chromosomal sites (Mullany et al., 2012; Roberts and Mullany, 2009). Each stategy of integration has its benefits. More promiscuous elements can acquire a wider range of genes adjacent to the integration sites, and their spread is not limited to organisms with a specific attachment site. On the other hand, site-specific elements are much less likely to disrupt important genes. The attachment site for these elements is typically in a conserved gene, often a tRNA gene (Burrus and Waldor, 2004; Williams, 2002). Frequently, gene function is not disrupted because the end of the integrating element encodes for the end of the gene disrupted, thereby maintaining functionality. Integration into conserved genes makes it likely that many organisms will have a site for these elements to integrate. We wished to learn more about the ability of site-specific ICEs to integrate into secondary integration (or attachment) sites when their primary site is not present in a genome. We wondered if an ICE could function normally in a secondary site and if there was any effect on the host.

We used ICE*Bs1* of *Bacillus subtilis* to analyze effects of integration into secondary attachment sites. ICE*Bs1* is a site-specific conjugative transposon that is normally found integrated into a tRNA gene (*trnS-leu2*) (Auchtung et al., 2005; Lee et al., 2007). ICE*Bs1* is approximately 20 kb (Fig. 1), and many of its genes are similar to genes in other ICEs, including those in Tn*916* (Auchtung et al., 2005; Burrus et al., 2002b), the first conjugative transposon identified (Franke and Clewell, 1981a, b). Induction of ICE*Bs1* gene expression leads to excision from the chromosome, autonomous rolling-circle replication of ICE*Bs1*, and initiation of mating (Auchtung et al., 2005; Lee et al., 2007; Lee et al., 2010). It is not known what properties or advantages ICE*Bs1* may confer to host cells. Naturally occurring ICE*Bs1* is not known to carry genes involved in antibiotic resistances, virulence, or metabolism. However, because of the conservation of many of its functions, the ease of manipulating *B. subtilis*, and the high efficiency of experimental induction of gene expression, ICE*Bs1* is extremely useful for studying basic and conserved properties of ICEs.

We are interested in the physiological consequences of integration of ICE*Bs1* into secondary attachment sites. Previous work showed that in the absence of its primary attachment site (*attB* in the gene for tRNA-leu2), ICE*Bs1* integrates into secondary attachment sites (Lee et al., 2007). Seven different sites were identified and provided insight into the chromosomal sequences needed for integration. Work presented here extends these findings by evaluating the ability of ICE*Bs1* to excise from these sites, and determining the effects on host cells. We isolated ICE*Bs1* insertions in 15 different secondary integration sites, including the seven previously identified sites and eight new ones. Some of these sites were similar to the primary attachment site and others were remarkably different. We found that, once integrated, excision of ICE*Bs1* from all secondary sites analyzed was reduced or undetectable compared to the rate at the primary site

(*attB*). In two of the four insertions that had reduced but detectable excision, the excised, circular form of ICE*Bs1*was present at lower levels than expected, indicating that circular ICE*Bs1* might be unstable. Defects in excision and stability of ICE*Bs1* severely limit its ability to spread to other cells. In addition, we found that induction of ICE*Bs1* gene expression in these secondary integration sites caused a defect in cell proliferation/viability and induction of the SOS response. Together, our results indicate that integration of ICE*Bs1* in secondary integration sites is deleterious to ICE*Bs1* and to the host cell.

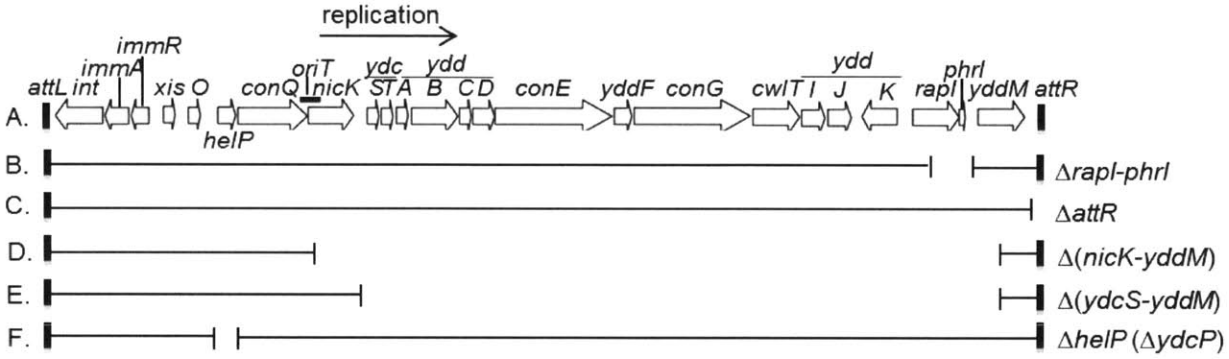



**Figure 1. Map of ICE*Bs1* and its derivatives.**
**A.** Linear genetic map of ICE*Bs1* integrated in the chromosome. Open arrows indicate open reading frames and the direction of transcription. Gene names are indicated above the arrows. The origin of transfer (*oriT*) is indicated by a thick black line overlapping the 3' end of *conQ* and the 5' end of *nicK*. *oriT* functions as both the ICE*Bs1* origin of transfer and origin of replication (Lee et al., 2007; Lee et al., 2010). The thin black arrow indicates the direction of ICE*Bs1* rolling-circle replication. The small rectangles at the ends of ICE*Bs1* represent the 60 bp direct repeats that contain the site-specific recombination sites in the left and right attachment sites, *attL* and *attR*, that are required for excision of the element from the chromosome.
**B-F.** Various deletions of ICE*Bs1* were used in this study. Thin horizontal lines represent regions of ICE*Bs1* that are present and gaps represent regions that are deleted. Antibiotic resistance cassettes that are inserted are not shown for simplicity. **B.** *rapI* and *phrI* are deleted and a kanamycin resistance cassette inserted. **C.** The right attachment site (*attR*) is deleted and a tetracycline resistance cassette inserted. **D.** The genes from the 5' end of *nicK* and into *yddM* are deleted and a chloramphenicol resistance cassette inserted. **E.** The genes from the 5' end of *ydcS* and into *yddM* are deleted and a chloramphenicol resistance cassette inserted. **F.** The entire coding sequence of *helP* (previously known as *ydcP*) and 35 bp in the *helP-ydcQ* intergenic region is removed. There is no antibiotic resistance cassette in this construct.

## Results

### Identification of secondary sites of integration of ICE*Bs1*

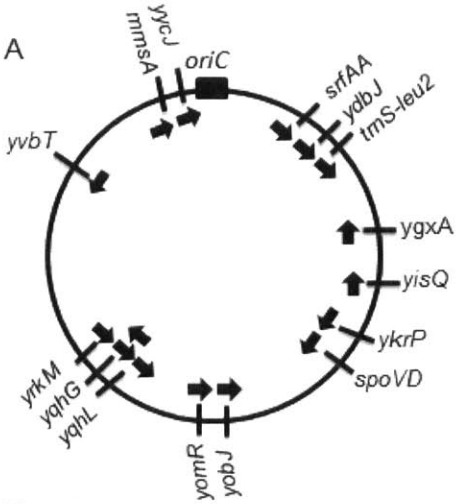
We examined 27 independent insertions of ICE*Bs1* into secondary integration sites in the *B. subtilis* chromosome. Briefly, these insertions were obtained by: 1) mating ICE*Bs1* into a recipient strain deleted for the primary attachment site *attB* (located in the tRNA gene *trnS-leu2*), 2) isolating independent transconjugants, and 3) determining the site of insertion in each of 27 independent isolates.

There were 15 different secondary integration sites for ICE*Bs1* among the 27 independent transconjugants (Fig. 2). Seven of the 15 sites were described previously (Lee et al., 2007), and eight additional sites are reported here. There appears to be no requirement for the alignment of the orientation of ICE*Bs1* insertions with the direction of the host replication, although 10 of the 15 insertions were oriented such that ICE*Bs1* was co-directional with chromosomal replication (Fig. 2A). Of the 27 independent transconjugants, 11 (41%) had ICE*Bs1* inserted in a site in *yrkM* (designated *yrkM*::ICE*Bs1*) (Fig. 2B), a gene of unknown function. Three of the 27 (11%) transconjugants had ICE*Bs1* inserted in a site in *mmsA* (encoding an enzyme involved in myo-inositol catabolism (Yoshida et al., 2008). The site in *yrkM* is the most similar to the primary attachment site *attB*, differing by two base pairs. The site in *mmsA* differs from *attB* by three base pairs (Fig. 2B). Two insertions were in a site in *yqhG*, although in opposite orientations. These are counted as two different sites since the sequence in each orientation is different (Fig. 2B). The remaining 11 insertions were in unique sites, either in genes or intergenic regions (Fig. 2B). None of the 15 identified insertions were in essential genes, and none caused a noticeable defect in cell growth while ICE*Bs1* was repressed. Furthermore, none of the insertions were in

tRNA genes (including redundant, nonessential tRNA genes) that are common integration sites for many ICEs (Burrus and Waldor, 2004; Williams, 2002).

Some of the secondary insertion sites were similar to, and others quite different from, the primary ICE*Bs1* attachment site (*attB*ICE*Bs1*, or simply *attB*). *attB* contains a 17 bp stem-loop sequence consisting of a 5 bp inverted repeat separated by 7 bp (Fig. 2C). We aligned and compared the sequences of the 15 different secondary attachment sites and searched for a common motif using WebLogo 3.3 (http://weblogo.threeplusone.com/) (Crooks et al., 2004). For each secondary attachment site, we provided an input of 26 bp that included the region of the stem-loop sequence (17 bp, inferred from the sequence of *attB*) and a few base pairs upstream and downstream. The conserved sequences were largely in the 17 bp that were originally proposed to define *attB* (Lee et al., 2007), including several positions in the loop region of the stem-loop sequence, the 5 bp inverted repeats, and perhaps 1-2 additional base pairs downstream of the stem-loop (Fig. 2C). There was considerable sequence diversity among the 15 secondary integration sites and the primary site *attB*, and no single position was conserved in all the secondary sites (Fig. 2B). In some cases (e.g., insertions in *yrkM*, *mmsA*, *yqhG*, and *srfA*) there are only 2 - 3 base pairs that are different between the secondary site and *attB*. In contrast, insertion sites in *yghL*, *yvbT*, and *ykrP* have 10-12 mismatches (out of 17 bp) from the sequence of *attB* (Fig. 2B). These results indicate that in the absence of the primary integration site in *trnS-leu2*, ICE*Bs1* can integrate into many different sites throughout the genome, albeit at a lower efficiency (Lee et al., 2007). Based on these data and the number of insertion sites identified only once, it is clear that we have not identified all of the possible secondary integration sites for ICE*Bs1*.

A




B

| gene | chromosomal target | mm | occurences |
|---|---|---|---|
| *attB* | ctaggttgagggcctag | N/A | N/A |
| *yrkM* | ctaTgttgagggActag | 2/17 | 11 (41%) |
| *mmsA* | cAaggttgaTg-cctag | 3/17 | 3 (11%) |
| *yqhG** | ctagg-tgaAgAcctag | 3/17 | 1 (4%) |
| *srfAA* | cCagAttgagggccAag | 3/17 | 1 |
| *ygxA* | ctagTttgaCTgActaC | 5/17 | 1 |
| *yqhG** | ctaggtCTTCA-cctag | 6/17 | 1 |
| *yobJ* | ctaggGAAaTggActaA | 6/17 | 1 |
| *ydbJ* | cGaggttgaAgAcTtGT | 6/17 | 1 |
| *yycJ* | TCaTgttgTTgTcctGg | 7/17 | 1 |
| *yomR* | ctTgTtATagAAActag | 7/17 | 1 |
| *spoVD* | TGagAGAgTgggccCGg | 8/17 | 1 |
| *yisQ* | GtGggGCATTggcctCC | 9/17 | 1 |
| *yqhL* | cAGggttTGTgCcGCGC | 10/17 | 1 |
| *yvbT* | GCCggGCAGAggAAtGg | 11/17 | 1 |
| *ykrP* | cAaTTtGTTTgCcGCGT | 12/17 | 1 |

C

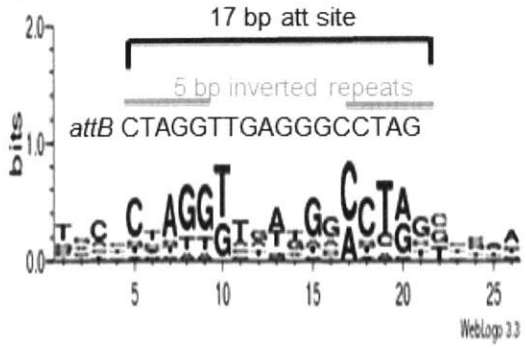


**Figure 2. Map and DNA sequence of the primary and 15 secondary integration sites for ICE*Bs1*.**

**A.** Approximate position of the primary and 15 secondary ICE*Bs1* integration sites on the *B. subtilis* chromosome. The circle represents the *B. subtilis* chromosome with the origin of replication (*oriC*) indicated by the black rectangle at the top. The slash marks represent the approximate location of the ICE*Bs1* insertion site. The name of the gene near which (*ygxA*) or into which (all other locations) ICE*Bs1* inserted is indicated on the outside of the circle. The arrows on the inside of the circle indicate the direction of ICE*Bs1* replication for each insertion. *trnS-leu2* contains the primary ICE*Bs1* integration site *attB*.

**B.** DNA sequence of the primary and 15 secondary integration sites. The gene name is indicated on the left, followed by the DNA sequence (chromosomal target). The primary attachment site (*attB*) is a 17 bp sequence with 5 bp inverted repeats (underlined) separated by a 7 bp spacer. Mismatches from *attB* are indicated in bold, capital letters. "mm" indicates the number of mismatches from the primary 17 bp *attB*. "occurrences" indicates the number of independent times an insertion in each site was identified. Percentages of the total (27) are indicated in parenthesis. The * next to *yqhG* indicates that two different ICE*Bs1* insertions were isolated in this gene, once in each orientation.

**C.** Sequence logo of the ICE*Bs1* secondary attachment sites. Using Weblogo 3.3 (Crooks et al., 2004), we generated a consensus motif of the 26 bases surrounding the insertion site of the 15 secondary insertion sites for ICE*Bs1*. For comparison, the primary attachment site for ICE*Bs1* is a 17 bp region with 5 bp inverted repeats and a 7 bp spacer region in the middle (Lee et al., 2007). The size of each nucleotide corresponds to the frequency with which that nucleotide was observed in that position in the secondary attachment sites.

### Excision of ICE*Bs1* from secondary integration sites is reduced

We wished to determine if there were any deleterious consequences of integration of ICE*Bs1* into secondary attachment sites. We found that although ICE*Bs1* could integrate into the secondary integration sites, excision was reduced compared to the primary site. We induced excision from seven of the secondary sites by overexpressing the activator of ICE*Bs1* gene expression, RapI, from a regulated promoter (Pxyl-*rapI*) integrated in single copy in the chromosome at the nonessential gene *amyE* (Materials and Methods). Overproduction of RapI induces ICE*Bs1* gene expression (Auchtung et al., 2007; Auchtung et al., 2005) and typically results in excision of ICE*Bs1* from *attB* in >90% of cells within 1-2 hrs (Lee et al., 2007; Lee et al., 2010). Following a similar protocol as described for monitoring excision from *attB* (Auchtung et al., 2007; Auchtung et al., 2005; Lee et al., 2010), we performed quantitative real

time PCR (qPCR) using genomic DNA as template and primers designed to detect the empty secondary attachment site that would form if the element excises. In a positive control, excision of ICE*Bs1* from *attB* occurred in >90% of cells by two hours after expression of the activator RapI (Fig. 3A, wt). In a negative control, excision of an ICE*Bs1* Δ*attR* mutant integrated in *attB* was undetectable (Fig. 3A, Δ*attR*). Excision from four of the sites tested, *yrkM*, *mmsA*, *srfAA*, and *yycJ*, was reduced yet still detectable, ranging from 4% to 15% of that of ICE*Bs1* from *attB*. Excision from the other three sites tested, *yvbT*, *spoVD*, and *ykrP*, was undetectable (Fig. 3A), as observed for ICE*Bs1* Δ*attR*, the excision-defective control. In general, the secondary integration sites that are most divergent from *attB* had the least amount of excision (Table 1).

These findings indicate that integration of ICE*Bs1* into sites other than *attB* causes a reduction, sometimes quite severe, in the ability of the element to excise. Because excision is required for transfer of a functional ICE, this reduced excision will limit the spread of ICE*Bs1*.

**Decreased conjugation of ICE*Bs1* from secondary sites**

We measured the mating efficiencies of ICE*Bs1* following excision from the four secondary attachment sites from which excision was reduced but detectable. Excision of ICE*Bs1* is required for transfer of the element to recipient cells. Thus, if the ICE*Bs1* circle is stable, then the mating efficiencies should be proportional to the excision frequencies. The mating efficiencies of ICE*Bs1* from *yrkM* and *srfAA* were ~2 - 5% of that of ICE*Bs1* from *attB*. Likewise, the excision frequencies of ICE*Bs1* inserted in *yrkM* and *srfAA* were ~5% of those of ICE*Bs1* in *attB*. These results indicate that for ICE*Bs1* integrated in *yrkM* and *srfAA*, the mating efficiencies were mainly a function of the reduced excision frequencies.

In contrast, the mating efficiencies of ICE*Bs1* that excised from *mmsA* or *yycJ* were much lower than what would be predicted based only on excision frequency. In both cases, the

excision frequencies were ~15% of that of ICE*Bs1* integrated in *attB*. However, the mating efficiencies were ~0.2% of that of ICE*Bs1* from *attB*, a 75-fold difference. Based on this result, we postulated that the reduced mating efficiency relative to the excision frequency was indicative of a reduction in the amount of circular ICE*Bs1*, since excised, circular ICE*Bs1* is necessary for efficient mating.

**Lower levels, than predicted based on repaired junction data, of circular ICE*Bs1* from secondary sites that generate a heteroduplex**

We measured the relative amounts of circular ICE*Bs1* after excision from *attB*, *yrkM, srfAA*, *mmsA,* and *yycJ*, the four insertions with reduced but detectable excision, using qPCR primers designed to detect only the circular form of ICE*Bs1*. The relative amounts of each circle were compared to the relative amount of the empty secondary attachment site (repaired junction) from which ICE*Bs1* excised. Measurements were made two hours after induction of ICE*Bs1* gene expression (overproduction of RapI).

As expected, the ratio of the amounts of the circular form to the empty attachment site was about the same for insertions in *yrkM* and *srfAA* as for an insertion in *attB* (Fig. 3B). In contrast, the ratio of the circle to the empty attachment site for *mmsA* and *yycJ* was significantly less than that for wild type (Fig. 3B). Comparing the total amount of the ICE*Bs1* circle from *mmsA* and *yycJ* to that from *attB* indicated that there was approximately 0.3% as much circle from each site as from *attB*. This decrease in the amount of ICE*Bs1* circle is consistent with and likely causes the drop in mating efficiency to approximately 0.2% of that of ICE*Bs1* from *attB*.

The decrease in the amount of circular ICE*Bs1* from *mmsA* and *yycJ* is likely due to the generation of a heteroduplex in the attachment site on the circular ICE*Bs1*. The ICE*Bs1* attachment site contains a 17 bp sequence with a 7 bp spacer region between 5 bp inverted

repeats. Integrase-mediated site-specific recombination occurs in the 7 bp spacer (the crossover region) (Lee et al., 2007) (Fig. 3C). If the 7 bp region in a chromosomal attachment site is different from that in ICE*Bs1*, as is the case for *mmsA* and *yycJ*, then integration and host replication will create left (*attL*) and right (*attR*) ends that have different crossover regions (Fig. 3D). Upon excision, these elements are predicted to contain a heteroduplex in the attachment site on the excised circular ICE*Bs1*. Of the four insertions that have readily detectable excision frequencies, two (*mmsA* and *yycJ*) are predicted to form a heteroduplex and two (*yrkM* and *srfAA*) are not. In the case of *mmsA*::ICE*Bs1*, the left and right ends are known to have different sequences (Lee et al., 2007).

Our results indicate that excision of ICE*Bs1* from secondary sites from which a heteroduplex is formed leads to lower levels of the circular ICE*Bs1* heteroduplex, indicating that circular ICE*Bs1* may be unstable. We do not yet know what causes lower levels of the circular ICE*Bs1* heteroduplex. Loss of the DNA mismatch repair gene *mutS* did not alter the levels of the ICE*Bs1* heteroduplex (unpublished results), indicating that mismatch repair is not solely responsible for this effect. Nonetheless, the overall reduction in conjugation is due to both decreased excision and further decreased levels of the excised element. Both of these defects provide barriers to the spread of ICE*Bs1* from secondary attachment sites.

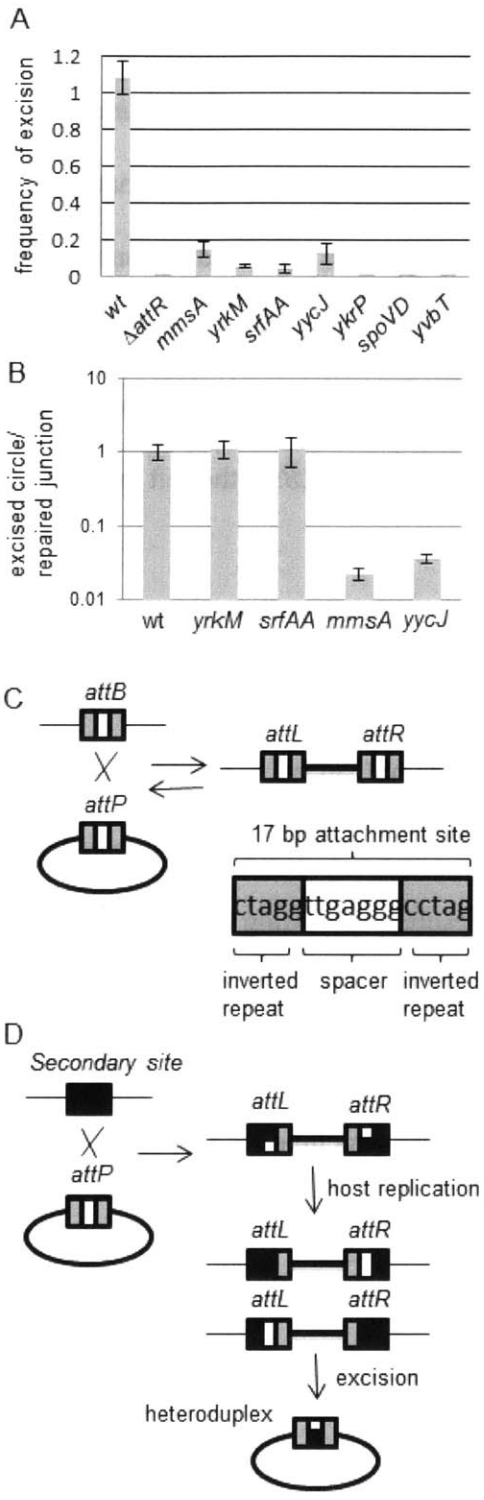

A
frequency of excision
1.2
1
0.8
0.6
0.4
0.2
0
wt
ΔattR
mmsA
yrkM
srfAA
yycJ
ykrP
spoVD
yvbT
B
excised circle/ repaired junction
10
1
0.1
0.01
wt
yrkM
srfAA
mmsA
yycJ
C
attB
attP
attL
attR
17 bp attachment site
ctaggttgagggcctag
inverted repeat
spacer
inverted repeat
D
Secondary site
attP
attL
attR
host replication
attL
attR
attL
attR
excision
heteroduplex

**Figure 3. Mobility of ICE*Bs1* from secondary attachment sites is reduced.**
**A-B.** Excision frequencies and relative amounts of the excision products (circular ICE*Bs1* and empty chromosomal site) were determined as described in Materials and Methods. Cells were grown in defined minimal medium with arabinose as carbon source. Products from excision were determined two hours after addition of xylose to induce expression of Pxyl-*rapI* to cause induction of ICE*Bs1* gene expression. Primers for qPCR were unique to each attachment site. Strains used include: wt, that is, ICE*Bs1* inserted in *attB* (CAL874); Δ*attR*, ICE*Bs1* integrated in *attB*, but with the right attachment site deleted and ICE*Bs1* unable to excise (**Fig.** 2) (CAL872); *mmsA*::ICE*Bs1* (KM70); *yrkM*::ICE*Bs1* (KM72); *srfAA*::ICE*Bs1* (KM141); *yycJ*::ICE*Bs1* (KM132); *ykrP*::ICE*Bs1* (KM77); *spoVD*::ICE*Bs1* (KM130); *yvbT*::ICE*Bs1* (KM94). Each strain was assayed at least three times (biological replicates) and qPCR was done in triplicate on each sample. Error bars represent standard deviation.
**A.** Frequency of excision of ICE*Bs1* from the indicated site of integration. The relative amount of the empty chromosomal attachment site was determined and normalized to the chromosomal gene *cotF*. Data were also normalized to a strain with no ICE*Bs1* (JMA222), which represents 100% excision.
**B.** Relative amount of circular ICE*Bs1* compared to the amount of empty chromosomal attachment site for the indicated insertions. The relative amount of the ICE*Bs1* circle, normalized to *cotF*, was divided by the relative amount of the empty attachment site, also normalized to *cotF*. These ratios were then normalized to those for wild type.
**C.** Cartoon of integration of ICE*Bs1* into its primary bacterial attachment site *attB*. *attB* is identical to the attachment site on ICE*Bs1*, *att*ICE*Bs1*. They consist of a 17 bp region with 5 bp inverted repeats (gray boxes) on each side of a 7 bp spacer region (white box). During integration and excision, a recombination event occurs in the 7 bp crossover region.
**D.** Cartoon of integration of ICE*Bs1* into secondary integration sites. A secondary integration site is indicated with a black box. When ICE*Bs1* integrates into a secondary site, the crossover regions in *att*ICE*Bs1* and that of the secondary site are not necessarily identical, potentially creating a mismatch. This mismatch, if not repaired, will be resolved by host replication, generating left and right ends with different crossover sequences. Excision would then create a circular ICE*Bs1* with a heteroduplex in the attachment site on ICE*Bs1*.

**Decreased proliferation and/or viability of strains in which ICE*Bs1* has decreased excision**

We found that induction of strains with ICE*Bs1* in secondary integration sites had decreased ability to form colony forming units (CFUs). We measured colony forming units (CFUs) of several strains with excision-defective ICE*Bs1* insertions, including ICE*Bs1* in secondary sites and ICE*Bs1* Δ*attR* (in *attB*), both with and without induction of ICE*Bs1* gene expression. We also measured CFUs of wild type, excision-capable ICE*Bs1* integrated at *attB* under similar

conditions (Fig. 4A). In the absence of RapI expression, when most ICE*Bs1* genes are repressed, CFU/mL of excision-defective strains were indistinguishable from that of excision-competent strains. In contrast, by three hours after induction of ICE*Bs1* gene expression in excision-defective ICE*Bs1* strains (Δ*attR* with ICE*Bs1* in *attB*, or insertions in *mmsA*, *yrkM*, *srfAA*, *yycJ*, *spoVD*, *yvbT*, and *ykrP*), the number of CFUs was reduced compared to that of the excision-competent ICE*Bs1* (in *attB*) (Fig. 4A). These results are consistent with previous observations that excision-defective *int* and *xis* null mutants have a proliferation/viability defect when RapI is overproduced (Lee et al., 2007).

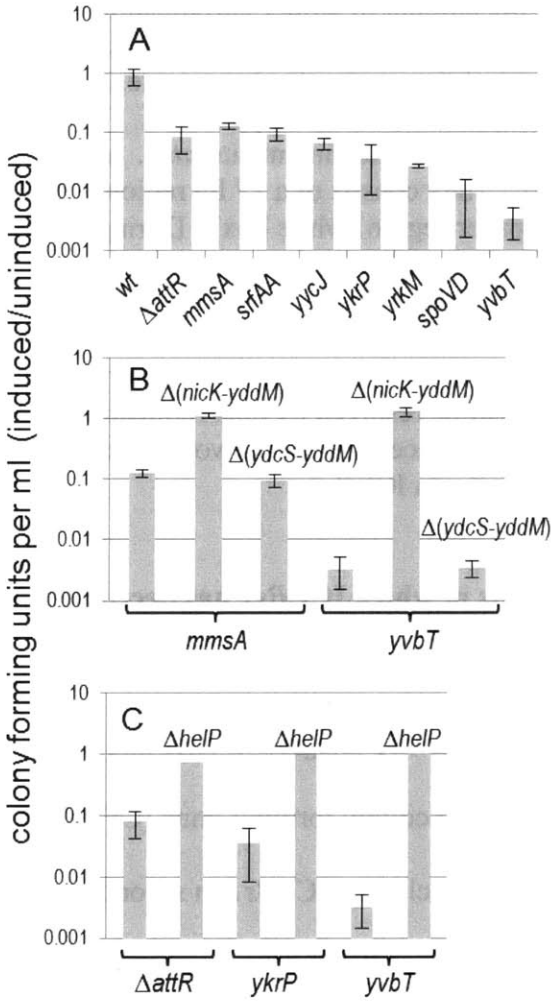



**Figure 4. Effects of induction of ICE*Bs1* gene expression on cell proliferation/viability.**

The effects of induction of ICE*Bs1* gene expression on cell proliferation/viability are shown for the indicated insertions and their derivatives. Cells were grown in defined minimal medium with arabinose to early exponential phase (OD600 ~0.05) and xylose was added to induce expression of Pxyl-*rapI*, causing induction of ICE*Bs1* gene expression. The number of colony forming units was measured three hours after induction and compared to cells grown in the absence of xylose (uninduced). All experiments were done at least three times, except for the *helP* mutants (panel C), which were done twice with similar results. Data presented are averages of the replicates. Error bars represent the standard deviation of at least three replicates.
**A.** Drop in CFU/mL of strains in which excision of ICE*Bs1* is defective. Strains used include: wt, that is, *attB*::ICE*Bs1* (CAL874); *attB*::ICE*Bs1* Δ*attR*::*tet* (CAL872); *mmsA*::ICE*Bs1* (KM70); *srfAA*::ICE*Bs1* (KM141); *yycJ*::ICE*Bs1* (KM132); *ykrP*::ICE*Bs1* (KM77); *yrkM*::ICE*Bs1* (KM72); *spoVD*::ICE*Bs1* (KM130); *yvbT*::ICE*Bs1* (KM94).
**B.** Data are shown for two secondary insertion sites (*mmsA*::ICE*Bs1* and *yvbT*::ICE*Bs1*). Similar results were obtained with *ykrP*::ICE*Bs1* and *srfAA*::ICE*Bs1* (data not shown). Derivatives of each insertion that delete *nicK* and all downstream ICE*Bs1* genes (Δ*nicK-yddM*) or that leave *nicK* intact and delete just the downstream genes (Δ*ydcS-yddM*) (Fig. 1) were tested. Strains used include: *mmsA*::ICE*Bs1* (KM70); *mmsA*::{ICE*Bs1* Δ(*nicK-yddM*)::*cat*} (KM366); *mmsA*::{ICE*Bs1* Δ(*ydcS-yddM*)::*cat*} (KM358); *yvbT*::ICE*Bs1* (KM94); *yvbT*::{ICE*Bs1* Δ(*nicK-yddM*)::*cat*} (KM370); *yvbT*::{ICE*Bs1* Δ(*ydcS-yddM*)::*cat*} (KM362). Data for KM70 and KM94 are the same as those shown above in panel A and are shown here for comparison.
**C.** The ICE*Bs1* helicase processivity protein encoded by *helP* is required for cell killing by ICE*Bs1*. Data are shown for two secondary integration sites (*ykrP* and *yvbT*) and the excision defective ICE*Bs1* Δ*attR*. The *helP* allele is a non-polar deletion (Thomas et al., 2013). Strains used include: *attB*::(ICE*Bs1* Δ*attR*::*tet*) (CAL872); *attB*::(ICE*Bs1* Δ*helP* Δ*attR*::*tet*) (KM437); *ykrP*::ICE*Bs1* (KM77); *ykrP*::(ICE*Bs1* Δ*helP*) (KM429); *yvbT*::ICE*Bs1* (KM94); *yvbT*::(ICE*Bs1* Δ*helP*) (KM459). Data for KM94, KM77, and CAL872 are the same as those shown above in panel A and are shown here for comparison

Several of the integration sites (ICE*Bs1* Δ*attR* in *attB*, and insertions in *mmsA*, *srfAA*, *yycJ*) caused a drop in CFU/mL to ~10% of that of strains without ICE*Bs1* induction or the strain with wild type ICE*Bs1* at *attB* (Fig. 4A). This difference in CFU/mL between induced and uninduced cells (after three hours induction) appears to be due to a mixture of both a defect in proliferation (cell division) and viability (cell death). At earlier time points (less than two hours induction), the number of CFU/mL for induced strains remained similar to the number of CFU/mL present before induction, indicating that cell division (proliferation) has been halted (data not shown).

However, at later time points (after ~2.5 hours), the CFU/mL dropped below levels observed pre-induction, indicating that cell viability has been compromised.

Induction of ICE*Bs1* in other insertion sites (*ykrP*, *yrkM*, *spoVD*, *yvbT*) caused a more severe drop in CFU/mL. These strains exhibited a marked decrease in CFU/mL below pre-induction levels even at early time points (data not shown), indicating that the difference in CFU/mL between induced and uninduced cells is largely due to cell death (viability defect).

The drop in CFU/mL of induced compared to uninduced cells in the various ICE*Bs1* insertions did not correlate with dissimilarity of the attachment sites to *attB*, or to the amount of residual excision in the excision-defective strains. For example, the ICE*Bs1* Δ*attR* mutant is completely unable to excise, and CFU/mL are ~10% of uninduced cells three hours after induction of ICE*Bs1* gene expression. In contrast, for the strain with ICE*Bs1* inserted into *yrkM*, excision of ICE*Bs1* is ~5%, and CFU/mL are ~3% of uninduced cells (Table 1). Together, these results indicate that something about the specific locations of the insertions is likely causing the more extreme viability defect observed in some of the excision-defective ICE*Bs1* strains.

One of the strains with the most extreme defect in viability is ICE*Bs1* inserted in *yvbT*. Within three hours after induction of ICE*Bs1* gene expression in the *yvbT*::ICE*Bs1* strain, viability was ~0.3% of that of strains without ICE*Bs1* induction or of the strain with excision-competent ICE*Bs1* (Fig. 4A). *yvbT* encodes a product of unknown function, but is predicted to be similar to alkanal monooxygenases (luciferases). Insertion of ICE*Bs1* in *yvbT* likely knocks out *yvbT* function, so it seems possible that loss of *yvbT* combined with induction of ICE*Bs1* gene expression was causing the severe drop in viability. To test this, we deleted *yvbT* in cells containing ICE*Bs1* inserted into *mmsA* and measured CFU/mL after induction of ICE*Bs1* gene expression. There was no additional drop in CFU/mL of the *mmsA*::ICE*Bs1* *yvbT* null mutant

compared to the *mmsA*::ICE*Bs1* secondary site alone (wild type *yvbT*), either with or without induction of ICE*Bs1* gene expression. Based on these results, we conclude that the severe defect in viability of the *yvbT*::ICE*Bs1* secondary site mutant was not due to the loss of *yvbT* function combined with induction of ICE*Bs1* gene expression.

We do not know what causes the more severe drop in viability in some insertions. However, the decrease in cell proliferation and viability caused by expression of ICE*Bs1* in secondary attachment sites should provide selective pressure against the long term survival of these strains. The more severe the loss in viability, the stronger the selective pressure against long term survival of strains with insertions in these sites.

**ICE*Bs1* replication functions are required for the defect in proliferation/viability of excision-defective insertions**

Because the drop in proliferation/viability in the first few hours after induction of ICE*Bs1* gene expression occurs in ICE*Bs1* excision-defective, but not in excision-competent strains, the decreased proliferation/viability is likely due to a cis-acting property of ICE*Bs1* and not a diffusible ICE*Bs1* product. One of the more dramatic changes following induction of ICE*Bs1* gene expression is initiation of multiple rounds of unidirectional rolling circle replication (Lee et al., 2010). This replication initiates from the ICE*Bs1* origin of transfer *oriT*, requires the ICE*Bs1* relaxase encoded by *nicK* and the helicase processivity factor encoded by *helP* (previously *ydcP*) (Thomas et al., 2013). Rolling circle replication of ICE*Bs1* occurs even when ICE*Bs1* is unable to excise from the chromosome (Lee et al., 2010). Therefore, we expected that induction of ICE*Bs1* gene expression in the secondary site insertions would lead to unidirectional rolling circle replication from *oriT* in the host chromosome (Fig. 5), as observed previously for ICE*Bs1* Δ*attR* (Lee et al., 2010). It seemed likely that this replication could interfere with chromosome

replication or gene expression and possibly cause damage to the chromosome, leading to the defect in cell proliferation and viability.

Preliminary experiments indicated that loss of *nicK* restored proliferation/viability after induction of ICE*Bs1* gene expression. However, this effect could have been due to polarity on downstream genes. Unfortunately, *nicK* null mutants are difficult to fully complement, perhaps because NicK might act preferentially in cis. In addition, complementation of other supposedly "non-polar" mutations in ICE*Bs1* are not complemented fully. Therefore, to test if loss of *nicK* was responsible for the suppression of the proliferation/viability defect, or if the suppression was due to loss of expression of a downstream gene, we compared two different deletions in ICE*Bs1*. In one deletion, we left *nicK* intact and removed most of the genes downstream from *oriT* and *nicK* {Δ(*ydcS-yddM*)} (Fig. 1E). In the second deletion, we removed *nicK* and the downstream genes (Fig. 1D). Although the *nicK* deletion {Δ(*nicK-yddM*)} is not known to alter sequences in *oriT*, it is possible that *oriT* function is affected. In either case, loss of *nicK* eliminates nicking and replication of ICE*Bs1* (Lee et al., 2007; Lee et al., 2010). We found that deletion of *nicK* alleviated the proliferation/viability defect of excision-defective secondary insertions, including *mmsA*::ICE*Bs1* and *yvbT*::ICE*Bs1* that caused the most severe drop in viability (Fig. 4B). Deletion of the genes downstream from *nicK* did not alleviate the drop in CFU/mL (Fig. 4B), indicating that expression of conjugation functions was not the cause of the decreased cell CFU/mL. Together, these results indicate that a NicK-dependent process is causing the defect in proliferation/viability of the excision-defective ICE*Bs1*.

NicK creates a nick at a specific site in ICE*Bs1 oriT* (Lee and Grossman, 2007), and nicking is required for ICE*Bs1* replication (and conjugation) (Lee et al., 2010). To determine if the drop in CFU/mL was due to nicking per se, or to replication, we used a recently defined ICE*Bs1* gene,

*helP*, which encodes a helicase processivity factor that is needed for ICE*Bs1* replication but not for nicking (Lee and Grossman, 2007; Thomas et al., 2013). Deletion of *helP* (Fig. 1F) completely alleviated the drop in CFU/mL associated with induction of ICE*Bs1* (Fig. 4C).

Based on these results, we conclude that unidirectional rolling-circle replication from *oriT* in the chromosome most likely caused the defect in proliferation/viability of the excision-defective ICE*Bs1*. The decrease in proliferation/viability could be due to breaks and degradation of chromosomal DNA around the site of insertion and/or disruptions in host chromosomal replication or gene expression from the multiple rounds of rolling circle replication from *oriT* (Fig. 5).

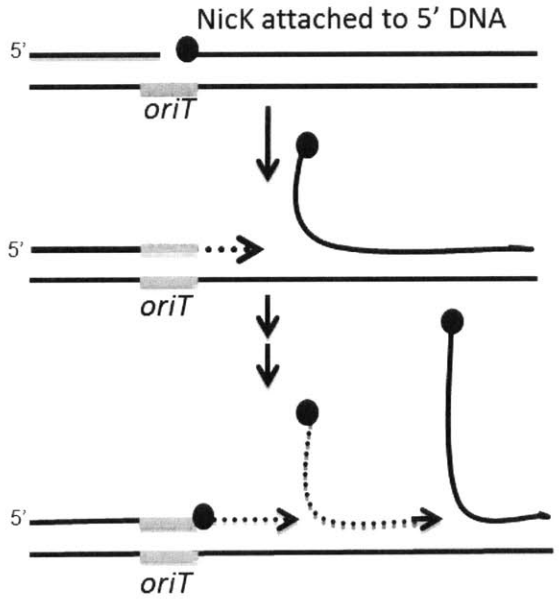



**Figure 5. Cartoon of repeated rolling-circle replication from the ICE*Bs1* *oriT* that is stuck in the chromosome.**
Rolling circle replication is induced in ICE*Bs1* insertions that are unable to excise from the chromosome. During this replication, the ICE*Bs1* relaxase NicK (black circles) nicks a site in *oriT*, the origin of transfer (gray bar) that also functions as an origin of replication (Lee et al., 2010; Lee and Grossman, 2007). NicK presumably becomes covalently attached to the 5' end of the nicked DNA. Replication extends (dotted line with arrow) from the free 3'-end, and regenerates a functional *oriT* that is a substrate for another molecule of NicK. The only other ICE*Bs1* product needed for ICE*Bs1* replication is the helicase processivity factor HelP (Thomas

et al., 2013). The rest of the replication machinery (not shown) is composed of host-encoded proteins.

**Induction of the SOS response in strains in which ICE*Bs1* is defective in excision**

We found that induction of ICE*Bs1* gene expression in the excision-defective insertions caused induction of the host SOS response. Like that in other organisms, the SOS response in *B. subtilis* results in increased expression of a large set of genes in response to DNA damage or replication stress (Goranov et al., 2005). We used a *lacZ* fusion to a damage-inducible gene, *dinC-lacZ* (Cheo et al., 1991; Ireton and Grossman, 1994), to monitor the SOS response in cells following induction of ICE*Bs1*. Without induction of ICE*Bs1* gene expression, there was no detectable ß-galactosidase activity above background levels, indicating that none of the insertions alone caused elevated SOS gene expression. In all of the excision-defective ICE*Bs1* strains analyzed (ICE*Bs1* Δ*attR* in *attB*, and insertions in *mmsA*, *yvbT*, *ykrP*, *srfAA*, and *yrkM*), there was a ≥3.5-fold increase in ß-galactosidase levels from the *dinC-lacZ* fusion 3 hrs after induction of ICE*Bs1* gene expression (Fig. 6). In contrast, there was no detectable increase in ß-galactosidase activity three hrs after induction of ICE*Bs1* gene expression in the excision-competent insertion in *attB* (Fig. 6). There was no apparent correlation between the amount of SOS induction and the severity of the viability defect. For example, one of the strains with the most severe viability defect (ICE*Bs1* in *ydcP*) had a relatively low amount of expression of *dinC-lacZ* (Fig. 6). However, the amount of SOS induction could be an underestimate since many cells in the population lose viability.

Induction of *dinC-lacZ* in the strains with ICE*Bs1* in secondary attachment sites was consistent with prior preliminary experiments using DNA microarrays that indicated induction of the SOS response in ICE*Bs1* *int* and *xis* mutants that are incapable of excision (N. Kavanaugh, C.

Lee, ADG, unpublished results). Based on these results, we conclude that induction of ICE*Bs1* gene expression in cells in which ICE*Bs1* is stuck in the chromosome causes DNA damage that induces the host SOS response. However, the SOS response per se is not what causes cell death.

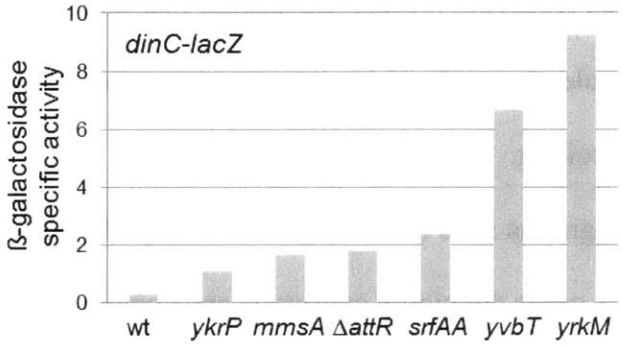



**Figure 6. Induction of the SOS response.**
The ß-galactosidase specific activity from the SOS transcriptional reporter fusion *dinC-lacZ* in strains with ICE*Bs1* in the indicated secondary attachment sites is presented. Strains were grown as described in Fig. 4 and samples for ß-galactosidase assays were taken 3 hours after induction of ICE*Bs1* gene expression. Data presented are the averages of at least two replicates. Strains used include: wt, *attB*::ICE*Bs1* (KM390); *ykrP::*ICE*Bs1* (KM402); *mmsA*::ICE*Bs1* (KM394); *attB*::ICE*Bs1* Δ*attR*::*tet* (KM392); *srfAA*::ICE*Bs1* (KM400); *yvbT*::ICE*Bs1* (KM396); *yrkM*::ICE*Bs1* (KM404).

## Discussion

We isolated and characterized insertions of the integrative and conjugative element ICE*Bs1* of *B. subtilis* into secondary integration (attachment or insertion) sites. We found that these insertions are detrimental for the propagation of ICE*Bs1* and detrimental to the proliferation and survival of the host cells. Below we discuss target site selection among ICEs, aspects of ICE*Bs1* biology that make insertions into secondary sites detrimental, and the more general implications for the evolution of ICEs.

### Target site selection and maintenance of tRNA genes as integration sites

We have identified 15 different secondary insertion sites for ICE*Bs1*. Some of these sites are similar to the primary attachment site, but some are quite different. Based on the diversity of

sites, and the isolation of only a single insertion in many of them, it is likely that we are nowhere near saturation for identifying all possible sites in non-essential regions. DNA sequence is clearly important in the potential function as an integration site. But, given the use of very dissimilar sites when more similar sites (such as *yrkM*) are present, we suspect that other factors also contribute. These could include possible roles for nucleoid binding proteins, other DNA binding proteins, transcription, and local supercoiling.

Many site-specific ICEs have preferred integration sites in tRNA genes. This preference is thought to occur, at least in part, because tRNA genes are highly conserved and contain inverted repeats that are typically used as integration targets for site-specific recombinases (Williams, 2002). We postulate that the selective pressure to maintain site-specific integration in a tRNA gene comes from a combination of: the conservation of tRNAs, the ability to efficiently excise from the primary attachment site, and the decreased cell viability and decreased ability of an ICE to spread when excision is reduced due to integration into a secondary site.

**Selective pressures against ICEs in secondary attachment sites**

Our results indicate that there are likely at least two main types of selective pressures against propagation of ICE*Bs1* that has inserted into a secondary integration site. First, there is probably pressure against the spread of that particular element due to both the defect in its ability to excise and also the lower detectable levels of circular ICE*Bs1* when it forms a heteroduplex. The excised circular form of an ICE is necessary for its complete transfer to a recipient cell. At least one other ICE has a reduced excision frequency from a secondary integration site. Excision of SXT from a secondary attachment site in *Vibrio cholerae* was reduced 3-4-fold relative to its ability to excise from the primary attachment site (Burrus and Waldor, 2003). In addition, lysogenic phage can also have reduced excision efficiencies from secondary attachment sites

(Shimada et al., 1972). Insertion of any type of mobile genetic element into a location from which it has trouble getting out will be deleterious to the further horizontal propagation of that element. Based on our results, this is particularly true for ICE*Bs1*.

In addition to the defect in ICE*Bs1* excision and transfer from secondary integration sites, there is a decrease in cell proliferation and viability following induction of ICE*Bs1* gene expression. ICE*Bs1* gene expression is normally induced under conditions of starvation or cell crowding when the activator RapI is expressed and active, or when the RecA-dependent SOS response is induced (Auchtung et al., 2007). Induction of ICE*Bs1* gene expression causes rolling circle replication from the ICE*Bs1* origin of transfer *oriT* (Lee et al., 2010; Thomas et al., 2013). Our results indicate that rolling circle replication from an element that is unable to efficiently excise from the chromosome causes a defect in cell proliferation and viability. This is likely due to chromosomal damage and stalling of the chromosomal replication forks when they reach the complex structure formed by repeated initiation of rolling circle replication from *oriT* in the chromosome (Fig. 5).

We suspect that autonomous replication is a common property of many ICEs but has not been generally observed because of the low frequency of induction and excision of most of these elements. There are indications that some other ICEs undergo autonomous replication (Carraro et al., 2011; Ramsay et al., 2006; Sitkiewicz et al., 2011; Wozniak and Waldor, 2010). If autonomous replication of ICEs is widespread, as we postulate (Lee et al., 2010; Thomas et al., 2013), then there should be selective pressure against proliferation and viability of cells in which an ICE is induced, replicates, and is unable to excise.

There were at least two different effects caused by replication of excision-defective elements. All caused a drop in CFU/mL of at least 10-fold, but some caused a severe drop, 100-1000-fold

in about 3 hrs. We do not know what causes this more severe drop in viability. It could be due to increased dosage of nearby genes or perhaps differential fragility of these chromosomal regions. In any case, the severe drop in viability provides even stronger selective pressure against propagation of the strains with insertions of ICE*Bs1* in these locations.

The proliferation and viability defect associated with the secondary insertions is most obvious when ICE*Bs1* gene expression is induced. Cells with ICE*Bs1* insertions in secondary attachment sites might be purged from the population under natural conditions of induction, providing selective pressure against maintenance of integrants in secondary sites and favoring site-specific integration.

We estimated the effects of insertions in secondary sites in populations without experimentally induced activation of ICE*Bs1*. The "spontaneous" activation and excision frequency of ICE*Bs1* in a population of cells is estimated to be approximately one cell in $10^4$ - $10^5$ (Auchtung et al., 2007; Lee et al., 2007; Smits and Grossman, 2010). Assuming a frequency of activation of ICE*Bs1* of ~$10^{-4}$ per generation, and that all activated cells with ICE*Bs1* in a secondary site die, we estimate that it would take ~23,000 generations for a population of cells with ICE*Bs1* in a secondary site to be 0.1 times the size of a population of cells with ICE*Bs1* in the primary site. The activation frequency increases under several conditions likely to be more relevant than growth in the lab, including: the presence of cells without ICE*Bs1*, entry into stationary phase, and during the SOS response (Auchtung et al., 2007; Auchtung et al., 2005; Lee et al., 2007). If activation of ICE*Bs1* actually occurs in 0.1% of cells, then it would take ~2,300 generations for the secondary site insertion population to be 0.1 times the population of cells with ICE*Bs1* in the primary site. These effects are difficult to measure experimentally, but easy to see when ICE*Bs1* is efficiently induced.

### ICEs with single versus multiple integration sites

ICEs of the Tn*916*/Tn*1545* family can integrate into multiple sites in many organisms, yet they are not known to cause a defect in cell proliferation or viability when gene expression is induced. Tn*916* and most family members contain *tetM*, a gene encoding resistance to tetracycline. Expression of *tetM* and Tn*916* genes is induced in the presence of tetracycline (Celli and Trieu-Cuot, 1998). Tn*916* has two *helP* homologues and we predict that it undergoes autonomous rolling circle replication (Thomas et al., 2013). Despite relatively low excision frequencies, tetracycline-induced Tn*916* gene expression is not known to cause a defect in cell proliferation or viability. Speculatively, this is most likely because the two *helP* homologues (*orf22* and *orf23*) and the Tn*916* relaxase (*orf20*) (and the conjugation genes) are not expressed until Tn*916* excises and circularizes. Therefore, the regulation of Tn*916* gene expression specifically prevents expression of replication functions until after excision. Consequently, rolling-circle replication of Tn*916* cannot occur while the element is integrated in the chromosome. We speculate that some of the evolutionary pressures to establish and maintain a high degree of site specificity is lost when expression of ICE replication functions does not occur until after excision from the host genome.

## Materials and Methods

### Media and growth conditions

*Bacillus subtilis* was grown at 37°C in LB or defined $S7_{50}$ minimal medium (Vasantha and Freese, 1980) with arabinose (1%) as carbon source. Antibiotics and other chemicals were used at the following concentrations: Isopropyl β-D-1-thiogalactopyranoside or IPTG (1 mM), chloramphenicol (*cat*, 5 μg/ml), kanamycin (*kan*, 5 μg/ml), spectinomycin (*spc*, 100 μg/ml),

erythromycin (0.5 μg/ml) and lincomycin (12.5 μg/ml) together, to select for macrolide-lincosamide-streptogramin B resistance (*mls* or *erm*).

***Bacillus subtilis* strains and alleles**

*B. subtilis* strains used are listed in Table 2. All except BTS14 are derived from AG174 (JH642) and contain mutations in *trpC* and *pheA* (not shown). Most of the strains were constructed using natural transformation or conjugation, as described below. Many alleles were previously described. *dinC18*::Tn*917lac* is an insertion in the damage-inducible gene *dinC* and creates a transcriptional fusion to *lacZ* (Cheo et al., 1991). Most ICE*Bs1* strains contained a kanamycin-resistance cassette {Δ(*rapI-phrI*)*342*::*kan*} (Auchtung et al., 2005). ICE*Bs1* was induced by overexpression of *rapI* from a xylose-inducible promoter using *amyE*::{(Pxyl-*rapI*), *spc*} (Berkmen et al., 2010) or from an IPTG-inducible promoter using *amyE*::{(Pspank(hy)-*rapI*), *spc*} (Auchtung et al., 2005). Δ*attR100*::*tet* deletes 216 bp spanning the junction between the right end of ICE*Bs1* and the chromosome (Lee et al., 2007). Δ*helP155* is an unmarked 413-bp deletion that removes the entire coding sequence and the 35 bp *helP-ydcQ* intergenic region (Fig. 1F) (Thomas et al., 2013).

*ΔattB mutant with a compensatory mutation in trnS-leu1.* Δ*attB*::*cat* is a deletion-insertion that is missing ICE*Bs1* and removes 185 bp that normally contains the primary chromosomal ICE*Bs1* attachment site, resulting in the loss of a functional *trnS-leu2* (Lee et al., 2007). Although *trnS-leu2* is non-essential (Garrity and Zahler, 1994; Lee et al., 2007), cells with Δ*attB* do not grow as well as wild type. To improve the growth of Δ*attB*::*cat*, we used a compensatory mutation in *trnS-leu1* that changes the anti-codon to that normally found in *trnS-leu2* (C. Lee, & ADG), analogous to the *leuF1* mutation previously described (Garrity and Zahler, 1994). The compensatory mutation was constructed by site-directed mutagenesis using the overlap-

extension PCR method (Ho et al., 1989). Because *trnS-leu1* and Δ*attB*::*cat* are genetically linked, we selected for chloramphenicol resistant colonies and screened for the single bp mutation in *trnS-leu1* by sequencing. In addition to the mutant *trnS-leu1* allele (*trnS-leu1-522*), the strain had an additional mutation, (5'-CAAAAAAACTAA**A** to 5'-CAAAAAAACTAA**G**) in the non-coding region between Δ*attB*::*cat* and *yddN*. Growth of the resulting strain, CAL522, was indistinguishable from that of wild type.

*Deletion of* nicK *and downstream genes*. We constructed two large deletion-insertion mutations in ICE*Bs1*, one removing *nicK* and all downstream genes, Δ(*nicK-yddM*)::*cat*, and the other leaving *nicK* intact, but removing the downstream genes, Δ(*ydcS-yddM*)::*cat*. Both deletions leave the ends of ICE*Bs1* intact (Fig. 1D, E), have *cat* (chloramphenicol resistance) from pGEMcat (Youngman et al., 1989), and were constructed using long-flanking homology PCR (Wach, 1996). The Δ(*nicK-yddM*)::*cat* allele contains the first 127 bp in the 5' end of *nicK*. The Δ(*ydcS-yddM*)::*cat* allele contains the first 29 bp in the 5' end of *ydcS*. Both deletions extend through the first 170 bp in *yddM* (Fig. 1). The alleles were first transformed into wild type strain AG174. Chromosomal DNA was then used to transfer the alleles into other strains, including KM70 (*mmsA*::ICE*Bs1*), KM94 (*yvbT*::ICE*Bs1*), KM77 (*ykrP*::ICE*Bs1*), KM141 (*srfAA*::ICE*Bs1*), and CAL874 (ICE*Bs1* at *attB*). In all cases, the incoming deletion associated with *cat* replaced the Δ(*rapI-phrI*)*342*::*kan* allele present in ICE*Bs1* in the recipient.

*Deletion of* yvbT *in* mmsA::*ICE*Bs1. We constructed a deletion-insertion that removes the 19 bases before *yvbT* and the first 808 bp of *yvbT*, leaving the last 200 bp intact. The sequence from *yvbT* was replaced with *cat*, from pGEMcat (Youngman et al., 1989), using long-flanking homology PCR (Wach, 1996). The insertion-deletion was verified by PCR and the mutation was introduced into strain KM70 (*mmsA*::ICE*Bs1*) by transformation.

**Isolation and identification of secondary ICE*Bs1* integration sites**

*Mating ICE*Bs1 *into ΔattB recipient.* Mating assays were performed essentially as described (Auchtung et al., 2005; Lee et al., 2007). Excision of a kanamycin resistant ICE*Bs1* (ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) was induced in the donor cells by overproduction of RapI from Pspank(hy)-*rapI*. Donors (resistant to kanamycin and spectinomycin) were mixed with an approximately equal number of recipients (resistant to chloramphenicol) and filtered on sterile cellulose nitrate membrane filters (0.2 μm pore size). Filters were cut into 8 pieces (so that transconjugants were independent isolates), placed on Petri plates containing LB and 1.5% agar, and incubated at 37°C for 3 hours. Cells from each piece of filter were streaked for independent transconjugants by selecting for the antibiotic resistance conferred by the incoming ICE*Bs1* (kanamycin) and the resistance unique to the recipient (chloramphenicol). The recipient used in this report {Δ*attB*::*cat trnS-leu1-522*} is different from the recipient {Δ*attB*::*cat*} used previously (Lee et al., 2007). The *trnS-leu1-522* confers normal growth to the Δ*attB* (Δ*trnS-leu2*) mutant (see above).

*Inverse PCR to identify the site of insertion of independent transonjugants.* We used inverse PCR to amplify the junction between the chromosome and the right (*yddM*) end of ICE*Bs1* integrated into various secondary sites. Chromosomal DNA was digested with HindIII and approximately 50 ng was ligated in a 100 μl reaction to favor circularization of DNA fragments. One-fourth of the ligation reaction was used in inverse PCR with either of two primer pairs (CLO17-CL058 or CLO50-oJMA97) designed to amplify the ICE*Bs1* and chromosomal sequences flanking *yddM*. PCR products were sequenced with primers CLO17, CLO50, oJMA207, and CLO114 (sequences available upon request). Comparison to the *B.subtilis* genome sequence indicated where ICE*Bs1* had integrated.

*Backcross of ICE*Bs1 *insertions*. Seven of the 15 different insertions of ICE*Bs1* in secondary attachment sites were initially chosen for further study. These were first backcrossed into a strain cured of ICE*Bs1* (JMA222). Pxyl-*rapI* (*amyE*::{(Pxyl-*rap*) *spc*}) was introduced into these strains by transformation and selection for spectinomycin resistance using chromosomal DNA from strain MMB869. We verified that ICE*Bs1* was still at the original secondary attachment site using PCR with site-specific primers. The final strains from these crosses include: KM70 (*mmsA*::ICE*Bs1*), KM94 (*yvbT*::ICE*Bs1*), KM72 (*yrkM*::ICE*Bs1*), KM77 (*ykrP*::ICE*Bs1*), KM130 (*spoVD*::ICE*Bs1*), KM141 (*srfAA*::ICE*Bs1*), and KM132 (*yycJ*::ICE*Bs1*).

**Assays for excision of ICE*Bs1***

Excision of ICE*Bs1* from a chromosomal attachment site creates an extrachromosomal ICE*Bs1* circle and an "empty" attachment site (also called "repaired chromosomal junction"). Each product was measured using specific primers for quantitative real time PCR (qPCR), using a LightCycler 480 Real-Time PCR system with Syber Green detection reagents (Roche), essentially as described (Lee et al., 2010). Cells were grown in defined minimal medium with arabinose as carbon source. Products from excision were determined two hours after addition of xylose to induce expression of Pxyl-*rapI* to cause induction of ICE*Bs1* gene expression.

The amount of each empty attachment site was compared to a chromosomal reference, *cotF* measured with primers CLO257-CLO258. The amount of empty attachment site from each of the secondary sites was normalized to strain JMA222, an ICE*Bs1*-cured strain that simulates 100% excision. Standard curves for qPCR with *cotF* and the repaired junction for each secondary insertion were generated using genomic DNA from JMA222. Primers (in parentheses) for empty secondary attachment sites were specific for: *yrkM* (CLO117-ABO17),

*mmsA* (CLO109-ABO18), *yycJ* (KM18-KM19), *srfAA* (KM22-KM23), *spoVD* (KM20-KM21), *yvbT* (ABO14-ABO15), *ykrP* (KM14-KM16), and *attB* (CLO261-CLO262).

The amount of ICE*Bs1* circle that forms after excision from the chromosome was measured with primers AB019/CLO114. The amount of excised circle was compared to the chromosomal reference *cotF* (primers CLO257-CLO258), and the amount of excised circle from each of the secondary sites was normalized to that from *attB* (strain CAL874). Standard curves for qPCR for *cotF* and the excised circle were generated using genomic DNA from RapI-induced CAL874. Primer sequences are available upon request.

**Cell growth assays**

Strains were grown in defined minimal medium with arabinose and expression of Pxyl-*rapI* was induced with 1% xylose at OD600 of 0.05. The number of colony forming units (CFU) was determined 3 hours after addition of xylose. For each strain, the number of CFU/ml 3 hrs after expression of Pxyl-*rapI* was compared to the number of CFU/ml without expression of Pxyl-*rapI*. All experiments were done at least twice.

**ß-galactosidase assays**

Cells were grown and treated as described for viability assays. Samples were taken 3 hours after induction of Pxyl-*rapI*. All experiments were done at least twice. ß-galactosidase assays were done essentially as described (Jaacks et al., 1989; Miller, 1972). Specific activity is expressed as the (ΔA420 per min per ml of culture per OD600 unit) x 1000.

**Acknowledgments**

We thank J. Wilks for preliminary isolation of some of the insertions in secondary attachment sites, C. Lee for helpful discussions and construction of the compensatory mutation

in *trnS-leu1* in the Δ*attB* (Δ*trnS-leu2*) strain, C. Bonilla and A. Babic for preliminary characterization of some of the secondary-site insertions, and N. Kavanaugh for preliminary analysis of mRNA levels in ICE*Bs1* mutants defective in excision. We thank C. Johnson, C. Lee, J. Thomas, and T. Washington for comments on the manuscript.

## Tables

**Table 1. Summary of properties of several ICE*Bs1* insertions in secondary attachment sites.**

| Insertion site (#mm)[a] | Excision frequency[b] | Viability[c] | *dinC-lacZ*[d] |
|---|---|---|---|
| *attB* | 1.0 | 1.0 | 1.0 |
| *yrkM* (2) | 0.06 | 0.030 | 34 |
| *mmsA* (3) | 0.15 | 0.14 | 6.1 |
| *srfAA* (3) | 0.04 | 0.10 | 8.7 |
| *yycJ* (7) | 0.12 | 0.073 | N.D. |
| *spoVD* (8) | $<10^{-4}$ | 0.010 | N.D. |
| *ykrP* (12) | $<10^{-4}$ | 0.040 | 4.1 |
| *yvbT* (11) | $<10^{-4}$ | 0.0038 | 24 |
| Δ*attR* | $<10^{-4}$ | 0.092 | 6.7 |

[a] site of insertion of ICE*Bs1*; #mm indicates the number of mismatches between the insertion site and *attB* (illustrated in Fig. 2).

[b] excision frequency measured as the empty attachment site 2 hrs after induction of ICE*Bs1* gene expression; normalized to wt; same data as in Fig. 3, except that here data is normalized to wt (*attB*). Excision frequency from *attB* was 1.

[c] cell viability normalized to ICE*Bs1* at *attB*; same data as in Fig. 4. Viability of ICE*Bs1* at *attB* was 0.9 of uninduced.

[d] expression of damage inducible gene *dinC-lacZ*, normalized to that of cells with ICE*Bs1* in *attB*; data from Fig. 6. ß-galactosidase specific activity of ICE*Bs1* at *attB* was 0.3.

**Table 2. *B. subtilis* strains used.**

| Strain | relevant genotype (comment and/or reference) |
|---|---|
| AG174 | *phe trp* (Perego et al., 1988) |
| AG1624 | *zbj-82*::Tn*917* (insertion at 65°) (Vandeyar and Zahler, 1986) |
| BTS13 | PY79 (*trp*$^+$ *phe*$^+$) Δ*mutSL*::*spc* (Smith et al., 2001) |
| CAL522 | *trnS-leu1-522* Δ*attB*::*cat* |
| CAL572 | *yomR572*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) Δ*attB*::*cat comK*::*cat*::*spc* (Lee et al., 2007) |
| CAL575 | *yvbT575*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) Δ*attB*::*cat comK*::*cat*::*spc* (Lee et al., 2007) |
| CAL576 | *yqhG576*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) Δ*attB*::*cat comK*::*cat*::*spc* (Lee et al., 2007) |
| CAL577 | *yobJ577*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) Δ*attB*::*cat comK*::*cat*::*spc* (Lee et al., 2007) |
| CAL578 | Intergenic *ygxA rrnD-16S-578*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) Δ*attB*::*cat comK*::*cat*::*spc* (Lee et al., 2007) |
| CAL872 | Δ*attR100*::*tet* Δ(*rapI-phrI*)*342*::*kan amyE*::{(Pxyl-*rapI*) *spc*} |
| CAL874 | Δ(*rapI-phrI*)*342*::*kan amyE*::{(Pxyl-*rapI*) *spc*} (Lee et al., 2010) |
| JMA168 | Δ(*rapI-phrI*)*342*::*kan amyE*::{(*Pspank(hy)-rapI*) *spc*} (Lee and Grossman, 2007) |
| J3 | *srfAA3*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *trnS-leu1-522* Δ*attB*::*cat* |
| J4 | *yycJ4*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *trnS-leu1-522* Δ*attB*::*cat* |
| J9 | *yrkM9*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *trnS-leu1-522* Δ*attB*::*cat* |
| J11 | *yqhG11*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *trnS-leu1-522* Δ*attB*::*cat* |
| J12 | *yisQ12*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *trnS-leu1-522* Δ*attB*::*cat* |
| J14 | *mmsA14*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *trnS-leu1-522* Δ*attB*::*cat* |
| J16 | *ykrP16*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *trnS-leu1-522* Δ*attB*::*cat* |
| JMA222 | ICE*Bs1*$^0$/cured of ICE*Bs1* (Auchtung et al., 2005) |
| KI1254 | *dinC18*::Tn*917lac*; allele originally from YB5018 (Cheo et al., 1991) |
| KM5 | *yghL5*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *trnS-leu1-522* Δ*attB*::*cat* |
| KM8 | *spoVD8*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *trnS-leu1-522* Δ*attB*::*cat* |
| KM10 | *ydbJ10*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *trnS-leu1-522* Δ*attB*::*cat* |
| KM70 | *mmsA15*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM72 | *yrkM9*(*J9*)::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM77 | *ykrP16*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM94 | *yvbT575*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM110 | ICE*Bs1*$^0$ *zbj*-82::Tn*917* |
| KM111 | *trnS-leu1-522* Δ*attB*::*cat zbj*-82::Tn*917* |
| KM130 | *spoVD8*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM132 | *yycJ4*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM141 | *srfAA3*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM252 | *mmsA15*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *cat*} |
| KM268 | *mmsA15*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *cat*} Δ*mutSL*::*spc* |
| KM304 | *mmsA15*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} Δ*yvbT*::*cat* |
| KM358 | *mmsA15*::(ICE*Bs1* Δ(*ydcS-yddM*)*356*::*cat*) *amyE*::{(Pxyl-*rapI*) *spc*} |

| | |
|---|---|
| KM362 | *yvbT575*::(ICE*Bs1* Δ(*ydcS-yddM*)*356*::*cat*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM366 | *mmsA15*::(ICE*Bs1* Δ(*nicK-yddM*)*354*::*cat*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM370 | *yvbT575*::(ICE*Bs1* Δ(*nicK-yddM*)*354*::*cat*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM384 | *srfAA3*::(ICE*Bs1* Δ(*ydcS-yddM*)*356*::*cat*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM386 | *srfAA3*::(ICE*Bs1* Δ(*nicK-yddM*)*354*::*cat*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM388 | *ykrP16*::(ICE*Bs1* Δ(*ydcS-yddM*)*356*::*cat*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM389 | *ykrP16*::(ICE*Bs1* Δ(*nicK-yddM*)*354*::*cat*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM390 | Δ(*rapI-phrI*)*342*::*kan amyE*::{(Pxyl-*rapI*) *spc*} *dinC18*::Tn*917*(*lacZ mls*) |
| KM392 | Δ*attR100*::*tet* Δ(*rapI-phrI*)*342*::*kan amyE*::{(Pxyl-*rapI*) *spc*} *dinC18*::Tn*917lac* |
| KM394 | *mmsA15*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} *dinC18*::Tn*917lac* |
| KM396 | *yvbT575*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} *dinC18*::Tn*917lac* |
| KM400 | *srfAA3*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} *dinC18*::Tn*917lac* |
| KM402 | *ykrP16*::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} *dinC18*::Tn*917lac* |
| KM404 | *yrkM9*(*J9*)::(ICE*Bs1* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} *dinC18*::Tn*917lac* |
| KM429 | *ykrP16*::(ICE*Bs1* Δ*helP* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| KM437 | Δ*attR100*::*tet* Δ*helP* Δ(*rapI-phrI*)*342*::*kan amyE*::{(Pxyl-*rapI*) *spc*} |
| KM459 | *yvbT575*::(ICE*Bs1* Δ*helP* Δ(*rapI-phrI*)*342*::*kan*) *amyE*::{(Pxyl-*rapI*) *spc*} |
| MMB868 | *amyE*::{(Pxyl-*rapI*) *cat*} |
| MMB869 | *amyE*::{(Pxyl-*rapI*) *spc*} (Smits and Grossman, 2010) |

# Appendix A

# ICE*Bs1* integrates into secondary sites at a low frequency when primary site present

**Background**: 15 secondary attachment sites for ICE*Bs1* were isolated by mating ICE*Bs1* into a recipient strain deleted for the primary attachment site (*attB*) (Chapter 2). One of these secondary attachment sites (*yrkM*) is strongly preferred over the others and was isolated 41% of the time. I was curious to know whether or not ICE*Bs1* integrates into these secondary attachment sites (particularly *yrkM*) under less artificial conditions, when the primary site is present.

**Result**: I found that ICE*Bs1* integrates into *yrkM* (the most frequently isolated secondary attachment site) at a frequency of ~$10^{-4}$ when the primary site is present in the *B.subtilis* genome. I did this by isolating a large pool of transconjugants and using qPCR primers to the junction formed after integration of ICE*Bs1* into *yrkM*. Additionally, as expected, ICE*Bs1* integrates at *attB* the vast majority of the time.

**Discussion:** While secondary attachment sites have been identified for other site-specific ICEs (by mating into a recipient where the *attB* site is deleted), this is the first report detailing the "natural" frequency of secondary sites utilization. I speculate that secondary sites of other site-specific ICEs are also utilized at a low frequency. Given that mating in the environment is considered to be quite frequent, perhaps even more frequent than in the laboratory (Aminov, 2011), it seems likely that integration into secondary attachment sites does occur readily in the environment. (This appendix is discussed further in Chapter 3).

**Method**: By mating ICE*Bs1* from a donor (KM250), where ICE*Bs1* is located at the primary attachment site (*attB*), with a recipient with an empty *attB* site (KM524), I collected a large pool of transconjugants (~50-130 million independent transconjugants) by scraping cells from petri plates with a glass pipette. (I followed the mating protocol outlined in Chapter 2 and did 4 matings, each with 2.5mL cells at OD of ~1 of both donor and recipient. The 4 filters, one from

each mating, were diluted in 10mL spizizen salts and 200µL placed on each plate.) Experiment was done in duplicate (aka. two pools were isolated), with similar results each time.

From this pool, I isolated DNA and performed qPCR using primers for the junction of *yrkM*:ICE*Bs1* (primers CLO116/KM76). Values were compared to a reference gene cotF. Also values were normalized to a strain (KM72), which simulates 100% integration at *yrkM*. While values for *yrkM*:ICE*Bs1* were low, they were in the linear range of the primers and higher than the negative control (JMA222, which is cured of ICE*Bs1*) by ~3-fold or greater. DNA used for standard curves was from strain KM72, where ICE*Bs1* is integrated at *yrkM*. Since the actual value is relatively close to the limit of detection, I did additional controls where known amounts of *yrkM* (in the range of $10^{-3}$ to $10^{-5}$) were mixed with a majority of DNA where ICE*Bs1* is integrated at *attB*. These controls verified that the qPCR assay is approximately accurate (within an order of magnitude).

To determine the frequency of integration at *attB,* in the transconjugant pool, I followed the same protocol as for *yrkM*. I used primers CLO273/264 and normalized to either strain AG174 or CAL874. DNA used for standard curves was either AG174 or CAL874.

Genotypes for strains JMA222, CAL874, KM72, and AG174 are listed in the strain table for Chapter 2. Genotype for strain KM250 is: Δ(*rapI-phrI*)342::*kan amyE*::{(Pxyl-*rapI*) *cam*}. KM250 was constructed by transforming strain IRN342 with DNA from strain MMB868. Genotype for strain KM524 is: ICE*Bs1*°/cured of ICE*Bs1 amyE*::promotorless *lacZ MLS*. Strain KM524 was constructed by transforming *amyE*::promotorless *lacZ MLS* from KM518 into JMA222 (cured of ICE*Bs1*). KM518 was contructed by transforming AG174 with plasmid pKS5 (in strain AG1001). Before transformation, plasmid pKS5 was cut with restriction enzyme XbaI.

# Appendix B:

# ICE*Bs1* returns to *attB* following excision and conjugation from secondary sites

When ICE*Bs1* is mated from a secondary site (donor either ICE*Bs1* at *yrkM*–KM72, *mmsA*–*KM70*, *yycJ*–*KM132*, or *srfAA*–*KM141*), it reintegrated into the *attB* site ≥90% of the time (recipient KM110). ~10 Independent transconjugants for each mating were isolated and purified and their DNA analyzed by PCR using primers to the junction of ICE and the chromosome. For each donor strain, either all or nearly all (90%) returned to *attB* (Table 1). In the two cases where ICE*Bs1* did not return to *attB*, it did not integrate into the site from which it excised. (Genotypes of all strains located in Table 2, Chapter 2).

Similarly, when an *attB* deletion recipient (KM111) is used, ICE*Bs1*'s site-specificity does not appear to be altered when mating from either *yrkM*, *mmsA*, or *srfAA* (*yycJ* not determined). ICE*Bs1* returned to *yrkM* at relatively high frequencies (5/18 or 27%), similar to when ICE*Bs1* is mated from *attB* (11/27 or 41%). The difference in the two efficiencies is likely due to the small sample sizes and probably does not reflect a difference in site-specificity; however, I cannot rule out a minor difference in frequency. Additionally, when mated from a secondary site into the *attB* deletion strain, ICE*Bs1* did not return to the site from whence it came from, except for the one instance where ICE*Bs1* returned to *yrkM* when coming from *yrkM*.

All in all, ICE*Bs1*'s site-specificity does not appear to be altered when mated from secondary sites versus *attB*. While I only looked at small sample sizes, there is no initial indication of any differences. This result is not surprising. When mated from sites that do not a form a heteroduplex (*yrkM* and *srfAA*), the circle product should be identical to the circle product formed from excision from *attB,* making it unlikely that site specificity would be altered. When mated from sites that form a heteroduplex after excision (*mmsA* and *yycJ*), the circle is now a heteroduplex and one strand of the crossover region has been altered, making it more likely that the product would have altered site specificity of integration. However, work with phage lambda

indicates that development of integrases with altered site-specificities requires more than simply an altered *attP* site and includes other changes, such as mutations to the integrase (Rutkai et al., 2003; Rutkai et al., 2006). This appears to be the case for ICE*Bs1*.

**Table 1. ICE*Bs1* returns to *attB* following excision and conjugation from secondary sites.**

| secondary site | *attB*+ recipient | Δ*attB* recipient |
|---|---|---|
| *yrkM* | 9/10 *attB*, none yycJ | 1/6 *yrkM,* no other sites tested |
| *mmsA* | 9/9 *attB* | 1/10 *yrkM*; none *mmsA* |
| *yycJ* | 9/10 *attB*, none *yycJ* | N/A |
| *srfAA* | 10/10 *attB* | 3/4 *yrkM*; none srfAA |

Independent transconjugants, from donors with ICE*Bs1* at different secondary integration sites, were analyzed for the location of ICE*Bs1* integration using primers to the junction of ICE and the chromosome. These primers were used: CLO116/CLO17 or oJMA141/CLO17 (*yrkM:*ICE*Bs1* junction), CLO109/oJMA141 (*mmsA*:ICE*Bs1*), CLO17/KM4 (*yycJ*:ICE*Bs1*), oJMA141/KM5 (*srfAA*:ICE*Bs1*), and CLO17/oJMA100 (*trnS-leu2*:ICE*Bs1*), and oJMA102/oJMA227 (ICE internal).

# Appendix C:

# ICE*Bs1* prefers exogenous *attB* over endogenous *attB*

Done in collaboration with Doug Cattie, a rotation student

ICE*Bs1* has one primary attachment site (*attB*) in a leucine tRNA gene (*trnS-leu2*). I was curious to know where ICE*Bs1* would integrate if another, identical attachment site was placed in the *B.subtilis* chromosome. To answer this question, a second, exogenous *attB* was placed at *thrC*; it contains the 17 bp *attB* plus ~1 kb surrounding *attB* (~200 kb on one side and ~800 kb on the other). Surprisingly, with two identical *attB*s in the chromosome, ICE*Bs1* prefers to integrate at the exogenous *attB* at *thrC* over the endogenous *attB* at *leu2*. ICE*Bs1* integrated at *leu2* only 27% of the time (average of three replicates, 22%, 31%, 27%) or ~ one-third of how frequently ICE*Bs1* integrates at *thrC*. This result strongly suggests that factors in addition to sequence specificity for *attB* play a role in target site selection.

**Method**: Mating assays were carried out as described in the Methods Section of Chapter 2. Transconjugants were scraped from the plate, DNA was isolated, and qPCRs (to each junction) performed. Experiment was performed in triplicate.

Primers used for the *leu2*:ICE*Bs1* junction are AB022/oKM34. Standard curve was DNA from strain AG174. Primers used for the junction *thrC*:ICE*Bs1* are ABO22/oKM35. Primers are able to distinguish between integration at *thrC* and *leu2* by one primer in the set being specific for DNA outside the ~1kb of similarity and one primer being specific for ICE*Bs1*. Standard curve was DNA from strain KM324. Known amounts of DNA were used for each standard curve. % integration at *leu2*:ICE*Bs1* determined by taking ratio of *leu2*:ICE*Bs1* to total amount of integrated ICE*Bs1* DNA. Total amount of integrated ICE*Bs1* DNA determined by adding together amounts for both *thrC*:ICE*Bs1* and *leu2*:ICE*Bs1*. To verify the method (especially since a reference gene control was not used), controls with known amounts of ICE*Bs1* at l*eu2* and *thrC* were mixed together at various ratios. The controls corresponded to actual values quite well

(88% for 9:1 *leu2*:ICE*Bs1* to *thrC*:ICE*Bs1*, 62%: for 2:1, 52% for 1:1, 36% for 1:2, and 10% for 1:9).

Donor strain was CAL874; its genotype is Δ(*rapI-phrI*)342::*kan amyE*::{(Pxyl-*rapI*) *spc*} (Lee et al., 2010). Recipient strain was KM306; its genotype is ICE*Bs1*° *thrC*213::*attB*-117 *mlsR*. Strain 306 was constructed by transformation of JMA222, ICE*Bs1* cured, (Auchtung et al., 2005) with chromosomal DNA from strain CAL213, *thrC*213::*attB*-117 *mlsR amyE*::Pspank *sacV specR*.

Genotype for strain KM324 (used for the standard curve) is *trnS-leu1*-522 Δ*attB*::*cat* *thrC*213::(ICE*Bs1* Δ(*rapI-phrI*)342::*kan*) *mlsR amyE*::{(Pxyl-*rapI*) *spc*}. Strain KM324 was constructed by mating donor CAL874 with recipient KM316, *trnS-leu1*-522 Δ*attB*::*cat* *thrC*213::*attB*-117 *mlsR amyE*::{(Pxyl-*rapI*) *spc*}. KM316 was constructed by transformation of strain KM308, *trnS-leu1*-522 Δ*attB*::*cat thrC*213::*attB*-117 *mlsR,* with chromosomal DNA from strain MMB869, *amyE*::{(Pxyl-*rapI*) *spc*} (Smits and Grossman, 2010). KM308 was constructed by transformation of CAL522, *trnS-leu1*-522 Δ*attB*::*cat,* (Table 2, Chapter 2) with chromosomal DNA from strain CAL213, *thrC*213::*attB*-117 *mlsR amyE*::Pspank *sacV specR*.

# Appendix D

# Mating efficiency depends on Δ*attB* recipient

Two different *attB* deletion recipients were used to isolate the 15 secondary attachment sites and, for reasons unknown, these two recipients (REM54 and CAL522) have significantly different mating efficiencies (appox. 40-fold different). In a direct comparison, mating efficiency for REM54 (Δ*attB*) was 4.25*10^-3 and for CAL522 (Δ*attB* compensatory) was 1.04*10^-4. While a direct comparison was only carried out once, similar mating efficiencies for both strains were obtained previously either by Lee (Lee et al., 2007) or me. Despite having different mating efficiencies, ICE*Bs1* appears to integrate into the same locations in either recipient since some of the same secondary attachment sites (*yrkM* and *mmsA*) were identified using both recipients.

Both recipients contain the same deletion (Δ*attB:cat*), which removes the primary attachment of ICE*Bs1* but also results in loss of a functional tRNA gene (*trnS-leu2*). Although *trnS-leu2* is non-essential (Garrity and Zahler, 1994; Lee et al., 2007), cells with Δ*attB* do not grow as well as wild-type. To improve growth of Δ*attB:cat*, C. Lee made a compensatory mutation in *trnS-leu1* that changes the anticodon to that normally found in *trnS-leu2* (C. Lee and ADG), analogous to the *leuF1* mutation previously described (Garrity and Zahler, 1994). In addition to the mutant *trnS-leu1* allele, the strain had an additional one base pair mutation in the non-coding region between Δ*attB:cat* and *yddN*. Growth of the resulting strain, CAL522, was indistinguishable from that of wild-type. Genotypes and further information on the construction of REM54, CAL522, and CAL874 (donor) can be found either in the strain table for Chapter 2 or Lee et al. (Lee et al., 2007). Mating assays were carried out as described in Chapter 2.

The difference in the mating efficiencies of these two strains is likely due one of 3 factors: 1) the compensatory mutation in *trnS-leu1*, 2) the additional mutation between Δ*attB:cat* and *yddN*, or 3) an unlinked mutation that entered this strain. It would be interesting to determine what is

accounting for this difference in mating efficiency as it could possibly provide insight into the integration/mating process or the role of tRNAs in integration.

# Appendix E

# Deletion of ICE*Bs1* integrase results in a worse viability defect than expected

**Background**: When ICE*Bs1* cannot excise from the chromosome, induction of ICE*Bs1* gene expression results in a proliferation/viability defect for the cell (Chapter 2). This proliferation/viability defect is due to ICE*Bs1* rolling-circle replication when ICE*Bs1* is stuck in the chromosome. ICE*Bs1* gets stuck in the chromosome when it is integrated into secondary attachment sites, but also when ICE*Bs1* is deleted for regions necessary for excision, such as the left and right attachment sites of ICE*Bs1* (*attL* and *attR*, respectively) or the ICE*Bs1* recombinase (*int*). I expected deletion of the ICE*Bs1* recombinase (*int*) to result in a similar proliferation/viability defect (~10% CFU/mL of induced cells compared to uninduced) as either deletion of *attR* or integration of ICE*Bs1* into the secondary attachment sites *mmsA* or *srfAA* (which have a more moderate growth defect than some of the secondary attachment sites) (Table 1, Chapter 2).

**Results**: Deletion of the ICE*Bs1* integrase (int) resulted in 3% induced CFU/mL compared to uninduced), 3.3 fold less than expected. Interestingly, deletion of several genes in the middle of ICE*Bs1*, Δ(*ydcS-yddM*), resulted in the originally expected phenotype of ~10% induced CFU/mL compared to uninduced. Also, as expected, deletion of the nickase (a gene necessary for ICE*Bs1* replication) in addition to the other deletions (*int* and *ydcS-yddM*) resulted in no viability defect (~100% cells still viable after induction).

**Discussion**: One explanation for these results is that the ICE*Bs1* integrase (Int) is mitigating the detrimental effects of a gene(s) in the middle of ICE*Bs1*, somewhere between *ydcS* and *yddM*. It seems plausible that the ICE*Bs1* integrase is acting as a repressor of ICE*Bs1* genes in that region (*ydcS-yddM*) and that when the integrase is deleted, expression of genes between *ydcS* and *yddM* incur an additional viability defect. In fact, integrases of other mobile elements have been shown to have transcriptional activity as a repressor (Kaufman and Rio, 1991;

MacRae, 2002). It would interesting to determine which gene(s) in the middle of ICE*Bs1* is causing the viability defect and whether or not Int is indeed a repressor of ICE*Bs1* gene expression.

**Method**: Viability (CFU/mL) assays were done as described in the Methods Section of Chapter 2. Experiments were only performed once.

Genotype for strain KM376 is: Δ(*rapI-phrI*)342::*kan amyE*::{(Pxyl-*rapI*) *spc*} Δ*int*::*cat*::*tet*. Strain KM376 was constructed by transformation of strain CAL874, Δ(*rapI-phrI*)342::*kan amyE*::{(Pxyl-*rapI*) *spc*}, with chromosomal DNA from strain JMA230, Δ*int*::*cat*::*tet*. Construction of strain JMA230 explained in Auchtung et al., 2007 and construction of strain CAL874 explained in Lee et al., 2010 (Auchtung et al., 2007).

Genotype for strain KM378 is: Δ(*ydcS-yddM*)356::*cat amyE*::{(Pxyl-*rapI*) *spc*} Δ*int*::*cat*::*tet*. KM378 was constructed by transformation of strain KM360 Δ(*ydcS-yddM*)356::*cat amyE*::{(Pxyl-*rapI*) *spc*} with DNA from strain JMA230, Δ*int*::*cat*::*tet*. Strain KM360 was constructed by transformation of CAL874, Δ(*rapI-phrI*)342::*kan amyE*::{(Pxyl-*rapI*) *spc*} with chromosomal DNA from strain KM356, Δ(*ydcS-yddM*)356::*cat*. Further details on construction of Δ(*ydcS-yddM*)356::*cat* available in Methods Section of Chapter 2.

Genotype for strain KM380 is Δ(*nicK-yddM*)354::*cat amyE*::{(Pxyl-*rapI*) *spc*} Δ*int*::*cat*::*tet*. Strain KM380 was constructed by transformation of strain KM368, Δ(*nicK-yddM*)354::*cat amyE*::{(Pxyl-*rapI*) *spc*} with chromosomal DNA from chromosomal DNA from strain JMA230, Δ*int*::*cat*::*tet*. Strain KM368 was constructed by transformation of chromosomal DNA from CAL874, Δ(*rapI-phrI*)342::*kan amyE*::{(Pxyl-*rapI*) *spc*} with chromosomal DNA from strain KM354, Δ(*nicK-yddM*)354::*cat*. Further details on construction of Δ(*nicK-yddM*)356::*cat* available in Methods Section of Chapter 2.

# Chapter 3

# Discussion

Many, but not all, ICEs appear to be site-specific, integrating into one or a few primary integration sites (Burrus and Waldor, 2004). This section discusses site-specificity of ICEs, including: 1) the possibility of host factors influencing site-specificity, 2) the frequency of use of non-primary sites, 3) the evolution of altered specificities, and 4) the selective pressures acting to limit the evolution of altered site-specificities. When appropriate, I address how the experiments reported in this thesis (Chapter 2 and Appendices A, B, and C) shed light on site-specificity in ICEs.

**1) Host factors that influence site-specificity**

Host factors are known to be required for integration of various mobile genetic elements, such as phages or transposons. For example, integration of phage lambda, which has an integrase similar to many ICE integrases, requires an integration host factor (IHF) (Azaro and Landy, 2002). The lambda IHF protein binds to DNA near the lambda attachment site and is thought to bend the DNA in such a way to allow the integration reaction.

Host factors are not only required for integration, they are also known to influence target site selection of mobile genetic elements (Leem et al., 2008; Parks et al., 2009). For example, the transposon Tn7, in *Escherichia coli*, uses two different pathways for target site selection, one of which involves integration into replicating DNA by interaction with the β processivity factor (the other pathway involves site specific integration into one chromosomal location) (Parks et al., 2009). In ICEs, little is known about host factors, besides sequence specificity, that influence target site selection. However, based on two results reported in this thesis (Chapter 2 and Appendix C), I suspect that host factors also play a role in site-specificity of ICEs.

The first line of evidence supporting the presence of host factors that influence target selection in ICE*Bs1* is based on the sequence of 15 secondary attachment sites identified by

mating ICE*Bs1* into a recipient deleted for the primary site (Chapter 2). Some of these sites are very similar to the primary attachment site (only 2-3 mismatches in the 17 bp attachment site), but some are quite different (14-15 mismatches in the 17 bp attachment site). The fact that ICE*Bs1* integrates into secondary attachment sites that are so dissimilar to the primary site, when many more similar sites exist, suggests that factors besides sequence specificity are influencing choice of secondary integration site.

The second line of evidence is based on the experiment reported in Appendix C. This experiment demonstrated that when two identical, primary attachment sites are both placed in the chromosome (one at the endogenous site *leu2* and one at an exogenous site *thrC*), both sites are not used to an equal extent. In fact, the exogenous site at *thrC*, was the preferred integration target by a ratio of ~3:1. Since the two attachment sites are identical over a region of ~1 kb, it is likely that other factors are at play.

Potential host factors might include effects of DNA binding proteins, transcription, or supercoiling. For example, I hypothesize that nucleoid associated proteins (NAPs) play a role in ICE*Bs1* site selection by modulating the accessibility of portions of the genome. It would be interesting to gain a better understanding of what host factors influence target site selection in ICEs.

**2) Frequency of use of non-primary integration sites.**

It is known that site-specific ICEs will integrate into non-primary sites when the primary site is deleted (Burrus and Waldor, 2003; Doublet et al., 2008; Lee et al., 2007). However, given the difficulty in detecting low frequency integration events, much less is known about whether or not they are able to integrate into secondary sites when the primary site is present and with what frequently this occurs. In work reported in Appendix A, I found that integration into secondary

attachment sites does occur when the primary site (*attB*) is present in the *B.subtilis* recipient. Integration into *yrkM*, the most frequently isolated secondary site, occurred at a frequency of ~$10^4$ conjugation events. While 1 in $10^4$ is not a particularly high frequency, mating in the environment is thought to occur quite readily, perhaps at even higher efficiencies than observed in the laboratory (Aminov, 2011), making it very likely that integration into secondary attachment sites occurs naturally in the environment.

The reason that the frequency of integration into secondary attachment sites (in the presence of the primary site) is as low I observed is very likely due to the various selective pressures that limit ICE's ability to proliferate in these locations. The integrases of ICEs have probably evolved to only very rarely integrate into secondary locations in order to avoid these detrimental effects. I will discuss the selective pressures at play further below.

While it appears to only occur infrequently, the ability of site-specific ICEs to integrate into secondary attachment sites, in the presence of the primary site, may actually be a very important step in the evolution of altered site-specificities (Rutkai et al., 2003; Rutkai et al., 2006). I will discuss the development of altered site-specificities in more detail in the next section.

**3) Development of altered site-specificities.**

Given that families of integrases have different site specificities but are presumed to have evolved from the same source, their altered site specificities must have evolved somehow (Rutkai et al., 2003). Work on phage lambda has led to a model for the development of altered site-specificities among site-specific tyrosine recombinases (Rutkai et al., 2003; Rutkai et al., 2006). This model is referred to as the "chromosome jumping model" and it begins with integration into a secondary attachment site. Then, abnormal excision from a secondary attachment site removes DNA extending beyond the attachment site on one side (for example,

*attR*). In this way, the entire *attR* sequence is removed and it will eventually become the element's new *attP* site. This removed *attR* contains significant similarity to the secondary attachment site, which will become the new *attB*. In addition to abnormal excision from secondary attachment sites, the model predicts that mutations to both the removed *attR* (aka. new *attP*) and to the integrase are necessary to alter site specificity. Initial mutations to the integrase likely relax the site-specificity of the enzyme and then later mutations likely restrict site-specificity to the new *attB*. If at times during this evolution process site-specific integration is difficult, homologous recombination with host chromosome might allow propagation of the element.

It is currently unclear to what extent this model is true generally or for ICEs, which often encode a tyrosine recombinase that is related to phage lambda's. One piece of evidence, presented in this thesis, sheds some initial light on this. I found that when ICE*Bs1* mates from a secondary attachment site to another *B.subtilis* recipient, the site-specificity of ICE*Bs1* does not appear to be altered (Appendix C). This is consistent with predictions of the "chromosome jumping model", wherein steps besides simply integration into secondary attachment sites are necessary to develop altered site specificity.

More work is needed to determine how specificity evolves, particularly in ICEs. While many similarities to integration between phage lambda and ICEs exist, there are also many key differences that are only just beginning to be sorted out (Rajeev et al., 2009).

**4) Selective pressures acting to limit altered site-specificities**

Several selective pressures are likely acting to limit the development of integrases with altered or promiscuous integration patterns. As mentioned in the Introduction to this thesis, these selective pressures include disruption of target genes, effects on host transcription, and

chromosomal deletions. To add to this, I have uncovered two additional potential selective pressures favoring a site-specific strategy of integration for ICE*Bs1*. This section will elaborate on these two additional forces of selective pressure: A) reduced mobility of ICE from secondary attachment and B) defect in proliferation and viability after induction of ICE gene expression, with a particular focus on the likelihood that these pressures on ICE*Bs1* apply to other ICEs more generally.

**A) Reduced mobility from secondary integration sites**. The mobility of ICE*Bs1* after integration into secondary integration sites is limited in two ways. First, the element is unable to efficiently excise from secondary attachment sites. At least one other ICE has a reduced excision frequency from a secondary integration site. Excision of SXT from a secondary attachment site in *Vibrio cholerae* was reduced 3-4-fold relative to its ability to excise from the primary attachment site (Burrus and Waldor, 2003). In addition, phage lambda also has reduced excision efficiencies from secondary attachment sites (Shimada et al., 1972). The fact that reduced excision has been observed in other tyrosine recombinases, including other ICEs, makes is very likely that this phenomenon extends to other site-specific tyrosine recombinase ICEs.

The second way that mobility of ICE*Bs1* is limited is by the lower than expected levels of excised, circular ICE*Bs1* after excision from secondary attachment sites that form a heteroduplex. Lower levels of the excised, circular form may indicate that ICE*Bs1* is unstable after excision from some secondary sites. To my knowledge, instability of circular, heteroduplex ICE has not been previously reported for other ICEs. However, I suspect that this does occur with other ICEs that form a heteroduplex, at least when present in certain hosts. It seems likely that the stability of heteroduplex ICEs depends on host factors, such as host nucleases and/or DNA repair mechanisms, that may be recognizing and degrading circular ICEs with mismatches.

Thus, instability of ICE heteroduplexes may not be a universal trait of heteroduplex ICEs but instead depends on the host in which it resides.

Both of these effects (inability to excise and loss of heteroduplex) are important for mobility because an excised, circular form of ICE is necessary for successful transfer to a recipient cell. Insertion of a mobile genetic element into locations where its mobility is limited is problematic for further propagation of the element.

**B) Defect in cell proliferation and viability after induction of ICE*Bs1* expression.** In addition to the defects in ICE*Bs1* mobility from secondary integration sites, there is a decrease in cell proliferation and viability following induction of ICE*Bs1* gene expression. My results indicate that the defect in cell proliferation/viability is due to rolling-circle replication from an element that is unable to efficiently excise from the chromosome. Rolling-circle replication likely causes cell death because of DNA damage and/or host replication blocks at the site of ICE*Bs1* replication.

The extent to which this defect in viability and proliferation applies to other ICEs, therefore, depends on how common autonomous rolling-circle replication is among ICEs. There are indications that some other ICEs undergo autonomous replication (Carraro et al., 2011; Ramsay et al., 2006; Sitkiewicz et al., 2011; Wozniak and Waldor, 2010), however, it is still unclear for other ICEs. It is possible that the low frequency of induction and excision of most of these elements makes replication difficult to observe in some ICEs. For ICEs that undergo autonomous replication, I would expect a proliferation and/or viability defect after induction of ICE replication genes.

The proliferation/viability defect, defect in ICE mobility, and other selective pressures previously known to exist provide many reasons for ICEs to integrate site-specifically rather than

promiscuously. These pressures may explain the prevalence and maintenance of ICEs that use a site-specific strategy of integration.

# References

Aminov, R.I. (2011). Horizontal gene exchange in environmental microbiota. Frontiers in microbiology *2*, 158.

Auchtung, J.M., Lee, C.A., Garrison, K.L., and Grossman, A.D. (2007). Identification and characterization of the immunity repressor (ImmR) that controls the mobile genetic element ICE*Bs1* of Bacillus subtilis. Molecular microbiology *64*, 1515-1528.

Auchtung, J.M., Lee, C.A., Monson, R.E., Lehman, A.P., and Grossman, A.D. (2005). Regulation of a Bacillus subtilis mobile genetic element by intercellular signaling and the global DNA damage response. Proceedings of the National Academy of Sciences of the United States of America *102*, 12554-12559.

Azaro, M.A., and Landy, A. (2002). λ Integrase and the λ Int Family. In Mobile DNA II (ASM Press), pp. 118-148.

Barlow, M. (2009). What Antimicrobial Resistance Has Taught Us About Horizontal Gene Transfer In Horizontal Gene Transfer, pp. 397-411.

Berkmen, M.B., Laurer, S.J., Giarusso, B.K., and Romero, R. (2011). The Integrative and Conjugative Element ICE*Bs1* of Bacillus subtilis. In Bacterial Integrative Mobile Genetic Elements, A.P. Roberts, and P. Mullany, eds. (Landes Bioscience).

Berkmen, M.B., Lee, C.A., Loveday, E.K., and Grossman, A.D. (2010). Polar positioning of a conjugation protein from the integrative and conjugative element ICE*Bs1* of Bacillus subtilis. Journal of bacteriology *192*, 38-45.

Bose, B., Auchtung, J.M., Lee, C.A., and Grossman, A.D. (2008). A conserved anti-repressor controls horizontal gene transfer by proteolysis. Molecular microbiology *70*, 570-582.

Brochet, M., Da Cunha, V., Couve, E., Rusniok, C., Trieu-Cuot, P., and Glaser, P. (2009). Atypical association of DDE transposition with conjugation specifies a new family of mobile elements. Molecular microbiology *71*, 948-959.

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

Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. (2002b). The ICESt1 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., and Waldor, M.K. (2003). Control of SXT integration and excision. Journal of bacteriology *185*, 5045-5054.

Burrus, V., and Waldor, M.K. (2004). Shaping bacterial genomes with integrative and conjugative elements. Research in microbiology *155*, 376-386.

Carraro, N., Libante, V., Morel, C., Decaris, B., Charron-Bourgoin, F., Leblond, P., and Guedon, G. (2011). Differential regulation of two closely related integrative and conjugative elements from Streptococcus thermophilus. BMC microbiology *11*, 238.

Celli, J., and Trieu-Cuot, P. (1998). Circularization of *Tn916* is required for expression of the transposon-encoded transfer functions: characterization of long tetracycline-inducible transcripts reading through the attachment site. Molecular microbiology *28*, 103-117.

Cheo, D.L., Bayles, K.W., and Yasbin, R.E. (1991). Cloning and characterization of DNA damage-inducible promoter regions from Bacillus subtilis. Journal of bacteriology *173*, 1696-1703.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome research *14*, 1188-1190.

Doublet, B., Golding, G.R., Mulvey, M.R., and Cloeckaert, A. (2008). Secondary chromosomal attachment site and tandem integration of the mobilizable Salmonella genomic island 1. PloS one *3*, e2060.

Duval-Valentin, G., Marty-Cointin, B., and Chandler, M. (2004). Requirement of IS911 replication before integration defines a new bacterial transposition pathway. The EMBO journal *23*, 3897-3906.

Franke, A.E., and Clewell, D.B. (1981a). Evidence for a chromosome-borne resistance transposon (Tn916) in Streptococcus faecalis that is capable of "conjugal" transfer in the absence of a conjugative plasmid. Journal of bacteriology *145*, 494-502.

Franke, A.E., and Clewell, D.B. (1981b). Evidence for conjugal transfer of a Streptococcus faecalis transposon (Tn916) from a chromosomal site in the absence of plasmid DNA. Cold Spring Harb Symp Quant Biol *45 Pt 1*, 77-80.

Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. Nature reviews. Microbiology *3*, 722-732.

Gaillard, M., Pernet, N., Vogne, C., Hagenbuchle, O., and van der Meer, J.R. (2008). Host and invader impact of transfer of the clc genomic island into Pseudomonas aeruginosa PAO1. Proceedings of the National Academy of Sciences of the United States of America *105*, 7058-7063.

Garrity, D.B., and Zahler, S.A. (1994). Mutations in the gene for a tRNA that functions as a regulator of a transcriptional attenuator in Bacillus subtilis. Genetics *137*, 627-636.

Goranov, A.I., Katz, L., Breier, A.M., Burge, C.B., and Grossman, A.D. (2005). A transcriptional response to replication status mediated by the conserved bacterial replication protein DnaA. Proceedings of the National Academy of Sciences of the United States of America *102*, 12932-12937.

Grindley, N.D., Whiteson, K.L., and Rice, P.A. (2006). Mechanisms of site-specific recombination. Annual review of biochemistry *75*, 567-605.

Guerillot, R., Da Cunha, V., Sauvage, E., Bouchier, C., and Glaser, P. (2013). Modular Evolution of TnGBSs, a New Family of Integrative and Conjugative Elements Associating Insertion Sequence Transposition, Plasmid Replication, and Conjugation for Their Spreading. Journal of bacteriology *195*, 1979-1990.

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

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene *77*, 51-59.

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

Ike, Y., Flannagan, S.E., and Clewell, D.B. (1992). Hyperhemolytic phenomena associated with insertions of Tn916 into the hemolysin determinant of Enterococcus faecalis plasmid pAD1. Journal of bacteriology *174*, 1801-1809.

Ireton, K., and Grossman, A.D. (1994). A developmental checkpoint couples the initiation of sporulation to DNA replication in Bacillus subtilis. The EMBO journal *13*, 1566-1573.

Jaacks, K.J., Healy, J., Losick, R., and Grossman, A.D. (1989). Identification and characterization of genes controlled by the sporulation-regulatory gene spo0H in Bacillus subtilis. Journal of bacteriology *171*, 4121-4129.

Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., and Crook, D.W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS microbiology reviews *33*, 376-393.

Kaufman, P.D., and Rio, D.C. (1991). Drosophila P-element transposase is a transcriptional repressor in vitro. Proceedings of the National Academy of Sciences of the United States of America *88*, 2613-2617.

Lee, C.A., Auchtung, J.M., Monson, R.E., and Grossman, A.D. (2007). Identification and characterization of int (integrase), xis (excisionase) and chromosomal attachment sites of the integrative and conjugative element ICE*Bs1* of Bacillus subtilis. Molecular microbiology *66*, 1356-1369.

Lee, C.A., Babic, A., and Grossman, A.D. (2010). Autonomous plasmid-like replication of a conjugative transposon. Molecular microbiology *75*, 268-279.

Lee, C.A., and Grossman, A.D. (2007). Identification of the origin of transfer (oriT) and DNA relaxase required for conjugation of the integrative and conjugative element ICE*Bs1* of Bacillus subtilis. Journal of bacteriology *189*, 7254-7261.

Lee, C.A., Thomas, J., and Grossman, A.D. (2012). The Bacillus subtilis conjugative transposon ICE*Bs1* mobilizes plasmids lacking dedicated mobilization functions. Journal of bacteriology *194*, 3165-3172.

Leem, Y.E., Ripmaster, T.L., Kelly, F.D., Ebina, H., Heincelman, M.E., Zhang, K., Grewal, S.I., Hoffman, C.S., and Levin, H.L. (2008). Retrotransposon Tf1 is targeted to Pol II promoters by transcription activators. Molecular cell *30*, 98-107.

Lin, T.L., Lee, C.Z., Hsieh, P.F., Tsai, S.F., and Wang, J.T. (2008). Characterization of integrative and conjugative element ICEKp1-associated genomic heterogeneity in a Klebsiella pneumoniae strain isolated from a primary liver abscess. Journal of bacteriology *190*, 515-526.

MacRae, A.F. (2002). Experimental evidence that the maize Activator (ac) transposase can act on a non-transposable element promoter to repress reporter gene expression in transient plant assays. Genetica *115*, 289-309.

Miller, J.H. (1972). Experiments in Molecular Genetics, Vol xvi (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

Moon, K., Sonnenburg, J., and Salyers, A.A. (2007). Unexpected effect of a Bacteroides conjugative transposon, CTnDOT, on chromosomal gene expression in its bacterial host. Molecular microbiology *64*, 1562-1571.

Moran, N.A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. Cell *108*, 583-586.

Mullany, P., Williams, R., Langridge, G.C., Turner, D.J., Whalan, R., Clayton, C., Lawley, T., Hussain, H., McCurrie, K., Morden, N., *et al.* (2012). Behavior and target site selection of conjugative transposon Tn916 in two different strains of toxigenic Clostridium difficile. Applied and environmental microbiology *78*, 2147-2153.

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

Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., *et al.* (2001). Genome sequence of Yersinia pestis, the causative agent of plague. Nature *413*, 523-527.

Parks, A.R., Li, Z., Shi, Q., Owens, R.M., Jin, M.M., and Peters, J.E. (2009). Transposition into replicating DNA occurs through interaction with the processivity factor. Cell *138*, 685-695.

Perego, M., Spiegelman, G.B., and Hoch, J.A. (1988). Structure of the gene for the transition state regulator, abrB: regulator synthesis is controlled by the spo0A sporulation gene in Bacillus subtilis. Molecular microbiology *2*, 689-699.

Rajeev, L., Malanowska, K., and Gardner, J.F. (2009). Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. Microbiology and molecular biology reviews : MMBR *73*, 300-309.

Ramsay, J.P., Sullivan, J.T., Stuart, G.S., Lamont, I.L., and Ronson, C.W. (2006). Excision and transfer of the Mesorhizobium loti R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. Molecular microbiology *62*, 723-734.

Redfield, R.J. (2001). Do bacteria have sex? Nature reviews. Genetics *2*, 634-639.

Rice, L.B., Carias, L.L., Marshall, S., Rudin, S.D., and Hutton-Thomas, R. (2005). Tn5386, a novel Tn916-like mobile element in Enterococcus faecium D344R that interacts with Tn916 to yield a large genomic deletion. Journal of bacteriology *187*, 6668-6677.

Roberts, A.P., and Mullany, P. (2009). A modular master on the move: the Tn916 family of mobile genetic elements. Trends in microbiology *17*, 251-258.

Rutkai, E., Dorgai, L., Sirot, R., Yagil, E., and Weisberg, R.A. (2003). Analysis of insertion into secondary attachment sites by phage lambda and by int mutants with altered recombination specificity. Journal of molecular biology *329*, 983-996.

Rutkai, E., Gyorgy, A., Dorgai, L., and Weisberg, R.A. (2006). Role of secondary attachment sites in changing the specificity of site-specific recombination. Journal of bacteriology *188*, 3409-3411.

Salyers, A.A., Shoemaker, N.B., and Li, L.Y. (1995). In the driver's seat: the Bacteroides conjugative transposons and the elements they mobilize. Journal of bacteriology *177*, 5727-5731.

Seth-Smith, H., and Croucher, N.J. (2009). Genome watch: breaking the ICE. Nature reviews. Microbiology *7*, 328-329.

Shimada, K., Weisberg, R.A., and Gottesman, M.E. (1972). Prophage lambda at unusual chromosomal locations. I. Location of the secondary attachment sites and the properties of the lysogens. Journal of molecular biology *63*, 483-503.

Shimoji, Y., Mori, Y., Sekizaki, T., Shibahara, T., and Yokomizo, Y. (1998). Construction and vaccine potential of acapsular mutants of Erysipelothrix rhusiopathiae: use of excision of Tn916 to inactivate a target gene. Infection and immunity *66*, 3250-3254.

Sitkiewicz, I., Green, N.M., Guo, N., Mereghetti, L., and Musser, J.M. (2011). Lateral gene transfer of streptococcal ICE element RD2 (region of difference 2) encoding secreted proteins. BMC microbiology *11*, 65.

Smidt, H., Song, D., van Der Oost, J., and de Vos, W.M. (1999). Random transposition by Tn916 in Desulfitobacterium dehalogenans allows for isolation and characterization of halorespiration-deficient mutants. Journal of bacteriology *181*, 6882-6888.

Smith, B.T., Grossman, A.D., and Walker, G.C. (2001). Visualization of mismatch repair in bacterial cells. Molecular cell *8*, 1197-1206.

Smith, M.C., and Thorpe, H.M. (2002). Diversity in the serine recombinases. Molecular microbiology *44*, 299-307.

Smits, W.K., and Grossman, A.D. (2010). The transcriptional regulator Rok binds A+T-rich DNA and is involved in repression of a mobile genetic element in Bacillus subtilis. PLoS genetics *6*, e1001207.

Thomas, J., Lee, C.A., and Grossman, A.D. (2013). A conserved helicase processivity factor is needed for conjugation and replication of an integrative and conjugative element. PLoS genetics *9*, e1003198.

Van Duyne, G. (2002). A Structural View of Tyrosine Recombinase Site-Specific Recombination. In Mobile DNA II (ASM Press), pp. 93-117.

Vandeyar, M.A., and Zahler, S.A. (1986). Chromosomal insertions of Tn917 in Bacillus subtilis. Journal of bacteriology *167*, 530-534.

Vasantha, N., and Freese, E. (1980). Enzyme changes during Bacillus subtilis sporulation caused by deprivation of guanine nucleotides. Journal of bacteriology *144*, 1119-1125.

Wach, A. (1996). PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in S. cerevisiae. Yeast *12*, 259-265.

Wang, H., and Mullany, P. (2000). The large resolvase TndX is required and sufficient for integration and excision of derivatives of the novel conjugative transposon Tn5397. Journal of bacteriology *182*, 6577-6583.

Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M., and Losick, R. (2008). Molecular Biology of the Gene, 6th edn (CSHL Press).

Williams, K.P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. Nucleic acids research *30*, 866-875.

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

Yoshida, K., Yamaguchi, M., Morinaga, T., Kinehara, M., Ikeuchi, M., Ashida, H., and Fujita, Y. (2008). myo-Inositol catabolism in Bacillus subtilis. The Journal of biological chemistry *283*, 10415-10424.

Youngman, P., Poth, H., Green, B., York, K., G, O., and al., e. (1989). Methods for Genetic Manipulation, Cloning, and Functional Analysis of Sporulation Genes in *Bacillus subtilis* (Washington, D.C.: ASM Press).