Delineation of the integrase-attachment and origin-of-transfer regions of the symbiosis island ICEMISymR7A


PII: S0147-619X(18)30152-5
DOI: https://doi.org/10.1016/j.plasmid.2019.102416
Article Number: 102416
Reference: YPLAS 102416
To appear in: Plasmid

Received date: 14 December 2018
Revised date: 12 April 2019
Accepted date: 7 May 2019

Please cite this article as: C.J. Verdonk, J.T. Sullivan, K.M. Williman, et al., Delineation of the integrase-attachment and origin-of-transfer regions of the symbiosis island ICEMISymR7A, Plasmid, https://doi.org/10.1016/j.plasmid.2019.102416

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Delineation of the integrase-attachment and origin-of-transfer regions of the symbiosis island ICEMISymR7A

Callum J. Verdonk1,2, John T. Sullivan3, Kate M. Williman3, Leila Nicholson3, Tahlia R. Bastholm2, Michael F. Hynes3, Clive W. Ronson3, Charles S. Bond1 and Joshua P. Ramsay2*

1 School of Molecular Sciences, University of Western Australia, Perth, WA, Australia
2 School of Pharmacy and Biomedical Sciences and Curtin Health Innovation Research Institute, Curtin University, Perth, WA, Australia
3 Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand
4 Department of Biological Sciences, University of Calgary, Calgary, Canada

* To whom correspondence should be addressed. Email: josh.ramsay@curtin.edu.au

Keywords: Integrative and Conjugative Element; origin of transfer; integrase; excisionase; ICEMISymR7A
ABSTRACT

Integrative and conjugative elements (ICEs) are chromosomally-integrated mobile genetic elements that excise from their host chromosome and transfer to other bacteria via conjugation. ICEMISymR7A is the prototypical member of a large family of "symbiosis ICEs" which confer upon their hosts the ability to form a nitrogen-fixing symbiosis with a variety of legume species. Mesorhizobial symbiosis ICEs carry a common core of mobilisation genes required for integration, excision and conjugative transfer. IntS of ICEMISymR7A enables recombination between the ICEMISymR7A attachment site attP and the 3' end of the phe-tRNA gene. Here we identified putative IntS attP arm (P) sites within the attP region and demonstrated that the outermost P1 and P5 sites demarcated the minimal region for efficient IntS-mediated integration. We also identified the ICEMISymR7A origin-of-transfer (oriT) site directly upstream of the relaxase-gene rlxS. The ICEMISymR7A conjugation system mobilised a plasmid carrying the cloned oriT to Escherichia coli in an rlxS-dependent manner. Surprisingly, an in-frame, markerless deletion mutation in the ICEMISymR7A recombination directionality factor (excisionase) gene rdfS, but not a mutation in ints, abolished mobilisation, suggesting the rdfS deletion tentatively has downstream effects on conjugation or its regulation. In summary, this work defines two critical cis-acting regions required for excision and transfer of ICEMISymR7A and related ICEs.
1. INTRODUCTION

Integrative and Conjugative Elements (ICEs) are chromosomally-integrated elements that transfer between bacteria by conjugation (1-3). ICEs are likely the most abundant conjugative element in prokaryotic genomes and carry diverse genetic cargo, which can confer phenotypes such as antimicrobial-resistance, virulence, metabolism and the capacity to enter symbiotic associations with eukaryotes (4-10). After entry of an ICE into a recipient cell, integration into the host chromosome occurs through the actions of an ICE-encoded site-specific recombinase; termed an integrase (11, 12). Integrases belong to either the tyrosine or serine recombinase families; defined by the conserved nucleophilic amino acid that becomes covalently bonded to a DNA strand during the recombination reaction (11).

The recombination events mediated by integrases have largely been defined by studies performed on phages (11-14). Integrases form nucleoprotein complexes consisting of a homodimer structures bound to cis-acting DNA regions called attachment (att) sites present in the phage and host chromosomes (11). During chromosomal integration, an attP site present on the circularised phage genome, is recombined with the host chromosome attachment site attB. Upon integration, the integration junctions termed attL and attR are formed. Each att site contains a “catalytic core” site where strand exchange occurs. att sites also contain “arm” or “P” sites which flank the core. The P sites facilitate structural organisation of the integrase-DNA complex during recombination (11, 15-17). Recombinases are heterobivalent in that the N-terminal region binds the P sites and the C-terminal region interacts with the att core sequence where recombination occurs. During recombination the nucleophilic residue within the C-terminus becomes covalently fused to a DNA strand. An incoming DNA strand then displaces the residue reforming an intact DNA strand resulting in formation of a double-stranded cruciform DNA structure known as a Holliday junction intermediate. A second round of cleavage, strand exchange and ligation resolves this structure to produce the recombination end products (3, 18, 19).

Integrated ICEs are flanked by hybrid attL and attR sites and the two core sites form a direct repeat bordering the inserted ICE. For both phages and ICEs, excision is catalysed by the integrase, but often an additional factor known as an excisionase or recombination directionality factor (RDF) is required. Excisionases have been shown to bind and bend att DNA, shifting the favoured direction of integrase-mediated recombination towards the excision reaction (attL + attR > attP + attB) (11, 20). Certain excisionases, particularly the subset cox family, have bifunctional roles as transcriptional regulators (12, 21-27). The P2 phage excisionase Cox stimulates the lytic cycle by blocking
transcription of repressor gene cl, in addition to its roles as an att-binding prophage excision stimulator (25, 26).

ICEMISym\textsuperscript{R7A} is a 502-kb ICE which encodes genes required to form a nitrogen-fixing symbiosis with legumes of the genus \textit{Lotus} (9, 28). Subsequent to the discovery of ICEMISym\textsuperscript{R7A}, numerous other mobile symbiosis ICES have been identified within mesorhizobial strains that confer the ability to form symbioses with a variety of legumes including \textit{Biserrula} and Chickpea (29-31). Recently, more complex tripartite symbiosis ICES have been discovered, which carry a similar core gene complement to ICEMISym\textsuperscript{R7A} but upon integration in a new host they fragment into three physically separate regions within the host chromosome through the action of two additional integrases (31, 32).

ICEMISym\textsuperscript{R7A} is a monopartite ICE that integrates into the single phe-tRNA gene of the \textit{Mesorhizobium loti} R7A chromosome (9, 20), a feature common to symbiosis ICES found within mesorhizobia including the tripartite ICES (31). Integration requires the integrase IntS, a tyrosine recombinase related to the P4 bacteriophage integrase family (20). The ICEMISym\textsuperscript{R7A} attP site contains a 17-bp DNA sequence identical to the last 17-bp of the phe-tRNA gene and integration of ICEMISym\textsuperscript{R7A} into this attB site reconstitutes the phe-tRNA gene within the newly-formed attL site (9, 20). We have previously demonstrated that a synthetic "mini-ICE" carrying only intS and attP integrates into the phe-tRNA gene in R7ANS, a derivative of R7A from which ICEMISym\textsuperscript{R7A} was cured (20). The ICEMISym\textsuperscript{R7A} RDF, RdfS, is a putative member of the MerR superfamily of winged helix-turn-helix (wHTH) DNA-binding proteins (12, 33, 34). Deletion of rdfS abolishes excision of ICEMISym\textsuperscript{R7A}, while overexpression of rdfS leads to loss of ICEMISym\textsuperscript{R7A}, producing the non-symbiotic derivative strain R7ANS (20). Expression of rdfS in R7ANS carrying an integrated copy of the mini-ICE stimulated its excision and subsequent loss, suggesting rdfS and intS together are sufficient for ICEMISym\textsuperscript{R7A} excision (20).

Unlike most RDF genes, which are generally positioned directly adjacent to their cognate integrase gene (12), rdfS is located within a highly conserved symbiosis-ICE gene cluster, containing conjugation-associated genes including \textit{traF} (encoding a putative prepilin peptidase), a lytic transglycosylase, a VirD2-like conjugative relaxase gene \textit{rlxS} and the putative ICEMISym\textsuperscript{R7A} origin-of-transfer (\textit{oriT}) region (20, 28). \textit{oriT} sites are \textit{cis}-acting regions that generally contain inverted repeats and a nic/core site targeted by a relaxase protein. Relaxases, often in association with other components, form a protein-DNA complex with the \textit{oriT} called the relaxosome. Within the relaxosome, the relaxase nicks DNA within the nic core site and remains covalently attached to the 5'
phosphate of the nicked strand, after which it pilots DNA to the type-IV secretion system for conjugative transfer (35, 36). Relaxosomes often include accessory proteins required for specific recognition of the oriT and/or stimulation of oriT cleavage (37-39). Mutation of ICEMISym^R7A rlxS abolishes transfer of ICEMISym^R7A. Moreover, rlxS is also required for maintenance of ICEMISym^R7A during excision, suggesting relaxase-mediated rolling-circle replication contributes to ICEMISym^R7A maintenance when in the excised form (20).

In this work, we further define two cis-acting DNA sites of ICEMISym^R7A, the IntS attachment site attP and the oriT site. The minimal attP site contained five copies of a directly repeated sequence motif in a pattern resembling the arm sites of other characterised attP regions. Removal of a single proximal or distal "P" site severely hampered IntS-mediated integration. Conjugative mobilisation experiments were used to confirm the ICEMISym^R7A origin-of-transfer (oriT) region was located upstream of the rlxS gene. Unexpectedly, an intact rdfS gene was required for mobilisation, suggesting this mutation has an effect on the expression of the ICEMISym^R7A conjugation system.
2. MATERIAL AND METHODS

2.1 Strains and growth conditions

Strains used in this study are listed in Supplemental Table 1. *M. loti* strains were cultured at 28°C or 30°C in Tryptone Yeast (TY) media (0.3% w/v yeast extract, 0.5% w/v tryptone, 0.6% v/v CaCl₂) or in glucose rhizobium defined medium (G/RDM) (40) supplemented with vitamins as described previously (20). *Escherichia coli* DC10B (41) cultures were grown in Lysogeny broth (LB) at 37°C. Media was supplemented with antibiotics where required at the following concentrations: 2.0 μg/mL (*M. loti*) and 5.0 μg/mL (*E. coli*) tetracycline, 50 μg/mL gentamicin and 50 μg/mL ampicillin.

2.2 Molecular biology and cloning

Plasmid construction is detailed in Supplemental Table 1 and oligonucleotides and QPCR probes used are described in Supplemental Table 2. QPCR was carried using the ViiA 7 Real-time PCR system (Applied Biosystems) in 20-μl reactions using the SensiFAST Probe Lo-ROX Kit (Bioline), using the standard-speed running conditions (i.e. non-fast). DNA was extracted from stationary-phase non-IPTG induced TY cultures (64 h) using PrepMan Ultra (ThermoFisher) reagent as previously described (20). Primers and probes used were purchased from Integrated DNA Technologies (IDT - Singapore) and probes were modified with a 5’ FAM fluorophore and 3’ ZEN quencher. The melR gene probe was lengthened to increase binding and compensate for the lack of a minor-groove-binder ligand on the IDT probes. The primer/probe combinations exhibited QPCR efficiencies of 2.04, 2.09 and 2.03 for attP, attB and melR, respectively. Reactions were normalised using the standard template as previously described (20).

2.3 Integration assays for attP-carrying plasmids

Two-hundred nanogram aliquots of plasmid DNA were electroporated into electrocompetent R7ANS(pJJ611) cells as previously described (42). Cells were recovered in 1 mL TY broth with shaking at 28°C for 3 hours. Serial dilutions were then plated on G/RDM containing vitamins (thiamin, biotin and nicotinate) (20) and gentamicin. Colonies were enumerated after 5 days of growth at 28°C. pFUS conjugation experiments of *E. coli* ST18 and *M. loti* R7ANS(pJJ611) were performed as described in section 2.4. Confirmation of pFUS vector integration sites were done by PCR and sequencing (primers used listed in Supplemental Table 2).

2.4 Conjugative mobilisation experiments
M. loti donor strains carrying pFAJ2013 (strains used are listed in Supplemental Table 1) (20) and the E. coli recipient DC10B were grown to stationary phase (OD_{600} 0.8-1.0) in 5 mL TY (R7A) or LB-broth (DC10B). Five-hundred microlitres of M. loti TY culture were pelleted and resuspended in 500 µL E. coli DC10B LB culture and centrifuged for an additional 2 minutes at 17,000 g. The pellet was resuspended in 100 µL TY broth. Thirty-microlitres of the cell mixture spotted in triplicate onto TY agar media and incubated at 30°C overnight. Strains containing pSRKKm or pSRKrdfs were spot plated on TY agar media. Mobilisation experiments with R7AΔrdfs(pFAJ2013)(pSRKrdfs) were carried out both with and without 1 mM IPTG, however no mobilisation was observed in either experiment. Spots were recovered and resuspended in 1 mL sterile water and serial diluted 10-fold to 1 x 10^{9}. One-hundred microlitre aliquots were plated onto LB agar containing tetracycline and incubated overnight at 37°C, and colonies were enumerated. Controls consisting of the donor strains only were prepared similarly, with 500 µL culture resuspended in 100 µL TY broth and 30 µl spots grown overnight at 30°C on TY agar media.

2.5 Bioinformatic analyses of attP and oriT regions.

The ICEMISym^{R7A} sequence is available from GenBank accession AL672111.1 (28). A 352-bp putative attP region (assembled from 501647-501801[attR] and 1-157[attL]) was used in MEME (43)(version 5.0.2) alignments with non-mesohizobial sequences identified by PSI-BLAST (44) as described in section 3.1. GenBank accessions used; Magnetospirillum sp. (FXXN01000020.1), Marinicaulis flavus (PJCH01000010.1), Nitrospirillum amazonense (AFBX01000767.1), Acidisphaera sp. (CP029176.1), Rhizobiales sp. (PIGG01000031.1), Prosthecomicrobium hirschii (LYYW01000001.1), Sphingomonas sp. (CH959307.1) and Rhodopseudomonas palustris (CP000301.1). The putative 352-bp attP region was also submitted to MEME with BLAST-identified Mesorhizobium spp. phe-trNA-adjacent intergenic regions (GenBank accessions: CCNB01000043.1, AYWT01000014.1, KI421454.1, MDLH0100001.1, CAAF01000028.1, AYVN0100010.1, CCNA01000045.1, MDFL01000146.1, NPKJ0100019.1, PZIX01000011.1, FNEE0100009.1, NNRI0100015.1, AHAM01000208.1 and MZXV0100055.1) to confirm the presence of the identified P-site motif.

ICEMISym^{R7A} putative oriT sequence was collected from the upstream region of rlxS (msi106) (133491-133680) and used in PSI-BLAST investigations described in section 3.2. GenBank accessions used; Sphingobium lactosutens (ATDP01000088.1), Oceanibaculum nanhaiense (MPOB01000010.1), Phaeospirillum fulvum (FNWO01000021.1), Rhodospirillum photometricum (HE663493.1), Thalassospira xiamenensis (LPXL0100008.1), Niveispirillum cyanobacteriorum (CP025612.1) and Erythrobacter xanthus (QXFM01000025.1). HMMER hmmscan described in section 3.2 used hmm
3. RESULTS

3.1 Identification of the minimal ICEMISymR7A attP site required for integration

IntS is located directly downstream of the ICEMISymR7A attL site, which also forms part of the attP site on the circularised element. To identify potential IntS P sites within the ICEMISymR7A att sites, we initially used PSI-BLASTP (44) searches to identify divergent IntS homologues outside the Mesorhizobium genus. Nucleotide sequences corresponding to each IntS-homologue gene were inspected to determine if the integrase was preceded by a tRNA gene. Intergenic regions between each tRNA and intS homologue were collected and analysed along with the ICEMISymR7A attL region using the MEME motif identifier software (43, 45). MEME was set to search for multiple repetitions (per sequence) of a 5-20-bp motif. A motif with consensus “TGKTGGTATC” was present as 2-3 direct repeats in each of the putative attL sequences (Fig 1A). Inspection of the ICEMISymR7A attP revealed two additional copies of this motif upstream of the core region (on attP / attR). The five sites were named P1-P5 (Fig. 1B). All five instances of the suspected P-site motif were conserved on the IntS attP regions of Mesorhizobium spp. strains: WSM1284, WSM1497, NZP2037 and WSM1271 (Fig. 1B).
Fig. 1. The ICEMISymR7A attP site. (A) Putative P-site sequence motif derived from an alignment of attL regions corresponding to divergent IntS homologues (see Section 2.5). (B) ICEMISymR7A attP sequence aligned (using MAFFT (46)) with the attP regions of Mesorizobium spp. WSM1284 (accession: CP015064.1), WSM1497 (accession: CP021070.1), NZP2037 (accession: CP016079.1) and WSM1271 (accession: CP002447.1) Asterisks (*) denotes alignment difference from consensus. Putative P-sites P1-P5 are boxed. ICEMISymR7A 17-bp core-containing region is underlined. The 5’ start site of primers used to PCR-amplify the attP regions used in pFUS2 integration assays are shown as arrows.
intS and attP can function together to recombine attP and attB when cloned on separate plasmids. An introduced suicide vector carrying a 628-bp attP region (pJJ610) is efficiently integrated into the phe-tRNA gene, but only in the presence of a functional intS expressed by a coresident plasmid pJJ611 (20). We extended this approach here to define the minimal attP region. A smaller 283-bp region of attP containing only the P1-P5 region was cloned into the gentamicin-resistance suicide plasmid pFUS2, producing pFUSP1-P5. pFUS plasmids lacking either P1, P5 or both P1 and P5, were additionally constructed. Each construct was then transferred by conjugation from E. coli ST18 (47) to R7ANS(pJJ611). Apparent transfer produced rates for the P1-P5 construct were on average 3.8 x 10^{-3} exconjugants per donor, but experiments using the pFUS2 clones lacking the outermost P1 or P5 sites were reduced 190- and 140-fold, respectively (Table 1). No colonies were isolated following transfer of the clone lacking both P1 and P5, suggesting removal of both sites abolished IntS-mediated integration. The attR sites formed by examples of the three integrated pFUS2 constructs were amplified by PCR and sequenced, which confirmed the constructs integrated downstream of the phe-tRNA gene (9, 20). We additionally electroporated the same pFUS2 constructs into R7ANS(pJJ611), which produced a similar pattern of results. Therefore, these experiments both defined the minimal attP site for IntS and suggested that the outermost P1 and P5 sites are required for efficient IntS-mediated recombination.

Table 1. Conjugation and electroporation assays for IntS-mediated attP integration.

<table>
<thead>
<tr>
<th>attP region†</th>
<th>Apparent conjugation rate§</th>
<th>Apparent electroporation efficiency cfu/ml/ng DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1-P5</td>
<td>3.8 x 10^{-3} ± 8.8 x 10^{-4}</td>
<td>935 ± 420</td>
</tr>
<tr>
<td>P1-P4</td>
<td>1.9 x 10^{-5} ± 3.0 x 10^{-6}</td>
<td>3.1 ± 1.3</td>
</tr>
<tr>
<td>P2-P5</td>
<td>2.5 x 10^{-5} ± 8.7 x 10^{-6}</td>
<td>4.0 ± 2.2</td>
</tr>
<tr>
<td>P2-P4</td>
<td>ND§</td>
<td>0</td>
</tr>
<tr>
<td>none</td>
<td>ND§</td>
<td>0</td>
</tr>
</tbody>
</table>

†attP regions cloned into non-replicative plasmid pFUS2 and conjugated or electroporated into R7ANS(pJJ611).

§Conjugation rates are expressed as exconjugants per donor and all data are the average of three replicates of the entire experiment.

§Limit of detection for conjugation approximately 3 x 10^{-8} exconjugants per donor.
3.2 Identification of ICEMISym\textsuperscript{R7A} origin-of-transfer region

The relaxase RlxS is essential for transfer of ICEMISym\textsuperscript{R7A} and extrachromosomal maintenance of ICEMISym\textsuperscript{R7A} when excised (20). A hmmsearch search (48) of relaxase protein family hmm profiles revealed RlxS was related to members of the MOB\textsubscript{P1} family (37, 38). oriT sites of MOB\textsubscript{P} relaxases generally contain an inverted-repeat sequence upstream of a conserved nic core region. oriT sites are often located upstream of their cognate relaxase gene (49, 50). To identify potential conserved sequence motifs upstream of the rlxS gene, the RlxS protein was used as a query in PSI-BLASTP searches (44) of non-mesorhizobial genomes. The NCBI-BLAST taxonomy viewer guided selection of a diverse collection of representatives from the PSI-BLASTP matches. Approximately 500-bp of DNA that preceded each relaxase was collected and analysed using MEME (43). MEME was set to search for 10-60-bp motifs present once on each sequence. Two large motifs were identified, which flanked an inverted-repeat which consisted of a variably-sized central region flanked by short regions of stronger conservation (Fig. 2A). One of the regions adjacent to the inverted-repeat bore resemblance to the consensus core region for MOB\textsubscript{P} relaxases and contained a conserved 'GC' nic sequence (Fig. 2B). Interestingly, an insertion sequence related to the IS1111 family is present within this site in the \textit{M. japonicum} MAFF303099 symbiosis-ICE sequence (Fig. 2A) (also noted in (28)), suggesting the MAFF303099 ICE may be non-mobile. It has been noted that members of the IS1111 family preferentially target regions adjacent to DNA stem-loop structures for their insertion (35, 51-53).

3.3 rlxS and rdfS are essential for mobilisation of a plasmid carrying the ICEMISym\textsuperscript{R7A} oriT

A 95-bp region encompassing the predicted oriT region was cloned into the broad host-range plasmid pFAJ1700, producing pFAJ2013. R7A(pFAJ2013) was used as a donor to mobilise pFAJ2013 by conjugation to \textit{Escherichia coli} DC10B. pFAJ2013, but not the vector-only control pFAJ1700, was mobilised to DC10B at an average rate of $1.2 \times 10^{-4}$ exconjugants per donor (Table 2). Next, pFAJ2013 was introduced into non-symbiotic derivative of R7A, R7ANS, and the rlxS-mutant R7ArlxS (20) and the resulting strains were used as donors in conjugation experiments. No transfer was observed from either of these strains. To genetically complement the rlxS mutation, a region carrying both the oriT and the rlxS gene was cloned into pFAJ1700, producing plasmid pFAJCVM. In contrast to pFAJ2013, this plasmid was mobilised to \textit{E. coli} from the R7ArlxS' background at an average frequency of $2.1 \times 10^{-4}$ exconjugants per donor (Table 2). These data confirmed that mobilisation was dependent on the ICEMISym\textsuperscript{R7A} conjugation machinery and rlxS.
rdfS is located directly upstream of the oriT and rlxS, suggesting it has evolved closely with the DNA transfer region of ICEMISymR7A. Therefore, we tested if a previously characterised strain carrying an in-frame deletion in rdfS (20) was able to mobilise pFAJ2013. Surprisingly, pFAJ2013 mobilisation was completely abolished in this background (Table 2). The lack of mobilisation was not related to a deficit in ICEMISymR7A excision per se, as pFAJ2013 transfer was observed from the intS mutant strain R7AΔintS (in which ICEMISymR7A does not excise (20)), albeit at a reduced frequency. In order to complement the rdfS deletion, a vector compatible with pFAJ2013 was constructed, using pSRKKm (54). The rdfS coding sequence was placed downstream of the lac promoter in pSRKKm and the resulting plasmid pSRKrdfS was introduced into R7AΔrdfS(pFAJ2013). To confirm pSRKrdfS was able to complement the ICEMISymR7A excision defect in R7AΔrdfS, QPCR was carried out to measure the abundance of the excision products attP and attB. As expected, the R7AΔrdfS(pFAJ2013)(pSRKKm) strain exhibited around 100-fold reduced abundance of attP and attB excision products compared to R7A carrying the same plasmids (Supplemental Fig. 1). Strain R7AΔrdfS(pFAJ2013)(pSRKrdfS) however, exhibited close to wild-type levels of excision. Attempts to complement the deficit in pFAJ2013 transfer using the same strains were unsuccessful.

Table 2. Conjugation assay data and colony counts for pFAJ1700 vectors.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Vector</th>
<th>Region cloned</th>
<th>Transfer frequency*</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>R7A</td>
<td>pFAJ700</td>
<td>ND*</td>
<td>ND*</td>
<td>ND*</td>
</tr>
<tr>
<td></td>
<td>pFAJ2013</td>
<td>oriT</td>
<td>1.24 x 10^4</td>
<td>2.62 x 10^5</td>
</tr>
<tr>
<td>R7ANS</td>
<td>pFAJ700</td>
<td>ND*</td>
<td>ND*</td>
<td>ND*</td>
</tr>
<tr>
<td></td>
<td>pFAJ2013</td>
<td>oriT</td>
<td>ND*</td>
<td>ND*</td>
</tr>
<tr>
<td>R7ArIxS</td>
<td>pFAJ700</td>
<td>ND*</td>
<td>ND*</td>
<td>ND*</td>
</tr>
<tr>
<td></td>
<td>pFAJ2013</td>
<td>oriT</td>
<td>ND*</td>
<td>ND*</td>
</tr>
<tr>
<td></td>
<td>pFAJCVM</td>
<td>oriT-rlxS</td>
<td>2.07 x 10^4</td>
<td>6.29 x 10^5</td>
</tr>
<tr>
<td>R7AΔintS</td>
<td>pFAJ1700</td>
<td>ND*</td>
<td>ND*</td>
<td>ND*</td>
</tr>
<tr>
<td></td>
<td>pFAJ2013</td>
<td>oriT</td>
<td>7.12 x 10^6</td>
<td>3.88 x 10^6</td>
</tr>
<tr>
<td></td>
<td>pFAJCVM</td>
<td>oriT-rlxS</td>
<td>6.64 x 10^6</td>
<td>2.97 x 10^6</td>
</tr>
<tr>
<td>R7AΔrdfS</td>
<td>pFAJ1700</td>
<td>ND*</td>
<td>ND*</td>
<td>ND*</td>
</tr>
<tr>
<td></td>
<td>pFAJ2013</td>
<td>oriT</td>
<td>ND*</td>
<td>ND*</td>
</tr>
<tr>
<td></td>
<td>pFAJCVM</td>
<td>oriT-rlxS</td>
<td>ND*</td>
<td>ND*</td>
</tr>
</tbody>
</table>

*Per-donor transfer frequencies are the average of three independent experiments.

*Not detected. Detection limit approximately 3.4 x 10^9
Fig. 2. The ICEMISymR7A origin-of-transfer (oriT).

(A) Organisation of rdfS-rlxS gene cluster, with relative location of oriT indicated (20, 28). Conserved sequences based on non-Mesorhizobium sp. relaxase-upstream regions (underlined) and inverted repeat location (arrows). Suspected nic core site bolded for ICEMISymR7A based on similarity to core sites shown in B. MEME-generated (43) 15-bp conserved motifs shown under respective underlined location. Location of IS1111-family transposon insertion within the related MAFF303099 ICE is indicated. (B) Inverted-repeat regions (arrows), partial nic core sites (bold) and nic sites (arrowhead) of oriT within MOB β-carrying elements; conjugative transposon Tn1549 (49), plasmid RP4 (IncPα) (55), pIP501 (56) and plasmid R64 (57).
4. DISCUSSION

Regulation of ICEMISym\textsuperscript{R7A} excision and transfer has been extensively studied (9, 20, 58-60), however, the cis-acting sites integral to these processes, attP and the oriT, have remained relatively uncharacterised. In this work, we defined the minimal ICEMISym\textsuperscript{R7A} attP region and showed that a 283-bp region containing five putative arm, or “P” sites, P1-P5, were sufficient for efficient integration by IntS. We also confirmed the location of the ICEMISym\textsuperscript{R7A} oriT and demonstrated that a plasmid carrying the oriT region was mobilised from R7A to \textit{E. coli} and this was dependent on the presence of ICEMISym\textsuperscript{R7A} and rlxS.

Integrase P-site sequences and their positions relative to the core region vary but they are typically short, loosely conserved repeats positioned 50-140-bp from the attP core (11, 61) (Fig. 3). We identified a conserved 10-bp motif sequence following bioinformatic comparisons of putative attL sites from distantly related IntS-gene homologues. The consensus motif sequence “TGKTGATC” was present five times in each of the attP regions of ICEMISym\textsuperscript{R7A}, ICEMcSym\textsuperscript{1271}, ICEMISym\textsuperscript{2037} (32) and on putative ICEs in \textit{Mesorhizobium} spp. strains WSM1284 and WSM1497. The arrangement of the likely P sites relative to the core site was similar to that observed for other well-characterised attP sites including Tn916 and \textlambda bacteriophage, despite the fact that these elements encode distantly related integrases (Fig. 3) (20, 62, 63). Removal of either of the outermost putative P sites significantly reduced the efficiency of integration, presumably due to reduced binding, or incorrect positioning of IntS molecules within the nucleoprotein complex during recombination (13, 14, 63). Integration was abolished when both outermost putative P sites were removed.
Fig. 3: Comparison of the ICEMISymR7A att sites with those of Tn916 and bacteriophage λ.

Schematic of ICEMISymR7A, Tn916 (34) and bacteriophage λ attP sites (63). Black lines representing each element are to scale, indicated below the DNA (in base-pairs relative to the core). The black rectangles show the attP site core regions (not to scale). Relative orientation of ICEMISymR7A putative P sites, and Tn916 and bacteriophage λ P sites shown as the black arrows (not to scale). Start location of ICEMISymR7A IntS CDS shown with grey arrow.
RlxS is a member of the MOB\(_r\) family, a large and diverse family of relaxases found in a diverse collection of plasmids and ICEs. ori\(T\) sites of MOB\(_r\) relaxases typically display considerable sequence diversity, however they often contain a moderately conserved nic core sequence adjacent to a less conserved inverted-repeat region (49, 64). The ICEMISym\(^{R7A}\) ori\(T\) region contained an inverted-repeat region and interestingly, strongly conserved motifs that flanked it were identified. One of these motifs resembled a MOB\(_r\) nic core site (Fig. 2B), however, it is possible that the core region could be positioned on the other side of the inverted repeat. Further work is required to fully dissect the functional components of the ICEMISym\(^{R7A}\) ori\(T\).

\(rdfS\) is peculiar for an RDF gene in that it is present upstream of genes involved in conjugation; \(traF, msi107\) (a lytic transglycosylase) and \(rlxS\), rather than adjacent to its cognate integrase \(intS\) (28). Here we demonstrated that deletion of \(rdfS\) abolished the ability of the ICEMISym\(^{R7A}\) conjugation system to mobilise a plasmid carrying the ori\(T\). Strangely, ectopic expression of \(rdfS\) restored the excision defect observed in the \(rdfS\) mutant, as previously demonstrated (20), but did not restore the mobilisation defect (Supplemental Fig. 1). \(rdfS\) is the first gene in a polycistronic operon encoding \(traF, msi107\) and probably \(rlxS\) (33, 58). The \(traF\) start site also overlaps with the \(rdfS\) stop codon suggesting they may be translationally coupled. We cannot exclude the possibility that loss of mobilisation was due to disruption of downstream expression or translation of \(traF\) and/or \(msi107\). However, the R7A\(\Delta rdfS\) deletion is markerless, in-frame and the deletion terminates 59-bp upstream of the \(traF\) open-reading frame (20).

It is possible that RdfS is required for transcriptional activation of other conjugation genes on ICEMISym\(^{R7A}\) such as the type-IV secretion system (\(trb\)) gene cluster or \(rlxS\). We have previously demonstrated that RdfS activates expression of both \(rdfG\) and \(rdfM\) on ICEMcSym\(^{1271}\) (33), so a role for RdfS as a transcriptional activator of conjugation genes is conceivable. The TraN protein of the \(Enterococcus\) sp. plasmid pIP501, like RdfS, is a wHTH protein and it binds upstream of the pIP501 ori\(T\) and represses transcription of the \(tra\)-operon, represing transfer (65, 66). Alternatively, RdfS could have a more direct role in \(oriT\) processing or relaxosome assembly. The \(Clostridium\) perfringens plasmid pCW3 encodes a wHTH transcription factor necessary for conjugation, TcpK, which binds the \(oriT\) utilising a novel binding mechanism for wHTH proteins (67). However, these possibilities aside, it remains unclear why mobilisation of pFAJ2013 was not restored by pSRKrdfS, unless the appropriate physiological levels of RdfS protein were not achieved using this vector. In future experiments investigating the downstream effects of the \(rdfS\) deletion, including possible polarity of the mutation on expression and translation of \(traF\) and \(msi107\), will be required to resolve this.
ICE\textit{Ml}Sym\textsuperscript{R7A} transfer requires coordination of IntS-mediated excision with expression of the type-IV secretion system and processing of the relaxosome. It is clear from the positioning of \textit{rdfS} as the first gene in a cluster containing essential transfer genes \textit{traF} and \textit{rlxS} that coregulation of these processes has become linked over the course of evolution and our results suggest that they may be mechanistically intertwined. The work described herein defines the \textit{attP} and \textit{oriT} sites ICE\textit{Ml}Sym\textsuperscript{R7A} and demonstrates a potentially complex link between the regulation of excision and transfer through expression of the \textit{rdfS-traF-msi107} region.

DATA AVAILABILITY

All data have been provided in-text and accession numbers listed where appropriate.

SUPPLEMENTARY DATA

Supplemental Table 1: Bacterial strains and plasmids utilised in this study.

Supplemental Table 2: Oligonucleotides used in this study.

Supplemental Fig. 1: Excision percentages of R7A and \textit{rdfS} mutant strains containing pFAJ2013 and pSRK or pSRK\textit{rdfS} vectors.

ACKNOWLEDGEMENTS

CJV and JPR acknowledge technical support from the Curtin Health Innovation Research Institute.

FUNDING

JPR is the recipient of an Australian Research Council Future Fellowship [Project ID FT170100235] funded by the Australian Government (http://www.arc.gov.au/grants). CJV is the recipient of an Australian Government Research Training Program (RTP) Scholarship administered by the University of Western Australia. CJV and JPR received funding from the School of Pharmacy and Biomedical Sciences, Curtin University. Part of this work was supported by a grant from the Marsden Fund administered by the Royal Society of New Zealand.

CONFLICT OF INTEREST
The authors have declared that no competing interests exist.
REFERENCES


HIGHLIGHTS

Verdonk, CJ, et al. Delineation of the integrase-attachment and origin-of-transfer regions of the symbiosis island ICEMISym\textsuperscript{R7A}

- We define the minimal ICEMISym\textsuperscript{R7A} integrase attachment site \textit{attP} region sufficient for efficient IntS-mediated integration
- We define the origin-of-transfer (\textit{oriT}) region for ICEMISym\textsuperscript{R7A}
- The ICEMISym\textsuperscript{R7A} conjugation system can mobilise an \textit{oriT}-carrying plasmid to \textit{Escherichia coli}