



OPEN

Isolation, nucleotide sequencing and genomic comparison of a Novel SXT/R391 ICE mobile genetic element isolated from a municipal wastewater environment

Shannon Slattery¹, J. Tony Pembroke¹, John G. Murnane² & Michael P. Ryan¹✉

Integrative Conjugative Elements (ICE's) of the SXT/R391 family have largely been detected in clinical or environmental isolates of Gammaproteobacteria, particularly *Vibrio* and *Proteus* species. As wastewater treatment plants accumulate a large and diverse number of such species, we examined raw water samples taken from a municipal wastewater treatment plant initially using SXT/R391 family integrase gene-specific PCR probes to detect the presence of such elements in a directed approach. A positive amplification occurred over a full year period and a subsequent Restriction Fragment Length Polymorphism (RFLP) analysis revealed a very limited diversity in the treatment plant examined. Samples demonstrating positive amplification were cultured using *Vibrio* and *Proteus* selective media and PCR amplification tracking was utilized to monitor SXT/R391-ICE family containing strains. This screening procedure resulted in the isolation and identification of a *Proteus mirabilis* strain harbouring an ICE. Whole-genome sequencing of this ICE containing strain using Illumina sequencing technology revealed a novel 81 kb element that contained 75 open reading frames on annotation but contained no antibiotic or metal resistance determinants. Comparative genomics revealed the element contained a conserved ICE core with one of the insertions containing a novel bacteriophage defence mechanism. This directed isolation suggests that ICE elements are present in the environment without apparent selective pressure but may contain adaptive functions allowing survival in particular environments such as municipal wastewater which are reservoirs for large bacterial phage populations.

Wastewater treatment plants (WWTP) accumulate large and diverse numbers of pathogens including viruses, bacteria, fungi, protozoans and helminths. The pathogens present may reflect the common diseases in a community and may be present in numbers as great as 100,000,000 per millilitre¹, with the most dominant bacteria being Alphaproteobacteria, Thermotogae, Deltaproteobacteria and Gammaproteobacteria². Wastewater influent may also contain many *Proteobacteria* which normally occur in the environment³. With the abundance and diversity of microorganisms contained in wastewater, many are carriers of Mobile Genetic Elements (MGEs)⁴.

Integrating Conjugative Elements (ICEs) are self-transmissible MGEs that have the ability to integrate into their host's chromosome, replicate and transfer by conjugation^{5,6}. ICEs present in the environment can rapidly spread in bacterial populations and often carry genes that give the ICE an adaptive advantage to that environmental niche. The SXT/R391 family of ICEs is one of the most studied types of MGE with various sub-types of SXT/R391 elements reported⁷. Type 1 elements have been found in a variety of Gammaproteobacteria species including *Vibrio*, *Providencia*, *Proteus*, *Shewanella*, *Actionbacillus*, *Alteromonas*, *Escherichia coli*, *Enterovibrio*, *Photobacterium damsela*, *Idiomarinaceae bacterium* and *Marinomonas*⁸. They are distinguished by possessing a Type 1 integrase, which allows site-specific integration into the 5' end of the essential *prfC* gene⁹, integration restores a functioning *prfC* gene and encodes a new hybrid PrfC protein once integrated⁹. Type 2, 3 and 4 ICEs all

¹Department of Chemical Sciences, School of Natural Sciences, Bernal Institute, University of Limerick, Limerick, V94 T9PX, Ireland. ²School of Engineering, University of Limerick, Limerick, V94 T9PX, Ireland. ✉e-mail: michaelpryan1983@gmail.com

| Bacterial strain | Genotype/Phenotype | Source |
|--------------------------------------|--|--|
| <i>E. coli</i> AB1157 | F ⁻ , <i>thr-1</i> , <i>araC14</i> , <i>leuB6</i> , <i>_(gpt-proA)62</i> , <i>lacY1</i> , <i>tsx-33</i> , <i>qsr'-0</i> , <i>glnV44</i> , <i>galK2</i> , λ -, <i>Rac-0</i> , <i>hisG4</i> , <i>rfbC1</i> , <i>mgl-51</i> , <i>rpoS396</i> , <i>rpsL31(StrR)</i> , <i>kdgK51</i> , <i>xylA5</i> , <i>mil-1</i> , <i>argE3</i> , <i>thi-1</i> | <i>E. coli</i> genetic stock centre (CGSC), Yale University, New Haven, Connecticut, USA |
| <i>E. coli</i> J53 | Rif ^r | |
| <i>E. coli</i> CC118(λ pir) | Δ (<i>ara</i> , <i>leu</i>)7697 <i>araD139</i> Δ <i>lacX74</i> <i>galE</i> <i>galK</i> <i>phoA20</i> <i>thi-1</i> <i>rpsE</i> <i>rpoB(Rf^r)</i> <i>argE(am)</i> <i>recA1</i> λ pir ⁺ | Addgene |
| Mobile genetic element/Plasmid | Genotype | Source |
| ICER391 ⁴⁰ | Km ^r , Hg ^r | Dr. R. W. Hedges, Royal Postgraduate, Medical School, London |
| ICER997 ³² | Am ^r | Dr. R. W. Hedges, Royal Postgraduate Medical School, London |
| ICESXT ⁴⁶ | Sm ^r , Su ^r , Tm ^r | Prof. M.M. Colombo, Università La Sapienza, Rome, |
| ICEpMERPH ⁵² | Hg ^r , As ^r | National Collection of Industrial Food and Marine Bacteria |
| pBAM1 ⁵³ | Am ^r , Km ^r | Addgene |

Table 1. List of all bacterial strains, bacterial mobile genetic elements and plasmids used in this study. As^r: Arsenic resistance, Am^r: Ampicillin resistance, Rif^r: Rifampicin resistance Hg^r: Mercury resistance, Km^r: Kanamycin resistance, Sm^r: Streptomycin resistance, Su^r: Sulfamethoxazole resistance, Tm^r: Trimethoprim resistance.

insert at the 3' end of the multi-copy tRNA-Ser gene and have been found solely in *Vibrio* species^{7,10}. Another type of SXT/R391 ICE integrates into the *pabA* (para-aminobenzoate synthase) gene as with ICESh95¹¹.

The integrase encoding gene of the tyrosine recombinase class is a disguising feature of Type 1 elements. Integrase genes from several different SXT/R391 ICEs were compared to design specific primers to amplify the gene¹². Type 1 SXT/R391 ICEs show site-specific integration into the 17 bp integration site at the 5' end of the *prfC* gene^{9,13} and contain a core of 52 conserved genes that make up the 'backbone' of the element with the key functions being integration, excision and regulation¹⁴. Type 1 elements contain five hotspot regions and five variable regions which allow for gene or gene segments to be inserted, giving the element an array of potential adaptive functions¹⁴. All ICES characterised to date have been isolated serendipitously or identified from genome sequences. There have been reports of SXT/R391-like ICEs from wastewater environments, ICEVchMex1 (GQ463143.1) isolated from sewage in San Luis Potosi, Mexico¹⁵ and ICESheChnS12 (AXZL01000060.1), isolated from a wastewater treatment plant in China, 2012¹⁵. However, these were discovered following analysis rather than using a directed approach. ICEVchMex1 contains genes in all its hotspots and in one variable region and encodes ampicillin resistance¹⁵. ICESheChnS12 was found to contain genes in all its hotspot but no genes were contained in the variable regions. The hotspots contained accessory genes such as AAA ATPase and a Restriction Modification System¹⁶.

In this study, a directed approach was used to detect and isolate ICEs of the SXT/R391 family from a municipal wastewater treatment plant and to characterise one of these isolates as the first ICE detected and isolated from Ireland.

Materials and Methods

Bacterial strains, elements and media. The bacterial strains and SXT/R391 elements utilised as part of this study are listed in Table 1. Strains were stored at -80°C in either Luria-Bertani (LB) broth or M9 minimal media containing 50% (v/v) glycerol.

Collection and processing of wastewater sample for detection of ICE containing strains.

Composite 24 hr wastewater samples were collected from the inlet works at a municipal wastewater treatment plant, Ireland in May and November 2018.

Bacterial screening and molecular identification of ICE containing strains.

Serial dilutions of the collected and processed samples isolated from wastewater were carried out using a 0.85% Saline solution. Serial dilutions from 10^{-2} to 10^{-8} were prepared and samples spread on selective media. CLED media was used for the detection of *Proteus* species and TCBS media for the detection of *Vibrio* species, plates were incubated at 37°C for 24 hours. Colonies were routinely isolated and inoculated using LB broth, incubated at 37°C for 24 hours at 200 rpm. The remaining overnight culture was used to prepare glycerol stocks and stored at -80°C . Replica plating was carried out to identify single colonies containing ICE-like elements via PCR. The streak plate technique was used to isolate a single pure colony from the replica plate.

PCR. DNA extractions from isolates were performed using the DNeasy Blood & Tissue extraction kit (Qiagen) under the Gram-Negative bacterial protocol for liquid cultures. A PowerSoil DNA extraction kit (Qiagen) was used for solid materials. PCR was carried out using specifically designed primers (Table 2) to amplify the characteristic integrase gene of SXT/R391 family ICE MGEs¹².

| Primer name | Gene amplified | Nucleotide sequence (5' to 3') |
|-------------|----------------|--------------------------------|
| IntFor1 | Integrase | AAACTAGGGCTGGGCTTATAACATGGCG |
| IntRev1 | Integrase | AAAGATGGCAGCTTGCCGCAACCTC |

Table 2. List of primers used in this study with the gene amplified and the nucleotide sequence¹².

Community analysis. PCR-RFLP (Restriction Fragment Length Polymorphism) was carried to determine whether more than one ICE type might be present in the isolated samples. NEB_Cutter¹⁷ (<http://www.labtools.us/nebcutter-v3-0/>) was used to select restriction endonucleases that could be used to digest modified integrase genes. The nucleotide sequences of sequenced SXT/R391 ICEs present in the ICEBerg database⁸ and the respective integrase gene sequences were selected to determine what enzymes might be utilised to determine RFLP variation. As a result, three restriction endonucleases were selected, *HpaII*, *ApoII* and *TauI*. An *ApoII* site was not present in ICER391 (AY090559.1) *int* amplicons while a *TauI* site was not present in ICER997 (KY433363), ICESXT or ICEpMERPH (MH974755).

PCR, as described above, was carried out on selected isolates containing an amplified integrase gene, ICER391, ICER997 and ICESXT amplicons were used as controls. Additional controls consisted of the use of two different molecular weight markers (NEB 1 kb and NEB 100 kb), uncut PCR product and a control no enzyme “mock” digest. After the PCR, the restriction digest was carried out using 5 µl PCR product, 2 µl Tango Buffer, 0.5 µl of relevant restriction endonuclease and 12.5 µl Milli Q water, bringing the total volume to 20 µl. A cocktail of 15 µl was aliquoted into the PCR tube and 5 µl of the PCR product was then added before the digestion, the tubes were incubated at the appropriate temperatures, *HpaII* and *ApoII* at 37 °C and *TauI* at 65 °C. The tubes were incubated for 45 minutes in the thermocycler. To visualize the samples, 3 µl of 5X loading dye was added directly into the digest, 10 µl was loaded onto a 1.5% (w/v) agarose gel in 1X TAE Buffer. The gel was set to run for 1 hour at 100 V and photographed under UV light.

ICE transfer via triparental mating. To determine if the identified ICE could undergo conjugation tri-parental mating was carried out using the isolated strain as a host strain, pBAM1 as a helper (pBAM1 reference) and *E. coli* J53 as a recipient. The idea was to mark the ICE with a mini Tn5 (Kanamycin resistance) via pBAM1 and then mobilise the marker ICE into *E. coli* J53, a Rif resistant strain. Conjugative mating was carried out following the method by Kristensen *et al.* with some modifications¹⁸. The donor, helper and recipient strains were grown overnight. Overnight cultures were then centrifuged for 5–10 minutes at 10,000 rpm to collect cells, and washed with fresh LB broth. The cultures were mixed in a proportion of 5:1:1 (donor population was increased) and 50 µl were applied onto 0.45 µm filter on LB Agar plates. The plates were incubated at 37 °C for 24 hours. The cells were then scraped off the surface of the filter and added to a 0.85% saline solution. The culture was diluted with antibiotics for counter selection of the donor and helper cells.

The dilutions were spread onto rifampicin (200 µg.ml⁻¹) and kanamycin (50 µg.ml⁻¹) supplemented plates and grown at 37 °C for 24 hours. Rifampicin had an MIC of 200 µg.ml⁻¹ for the host strain as it was susceptible at this concentration while *E. coli* remained resistant. Controls for the donor, helper and recipient were also prepared. Transconjugants were then tested for the presence of the ICE as described above.

Genome sequencing & genome annotation. Full genome sequencing was carried out by MicrobesNG (University of Birmingham, UK) using 2 × 250 bp paired-end reads HiSeq Illumina Miseq technology giving 30X genome coverage. The genome sequences were analysed using the MicrobesNG’s automated analysis pipeline, the closest available reference was found using Kraken software (Taxonomic Sequence Classification System)¹⁹ and reads were mapped using BWA (Burrows-Wheeler Aligner) mem²⁰ software to assess the quality of the data. *De novo* assembly of the reads was carried out using SPAdes genome assembler software²¹ and the reads were mapped back to the resulting contigs using BWA mem software. The genome was identified amongst the contigs by using the BLAST tool to investigate the presence of several different ICER391 (AY090559) and ICESXT (AY055428) core scaffold genes (*int*, *jef*, *traLEKBVA*, *setCD*). The sequence data was edited on notepad ++ (<https://notepad-plus-plus.org/downloads/>), submitted to RAST (Rapid Annotation using Subsystem Technology) for annotation (see below), exported as a Genbank and Fasta file and reannotated manually with predicted Orfs blasted via BLASTP to confirm ICE association. Molecular maps were created by inputting the ICE sequence files in GenBank (.gmk) or FASTA (.faa) format into RAST²² (<http://rast.nmpdr.org/>). Gene Graphics was used to create gene representations of hotspot and variable regions by inputting the ICE sequence in GenBank (.gmk) format²³.

Phylogenetic analysis. Phylogenetic analysis of fully sequenced SXT/R391 ICEs was performed based on comparison with the concatenated amino acid sequences of 48 to 52 SXT/R391 core ICE gene-encoded proteins on all previously sequenced whole SXT/R391 elements. An unrooted phylogenetic tree was constructed by the maximum-likelihood method based on the Poisson correction model using the MEGAX²⁴. Bootstrap analysis with 1000 replications was performed to test the reliability of the tree.

Results & Discussion

Targeted isolation of SXT/R391 ICEs. Composite 24 hr raw samples isolated from wastewater (from a municipal wastewater treatment plant) were collected at two different times of the year, May 2018 (summer) and November 2018 (winter).

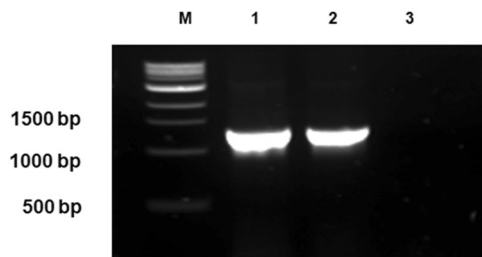


Figure 1. PCR detection of Irish ICE, M: NEB 1 kb Molecular weight marker, 1: ICEPmiIre01, 2: Positive control (AB1157ICER391) 3: Negative control (dH₂O as template). SYBR Safe stained on 1.2% agarose gel using primers IntFor1 and IntRev1.

Samples were plated on genera specific selective media as a significant number of SXT/R391 ICE MGEs have been detected in *Proteus* and *Vibrio* species⁷. After plates were incubated overnight, colonies of interest were swabbed and inoculated in LB Broth, the overnight cultures were then used for DNA extraction and PCR was performed to detect the presence or absence of an ICE specific integrase gene¹² indicating the presence of an SXT/R391 ICE MGE containing host. Screening and re-plating (once PCR positive amplicons were detected) were carried out over several weeks.

Following multiple screening, and replica plating (see Fig. 1 in Supplementary Material) followed by repeated PCR analysis, the integrase gene was detected by PCR in three translucent blue extracts, ULP001, ULP004 and ULP014 identified on CLED selective media. Translucent blue colonies were screened for the integrase gene and, if detected, samples were streaked out to isolate a single pure colony. A potential SXT/R391-like ICE was detected following further screening from one *Proteus* species, selectively grown on CLED media. A PCR amplicon was detected at 1378 bp matching the positive control, AB1157ICER391, see Fig. 1. From the screened single colony, full sequence analysis of this isolate determined that the host was a *Proteus* species based on analysis of its 16S rDNA sequence. The novel ICE was named ICEPmiIre01 as per standard ICE nomenclature²⁵. *Proteus* isolates isolated in the winter also showed positive amplification for the integrase gene, indicating that the presence of SXT/R391 ICE-like elements is apparently not season dependent.

Community analysis. RFLP was used to determine whether there might be more than one SXT/R391 ICE type present in the wastewater. The integrase specific PCR amplicon generated by IntFor1 and IntRev1 is normally 1378 bp in length. Alignments of numerous ICE integrase genes sequenced from SXT/R391 family members within the ICEberg database⁸ revealed some heterogeneity in gene sequence. Analysis of this heterogeneity using NEBcutter restriction site analysis suggested that using the restriction enzymes *Hpa*II, *Tau*I and *Apo*II could reveal at least some of this RFLP heterogeneity. Therefore, a number of samples were analysed by amplifying putative *int* gene amplicons from wastewater isolates and then subjecting the amplicons to restriction analysis via *Hpa*II, *Tau*I and *Apo*II. We demonstrated that *Hpa*II and *Tau*I were able to cut amplicons generated from samples isolated from wastewater (but not *Apo*II). *Hpa*II cleaved two types of environmental SXT/R391 ICE integrase genes amplified, one without an *Hpa*II site and one with an *Hpa*II site (each amplicon 1378 bp in length), *Tau*I cleaved one SXT/R391 ICE (two amplicons equal to 1378 bp in length) as it contained a *Tau*I site and *Apo*II did not show any cleavage of amplicons from isolates in our case.

This limited RFLP yield suggested that only a limited number of amplicon types were present, and indeed possibly only two different types (see Fig. 2, lane 1, where isolate ULP014 when amplified shows two ICE types each at 1378 bp in size, one that contains the *Hpa*II site and one that does not.

Full sequence analysis. ICEPmiIre01 (was identified among 271 contigs with a mean fold coverage of 14.98 by searching for the integrase gene and blasting against the prototype ICER391 (AY090559.1) The ICE was found on node 1 (287508 bp in length), this node revealed a novel SXT/R391-like ICE element associated with *Proteus mirabilis*. 16S rDNA analysis of the host-based on full genome analysis revealed a 100% match to *Proteus mirabilis*. Previously SXT/R391 ICE elements have been detected in *Proteus mirabilis* from China but isolated from clinical and animal samples generally associated with illness^{26,27}. *Proteus* species are known to be human and animal opportunistic pathogens, and frequently isolated from urine, wounds and other clinical sites and is present in soil and water habitats possibly from spreading from clinical sources²⁸.

This novel SXT/R391 ICE is of the Type 1 ICE variety based on comparative genomics⁶ and is 81 kb in length with prediction of 75 orfs based on RAST annotation. A full genome map can be seen in Fig. 3 Comparative genomics revealed that ICEPmiIre01 contains 51 core genes, sharing the core molecular backbone and synteny of other SXT/R391 ICEs. A full list of *orfs* detected is presented in Supplementary Table 1 with % similarity to ICER391 and ICESXT illustrated. The core genes of ICEPmiIre01 are predicted to encode activities for basic ICE functionality such as integration, excision and conjugative transfer²⁹. ICEPmiIre01 is one of the smallest SXT/R391-like ICE MGEs to be discovered thus far, the smallest being ICEPmiChn3 (KY437727) at 57 kb in length also from a *Proteus mirabilis*, isolated from broiler carcasses in China²⁶.

Structural analysis of ICEPmiIre01 revealed that it contains five hotspot regions and further insertions into one variable region. Some of the hotspot regions contain genes putatively encoding resistance to bacteriophage attack. This is not surprising as there is a continuous arms race between bacteria and phage and such resistance

I precursor protein (AQT24161.1) present in ICER997 (KY433363.1). Endonucleases are thought to generate double-strand breaks in DNA. The biological role associated with this hotspot is unknown³⁷.

HS5 contains eight predicted genes, six of which make up the novel putative BREX (bacteriophage exclusion) system found in ICE*PmiIre01*. This system has characterised homologs that allow for phage adsorption but blocks phage DNA replication. A similar BREX system was first identified in the genome of *Bacillus subtilis* H3081.97 (NZ_ABDL01000007.1) and there confers resistance to a broad range of phage both virulent and temperate³⁰. The BREX six gene cassette in *Bacillus subtilis* includes a putative Lon-like protease, a putative alkaline phosphatase protein, a putative RNA-binding protein, a putative DNA methylase, a putative ATPase domain-containing protein and a protein of unknown function³⁰. The six genes that make up the BREX system in ICE*PmiIre01* are *orf15*, a putative inner membrane protein (BrxA) that is 100% identical to *orf23*, a hypothetical protein (AAM08040.1) found in ICER391. BrxA shares structural homology with the RNA-binding antitermination protein NusB³¹. The second gene *orf16* is a putative DUF1788 domain-containing protein (BrxB) which is 100% identical to *orf24*, a hypothetical protein (AAM08046.1) contained in ICER391, the function of this gene is unknown³⁰. The third gene, *orf17* a putative BREX system P-loop protein (BrxC), is 100% identical to *orf25* a putative ATPase (AAM07998.1) present in ICER391. BrxC contains a putative large ATP binding motif³⁰. The fourth gene *orf19* encodes a putative BREX-1 system adenine-specific DNA-methyltransferase (PglX), showing 100% identity to *orf26*, a hypothetical protein (AAM08017.1) found in ICER391. This gene homolog encodes a protein that has putative DNA modifying functions that enable BREX to differentiate between host and phage DNA³⁰. The fifth gene is *orf20* encoding a putative BREX-1 system phosphatase protein (PglZ, type A), showing 93% identity to *orf30*, a hypothetical annotated protein (AAM07999.1) present in ICER391. The final gene within the system *orf21* encodes a putative a Lon-related BREX system protease (BrxL), showing 97% identity to *orf31*, a putative ATP-dependent Lon protease protein (AAM08002.1) found in ICER391. Lon proteases are ATP-dependent serine proteases that mediate the selective degradation of mutant and abnormal proteins, as well as certain short-lived regulatory proteins and is required for cellular homeostasis and for survival from DNA damage and development changes induced by stress³⁸. Other genes present in the HS5 include *orf18*, encoding a hypothetical protein 100% identical to a hypothetical protein (EKA96589.1) found in *Proteus mirabilis* WGLW6 and *orf22* a putative DNA repair protein showing 100% identity to *orf32*, a hypothetical protein (AAM08000.1) present in ICER391. A similar BREX system appears to be present in ICER391 (AY090559)³⁹ and ICE *VpaCan1* (CP028481)⁷ based on homology reported above but to our knowledge has not been annotated as a putative BREX system previously. Based on the observations above and based on comparative analysis with other SXT/R391 family members it is clear that ICE*PmiIre01* is a mosaic of core ICE genes and variable genes located in hotspots and variable regions of the element. These genes are most likely accumulated by ICE encoded recombination systems which generated a mosaic-like genetic structure³⁹. It is interesting to observe elements of this mosaic structure being found in disparate ICE-like elements isolated from different global locations as outlined above.

ICE*PmiIre01* has an insertion in Variable Region 1 (VR1). Two predicted genes are present *orf01*, encoding a putative transcriptional regulator of the XRE family, which shows 100% identity to a hypothetical protein (AAM08068.1) present in ICER391 (AY090559.1) and *orf02*, encoding a putative Serine/threonine-protein kinase HipA, which is 100% identical to a hypothetical protein (AAM08010.1) also present in ICER391 (AY090559.1). The putative transcriptional regulator of the XRE family, *orf01* is also present in ICEBs1 and is proposed to repress the transcription of genes required for excision and transfer and prevents the acquisition of additional ICE copies, therefore, conferring immunity⁴⁰. HipA-like proteins are expressed by various bacterial species and are involved in high-frequency persistence to the lethal effects of inhibition of either DNA or peptidoglycan. When expressed in high amounts alone, HipA is toxic to bacterial cells⁴¹ and may also be involved in multidrug tolerance⁴². These genes are commonly associated with VR1 of several other SXT/R391-family ICEs (see comparative genomics section).

ICE*PmiIre01* does not have genes present in variable regions VR1, VR2, VR3 and VR4 as occurs in other SXT/R391 elements and this may be reflect its small size relative to other SXT/R391 ICEs.

Accession number. The designated GenBank accession number for the nucleotide sequence ICE*PmiIre01* is MN520463.

Transfer of ICE*PmiIre01*. Due to not detecting a selectable marker associated with the ICE it was not possible to demonstrate conjugation of ICE*PmiIre01* in the normal way. Therefore a tri-parental mating was undertaken using the ICE*PmiIre01* containing *P. mirabilis* host (susceptible to rifampicin), *E. coli* CC118λ pBAM1, containing a miniTn5 with kanR and *E. coli* J53 as described above. We isolated putative strains of *E. coli* J53 containing ICE*PmiIre01* containing the transposed miniTn5. Transconjugants were then analysed to validate and confirm the conjugative transfer. Transfer frequency of the element can be seen in Table 3.

Phylogenetic analysis. A phylogenetic tree (Fig. 4) was constructed based on the concatenated amino acid sequences of all SXT-R391 core proteins for all published core genome sequences of these elements. ICE*PmiIre01* is clustered with ICE*PmiIpn1*, ICER391, ICE*IdBUS*A1 and ICE*PrSt33672*. These results illustrate the wide geographic spread of Type 1 SXT/R391 ICE-like elements with these near neighbours being isolated in Japan, South Africa and the USA respectively. Information on SXT/R391 element in Fig. 4 can be found in Supplementary Table 4.

Comparative analysis: bacteriophage defence mechanisms. As a result of the abundance of bacteriophage present in wastewater and other environmental niches, bacteria have developed various mechanisms to resist their attack. These resistance mechanisms include RMS, toxin-antitoxin (T-A) system, Abi systems³⁵ and

| Donor | Helper | Recipient | Frequency of transfer |
|---|----------------------------------|----------------------|------------------------|
| <i>Proteus mirabilis</i> (ICE <i>PmiIre01</i>) 10 ⁷ | <i>E. coli</i> CC118λpir (pBAM1) | <i>E. coli</i> (J53) | 2.8 × 10 ⁻⁵ |

Table 3. Transfer of frequency- calculated per donor cell.

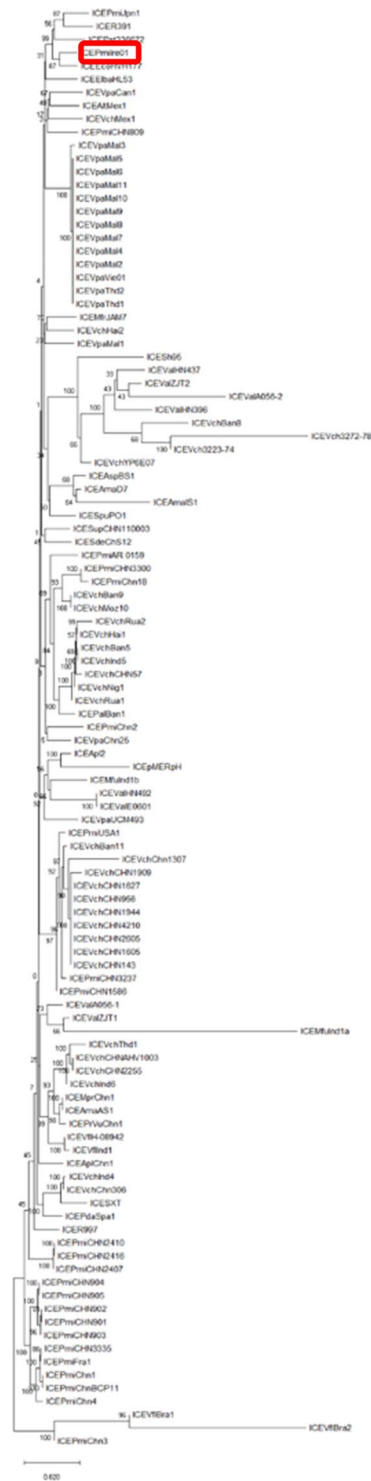


Figure 4. Phylogenetic tree from the maximum-likelihood analysis of the core concatenated proteins of 112 SXT/R391 ICEs including the novel isolate ICEPmiIre01 illustrating its relationship with other family members.

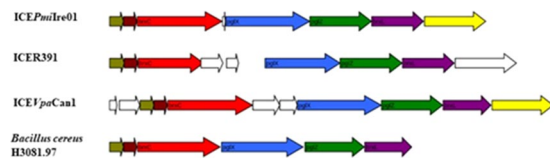


Figure 5. Comparative analysis of the BREX systems present in ICEPmiIre01, ICER391 and ICEVpaCan1, located in HS5 of these ICE elements. Genes coloured moss are associated with brxA (predicted RNA-binding domain), genes coloured brown associated with brxB (unknown function), genes coloured light blue are associated with pglX (DNA methylation domain), genes coloured dark green are associated with pglZ (alkaline phosphatase domain) and brxL (lon-like protease domain), genes coloured yellow are associated with a putative DNA repair and genes coloured white of unknown function.

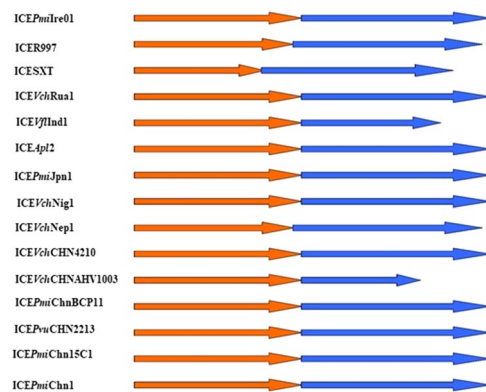


Figure 6. Examples of the genetic organization of HS2 of SXT/R391 ICE elements containing a T-A system, including ICEPmiIre01, ICER997³¹, ICESXT⁴⁵, ICEVchRua1⁴⁶, ICEVflInd1¹⁴, ICEApl2⁴⁷, ICEPmiJpn1⁴⁸, ICEVchNig1⁴⁹, ICEVchNep1⁴⁹, ICEVchCHN4210⁵⁰, ICEVchCHNAHV1003⁵⁰, ICEPmiChnBCP11⁵¹, ICEPvuCHN2213²⁷, ICEPmiChn15C1 (direct submission) and ICEPmiChn1⁴⁸. Genes coloured orange are associated with ynd and genes coloured blue are associated with ync. Both genes encode a toxin-antitoxin system.

BREX systems³⁰. The putative BREX system located in HS5 of ICEPmiIre01 is related to a novel bacteriophage defence system first reported by Goldfarb *et al.* (2015)³⁰. It consists of a six gene cassette which includes a putative Lon-like protease, an alkaline phosphatase domain protein, a putative RNA-binding protein, a DNA methylase, an ATPase-domain protein and a protein of unknown function. This system shows high levels of homology and synteny with the putative ICEPmiIre01 system.

By comparing each gene in the putative ICEPmiIre01 BREX system it was discovered that a near-identical unannotated system was also located in HS5 of ICER391 the prototype SXT/R391 ICE (showing between 93–100% homology)³⁹. A recently discovered SXT/R391 ICE, ICEVpaCan1, was also found to contain this putative system⁷ with homology levels between 68–99%. A comparison between the putative HS5 BREX system of ICEPmiIre01 demonstrated high (84–100%) homology to the BREX system found in *Bacillus cereus* H3081.9. This is most surprising as SXT/R391 ICEs are exclusively found in Gram-negative bacteria and finding a homologous system in a Gram-positive organism with such high levels of homology is unusual (Fig. 5).

The insertion in HS2 was also compared to ICER391 as it contains a putative T-A system and is commonly associated with this hotspot in other ICEs. T-A systems consist of a stable toxin and its labile antitoxin⁴³ that promote the maintenance of SXT/R391 ICEs, genes associated with this function are *mosT* and *mosA* (*ynd*, *ync*)⁴⁴. This system is found in numerous SXT/R391 ICE elements and is generally also found inserted in HS2¹⁵. Examples of this system are displayed in Fig. 6 and compared with ICEPmiIre01 found in this study.

Conclusion

Using a directed approach, a novel environmental SXT/R391 Type 1 family ICE, ICEPmiIre01 was isolated from a local municipal wastewater by using selective media for *Proteus* and *Vibrio* strains (common hosts for SXT/R391 ICEs) and monitoring the presence of ICEs using ICE specific *int* primers amplified by PCR. A *Proteus mirabilis* ICEPmiIre01 containing strain isolated in this way was subjected to genome sequencing and revealed an 81 kb element with 75 putative *orf*s. The element contained the 51 core ICE genes and contained insertions in some of its hotspots which indicated the presence of an encoded bacteriophage resistant mechanisms. A novel BREX system was identified in HS5, not previously identified or annotated in other SXT/R391 ICEs. Comparative analysis revealed that this system is also present in ICER391 but was not previously annotated there⁴⁰ and in ICEVpaCan1 again not previously annotated⁷. Unusually the nearest BREX homolog to ICEPmiIre01 was reported in a *Bacillus subtilis* isolate which given the differences between Gram-positive and Gram-negative hosts is surprising. ICEPmiIre01 is a novel SXT/R391 ICE-like element with interestingly no antibiotic or metal resistance, the first to be identified in Ireland.

Received: 3 December 2019; Accepted: 29 April 2020;

Published online: 26 May 2020

References

- Gerardi, M.H. & Zimmerman, M.C. Wastewater pathogens, John Wiley & Sons. (2004).
- Ye, L. & Zhang, T. Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. *Appl. Microbiol. Biotechnol.* **97**, 2681–2690 (2013).
- McLellan, S. L., Huse, S. M., Mueller-Spitz, S. R., Andreishcheva, E. N. & Sogin, M. L. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environ. Microbiol.* **12**, 378–392 (2010).
- Guo, J., Li, J., Chen, H., Bond, P. L. & Yuan, Z. Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water Res.* **123**, 468–478 (2017).
- Burrus, V. & Waldor, M. K. Shaping bacterial genomes with integrative and conjugative elements. *Res. Microbiol.* **155**, 376–386 (2004).
- Ryan, M. P., Pembroke, J. T. & Adley, C. C. Novel Tn 4371-ICE like element in *Ralstonia pickettii* and Genome mining for comparative elements. *BMC Microbiol.* **9**, 242 (2009).
- Bioteau, A., Durand, R. & Burrus, V. Redefinition and unification of the SXT/R391 family of integrative and conjugative elements. *Appl. Environ. Microbiol.* **84**, e00485–18 (2018).
- Liu, M. *et al.* ICEberg 2.0: An updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* **47**, D660–D665 (2019).
- Burrus, V., Marrero, J. & Waldor, M. K. The current ICE age: Biology and evolution of SXT-related integrating conjugative elements. *Plasmid* **55**, 173–183 (2006).
- McGrath, B. M. & Pembroke, J. T. Detailed analysis of the insertion site of the mobile elements R997, pMERPH, R392, R705 and R391 in *E. coli* K12. *FEMS Microbiol. Lett.* **237**, 19–26 (2004).
- Di Noto, G. P., Jara, E., Iriarte, A., Centrón, D. & Quiroga, C. Genome analysis of a clinical isolate of *Shewanella* sp. Uncovered an active hybrid integrative and conjugative element carrying an integron platform inserted in a novel genomic locus. *Microbiol. (United Kingdom)* **162**, 1335–1345 (2016).
- McGrath, B. M., O'Halloran, J. A., Piterina, A. V. & Pembroke, J. T. Molecular tools to detect the IncJ elements: A family of integrating, antibiotic resistant mobile genetic elements. *J. Microbiol. Methods* **66**, 32–42 (2006).
- Ryan, M. P., Armshaw, P. & Pembroke, J. T. SXT/R391 integrative and conjugative elements (ICEs) encode a novel 'trap-door' strategy for mobile element escape. *Front. Microbiol.* **7**, 829 (2016).
- Wozniak, R. A. F. *et al.* Comparative ICE genomics: Insights into the evolution of the SXT/R391 family of ICEs. *Plos Genet.* **5**, e1000786 (2009).
- Burrus, V., Quezada-Calvillo, R., Marrero, J. & Waldor, M. K. SXT-related integrating conjugative element in new world *Vibrio cholerae*. *Appl. Environ. Microbiol.* **72**, 3054–3057 (2006).
- Xu, M. *et al.* Draft genome sequence of *Shewanella decolorationis* S12, a dye-degrading bacterium isolated from a wastewater treatment plant. *Genome Announc.* **1**, e00993–13 (2013).
- Vincze, T., Posfai, J. & Roberts, R. J. NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Res.* **31**, 3688–3691 (2003).
- Kristensen, C. S. *et al.* Site-specific deletions of chromosomally located DNA segments with the multimer resolution system of broad-host-range plasmid RP4. *J. Bacteriol.* **177**, 52–58 (1995).
- Wood, D. E. & Salzberg, S. L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
- Aziz, R. K. *et al.* The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
- Harrison, K. J., Crécy-Lagard, V. & Zallot, R. Gene Graphics: A genomic neighbourhood data visualization web application. *Bioinformatics* **34**, 1406–1408 (2018).
- Kumar, S., Stecher, G., Li, M., Niyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
- Burrus, V., Marrero, J. & Waldor, M. K. The current ICE age: Biology and evolution of SXT-related integrating conjugative elements. *Plasmid* **55**, 173–183 (2006).
- Bie, L., Wu, H., Wang, X. H., Wang, M. & Xu, H. Identification and characterization of new members of the SXT/R391 family of integrative and conjugative elements (ICEs) in *Proteus mirabilis*. *Int. J. Antimicrob. Agents* **50**, 242–246 (2017).
- Li, X. *et al.* SXT/R391 integrative and conjugative elements in *Proteus* species reveal abundant genetic diversity and multidrug resistance. *Sci. Rep.* **6**, 37372 (2016).
- Drzewiecka, D. Significance and Roles of *Proteus* spp. Bacteria in Natural Environments. *Microb. Ecol.* **72**, 741–758 (2016).
- Armshaw, P. & Pembroke, J. T. Integrative conjugative elements (ICEs) of the SXT/R391 group as vehicles for acquisition of resistance determinants, stable maintenance and transfer to a wide range of enterobacterial pathogens. In: *Microbial Pathogens and Strategies for Combating Them: Science, Technology and Education*, Ed. Méndez-Vilas, A. p439–446 (Formatex Research Center, 2013).
- Goldfarb, T. *et al.* BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* **34**, 169–183 (2015).
- Ryan, M. P., Armshaw, P., O'Halloran, J. A. & Pembroke, J. T. Analysis and comparative genomics of R997, the first SXT/R391 integrative and conjugative element (ICE) of the Indian Sub-Continent. *Sci. Rep.* **7**, 8562 (2017).
- Chopin, M. C., Chopin, A. & Bidnenko, E. Phage abortive infection in lactococci: Variations on a theme. *Curr. Opin. Microbiol.* **8**, 473–479 (2005).
- Dila, D., Sutherland, E., Moran, L., Slatko, B. & Raleigh, E. A. Genetic and sequence organization of the *mcrBC* locus of *Escherichia coli* K-12. *J. Bacteriol.* **172**, 4888–4900 (1990).
- Tock, M. R. & Dryden, D. T. F. The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* **8**, 466–472 (2005).
- Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327 (2010).
- Constantinesco, E., Forterre, P., Koonin, E. V., Aravind, L. & Elie, C. A bipolar DNA helicase gene, *herA*, clusters with *rad50*, *mre11* and *nurA* genes in thermophilic archaea. *Nucleic Acids Res.* **32**, 1439–1447 (2004).
- Jekel, M. & Wackernagel, W. The periplasmic endonuclease I of *Escherichia coli* has amino-acid sequence homology to the extracellular DNases of *Vibrio cholerae* and *Aeromonas hydrophila*. *Gene* **154**, 55–59 (1995).
- Lee, I., Berdis, A. J. & Suzuki, C. K. Recent developments in the mechanistic enzymology of the ATP-dependent Lon protease from *Escherichia coli*: Highlights from kinetic studies. *Mol. Biosyst.* **2**, 477–483 (2006).
- Böltner, D., MacMahon, C., Pembroke, J. T., Strike, P. & Osborn, A. M. R391: A conjugative integrating mosaic comprised of phage, plasmid, and transposon elements. *J. Bacteriol.* **184**, 5158–5169 (2002).
- Tirumalai, M. R. & Fox, G. E. An ICE Bsl-like element may be associated with the extreme radiation and desiccation resistance of *Bacillus pumilus* SAFR-032 spores. *Extremophiles* **17**, 767–774 (2013).

41. Karkaria, C. E., Steiner, R. F. & Rosen, B. P. Ligand Interactions in the ArsA Protein, the Catalytic Component of an Anion-Translocating Adenosinetriphosphatase. *Biochemistry* **30**, 2625–2628 (1991).
42. Schumacher, M. A. *et al.* Role of Unusual P Loop Ejection and Autophosphorylation in HipA-Mediated Persistence and Multidrug Tolerance. *Cell Rep* **2**, 518–525 (2012).
43. Van Melderen, L. Toxin-antitoxin systems: Why so many, what for? *Curr. Opin. Microbiol.* **13**, 781–785 (2010).
44. Wozniak, R. A. F. & Waldor, M. K. A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genet.* **5**, e1000439 (2009).
45. Beaver, J. W., Hochhut, B. & Waldor, M. K. Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *J. Bacteriol.* **184**, 4259–4269 (2002).
46. Kuleshov, K. V. *et al.* Comparative genomic analysis of two isolates of *Vibrio cholerae* O1 Ogawa El Tor isolated during outbreak in Mariupol in 2011. *Infect. Genet. Evol.* **44**, 471–478 (2016).
47. Li, Y. *et al.* Characterization of the *Actinobacillus pleuropneumoniae* SXT-related integrative and conjugative element ICE *Apl2* and analysis of the encoded FloR protein: Hydrophobic residues in transmembrane domains contribute dynamically to florfenicol and chloramphenicol efflux. *J. Antimicrob. Chemother.* **73**, 57–65 (2018).
48. Lei, C. W. *et al.* Characterization of SXT/R391 integrative and conjugative elements in *Proteus mirabilis* isolates from food-producing animals in China. *Antimicrob. Agents Chemother.* **60**, 1935–1938 (2016).
49. Marin, M. A., Fonseca, E. L., Andrade, B. N., Cabral, A. C. & Vicente, A. C. P. Worldwide occurrence of integrative conjugative element encoding multidrug resistance determinants in epidemic *Vibrio cholerae* O1. *PLoS One* **9** (2014).
50. Wang, R., Yu, D., Yue, J. & Kan, B. Variations in SXT elements in epidemic *Vibrio cholerae* O1 El Tor strains in China. *Sci. Rep.* **6**, 22733 (2016).
51. Lei, C. W., Chen, Y. P., Kang, Z. Z., Kong, L. H. & Wang, H. N. Characterization of a novel SXT/R391 integrative and conjugative element carrying *ctr*, *bla* CTX-M-65, *fosA3*, and *aac(6)-Ib-cr* in *Proteus mirabilis*. *Antimicrob. Agents Chemother.* **62** (2018).
52. Ryan, M. P., Slattery, S. & Pembroke, J. T. A Novel Arsenate-Resistant Determinant Associated with ICEpMERPH, a Member of the SXT/R391 Group of Mobile Genetic Elements. *Genes*. **10**, 1048 (2019).
53. Martínez-García, E., Calles, B., Arévalo-Rodríguez, M. & de Lorenzo, V. pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol.* **11**, 38 (2011).

Acknowledgements

We would like to thank the Department of Chemical Sciences, University of Limerick for their support. Genome sequencing was provided by MicrobesNG [<http://www.microbesng.uk>], which is supported by the BBSRC [Grant Number BB/L024209/1].

Author contributions

S.S., M.P.R. and J.T.P. designed and performed the experiments, analysed the data, and drafted the article. J.G.M. assisted in performing the experiments and drafted the article.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-65216-5>.

Correspondence and requests for materials should be addressed to M.P.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020