Universidad de Cantabria

Facultad de Medicina

Departamento de Biología Molecular

Evaluación de la actividad recombinasa e integrasa sitio-específica de relaxasas conjugativas en bacterias y en células humanas

Coral González Prieto

Santander, 2015

University of Cantabria

School of Medicine

Molecular Biology Department

Evaluation of site-specific recombinase and integrase activity of conjugative relaxases in bacterial and human cells

Coral González Prieto

Santander, 2015

Dña. **Matxalen Llosa Blas**, Profesora Titular de Genética de la Facultad de Medicina de la Universidad de Cantabria,

CERTIFICA: que Dña. **Coral González Prieto** ha realizado bajo mi dirección el trabajo que lleva por título "Evaluación de la actividad recombinasa e integrasa sitio-específica de relaxasas conjugativas en bacterias y en células humanas".

Considero que dicho trabajo se encuentra terminado y reúne los requisitos necesarios para su presentación como Memoria de Doctorado, al objeto de poder optar al grado de Doctor.

Santander, a 5 de Junio de 2015

Fdo. Matxalen Llosa Blas

El presente trabajo ha sido realizado entre el Departamento de Biología Molecular de la Universidad de Cantabria y el Departamento de Microbiología y Genómica del Instituto de Biomedicina y Biotecnología de Cantabria, bajo la dirección de la Dra. Matxalen Llosa Blas y gracias a una beca predoctoral de la Universidad de Cantabria.

Durante este periodo se han realizado estancias de tres meses en los laboratorios del Dr. Christoph Dehio (Biozentrum de la Universidad de Basilea, Basilea; Suiza), gracias a una ayuda de la Universidad de Cantabria; y del Dr. Manfred Schmidt (DKFZ, Heidelberg; Alemania), gracias a una beca EMBO Short-Term.

A mi familia

Aquí va, la que es, sin duda, la parte más leída de cualquier tesis. Tengo mucho que agradecer a tanta gente que de una forma u otra a contribuido a ella que, por favor, perdonadme si me olvido de alguien (¡espero que no!).

En primer lugar, mi agradecimiento es para mi directora, Matxalen. Por aceptarme en el laboratorio allá por el 2008, y todavía más, por querer repetir al año siguiente y abrirme después las puertas del laboratorio para hacer la tesis. Durante estos años, has confiado en mí y has esperado pacientemente a que yo llamara a la puerta del despacho tras haber intentado las cosas por mi misma. Creo que así he aprendido mucho más. Y no menos importante, gracias por ser jefa pero también persona, por preocuparte y tener siempre unas palabras de ánimo.

A mis compañeros de laboratorio. A Leti, de ti aprendí lo que significa el trabajo bien hecho, gracias por ello. A Anabel, por los buenos tiempos. A Esther, mi Esterina. No tengo hojas para agradecerte todo lo que debería. Gracias por tu optimismo, por tu infinita ayuda y, sobre todo, gracias por tu amistad. Gracias por seguir ahí cada día, incluso en la distancia. Te echo de menos. A Elena, gracias por ese año compartido. Lo disfruté un montón, tanto dentro como fuera del labo. Saúl, ha sido un placer compartir este poquito tiempo contigo, ¡mucha suerte!. Gracias también a las técnicos que en un momento u otro me han ayudado tanto. A Patri, gracias por dedicar aquel primer mes a enseñarme, aun cuando no sabía ni lo que era hacer un clon. A Elena, gracias por tu enorme ayuda con todas las conjugaciones que hicimos a medias; y por la alegría que traías cada día al laboratorio.

A toda la gente de Genética, gracias por acogerme cada día como una más. A Mapi, por tu bondad y tu sabiduría, por compartirlas conmigo. Gracias a todos aquellos con los que he compartido tantas comidas y cafés, tantos planes fuera del labo. A Jorge, a Lillo, a Yolanda, a Carolina, por cada detalle. A Alej, por ser tan buenazo. Aun desde Inglaterra eres quien nos mantiene unidos. A Omar, porque tus locuras hacen mucho más ameno el día a día. A María y a Maris, por todas vuestras visitas al labo vecino, por compartir los buenos días y escucharme en los malos. Porque aun siendo tan distintas, me complementáis ambas tan bien. En definitiva, por estar ahí.

Gracias también a la gente del laboratorio de Micro, por estar siempre tan dispuestos a ayudar, por las ideas en los seminarios. Gracias especialmente a Félix, por salvarme siempre que voy al despacho pidiendo ayuda, lo que ocurre frecuentemente. Y como no, por todos los pedidos firmados.

A toda la gente de “cafeses”, por los planes compartidos. A Iago, por preguntarme siempre qué tal voy, por escucharme, ¡y por tu afición a la repostería!. Eres un amor.

A Juanje y Maite, por vuestra amistad desde el inicio del máster. Gracias por preocuparos siempre por cómo estoy y cómo van los experimentos. Gracias por hacerme partícipe de grandes momentos. ¡Y los que nos quedan!

En general, a todos los compañeros del departamento de Biología Molecular y del IBBTEC. Ha sido un placer trabajar con vosotros. Gracias a los jefes, por abrirme las puertas de los laboratorios siempre que lo he necesitado, y a los becarios, por estar siempre dispuestos a ayudar desinteresadamente. Gracias en especial a “los bioquímicos” y “los Pieros”, por vuestra ayuda con los cultivos celulares y la qPCR.

To Christoph, thank you for opening your lab to me. I learnt a lot from all of you and spent three terrific months in Basel. Thank you to Maxime, for spending your time teaching me while writing your thesis. To Alexander, for being such a good labmate.

To Manfred, thank you for the opportunity to join your lab. Thanks to all the people in the lab for the help every day. Thank you Richard, for teaching me and making my life in the lab easier. To Eliana, thank you for welcoming me and being so kind to me. A Irene, por aportar el toque español. Gracias por las comidas y las tardes de poyata compartidas.

A las bioquímicas, Bea, Martius, Maria (L) y Eva. Porque fuisteis las culpables de que ir a clase fuese TAN divertido, por las manzanillas de sobremesa, los ornitorrincos y tantas otras cosas. Fueron unos años geniales gracias a vosotras. Y, sobre todo, por seguir ahí; por los ánimos, las comidas navideñas y las reuniones en Asturias, en Madrid o en Berlín.

A mi familia, porque sois, sin duda, los que más me habéis sufrido durante esta tesis. Gracias por vuestra ayuda incondicional y vuestra paciencia. Gracias también por darme el espacio que necesito cuando no he tendido un buen día, y por comprenderlo. El haber llegado hasta aquí ha sido sin duda posible por el apoyo que me habéis brindado siempre. Papá, mama, gracias por darlo siempre todo por nosotras. Noe, gracias por ser tan buen ejemplo y por estar siempre tan unida a mi. Eres la mejor hermana que pudiera imaginar. Os quiero. Gracias también a Alicia y Manolín, ya que sin vosotros y todo lo que habéis hecho por mi no habría ni empezado esta tesis. Gracias por hacerme sentir una más en casa desde el primer día.

Coral

Index

Abbreviations

aa	Amino acid
AAV	Adeno-associated virus
Ap	Ampicillin
Bep	*Bartonella* effector protein
BID	Bep intracellular delivery
bp	Base pair
cat	Chloramphenicol acetyltransferase gene
cfu	Colony-forming unit
Cm	Chloramphenicol
CP	Coupling protein
CRISPR	Clustered regularly interspaced short palindromic repeats
DAP	Diaminopimelic acid
DNA	Deoxyribonucleotide acid
dNTP	Deoxyribonucleotide triphosphate
DR	Directed repeat
DSB	Double-strand break
dsDNA	Double stranded DNA
Dtr	DNA transfer and replication
Em	Erythromycin
G418	Geneticin
gDNA	Genomic DNA
Gm	Gentamicin sulphate
HR	Homologous recombination
HUH	Histidine-hydrophobic-histidine
Hyg	Hygromycin B
IHF	Integration host factor
Inc	Incompatibility group
IPTG	Isopropyl-β-D-thiogalactopyranoside
IE	Integration event
IJ	Integration junction
IR	Inverted repeat

IS	Integration site
Kb	Kilobase
kDa	Kilodalton
Km	Kanamycin monosulphate
LAM-PCR	Linear-amplification mediated PCR
MOI	Multiplicity of infection
Mpf	Mating pair formation
Mps	Mating pair stabilization
nic	Nicking site in the *oriT*
nt	Nucleotide
Nx	Nalidixic acid
OD	Optical density
o/n	Overnight
ORF	Open reading frame
oriT	Origin of transfer
$oriT_F$	F *oriT*
$oriT_P$	RP4 *oriT*
oriV	Origin of replication
$oriT_W$	R388 *oriT*
PCR	Polimerase chain reaction
RCR	Rolling circle replication
Rf	Rifampicin
RHH	Ribbon-helix-helix
r.p.m.	Revolutions per minute
scDNA	Supercoiled DNA
Sm	Streptomycin
Sp	Spectinomycin
ssDNA	Single-stranded DNA
SSR	Site-specific recombinase
Su	Sulfonamide
T4S	Type IV secretion
T4SS	Type IV secretion system
TALEN	Transcription activator-like effectors nuclease

Thy	Thymidine
TKC	Trans-kingdom conjugation
Tp	Trimethoprim
TRA	Conjugative transfer region
Tyr	Tyrosine
U	Enzyme unit
wt	Wild-type
X-gal	5-bromo-4-chloro-3- indolyl-beta-D-galactopyranoside
ZFN	Zinc-finger nuclease

Introduction

1. Type IV Secretion Systems

Bacterial Type IV Secretion Systems (T4SS) constitute a family of molecular transporters able to deliver a high variety of molecules through the bacterial membrane (Alvarez-Martinez & Christie, 2009; Zechner *et al.*, 2012). They show remarkable plasticity in terms of the nature of the substrate to be secreted because proteins, DNA, or nucleoprotein complexes can be transferred. The final destination for that cargo is also variable because Type IV substrates can be targeted to the extracellular milieu, into other bacteria, or into eukaryotic cells (either of plant or animal origin). This versatility allows them to be involved in a variety of biological processes such as DNA transfer among bacteria, known as bacterial conjugation (Schroder & Lanka, 2005), or effector protein translocation into human cells (Llosa *et al.*, 2009).

According to their main function, T4SS can be classified in three subfamilies (**Figure 1**) (Backert & Meyer, 2006):

a) T4SS involved in conjugation. They are found in gram-negative and gram-positive bacteria and form part of the machinery that mediates DNA transfer from a donor to a recipient bacterium by a contact-dependent process. Some of them are also able to transport DNA to fungi and plants (Bundock *et al.*, 1995; Zupan *et al.*, 2002). Conjugation is an important mechanism of horizontal gene transfer, contributing to genome plasticity. For modern medicine, it is becoming an important problem because of the dissemination of conjugative plasmids and mobile elements, which usually constitute a reservoir for antibiotic resistance genes, what contributes to the emergence of resistant bacterial populations (Davies & Davies, 2010).

b) T4SS that mediate DNA uptake and release. The members of this family mediate the transfer of DNA between the bacteria and the external medium. It includes the ComB T4SS of *Helicobacter pylori* that takes DNA from the extracellular medium and Tra system of *Neisseria gonorrhoeae* that secretes DNA to the milieu. As happens with the conjugative systems, this subfamily also contributes to genetic exchange between bacteria.

c) Effector translocation T4SS. They are involved in the secretion of effector proteins during the infection process of several mammalian and plant pathogens. Many pathogenic bacteria are dependent on the translocated effector proteins to survive or thrive in the host organism. This family includes T4SS of several human pathogens such as *H. pylori*, *Bartonella* spp and *Brucella* species.

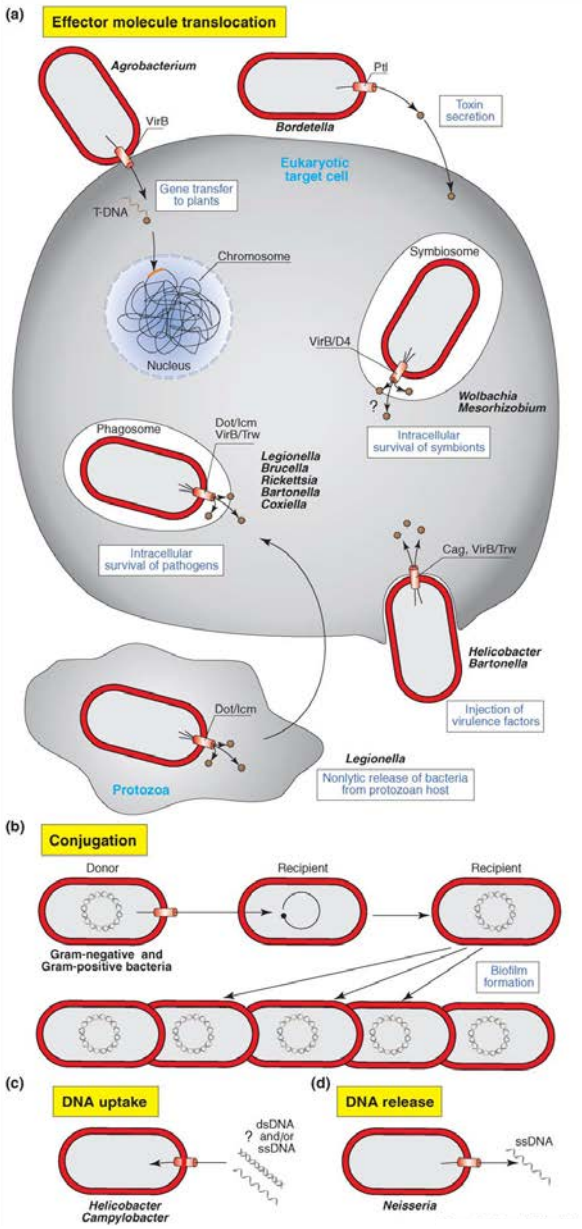

Figure 1. Different T4SS functions in gram-negative bacteria. (a) T4SS from pathogenic bacteria can either inject their effectors directly into the host cell or secrete them into the medium, thereby exerting remarkably different effects on host cell functions during infection. **(b)** Conjugative T4SS translocate DNA from a donor bacterium into various recipients, including other bacteria or eukaryotic cells. **(c)** DNA uptake or **(d)** DNA release systems facilitate exchange of DNA with the extracellular space. Taken from (Backert & Meyer, 2006).

However, some T4SS, which are not conjugative, are able to secrete both proteins and DNA, like VirB of *Agrobacterium tumefaciens* and Dot/Icm of *Legionella pneumophila*. The *A. tumefaciens* VirB T4SS can be considered as part of a) and c) groups at the same time in the above classification, due to the fact that it is involved in

T-DNA transfer from the bacteria to the plant cell but also in the translocation of effector proteins during the infection process. This system is the best studied of all of them, so it has become the paradigm of T4SS.

Recently, a new function of T4SS has been described. Souza and co-workers found that the *Xanthomonas citri* T4SS provides the capacity to kill other gram-negative bacteria in a contact-dependent manner, demonstrating for the first time the involvement of a T4SS in bacterial killing (Souza *et al.*, 2015). Until now, this activity had been only described for Type VI Secretion Systems (T6SS), known for efficiently targeting competitor bacterial cells by the injection of antibacterial toxins (Coulthurst, 2013).

1.1. T4SS involved in virulence

A high number of bacterial pathogens use T4SS to deliver virulence factors into their eukaryotic target cells or the extracellular milieu. Among the human pathogens that code for T4SS that are required for their infection processes we can find *H. pylori*, *L. pneumophila, Brucella* spp or *Bartonella* ssp. Most of them deliver toxigenic proteins to the cytoplasm of the host cell at different time points along the infection process.

1.1.1. T4SS of the genus *Bartonella*

The Gram-negative genus *Bartonella* is composed by 24 species of facultative intracellular pathogens of mammals, where they cause a long-lasting intra-erythrocytic bacteraemia (Schulein & Dehio, 2002). Three different T4SS have been identified among the *Bartonella* species. These T4SS are named Trw, VirB/D4, and Vbh, and they are categorized as pathogenicity factors (Saenz *et al.*, 2007). The Trw T4SS is present in a large number of *Bartonella* species able to infect different mammalian reservoirs. The acquisition of Trw seems to have contributed to the capacity of modern bartonellae to adopt novel reservoir hosts within short evolutionary distances (Dehio, 2008). The other two T4SS, VirB/D4 and Vbh, are closely related and may be functionally redundant; each modern species encodes at least for one of these T4SS.

Within the genus *Bartonella*, *B. henselae* is a worldwide zoonotic pathogen of growing medical importance. It was firstly isolated from patients with Cat scratch disease (CSD). Cats are the natural host for *B. henselae*, from which the bacteria can infect humans. They promote a wide range of symptoms that differ according to the immune status of the human host, from CSD in immunocompetent patients to bacillary

angyomatosis in immunocompromised patients (Dehio, 2005). As most of the modern species, *B. henselae* codes for two T4SS, VirB/D4 and Trw (Dehio, 2008).

The VirB/D4 T4SS of *B. henselae*

The *Bartonella* VirB/D4 T4SS was discovered while looking for the gene coding an immunogenic protein (VirB5) identified in the serum of a patient infected by *B. henselae* (Padmalayam *et al.*, 2000; Schmiederer & Anderson, 2000). As the prototypical VirB/D4 T4SS of *A. tumefaciens, B. henselae* VirB/D4 is encoded by an operon of 10 genes (*virB2-B11*) plus *virB4* (Schulein & Dehio, 2002). Both nucleotide sequence of *virB* genes and genetic organization are highly conserved between all the modern *Bartonella* species (Saenz *et al.*, 2007).

Further analysis revealed the requirement for the VirB/D4 T4SS at the early infection stages, before invasion of the bloodstream (Schulein and Dehio, 2002). *B. henselae virB/D4* operon is induced during *in vitro* infection of human vascular endothelial cells (HUVECs) (Schmiederer *et al.*, 2001) indicating that the VirB/D4 T4SS could be involved in the interaction between the pathogen and the endothelial cell in the infection process. In fact, analysis of the genes encoded downstream the *virB* operon in *B. henselae* led to the identification of seven VirB/D4-translocated substrates, termed *Bartonella* effector proteins (BepA to BepG), which mediate all the known VirB/D4-mediated endothelial cell changes (including invasion, antiapoptotic protection, proinflammatory activation, and control of proliferation of host cell) (**Figure 2**) (Schulein *et al.*, 2005).

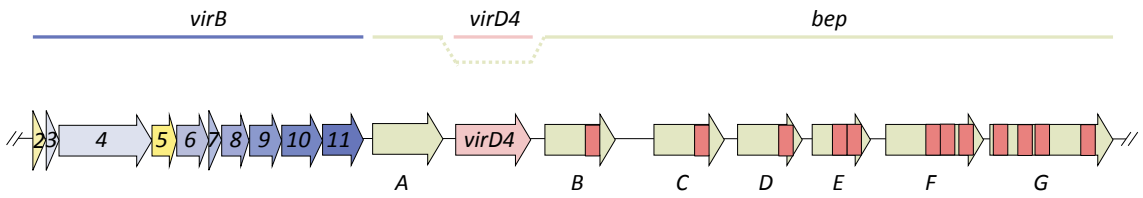

Figure 2. Structure of the *virB/virD4/bep* locus. It encodes the VirB components (VirB2–VirB11; in yellow: pilus-associated components; in blue: rest of the components of the T4SS), the T4S coupling protein (VirD4, in pink), and seven effector proteins (BepA-G, in green with the BID domains represented as red boxes). Modified from (Schroder & Dehio, 2005).

For the substrates to be translocated, some kind of recognition between the substrate and the T4SS is required. It is known that this recognition is mediated by a combination of secretion signals present mainly in the C-terminal region of the substrate, intrinsic motifs of the substrate, and other cellular factors (Cambronne & Roy, 2006). In the C-terminal region, Bep proteins contain a bipartite translocation signal composed of the 142-aa BID (Bep intracellular delivery) domain and an unconserved positively charged tail sequence. This translocation signal is essential for the intracellular delivery of Bep proteins (Schulein *et al.*, 2005). Additional BID domains that are apparently not required for protein translocation via the VirB/D4 system are present in some of the Bep proteins. It has been suggested that they have adopted effector functions within the host cell (Schulein *et al.*, 2005).

1.2. T4SS involved in conjugation

The conjugative systems are the largest and most widely distributed subfamily of T4SS, with systems described for most species of the *Bacteria* and some members of the *Archaea* (Alvarez-Martinez & Christie, 2009). Bacterial conjugation is described in detail in the next section.

2. Bacterial conjugation

Bacterial conjugation is a mechanism of horizontal DNA transfer from a donor cell to a recipient cell through a protein complex known as conjugative apparatus. This process requires cell contact and it is usually mediated by plasmids or conjugative transposons (Zechner *et al.*, 2000). Under laboratory conditions, DNA transfer by means of conjugation has also been described to occur between different kingdoms: from *Escherichia coli* to the budding yeast *Saccharomyces cerevisiae* (Heinemann & Sprague, 1989); from bacteria to plants (Buchanan-Wollaston *et al.*, 1987); and even from bacteria to mammalian cells (Waters, 2001). Naturally occurring DNA transfer between bacteria and plants occurs in certain species of *Agrobacterium* that have the ability to mobilize a DNA segment (T-DNA) to plants in a very similar way to bacterial conjugation, but distinct from it (Zupan *et al.*, 2002).

Bacterial conjugation has intrinsic interest as a mechanism that generates genetic variability (de la Cruz & Davies, 2000), contributing to the spread of virulence factors (Christie & Vogel, 2000) and antibiotic resistances (Mazel & Davies, 1999). But also, the

conjugative components could be used as a biotechnological tool for genetic modification (Gonzalez-Prieto *et al.*, 2013; Llosa & de la Cruz, 2005).

In 1947, Tatum and Lederberg discovered bacterial conjugation while assaying the prototrophic recovery of double nutritional mutants in *E. coli* strain K12 (Tatum & Lederberg, 1947). The real mechanism underlying this mating process was not elucidated until 1953, when Watson and Hayes proposed that the mating was unidirectional and that the viability of the mating depended upon the presence of a recipient strain (Watson & Hayes, 1953). It was proposed that the transfer was mediated by the fertility factor F, from F^+ donor cells, to F^- recipient cells.

In the early 1970s, Datta and Hedges developed a formal scheme of plasmid classification based on incompatibility (Inc) groups due to the inability of different plasmids to be propagated stably in the same cell line because of a common use of elements involved in various functions such as plasmid replication control (Couturier *et al.*, 1988; Datta & Hedges, 1971). They can also be classified according to the range of host bacteria where they can be stably maintained. Broad host range plasmids (IncN, -P, -W) show low dependency of the host cellular functions and can replicate in a wide range of bacteria. In contrast, narrow host range plasmids (IncF, -I) are stable in a limited number of highly related bacterial species.

Transmissible plasmids can be also classified according to their mobilization ability in conjugative (self-transmissible, containing all the information needed for conjugative transfer) and mobilizable (not self-transmissible, they can only be transferred by conjugation in the presence of a helper conjugative plasmid). Another characteristic that can be used with classification purposes of conjugative plasmids is the type of pilus. A flexible pilus allows bacteria to conjugate in both liquid and solid media. A stiff pilus allows bacteria to conjugate only on solid surfaces.

Although the study of conjugation covered plasmids from gram-positive bacteria (Grohmann *et al.*, 2003), the vast majority of the transfer systems studied belong to gram-negative bacteria. In the Enterobacteriaceae family, six groups of conjugative plasmids have been comprehensively analysed and used to establish a general model for conjugation with five different variants corresponding to groups: IncF (F, R1, R100), IncI (R64), IncW (R388) & IncN (pKM101, R46), IncP (RP4, RK2), and IncX (R6K) (Couturier *et al.*, 1988). Mobilizable plasmids have also been described in detail. These include the broad host range RSF1010 and R1162 IncQ plasmids and ColE1 and CloDF13 plasmids (Francia *et al.*, 2004).

Despite the clear differences in the conjugative strategy of different classes of plasmids, they all share some mechanistic components: the need for the synthesis of a conjugative pilus or some other mechanism to promote cell surface contact; the presence of DNA processing enzymes for the initiation of DNA transfer; a mechanism for guaranteeing the establishment of the incoming plasmid in the recipient cell; a regulatory system for transfer control and induction, and an origin of transfer (*oriT*), the unique element required in cis for a plasmid to be mobilizable (Zechner *et al.*, 2000).

2.1. A model for bacterial conjugation

Bacterial conjugation systems can be considered as the merge of two ancient bacterial systems: a rolling-circle replication (RCR) mechanism and a Type IV Secretion System (T4SS) (Llosa *et al.*, 2002). This assumption is based on the facts that the proteins implicated in DNA metabolism and their target sequences (*oriT* or oriV) present high sequence similarities in both conjugation and RCR (Waters & Guiney, 1993) and that the set of conjugative proteins that assembles the membrane transporter belong to the Type IV Secretion System family (Christie, 2001).

From a functional point of view, the conjugative machinery in gram-negative bacteria is composed by three different modules that are assembled to conform a complete system (Llosa & de la Cruz, 2005):

- The relaxosome, a nucleoprotein complex formed by an *oriT*, a relaxase and one or more accessory nicking proteins. It acts as substrate *selector* and is responsible for DNA processing.
- The T4SS, the multiprotein complex organized in a transmembranal *conduit* that spans both the inner and outer membrane.
- The coupling protein (CP), which brings together the selector and the conduit, thus approaching both parts of the transfer machinery.

In the general scheme for conjugative DNA transfer the following steps could be differentiated (**Figure 3**):

a) A donor and a recipient cell get in contact by a protein structure synthesized by the donor and known as conjugative pilus. This contact is mediated by the proteins of the mating pair formation (Mpf) system.
b) The relaxase localizes its target (*oriT*) in the DNA that is going to be transferred. It performs a single-strand cleavage and remains covalently bound to the 5' end.

The unwinding of the DNA is produced, giving rise to a single-stranded DNA (ssDNA) that will be mobilized to the recipient cell.

c) Once in the recipient, the DNA must be established. To this end, the ssDNA molecule is recircularized, converted into double-stranded DNA (dsDNA), and supercoiled.

d) Once the conjugative process has finished, the cells split up and entry-exclusion determinants are expressed in the recipient to avoid the entrance of more copies of the same plasmid. Now, the recipient strain can act as a donor for a new cycle in conjugation.

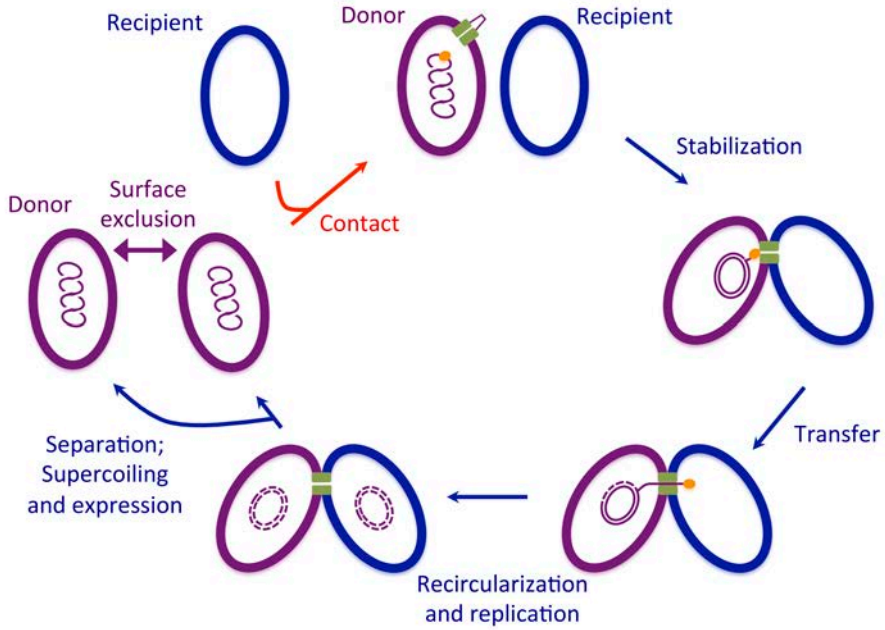

Figure 3. General scheme of bacterial conjugation. It is a process whereby plasmid DNA is transferred from a donor to a recipient cell by a mechanism that requires cell-to-cell contact through a T4SS. In this process, a strand of DNA is transferred and then, both strands are replicated. At the end, two donors are ready for new cycles of conjugation. Green, T4SS. Orange, relaxase.

Llosa and collaborators proposed a model for the transport of DNA in bacterial conjugation using the plasmid R388 system as a paradigm. This is called the shoot and pump model (Llosa *et al.*, 2002) (**Figure 4**). DNA transfer is proposed to take place in two sequential steps. In the first step, the pilot protein would be actively transferred by the T4SS and the DNA would be dragged by this protein in a passive way. In a second step, the DNA would be actively pumped out through the T4SS, presumably by the CP.

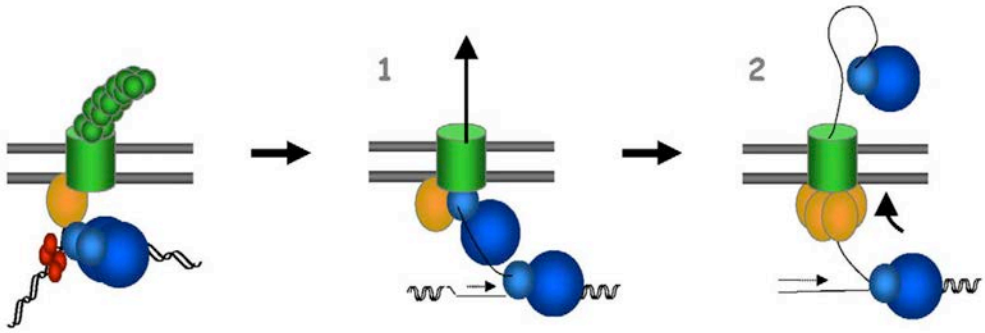

Figure 4. Shoot and pump model for conjugative DNA transfer. Step 1: the relaxase is secreted through the T4SS with the trailing covalently bound DNA strand. **Step 2:** the remaining DNA is pumped out by the CP. Blue, relaxase. Red, relaxosome accessory protein. Yellow, CP. Green, T4SS. Modified from (Llosa & de la Cruz, 2005).

2.2. Plasmid R388

Plasmid R388 was first described in 1972 when it was isolated from *E. coli* (Datta & Hedges, 1972). It belongs to the IncW incompatibility group and it is a conjugative plasmid of broad host range. Its transfer to different species of proteobacteria, like *Salmonella typhimurium*, *Shigella flexneri*, *Pseudomonas aeruginosa* (Bradley & Cohen, 1976), and *A. tumefaciens* (Loper & Kado, 1979) has been reported. It encodes resistance to trimethoprim and sulphonamides (Datta & Hedges, 1972). Its constitutive pilus is stiff and thin (Bradley, 1980) so conjugation only occurs when mating takes place on solid surfaces (Bradley *et al.*, 1980). R388-carrying bacteria are sensitive to phage PRD1 (Olsen *et al.*, 1974), which attaches to the R388 conjugative pilus.

Conjugative plasmids of the IncW group are among the smallest (30-40 Kb) self-transmissible plasmids studied so far. The complete sequence of plasmid R388 is known. It is 33,926 bp in size, and contains 43 detected ORFs. It is organised in functional blocks, as detailed in **Figure 5**, corresponding to all the basic functions implicated in survival and spreading of a plasmid: replication, stable inheritance, establishment, and conjugative transfer (Fernandez-Lopez *et al.*, 2006).

Almost half of the plasmid is devoted to code for the genes involved in conjugative transfer. This 14.9 kb region called TRAw, is divided into two units that can be separately cloned in trans, and they complement for efficient conjugation (Bolland *et al.*, 1990). As shown in **Figure 6**, these regions are: Mpf (mating pair formation), which contains the genes responsible for the synthesis and assembly of the T4SS that connects donor and recipient cell during conjugation; and Dtr (DNA transfer replication)

containing those genes responsible for DNA processing and mobilisation (Fernandez-Lopez *et al.*, 2006).

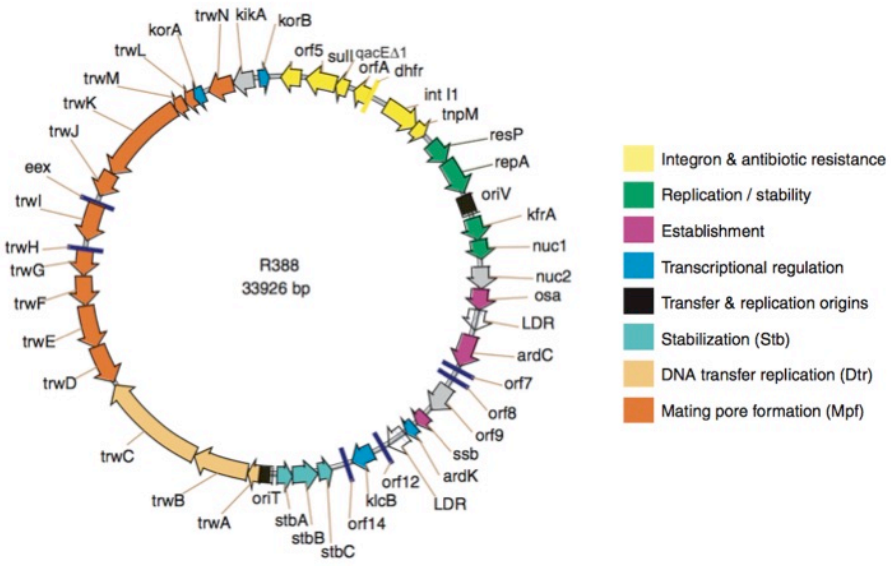

Figure 5. Genetic map of plasmid R388. The plasmid is organized in functional blocks, as shown by the colour-code. All the genes involved in conjugative transfer are indicated in light and dark orange. Modified from (Fernandez-Lopez *et al.*, 2006).

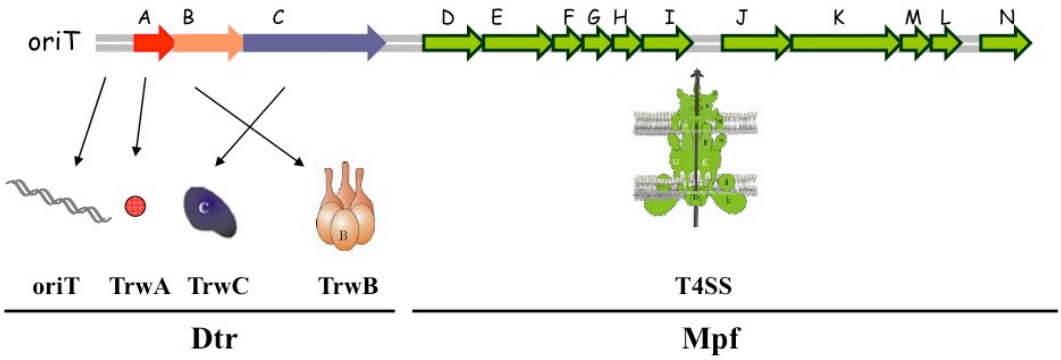

Figure 6. Genetic organization of R388 transfer region (TRAw). The Mpf region consists of eleven genes organized in four transcripts (not shown) and encodes the genes responsible for T4SS and pilus formation, regulators of transcription, and entry-exclusion determinants. The Dtr region is composed by three genes organized in a single operon and a non-coding sequence, corresponding to the *oriT*.

2.2.1. Mpf region

The Mpf region expands from *trwD* to *trwN* and is involved in the formation of the T4SS and the conjugative pilus. All these genes are essential for R388 conjugation except for *trwN*. When carrying mutations in these *trw* genes, bacteria are resistant to phage PRD1 (Bolland *et al.*, 1990). This region also contains *eex* gene, which codes for the entry exclusion determinant and *kil-kor* genes that code for regulation and stability functions (Fernandez-Lopez *et al.*, 2006).

Most Trw proteins of the Mpf region form a Type IV Secretion System (T4SS). The R388 T4SS has been recently reconstructed using electron microscopy (Low *et al.*, 2014) (**Figure 7**). Its individual components have from 42 to 58 % of similarity with the proteins of the prototypical T4SS, the VirB/D4 T4SS of *A. tumefaciens*. They also have a similar genetic organization (Christie & Cascales, 2005; de Paz *et al.*, 2005).

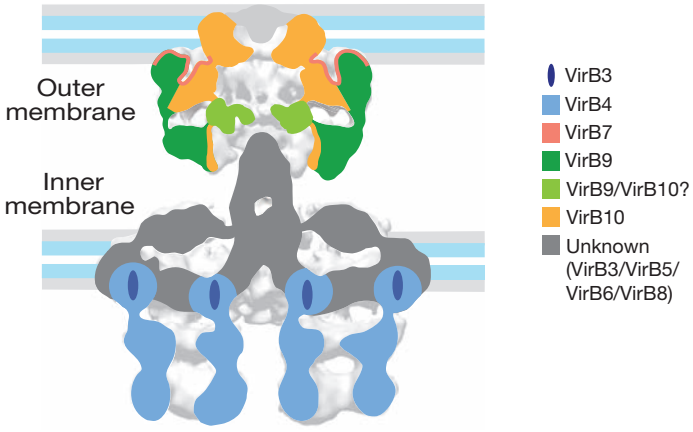

Figure 7. R388 T4SS reconstruction. The scheme shows the localization of known components and position of cell membranes. Only VirB nomenclature is shown. VirB3/TrwM, VirB4/TrwK, VirB7/TrwH, VirB9/TrwF, VirB10/TrwE, VirB5/TrwJ, Virb6/TrwI, VirB8/TrwG. Modified from (Low *et al.*, 2014).

A possible assembly pathway for the T4SS of R388 has been proposed (Larrea *et al.*, 2013). In a first step, the cytoplasmic ATPase TrwK (VirB4) would be anchored to the inner membrane establishing interactions with components of the core complex. As the absence of TrwE (VirB10) was observed to not affect stability of other T4SS components, its assembly after the rest of the core components was suggested. The pilus was proposed to be assembled in the last place, since the integrity of the minor

pilus component TrwJ (VirB5) is dependent on all T4SS components except for the CP TrwB (VirD4).

The CP was found to be structurally independent from other T4SS proteins. The authors suggested that it would be anchored to the inner membrane as a monomer, and when a mating signal is detected, it would locate at the T4SS to recruit the substrate. After protein translocation, the CP would oligomerize around the single stranded DNA and pump it out of the cell using the energy obtained from ATP hydrolysis.

2.2.2. Dtr region

The R388 Dtr region is a 5.2 kb segment lying next to the Mpf region and comprising three genes transcribed from a single operon, *trwABC*, and a 330 bp *oriT* (Llosa *et al.*, 1994a). The *trwABC* operon codes for three proteins: TrwA, the relaxosome accessory protein; TrwB, the coupling protein that connects the relaxosome and the DNA transfer apparatus; and TrwC, the relaxase-helicase.

Except for TrwB, all the components of the Dtr region are involved in the formation of the relaxosome, a nucleoprotein complex necessary for DNA mobilization. **Figure 8** shows the proposed architecture of the relaxosome of R388, composed by the *oriT*, the relaxase TrwC, and the helper proteins TrwA and IHF.

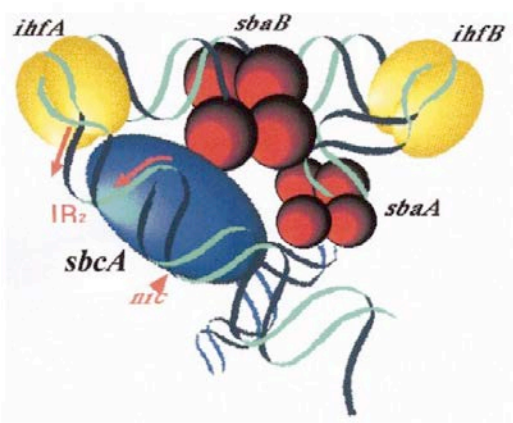

Figure 8. Architecture of the R388 relaxosome. TrwC is represented in blue. TrwA (in red) is shown as tetramers bound to their specific *sbaA*, *-B* binding sites. IHF (in yellow) sits on its binding sites, *ihfA* and *ihfB*. *nic*, *oriT* nicking site. IR_2, inverted repeat 2. Taken from (Moncalián *et al.*, 1999b).

IHF (integration host factor) is a small heterodimeric protein coded by the host cell chromosome. It is known that the DNA is wrapped around the protein and bent by more than 160º (Rice *et al.*, 1996). This bending stabilizes distinct DNA conformations that are required during several bacterial processes, like transcription or recombination. IHF binds specifically to two sites in the R388 *oriT*, *ihfA* and *ihfB*. Its binding inhibits the nicking reaction mediated by the relaxase TrwC, both *in vivo* and *in vitro*, but not TrwC binding. It is thought that initiation of the conjugative process in R388 could start when IHF leaves the *oriT* (César & Llosa, 2007; Moncalián *et al.*, 1999b). No significant difference in the conjugation frequency of R388 using IHF deficient strains either as donors or recipients has been observed (Llosa *et al.*, 1991).

The components of the Dtr of R388 are detailed below:

***oriT*.** The *oriT* is defined as the minimal sequence required in a DNA molecule to be transferred efficiently by conjugation, when the other components of the system are presented in trans. The R388 *oriT* ($oriT_W$) was first described as a sequence of 402 bp (Llosa *et al.*, 1991), and later delimited to bp 63-330 (César *et al.*, 2006). It contains the typical features of *oriTs*: a high A+T content (49% G+C versus 57% in whole R388) and the presence of several direct and inverted repeats (DRs and IRs), which are probably protein binding sites. The following features of the R388 *oriT* have been described (**Figure 9**):

- The CACTAC hexamer is found four times, constituting three direct repeats, and the inverted repeat IR_4. Each hexamer is proposed as the 'monomer' recognition site for TrwA binding, forming two distinct binding sites, *sbaA* and *sbaB* (Moncalián *et al.*, 1997).

- Four other inverted repeats: IR_1 (CCCAATGCGC), $IR_{1.1}$ (TGACCCT), IR_3 (AAGTCATTC) and IR_2 (CGCACC). IR_2 corresponds to the relaxase binding site *sbcA* (Guasch *et al.*, 2003; Llosa *et al.*, 1995).

- Two IHF binding sites. The first one is located in IR_3 (bp 203-215) and has the consensus sequence 5'-TAAgtcATTGATT-3' (Craig & Nash, 1984). The second one is located in IR_4 (bp 278-290) and does not contain an obvious consensus site (Moncalián *et al.*, 1999b).

- The nicking site. TrwC introduces a nick at nucleotide 176 (bottom strand) (Llosa *et al.*, 1995). Due to the fact that the transferred strand is the nicked one and the 5' end is transferred first, *trw* genes enter the recipient in the last place.

A deletion of bp 1-170, leaving only 6 bp 3' to the *nic* site, rendered an *oriT* 100% functional for nicking by TrwC and 1% functional for conjugal mobilization. The

sequence 5' to the *nic* site, comprising IR_2, IR_3 and IR_4, is required for mobilization, whereas only 17 bp, spanning IR_2 and the *nic* site, are required for efficient TrwC binding and nicking (Guasch *et al.*, 2003; Llosa *et al.*, 1991).

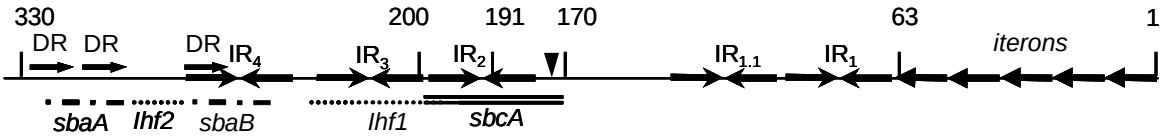

Figure 9. R388 *oriT*. Arrows indicate the presence of iterons, DRs and IRs. *nic* site is indicated as a vertical arrowhead. *sbaA* and *sbaB*, TrwA binding sites. *ihf1* and *ihf2*, IHF binding sites. *sbcA*, TrwC binding site. Coordinates are given as in (Llosa *et al.*, 1991). Modified from (César *et al.*, 2006).

TrwA. It forms part of R388 relaxosome. This protein is not indispensable for conjugation, although its absence decreases drastically the conjugation frequency (Llosa *et al.*, 1994a). TrwA is a protein of 121 amino acids and 13 kDa, which oligomerizes making tetramers of 53 kDa in solution. Two structural domains of TrwA have been described: an N-terminal domain, involved in DNA binding, and a C-terminal domain, responsible for its tetramerization ability (Moncalián & de la Cruz, 2004).

Based on the conservation of a series of polar residues in the N terminus of the protein, also present in the ribbon-helix-helix (RHH) family of proteins, TrwA was putatively assigned as a member of RHH protein family. Mutations in these residues supported the inclusion of TrwA in this family of transcriptional regulators (Moncalián & de la Cruz, 2004), represented by the Arc repressor (Breg *et al.*, 1990).

TrwA binds specifically to two regions within the R388 *oriT*, *sbaA* and *sbaB* (Figure 9). Binding at *sbaA* occurs at 10 fold higher affinity than at *sbaB*. TrwA binding to *oriT* results in transcriptional repression of the *trwABC* operon, probably because *sbaA* site overlaps with the putative -10 region of the operon promoter. It was also observed that TrwA enhances TrwC relaxation activity *in vitro*, providing evidence of a direct role in the *oriT*-processing reaction (Moncalián *et al.*, 1997). It has been shown that TrwA increases TrwC-mediated site-specific recombination (César et al., 2006) and integration (Agundez *et al.*, 2012), and the ATPase activity of TrwB (Tato *et al.*, 2007).

TrwB. It is the CP of the R388 conjugative system, responsible for the contact of the relaxosome and the T4SS during conjugative DNA transfer of plasmid R388 (Llosa *et al.*, 2003). TrwB has 507 aa and a molecular weight of 56 kDa. It is anchored to the inner membrane by two N- terminal transmembrane segments. In the cytosolic domain, the protein presents Walker A and Walker B motifs for nucleotide binding (Llosa *et al.*, 1994a). A TrwB mutant with a single amino acid alteration in the ATP-binding Walker A motif (K136T) was totally deficient in DNA transfer (Moncalián *et al.*, 1999a). The ATPase and DNA binding activities of its soluble cytoplasmic domain (TrwBΔN70) have been characterised (Tato *et al.*, 2005; Tato *et al.*, 2007).

The crystal structure of TrwBΔN70, has been solved by X-Ray diffraction (Gomis-Rüth *et al.*, 2001). It shows a quaternary structure consisting of hexamers, similar to that of the F1-ATPase, that form an almost spherical, orange-shape structure with a central channel of 20 Å in diameter. The structural similarity to the F1-ATPase and other proteins that pump or move DNA (SpoIIIE, FtsK) lead to the proposal of TrwB as an F1-ATPase-like molecular motor involved in DNA transport during bacterial conjugation (Cabezón & de la Cruz, 2006).

Apart from its presumed role in DNA transport, TrwB interacts with TrwA (Llosa *et al.*, 2003) and it is also required for the transfer of TrwC relaxase in the absence of DNA (Draper *et al.*, 2005), suggesting that it plays part in the recognition of the protein substrate of the T4SS. The domains of the protein directly involved in the interaction with the T4SS and the relaxosome, in DNA binding, and in the ATPase activity have been described (de Paz *et al.*, 2010).

It has been shown that the amount of TrwB in the cell is not affected by the absence of other T4SS components, nor did its absence alter significantly the levels of integral components of the T4SS, underscoring the independent role of the coupling protein on the T4SS architecture and supporting its role as a third functional module of the conjugative machinery (Larrea *et al.*, 2013).

TrwC. It is the relaxase-helicase of plasmid R388. It is described in detail in section 3.1.

2.3. Plasmid F

The F (fertility) factor of *E. coli* K-12 was the first plasmid to be described (Cavalli *et al.*, 1953) and has been subject of study since then, becoming a paradigm for

conjugative plasmid systems. It belongs to the IncFI incompatibility group and it is a narrow host range conjugative plasmid. IncF plasmids are limited by host range to the family Enterobacteriaceae (Villa *et al.*, 2010). Its constitutive pilus is thick and flexible (Bradley, 1980). This type of pili can dynamically extend and retract, enabling donor cells to bind and draw recipient cells into physical contact to conjugate in liquid as efficiently as on solid surfaces (Bradley *et al.*, 1980). F-carrying bacteria are sensitive to several male-specific phages, such as R17 (Crawford & Gesteland, 1964), which bind to F pilus.

IncF plasmids are usually low copy number plasmids and bigger than 100 kb in size. In 1994, the sequence of the F transfer region was published (Frost *et al.*, 1994). Nowadays its complete sequence is known and it has 99,159 bp coding for about 100 genes (**Figure 10**).

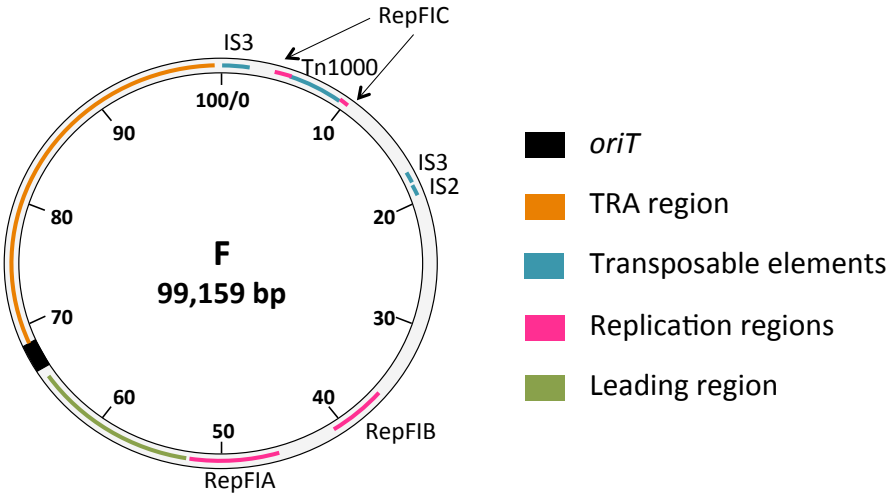

Figure 10. Genetic map of plasmid F. Numbers indicate the kilobase coordinates. Functional regions are indicated in a colour-code. The regions implicated in conjugative transfer of the plasmid are indicated in black and orange (*oriT* and TRA genes). Modified from (Firth *et al.*, 1996).

Plasmid F has three replication regions. RepFIA is believed to be mainly responsible for the typical replication properties of F and contains both unidirectional (*oriS*) and bidirectional (*oriV*) replication origins (Lane, 1981). RepFIB is independently functional and can sustain plasmid replication in the absence of RepFIA. RepFIC is not functional due to a Tn1000 insertion.

It contains four significant insertion sequences: Tn1000, inserted in RepFIC (Saadi *et al.*, 1987); IS3a, inserted in *finO* (Cheah & Skurray, 1986); IS3b (Deonier *et al.*, 1979); and IS2 (Hu *et al.*, 1975) these two inserted outside any genes coding for proteins of known function. It is thought that these ISs mediate F chromosomal integration events that can subsequently generate F´ species by inaccurate excision (Deonier & Mirels, 1977).

The 33.3 kb F transfer region contains all the sequences required for conjugative transfer (**Figure 11**). There are 36 detected ORFs, including those genes responsible for DNA processing (Dtr), T4SS synthesis, assembly, and retraction (Mpf), and mating contact stabilization (Mps region, mating pair stabilization) (Arutyunov & Frost, 2013).

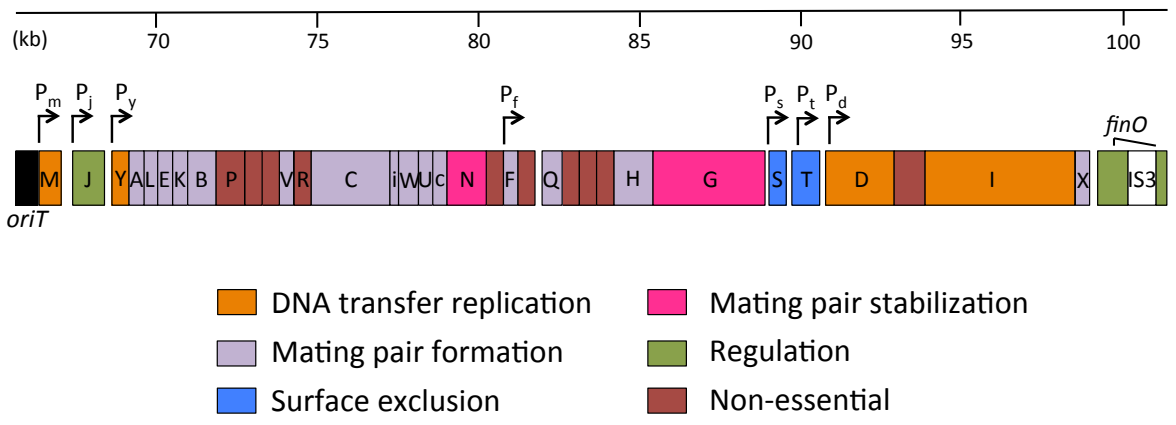

Figure 11. Genetic organization of F transfer region. Kilobase coordinates are indicated at the top. Arrows indicated the different promoters. Functional gene groups are indicated in colour-code. *tra* genes are indicated in capital letter. Apart from them, only *trbi and -c* genes are indicated, labelled as lowercase letters. Modified from (Firth *et al.*, 1996).

2.3.1. Mpf and Mps genes

The F Mpf region contains the genes responsible for the biogenesis of the T4SS (*traA, -B, -C, -E, -K, -L, -Q, -V, X, and trbC*) (Lawley *et al.*, 2003), and pilus outgrowth and retraction (*traF, -H, -U, -W, and trbI*) (Arutyunov & Frost, 2013; Maneewannakul *et al.*, 1992). TraN and TraG proteins are responsible for the mating pair stabilization, allowing equally efficient mating in liquid and solid media. This Mps is believed to represent the conversion of initial unstable contacts between donor and recipient cells to a form which is resistant to disruption by shear forces (Manning *et al.*, 1981).

This region of F-TRA also contains two genes responsible for surface exclusion (Sfx). The *traS* and *traT* gene products limit the host cell capacity to act as a recipient for the same or a closely related plasmid (Achtman *et al.*, 1980). Finally, *finO* and *traJ* genes code for proteins involved in the regulation of TRA region (Wong *et al.*, 2012).

2.3.2. Dtr genes

This region comprises a 289 bp sequence, corresponding to the *oriT*, and four *tra* genes. These genes (*traM, -Y, -D*, and *-I*) are spaced within the TRA region (Figure 11) and are responsible for DNA metabolism during F conjugative transfer. The components of the Dtr region are detailed below.

***oriT*.** It was first described as the region between the *BglII* site at nt 1 (coordinates as in (Frost *et al.*, 1994)) and the beginning of the *traM* gene (Frost *et al.*, 1994). The F *oriT* is complex, containing several protein-binding sites (*sbi*, for the relaxase TraI; *ihfA* and *-B*, for IHF; *sbyA* and *-C*, for TraY; and *sbmA, -B,* and *-C,* for TraM) and structural elements, including intrinsic bends, DRs, IRs, and A-tracts (**Figure 12**) (Wong *et al.*, 2012). All protein-binding sites and the majority of the structural elements are located 5' to *nic* (Di Laurenzio *et al.*, 1992; Thompson *et al.*, 1984; Tsai *et al.*, 1990). Sites *sbi*, *ihfA* and *sbyA* along with their associated proteins are implicated in the nicking reaction (Nelson *et al.*, 1995).

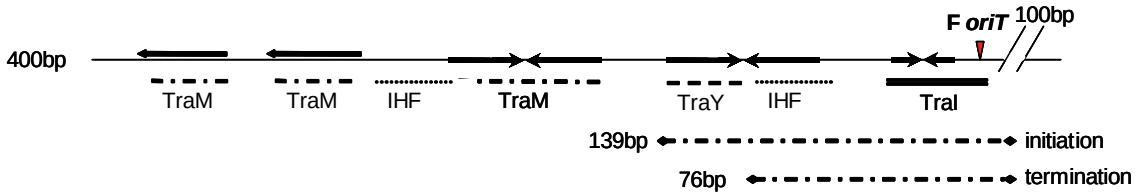

Figure 12. F *oriT* region. Arrows indicate the presence of DRs and IRs. The binding sites for host and plasmid accessory proteins are indicated. *nic* site is indicated by a vertical red arrowhead. Requirements for initiation and termination of DNA transfer are indicated by dashed lines.

Gao and collaborators showed that F conjugation termination required only of a 76 bp sequence containing the relaxase binding site (*sbi*), and adjacent IHF binding site (*ihfA*), showing independence from accessory protein TraY and TraM binding sites.

Initiation was found to require a 139 bp sequence including also the accessory protein TraY binding sites (Gao *et al.*, 1994).

TraM. It is a cytoplasmatic tetrameric protein that binds cooperatively to three different sites at F *oriT* (Fekete & Frost, 2002). Its regulatory activity is essential for conjugation although it is has not been found to be necessary for neither pilus synthesis, mating pair formation or nicking at *oriT in vivo*. Interaction between TraM and the CP TraD has been found *in vivo* (Lu & Frost, 2005; Lu *et al.*, 2008). Moreover, TraM-TraD (Beranek *et al.*, 2004; Disque-Kochem & Dreiseikelmann, 1997) and TraM-TraI (Ragonese *et al.*, 2007) interactions have been also described. TraM has been found to stimulate relaxosome-mediated cleavage at *oriT* (Ragonese *et al.*, 2007).

TraY. It is a member of the Arc repressor family (Bowie & Sauer, 1990), showing a RHH structure with the the β-strands or 'ribbons' that interact in an antiparallel fashion and contact the major groove of the DNA (Lum & Schildbach, 1999). Its monomer form yields a similar structure to the Arc dimer, and similarly to Arc, which tetramerises upon binding to DNA, TraY binds DNA as a monomer and subsequently dimerises (Nelson & Matson, 1996). TraY proteins of different IncF plasmids have been found to stimulate TraI catalysed *nic*-cleavage *in vivo* and *in vitro* (Karl *et al.*, 2001). TraY has been proved to play an essential role in the formation of the relaxosome. It is a dsDNA-binding accessory protein that bends DNA (Luo *et al.*, 1994; Rice *et al.*, 1996), possibly creating a ssDNA conformation suitable for TraI binding (Nelson *et al.*, 1995; Williams & Schildbach, 2007).

TraD. The coupling protein is responsible for the recruitment of the plasmid to the T4SS through interactions with components of the relaxosome. Although *traD* mutants elaborate apparently normal pili, form aggregates with recipients, and trigger conjugal DNA metabolism, they fail to transfer DNA (Panicker & Minkley, 1985). Purified TraD was originally reported to possess DNA-dependent ATPase activity (Panicker & Minkley, 1992), although it was then assigned to an impurity (Schroder *et al.*, 2002). However, Walker A and B motifs for nucleotide binding are present in its sequence (Lessl *et al.*, 1992). *In vivo* TraD oligomerization by coimmunoprecipitation has been demonstrated and the cytoplasmic domain of TraD has been structurally modelled based on its homology to TrwBΔN70 (Haft *et al.*, 2007).

A stimulatory effect of the coupling protein and relaxosome accessory proteins has been reported on T-strand unwinding by TraI (Mihajlovic *et al.*, 2009), implying that the CP is also a component of the relaxosome. Moreover, a model in which protein-protein and DNA ligand interactions at the CP interface coordinate the transition initiating production and uptake of the nucleoprotein secretion substrate has been proposed (Sut *et al.*, 2009).

TraI. It is the relaxase of F plasmid. TraI is explained in detail in section 3.2.

As in R388, host encoded IHF has been found to play an essential role in the formation of the relaxosome, possibly helping for TraI binding (Nelson *et al.*, 1995; Williams & Schildbach, 2007). Single and/or double mutations in either of the IHF coding genes (*himA*, *hip*) show a 100-500 fold reduction in the transfer efficiency of F and F-like R100 plasmids. It has been suggested that IHF binding to the *oriT* stimulates the expression of *tra* genes (Dempsey, 1987).

The three accessory proteins (TraM, TraY, and IHF) bind the F plasmid *oriT*, and in concert with the relaxase TraI are involved in the processing of DNA during F conjugation (**Figure 13**).

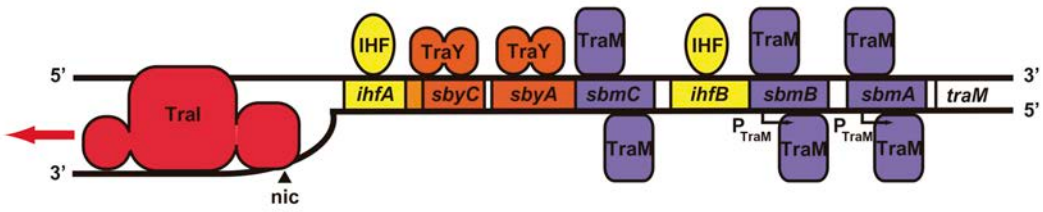

Figure 13. F plasmid relaxosome. *nic* site is represented as a black arrowhead. The relaxase TraI is bound to its binding site, *sbi* (not shown). The direction of TraI unwinding of DNA following cleavage at the *nic* site and covalent attachment to the 5' end of DNA is indicated by a red arrow. *sbmA, -B, -C*, TraM binding sites. *sbyA*, *-C*, TraY binding sites. *ihfA, -B*, IHF binding sites. Taken from (Wong *et al.*, 2012).

Assembling of the F relaxosome occurs in a stepwise manner in which both IHF and TraY bind prior to the relaxase TraI, and this binding facilitates TraI binding to *oriT* (Howard *et al.*, 1995). A model including indirect and possibly direct interactions between relaxosome proteins during the initiation of transfer has been proposed (Williams & Schildbach, 2007). The changes in DNA topology at the relaxase-binding site caused by TraY and IHF bending do not fully explain the mechanism required for the recruitment of TraI to its binding site. It seems that conformational changes of F *oriT* DNA during the initiation of transfer also position TraY closer to TraI-binding site, highlighting the possibility of direct TraY-TraI interactions to promote TraI binding (Williams & Schildbach, 2007).

In the F-related plasmid R1, TraY is dispensable for mobilisation in the presence of TraM, but the latter is essential, even in the presence of TraY. It has been suggested that while TraY and TraM function may overlap to a certain extent in the R1 relaxosome, TraM additionally performs a second function that is essential for successful conjugative transmission of plasmid DNA (Karl *et al.*, 2001). The nature of this activity is unknown but Karl and co-workers suggested that it may also involve the CP TraD, since both proteins are known to interact (Disque-Kochem & Dreiseikelmann, 1997).

3. Conjugative relaxases

The relaxase is the key protein in conjugative DNA transfer responsible for initiating and terminating DNA processing. They are site and strand-specific endonucleases that recognise and cleave one strand of their cognate *oriT* in a particular site, the *nic* site, accomplished through a transesterification reaction (Byrd & Matson, 1997), mechanistically similar to the reaction mediated by topoisomerase I (Forterre & Gadelle, 2009). They were first isolated in 1969 by Clewell and Helinski, who observed that when they incubated these proteins with supercoiled DNA (scDNA) and treated with proteinase K and SDS, a nucleoprotein complex that relaxed DNA was formed (Clewell & Helinski, 1969).

The sequence and properties of the relaxase proteins have been compared, allowing the classification of conjugative transfer systems in six MOB families (Garcillán-Barcia *et al.*, 2009): MOB_F (F_TraI, R388_TrwC); MOB_H (R27_TraI); MOB_C (CloDF13_MobC, pAD1_TraX); MOB_Q (R1162_MobB, RSF1010_MobB); MOB_P (RP4_TraI, R64_NikB, R6K_TaxC); MOB_V (pMV158_MobM) (**Figure 14**).

In a recent work, Francia and collaborators found MOB_C pAD1_TraX relaxase to share structural homology with the PD-(D/E)XK family of restriction endonucleases (REs) (Francia *et al.*, 2013). According to their results, they propose that not all the relaxases involved in plasmid conjugation converged into the same structural fold. MOB_F, MOB_P, MOB_Q, and MOB_V families showed the HUH (histidine-hydrophobic-histidine) fold (Chandler *et al.*, 2013; Garcillán-Barcia *et al.*, 2009), but MOB_C presented the RE fold. This observation suggests an adaptive functional convergence of these two completely separate protein folds.

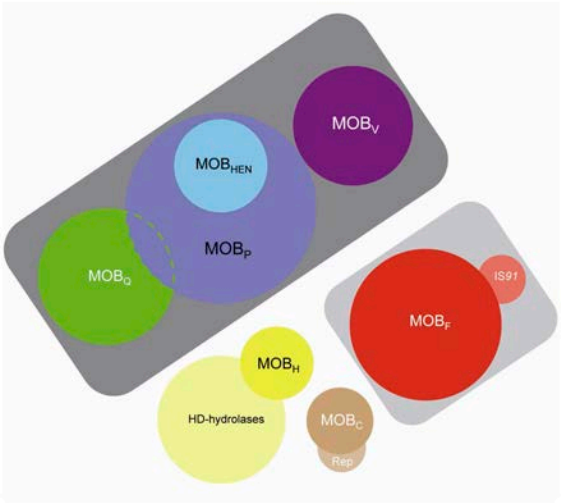

Figure 14. Scheme of the relationships between the main relaxase protein families. Classification was based in amino acid sequence similarity. A first relaxase cluster (in dark grey) collects the relaxase groups with one active Tyr in the catalytic centre. A second cluster (light grey) contains relaxases with two Tyr in the active centre. For the remaining groups (white) not enough was known about the biochemistry of the respective relaxases. Some relaxase protein families overlap other protein families such as RCR proteins (Rep), IS*91*-like transposases (IS*91*) or HD hydrolases. The area of the circles is proportional to relaxase number. Taken from (Garcillán-Barcia *et al.*, 2009).

Three amino acid sequence motifs are present in HUH relaxases from different conjugative systems (Zechner *et al.*, 2000):

- Motif I contains the active tyrosine(s) residue(s) involved in the DNA nicking / strand transfer reactions.
- Motif II contains a conserved serine and it is involved in maintaining tight DNA-protein contact.

– Motif III is characterised by the presence of a conserved histidine triad and the HUH signature. It is proposed to aid coordinating the divalent metal ion required for the nucleophilic attack by the active tyrosine residue.

The reaction mediated by relaxases (**Figure 15**) entails the formation of a covalent intermediate for all the relaxases studied so far, except for the CloDF13 plasmid (Núñez & de la Cruz, 2001).

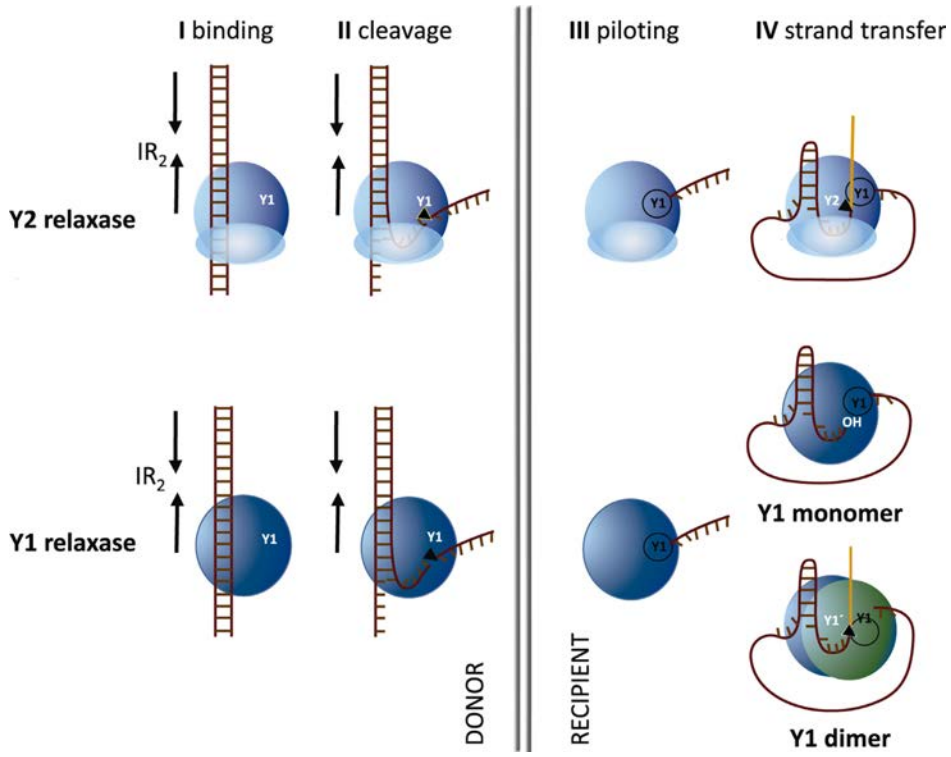

Figure 15. Models of conjugative DNA processing by relaxases containing one (Y1) or two (Y2) catalytic tyrosines. I) Conjugation is initiated by the relaxase recognizing the proximal arm of IR_2 adjacent to the *nic* site. **II)** Relaxase binding allows the formation of a ssDNA U-turn that positions the *nic* site at the relaxase active site and *nic*-cleavage takes place. **III)** Subsequent DNA strand displacement generates the DNA single strand that, piloted by the relaxase, is transferred into the recipient cell. **IV)** In Y2 relaxases, a second tyrosine present in the same molecule attacks the newly formed *nic* site to generate a free 3′OH end able to recircularize the transferred plasmid DNA. In Y1 relaxases, either the free 3′OH is released in the donor cell (monomeric Y1 model) or a second relaxase molecule provides the free tyrosine that attacks the newly formed *nic*-site (dimeric Y1 model). Taken from (Carballeira *et al.*, 2014).

The first step involves a transesterification reaction, the product of which is a nicked dsDNA molecule with a free 3' hydroxyl group and a 5' end covalently bound to

the relaxase via a phosphotyrosil linkage. This is a reversible reaction, since the free 3' hydroxyl is capable of carrying a second transesterification by means of a nucleophile attack to the covalently bound 5' end, hence re-joining the nicked DNA. Interestingly, the equilibrium between the nicked and ligated state of the relaxase-DNA complex is maintained, and cannot be altered by the addition of protein in excess or by long incubations (Matson & Morton, 1991). Nevertheless, the covalent DNA-protein complex has a rather long half-life, believed to facilitate the involvement of the nucleoprotein complex as a key intermediary of the conjugative transfer process. The nicked DNA is used to initiate the conjugative DNA transfer. In those families of relaxases whose biochemical mechanism is known, the nicking of the DNA could be performed for one or two catalytic tyrosines (Figure 15). DNA processing during conjugation terminates with the re-sealing of the transferred DNA by a second transesterification reaction.

3.1. The relaxase TrwC

TrwC is the relaxase responsible for the initiation and ending of the processing of DNA during conjugative transfer of plasmid R388. It is included in the MOB_F family of relaxases (Garcillán-Barcia *et al.*, 2009). TrwC is a 966 aa enzyme, with a molecular weight of 108 kDa. Its pI is 10.1 and it forms dimers in solution in the presence of NaCl 150 mM (Grandoso *et al.*, 1994). It has two catalytic tyrosines, Y18 and Y26. Grandoso and co-workers proposed a model for DNA processing in which Y18 performs the initial cleavage reaction, while Y26 is responsible for the second strand transfer reaction (Grandoso *et al.*, 2000).

In order to perform its role in conjugation, TrwC has to bind to the *oriT*, cleave the DNA strand to be transferred at the *nic* site, and make a covalent bond with the 5´end of this *nic*. Then, the covalent TrwC-DNA complex must be recruited to the T4SS and transported to the recipient cell, where TrwC will catalyse the recircularization of the transferred DNA strand. In addition, TrwC is able to catalyse site-specific recombination and integration reactions unrelated to the conjugative DNA transfer process. We will now survey the different activities and TrwC domains involved in each of the above steps.

3.1.1. TrwC activities involved in conjugative DNA transfer

Relaxase activity

This activity is responsible for sequence-specific cut and strand-transfer reactions. *In vitro*, TrwC can relax *oriT*-containing supercoiled DNA, and it also performs cut and

strand transfer activities on *nic*-containing oligonucleotides. Several works have addressed the protein and DNA requirements for these reactions.

TrwC presents sequence-specific endonuclease activity *in vitro*. The purified protein is able to nick an *oriT*-containing scDNA in the absence of other accessory proteins, while it is unable to nick linear dsDNA (Llosa *et al.*, 1995). The *nic* site and IR_2 (Figure 9) are the DNA requirements for this activity. Once the nicking is produced, the relaxase remains covalently bound to the 5' end of the DNA. The 3' end is also bound, although not covalently, as proteinase K and SDS are required to observe relaxation of the DNA. The reaction reaches the equilibrium in 5 minutes and the activity is enhanced when TrwA or TrwBΔN70 are added to the reaction, while IHF decreased it (Moncalián, 2000).

The minimal core sequence of the R388 *oriT* for TrwC relaxase activity is 17 bp. It comprises the *nic* site and a 5' region involving the recognition hairpin formed by the inverted repeat IR_2 (Guasch *et al.*, 2003). This core is essential for *in vivo* mobilization (Lucas *et al.*, 2010) and recombination (César *et al.*, 2006). *In vitro*, it has been shown that the sequence (6+2), 6 nt 5' and 2 nt 3' to the *nic* site, is absolutely required for TrwC binding, single-strand nicking and strand transfer reactions (Lucas *et al.*, 2010). The segment (14+3) confers basic *oriT* functions (Figure 9) (Guasch *et al.*, 2003; Lucas *et al.*, 2010).

A recent report has analysed the sequence information used by TrwC in *nic* site recognition using full and partially randomized libraries (Carballeira *et al.*, 2014). They have obtained a two-dimensional representation of the TrwC-*nic* interaction, identifying the *nic* nucleotides that are critical and the ones where mutations are allowed for TrwC function (**Figure 16**). The authors propose a model in which TrwC recognition of its cognate *nic* site occurs by a highly specific lock (TrwC) and key (*nic*) interaction (Carballeira *et al.*, 2014).

TrwC is also capable of nicking oligonucleotides containing the *nic* site *in vitro* and performing strand transfer to a second oligonucleotide also containing the *nic* site (Llosa *et al.*, 1996). TrwC binding assays with oligonucleotides reveal that, although the presence of a complete IR_2 is needed for efficient binding, binding affinity is decreased only three fold when either half of the IR_2 is removed, while the stretch of 6 nt adjacent to the nic site is essential for TrwC binding. With respect to single-strand nicking and strand transfer reactions, the sequence (6+2) (6 nt 5' to the nic site and 2 nt 3' to the nic site) is absolutely required for both reactions (Lucas *et al.*, 2010).

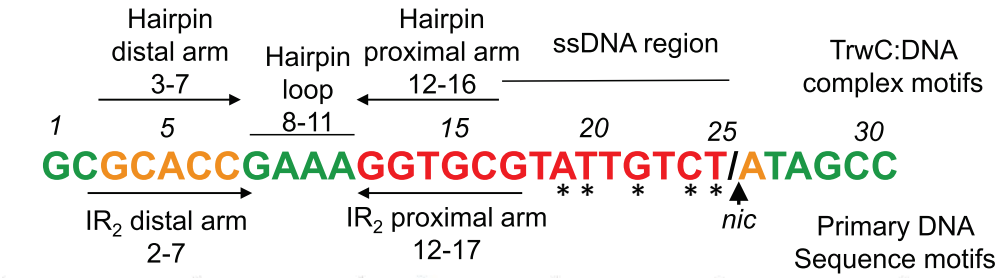

Figure 16. Mutations permitted in the *oriT* for TrwC recognition, obtained using full randomized libraries. Essential nucleotide positions are coloured in red, positions that admits only some mutations in orange, and positions where any nucleotide can be located appear in green. Nucleotides marked with an asterisk were found to admit some mutations using partially randomized libraries. Modified from (Carballeira *et al.*, 2014).

The relaxase domain is located in the N-terminal 293 aa of the protein. To catalyse cleavage and strand transfer activities on oligonucleotides, a fragment containing the N-terminal 275 aa is sufficient. However, fragment N293 is required to cleave scDNA (Guasch *et al.*, 2003; Llosa *et al.*, 1996). The relaxase fragment is purified as a monomer (Guasch *et al.*, 2003; Hernando, 2000). *In vitro*, this fragment shows scDNA nicking activity, ssDNA nicking, and strand-transfer ability.

The three-dimensional structure of the relaxase domain of TrwC (Guasch *et al.*, 2003), as well as the relaxase domain of F_TraI (Datta *et al.*, 2003), has a similar folding to DNA polymerases and RCR initiation proteins such as the initiator protein of tomato yellow leaf curl virus, TYLCV (Campos-Olivas *et al.*, 2002) and AAV_Rep (Hickman *et al.*, 2002) (**Figure 17**).

TrwC relaxase domain is built on a two-layer α/β plate. One of the layers is formed by an antiparallel five-stranded β-sheet with topology β1-β3-β7-β6-β2, and the other layer is constituted by two long helices, α5 and α7. The active centre is located in the β-sheet, on the other side to the α-helices.

In the relaxase domain sequence, the three motifs described for relaxases (Ilyina & Koonin, 1992) are included:

- Motif I: it comprises the four active tyrosil residues (Y18, Y19, Y26 and Y27), involved in the nucleophilic attack to the *nic* site (Grandoso *et al.*, 2000). Relaxases conserving the 4-Tyr motif have been suggested to involve a two tyrosine mechanism for DNA processing such as that described for ΦX174 DNA strand transferase gpA (Hanai & Wang, 1993). By mutagenesis, Y18 and Y26

were determined to be directly involved in the cleavage and strand transfer reactions, being essential for the role of TrwC during conjugative DNA processing. Both Y18 and Y26 promote cleavage and strand transfer reactions of oligonucleotides containing a *nic* site. However, only Y18 can support cleavage of supercoiled *oriT* containing DNA (Grandoso *et al.*, 2000). Mutagenesis in the catalytic residues showed that mutant in Y18 decreases conjugation frequency in 500-fold. Y26 mutant decreases conjugation frequency in 10 times. The double mutant, Y18+Y26, completely abolished conjugation (Grandoso *et al.*, 2000).

- Motif II: it presents an aspartic residue, D85, which is believed to activate the tyrosine hydroxyl group by proton abstraction (Boer *et al.*, 2006).
- Motif III: it contains the characteristic histidine triad (H150, H161 and H163) that coordinates the metal ion for the nicking reaction (Guasch *et al.*, 2003).

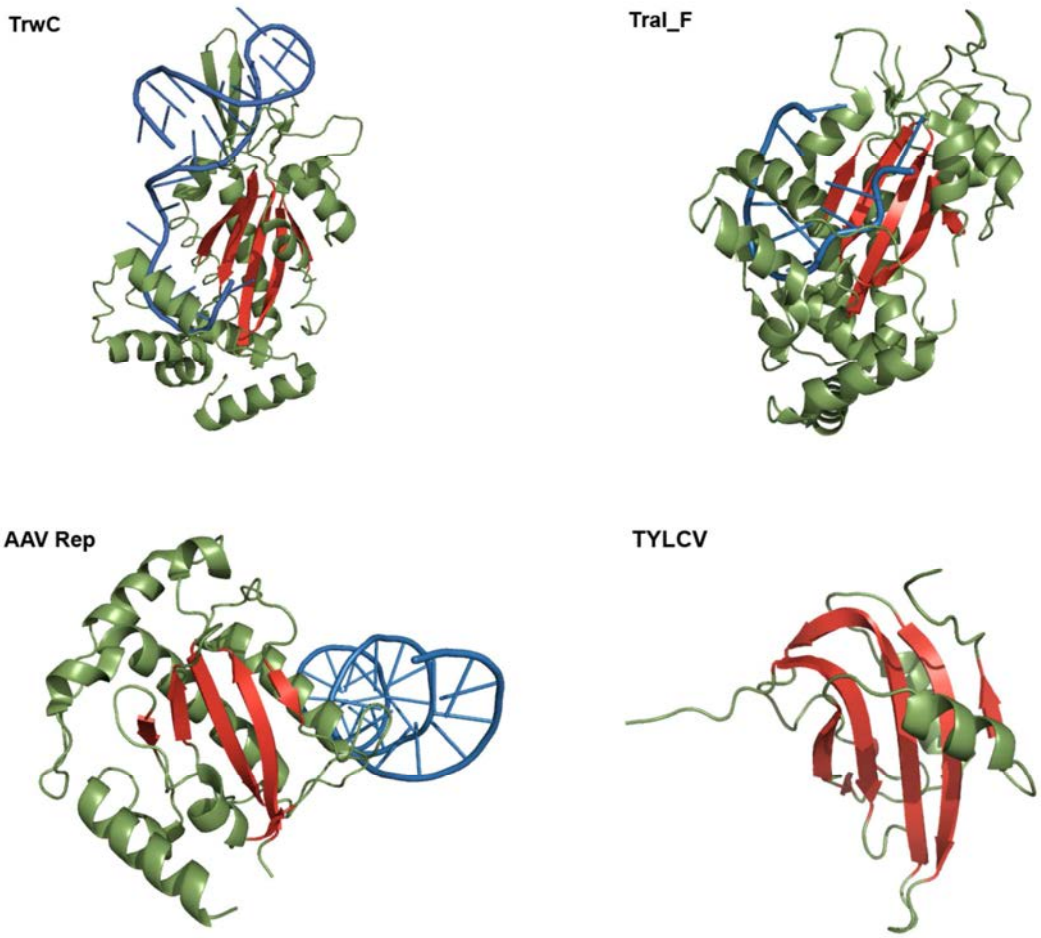

Figure 17. Cartoon representation of R388_TrwC (1QX0), F_TraI (2AO1), Rep AAV (IRZ9) and Rep TYLCV (1YLCV). The conserved central domain, corresponding to β-sheets, is shown in red; DNA chain is shown in blue.

DNA helicase activity

TrwC was shown to have DNA unwinding activity *in vitro* (Grandoso *et al.*, 1994). This activity is presumably required to unwind the DNA strand to be transferred to the recipient cell. This DNA helicase activity of TrwC is dependent on ATP, Mg^{2+} and ssDNA. Unwinding was found to be unidirectional in the 5' to 3' direction. *In vitro* it was shown that the efficiency of the unwinding reaction of fragments bigger than 100 bp decreases, meaning that purified TrwC is not a processive helicase.

The DNA helicase domain is located in the C-terminal region of the protein (192-966 aa). It contains the seven motifs characteristic of the family of DNA helicases (Matson *et al.*, 2001), including Walker A and Walker B nucleotide binding motifs. As many other DNA helicases, TrwC behaves as a dimer in solution (Grandoso *et al.*, 1994). The putative dimerization domain was assigned to the 495 C-terminal residues of the protein (472-966) (Llosa *et al.*, 1996).

3.1.2. TrwC as a site-specific recombinase

In 1994, Llosa and collaborators first described that TrwC could catalyse site-specific recombination between two *oriT* copies repeated *in tandem* (Llosa *et al.*, 1994b). This was a surprising finding, since it happened independently of conjugation and thus in the absence of single-stranded intermediates. Later on, a new recombination assay was set up to further describe this reaction (César *et al.*, 2006; César & Llosa, 2007). A recombination cassette which contains two R388 *oriT* copies separated by *nptII* and $lacI^q$ genes was constructed. Recombination was measured based on *lacZ* gene expression (**Figure 18a**). Recombination was also checked by restriction analysis of the substrate plasmid (Figure 18b). This new system permitted an exhaustive study of the requirements of the reaction (César *et al.*, 2006; César & Llosa, 2007).

TrwC-mediated site-specific recombination is influenced by host cell mechanisms such as DNA replication. When the recombination substrate had the *nic* sites on the leading strand, recombination dropped to 19 %, compared with 100 % recombinants obtained when the *nic* sites were on the lagging strand, suggesting that the exposure of ssDNA during longer times in the lagging strand favours the reaction. The first *oriT* encountered by the replication fork was named *oriT1*. The reaction was found to occur not only in *E. coli* host but also in *S. typhimurium* with the same efficiency. Using *Salmonella* mutants as host, it was confirmed that the reaction was promoted by TrwC and not by other ways of recombination such as RecA (as previously shown in *E. coli*), RecBCD and RecJ, what discarded the role of other recombination systems.

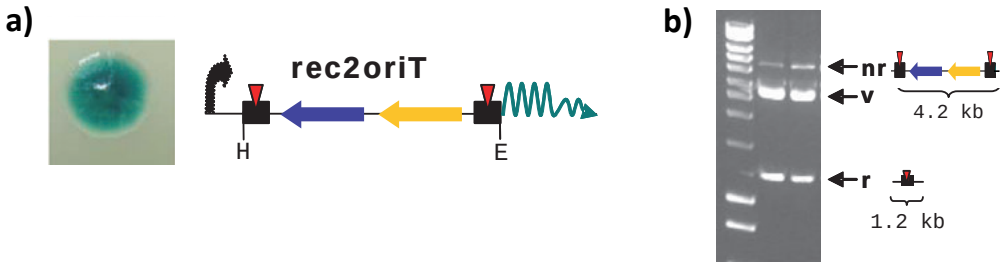

Figure 18. *oriT*-specific recombination assay. a) Structure of the recombining cassette, named *rec2oriT*. A representative colony is shown, obtained after co-trasformation with R388 into DH5α and plated on X-Gal containing plates. *oriTs* are represented by black squares with *nic* sites as red triangles. *LacI*^{q}, blue. *ntpII*, yellow. Arrows point in the direction of transcription. Black arrow, lactose promoter. H, *HindIII*. E, *EcoRI*. **b)** *EcoRI+HindIII* digestion of recombinant plasmids. nr, not recombinant. v, vector. r, recombinant. Modified from (César *et al.*, 2006).

A full *oriT* contains the putative promoter sequence for the *trwABC* operon. The role of transcription on recombination was checked. It was found that the formation of a transcription elongation complex at this locus hampers recombination. It was hypothesized that the formation of an active transcription complex could be affecting the DNA topology at *oriT1* and therefore impeding TrwC-mediated recombination.

It was also demonstrated that TrwC as a recombinase is able to catalyse site-specific recombination but it is incapable of producing an inversion of two inversely repeated *oriT* copies.

The recombinase activity of TrwC was described as not a general feature of relaxases since the ability of F_TraI (see section 3.2.2) and pKM101_TraI to support recombination was tested and the results showed that they are not capable of performing this reaction (César *et al.*, 2006).

Protein requirements

The site-specific recombination reaction is dependent on TrwC, since a R388 TrwC mutant does not catalyse recombination (Llosa *et al.*, 1994b). The relaxosome accessory protein TrwA was found to markedly favour the site-specific recombination reaction (**Figure 19**). This protein assists the TrwC nicking reaction (Moncalián *et al.*, 1997) (see section 2.2.2). The N-terminal 84 residues of TrwA, which maintain DNA binding ability (Moncalián & de la Cruz, 2004), were sufficient to enhance recombination. TrwA was

found to exert its influence on the recombination reaction independently from its putative *oriT* binding sites, which are dispensable at both recombination loci.

To directly test that the recombination reaction was reflecting the strand-transferase activity of TrwC, a set of point mutants affecting the relaxase active site that were known to abolish conjugation (Grandoso *et al.*, 2000; Guasch *et al.*, 2003) were checked for their capacity to promote recombination. The double mutant in the catalytic Tyr residues was totally deficient in recombination, reflecting that both Tyr play a role in the reaction.

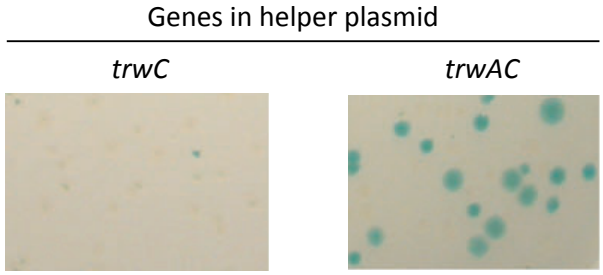

Figure 19. Effect of TrwA in TrwC-mediated recombination. Colonies obtained after incubation of the recombination substrate with the indicated helper plasmid. Each recombination event is seen as a blue sector in the colonies. In the presence of TrwA, all the colonies underwent recombination. Modified from (César *et al.*, 2006).

Although the relaxase activity is essential for recombination, the relaxase domain alone was not enough to promote efficient recombination. It was proved that the DNA helicase activity is dispensable for recombination. With a series of deletion and insertion mutants, the minimal recombinase domain was mapped to the N-terminal 600 residues.

This means that TrwC contains a third functional domain (**Figure 20**), which is the smallest expressed polypeptide with efficient recombination activity. The N600 domain behaves as a monomer in gel filtration assay, suggesting that oligomerization is not required to promote recombination (César *et al.*, 2006).

Apart from the R388 proteins, it was found that the host protein IHF (section 2.2.2) also influenced the reaction (César *et al.*, 2006; César & Llosa, 2007). IHF binding to its sites within *oriT1* exerts a negative regulatory role in TrwC-mediated recombination. The reaction was favoured in an IHF deficient strain (César & Llosa, 2007).

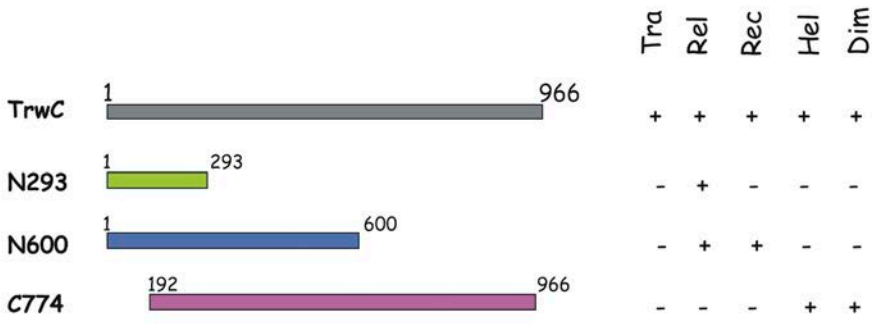

Figure 20. Functional domains of TrwC. N293, relaxase domain. N600, recombinase domain. C774, DNA helicase domain. Tra, conjugative DNA transfer. Rel, *in vitro* relaxase activity. Rec, *in vivo* site-specific recombinase activity. Hel, DNA helicase activity. Dim, dimerization ability. Data taken from (César *et al.*, 2006; Llosa *et al.*, 1996). Taken from (Agundez, 2011).

DNA requirements

In order to determine the DNA sequence requirements for *oriT*-specific recombination, the reaction was tested with different deletions of the 330 bp of R388 *oriT* in both copies in the recombination substrate. It was found that at *oriT1*, a sequence of only 33 bp, comprising the *nic* site and TrwC binding site, was enough for an efficient reaction. In contrast, the DNA required at *oriT2* showed an absolute dependence on the complete *oriT* region 3′ to the *nic* site. Deletion of the set of five iterons (bp 1–63) abolished recombination (César *et al.*, 2006) (**Figure 21**).

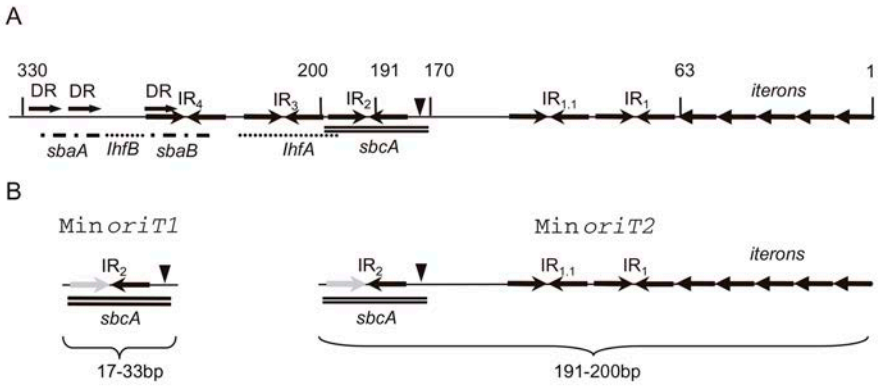

Figure 21. *oriT* sequence requirements. A) Sturcture of R388 *oriT*. The 1-330 bp *oriT* segment is numbered as in (Llosa *et al.*, 1991). Symbols as in Figure 9. **B)** Extent of DNA essential for recombination at each *oriT* copy. The range of bp indicated at the bottom are the extent of essential bp ± the distal arm of IR_2, shown in light grey to indicate that it is not an essential feature but it is required for efficient recombination. Taken from (César *et al.*, 2006).

Recombination also occurred, although with lower efficiency, when *oriT1* was narrowed to 17 bp, deleting the outer arm of the IR that comprises the TrwC binding site (IR_2). The role of the IR_2 at *oriT1* to boost recombination is probably due to binding of TrwC to the double chain formed by base pairing at IR_2, favouring processing of *oriT1* and triggering site-specific recombination (César *et al.*, 2006).

Model for TrwC-mediated *oriT*-specific recombination

Based on the above observations, César and collaborators proposed a model for TrwC-mediated site-specific recombination (César *et al.*, 2006) (**Figure 22**).

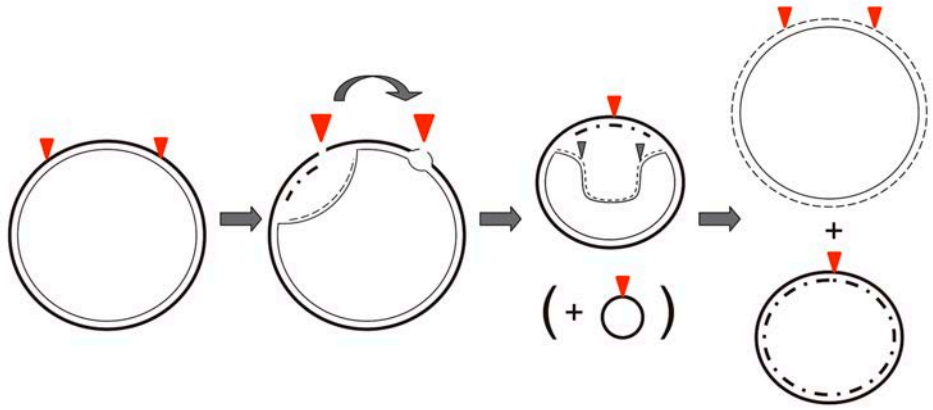

Figure 22. Model for TrwC-mediated site-specific recombination. The two plasmid DNA strands are depicted. The lagging strand is shown by a thicker line. DNA replication is shown by dashed lines. *nic* sites are shown by red vertical arrowheads. Initiation of recombination occurs when the replication fork generates a single-stranded region around *nic*. TrwC cleaves the ssDNA at site 1 and captures the DNA. A second cleavage reaction occurs at site 2. TrwC then catalyses the strand-transfer reaction (curly arrow) on the strand containing both *nic* sites. The intermediate generated is then resolved by concomitant DNA replication. Taken from (César *et al.*, 2006).

Since situations that increase the chance of the first *nic* site being as ssDNA (replication in the lagging strand, transcription) were observed to enhance recombination, it was suggested that the limiting step for recombination initiation would be the generation of a single strand at *oriT1*. Single-strand exposure would facilitate the formation of a hairpin at IR_2 to which TrwC firmly binds.

On the second stage of the reaction, after TrwC-mediated nicking and strand-transfer reactions took place between two direct repeats of *oriT*, an intermediate would be generated that could be resolved by host-encoded functions. Since no main recombination genes seem to be involved in TrwC-mediated recombination, a plausible scenario is that concomitant plasmid replication gives rise to two daughter molecules, one like the original plasmid and the other being the recombined product. Adopting this model would actually explain the inability of TrwC to resolve recombination between two inverted *oriT* copies and promote DNA inversion, which probably reflects the fact that TrwC is acting as a single stranded endonuclease at both *oriT* copies, leaving the other strand uncut. The reaction is affected by the presence or absence of accessory proteins at *oriT1* (**Figure 23**).

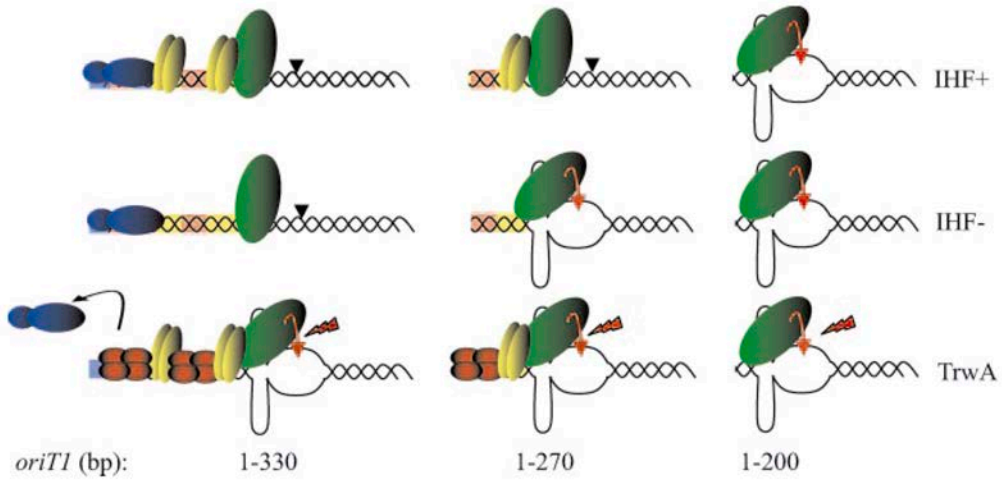

Figure 23. Model for the role of accessory proteins in the modulation of TrwC-mediated recombination. IHF dimers are shown in yellow, RNA polymerase in blue, TrwC monomer in green, and TrwA is represented as a tetrameric protein in red. Corresponding coloured shadows indicate protein-binding sites. Lightning bolts represent the enhancing effect of TrwA on TrwC activity, a black curved arrow shows the ejection movement of RNA polymerase from the *oriT* caused by the action of TrwA, and red curved arrows represent TrwC catalytic nicking of the *oriT* at the *nic* site. The lengths of the *oriTs* at recombination locus 1 are indicated below the columns. Binding of IHF at the *oriT* represses nicking by TrwC. In the absence of IHF, formation of a transcription elongation complex exerts a similar inhibition. TrwA relieves the inhibition imposed both by IHF and RNA polymerase. Short sites (column 3) are constitutively in an "open" conformation, facilitating TrwC nicking. Taken from (César & Llosa, 2007).

Binding of IHF and RNA polymerase at *oriT1* would impose a topological constraint, by IHF-mediated DNA bending and transcription-coupled DNA supercoiling. It would maintain the relaxosome in a "closed" conformation, preventing ssDNA exposure and hence nicking by TrwC. In turn, TrwA would act as a positive regulator, leading to the displacement of RNA polymerase from *oriT* and to a switch from a "closed" to an "open" relaxosome formation, facilitating the stabilization of the cruciform at IR_2 so that TrwC can firmly bind to catalyse cleavage at the *nic* site (César & Llosa, 2007).

3.1.3. Activity of TrwC in the recipient cell

Relaxases have been proposed to catalyse the final recircularization of the transferred DNA strand in the recipient cell. To this end, they must reach the recipient in an active form (Lanka & Wilkins, 1995). This was demonstrated for the first time for TrwC (Draper *et al.*, 2005). TrwC activity in the recipient cell was detected by the ability of the transferred protein to complement in conjugation a R388Δ*trwC* plasmid present in the recipient. Later, it was demonstrated that TrwC plays an essential role in the recipient expressing anti-TrwC antibodies in that cell, what produced a specific inhibition of R388 conjugation. In fact, the epitope recognized by one of the antibodies was mapped to a region of TrwC containing Y26 and involved in the conjugative DNA-processing termination reaction. From these and previous results, it was inferred that Y26 was catalysing the termination reaction in the recipient cell (Garcillan-Barcia *et al.*, 2007; Grandoso *et al.*, 2000).

Using suicide oligonucleotide substrates, the activity of both catalytic Y18 and Y26 residues was analysed in detail (Gonzalez-Perez *et al.*, 2007). They found that only Y18 was used for initiation while Y26 was specifically used in the termination of the reaction that occurs in the recipient cell. Based on that, they proposed a model for the mechanism of DNA processing during conjugation (**Figure 24**).

Once it gets in the recipient cell, TrwC is also capable of promoting site-specific recombination between two *oriT* sequences, as it does in the donor. Furthermore, TrwC can catalyse site-specific integration of the transferred DNA strand into an *oriT* copy present in the recipient cell (Draper *et al.*, 2005).

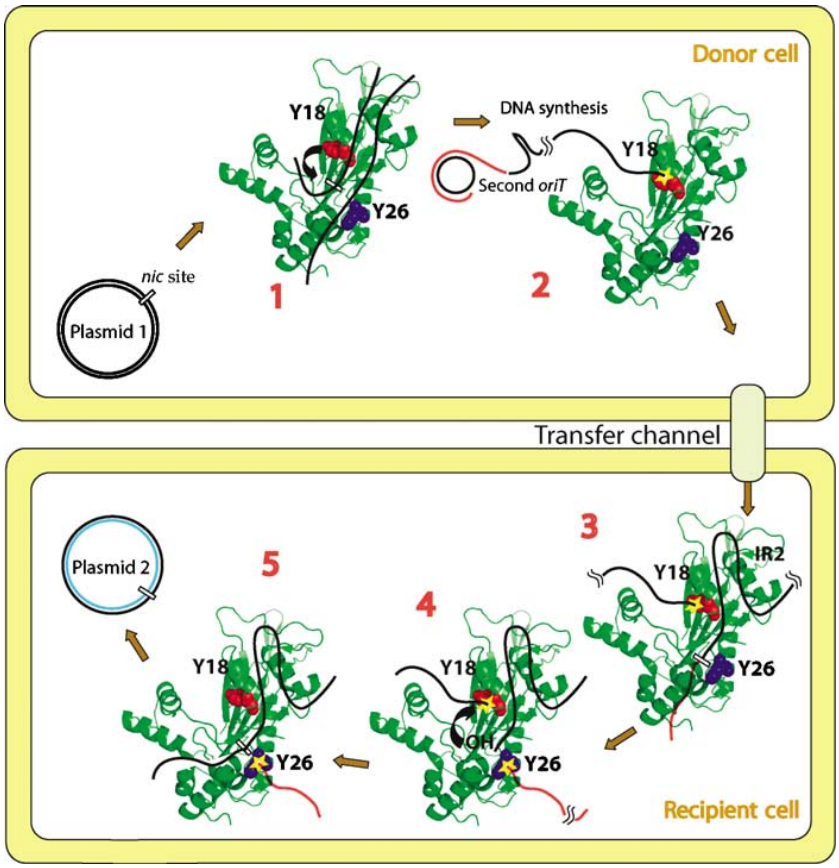

Figure 24. Model for the mechanism of DNA processing during R388 plasmid conjugation. 1) First *oriT* binding: TrwC (the protein structure is represented as a cartoon, colored in green; Y18 and Y26 residues are shown as spacefill representations in red and blue, respectively) binds supercoiled R388 plasmid DNA (black lines) around the *nic* site in *oriT*. **2)** Y18 cleavage: TrwC tyrosine Y18 cleaves R388 at *nic* (curved arrow in 1). Upon cleavage, TrwC becomes bound to the 5´ end of *nic* (yellow star). RCR in the donor plasmid DNA (new DNA represented as a red line) displaces the T-strand. TrwC bound to the T-strand pilots the DNA to the recipient cell by the transfer channel. **3)** Second *oriT* binding: TrwC specifically binds *nic* of the second *oriT*. **4)** Y26 cleavage: TrwC tyrosine Y26 cleaves the second *oriT* at *nic*, forming a second phosphotyrosine bond (yellow star bound to red line). A free 3´-OH is generated that attacks the Y18–DNA complex (curved arrow in 4). **5)** Plasmid release: DNA-strand ligation occurs by a second transesterification event, when the newly generated 3´-OH cleaves the phosphotyrosyl bond and the covalent intermediate is resolved. This last step allows recircularization and release of the transferred plasmid DNA. Taken from (Gonzalez-Perez *et al.*, 2007).

Site-specific recombinase activity in the recipient

TrwC can also catalyse this site-specific recombination reaction when the two *oriT* copies are on a plasmid in the recipient cell after being transferred itself by conjugation

(Draper *et al.*, 2005). To test this reaction in the recipient cells, matings were performed with a recipient cell containing the recombination substrate plasmid. The plasmid mobilized from the donor cell did not carry a functional *trwC* gene, so only the TrwC protein, with or without the transferred DNA, was transported to the recipient cell. 10% of the transconjugants contained the recombined substrate. To rule out the possibility that recombinants may arise in the recipient by a TrwC-independent recombination process, a CloDF13 derivative carrying R388 *oriT* was mobilized by MobC to the same recipient and in the absence of TrwC (**Figure 25**). The background recombination was 0.1%. This confirmed that the reaction is dependent on TrwC (Draper *et al.*, 2005). TrwA was found to improve the reaction in 3 times when present in the recipient cell (César *et al.*, 2006).

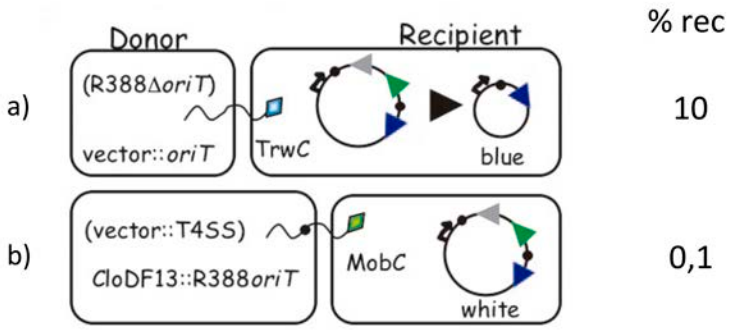

Figure 25. *oriT*-specific recombination mediated by TrwC in the recipient cell. a) TrwC protein arrives to the recipient cell by conjugation and recombination takes place. **b)** Negative control. Mobilization by MobC to the recipient cell does not promote recombination. Relaxases are shown as diamonds leading the DNA. Black circles, R388 *oriT*. Triangles: grey, *lacI*^{q}; green, *ntpII*; blue, *lacZ*. Modified from (Draper *et al.*, 2005).

Site-specific integrase activity

TrwC can also catalyse site-specific integration of the transferred DNA strand into an *oriT*-containing plasmid in the recipient cell. To test this reaction, matings were carried out between a *pir*-expressing donor strain and an *oriT*-containing recipient strain (**Figure 26a**). The *oriT*-containing mobilizable plasmid used was Pi-dependent for replication, so it could not be maintained in the recipient strain unless it became integrated. That plasmid also contained a RP4 *oriT*, used as a negative control. Integrants were obtained by selecting the mobilizable plasmid in the recipient strain what could only be found by integration of the incoming DNA into the *oriT* copy present

in the recipient cell. Integration events were found to occur at a frequency of 5.4×10^{-6} integrants per donor (Draper *et al.*, 2005). It was also demonstrated that the reaction is totally dependent on the presence of TrwC since no integrants were detected after transfer of the mobilizable plasmid by TraI-RP4 in the absence of TrwC.

The reaction (Figure 26) was later optimized and further characterized (Agundez, 2011). By using Π1 as donor strain and pSU36::*oriT* (p15A replicon) as the *oriT*-containing plasmid located in the recipient cell, the integration frequency was increased 1 log. Relaxase RP4_TraI integrase activity was also tested by mobilization of the suicide plasmid into a recipient cell containing a copy of its cognate *oriT*. No integration was detected, meaning that the reaction is specific for TrwC.

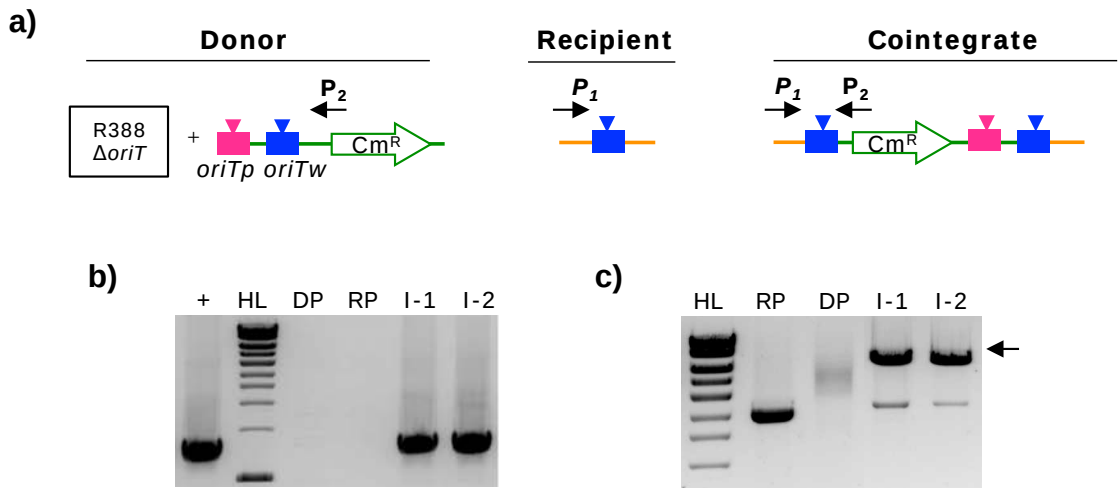

Figure 26. Site-specific integration assay. a) Scheme of the assay and the cointegrate molecule obtained. The suicide plasmid is represented with a green line and the recipient plasmid with an orange one. *oriTs* are shown as rectangles, with the *nic* sites indicated by arrowheads. P1 and P2, oligonucleotides used in the PCR reaction to detect the cointegrates. **b)** PCR-mediated detection of conintegrates using P1 and P2. **c)** Restriction analysis with XcmI, which cuts only once in the recipient plasmid. The cointegrate is indicated with an arrow. DP, donor plasmid; RP, recipient plasmid. I1-I2, DNA from two independent integrants obtained. HL, Hyperladder. Modified from (Agundez *et al.*, 2012).

The integration assay was also performed when TrwC was expressed in the recipient cells and the suicide plasmid was mobilized with RP4_TraI. An integration frequency of $1.4\text{x}10^{-7}$ was found, meaning that *trwC* expressed in the recipient cell was able to locate both *oriT*-containing plasmids and catalyse integration. However, the

frequency of the reaction decreases 2-3 logs compared to that obtained when the suicide plasmid was mobilized by TrwC.

Role of TrwA and IHF accessory proteins. To test the effect of TrwA and host-encoded IHF relaxosome proteins, the suicide plasmid was mobilized to isogenic IHF^+ and IHF^- recipient strains containing the recipient plasmid pSU::*oriT* in the presence or absence of pET3a::*trwA*. Results showed that both TrwA and IHF are enhancers of the reaction, producing an increase of about 3.5 fold and more than 30 fold, respectively, in the integration frequency. The lack of both proteins in the recipient cell decreased by two logs the frequency of integration mediated by TrwC. When integration was assayed with TrwC and TrwA produced in the recipient cell, TrwA was found to enhance TrwC-mediated integration but also to promote resolution of the cointegrates.

TrwC requirements. The integration assay expressing TrwC in the recipient allowed testing different TrwC variants for their integrase activity. The point mutants in the catalytic tyrosines Y18F and Y26F as well as the double mutant Y18F Y26F were tested. Y18F is known to strongly affect conjugation and recombination reactions, while Y26F has only a mild effect. The double mutant abolishes both processes (César *et al.*, 2006). In the integration assay, no integration was found with either TrwC Y18F or Y26F, indicating that the reaction is dependent on both catalytic residues. The helicase-deficient point mutant K502T is deficient in conjugation and proficient in recombination (César *et al.*, 2006). It showed a decrease of less than 3 fold in the integration frequency when compared to wild-type TrwC. To delimit the minimal TrwC domain sufficient to catalyse integration, N450 and N600 peptides were assayed. The recombinase domain N600 was found to catalyse also the integration reaction efficiently.

Integration into chromosomal DNA. To test integration into chromosomal DNA, two recipient strains were constructed containing an *oriT* copy in place of the chromosomal *lacZ* gene, with the *oriT nic* site lying in the lagging or leading strands. Integration was detected when mobilizing the suicide plasmid with TrwC and using the recipient containing the *oriT* in the lagging strand. The integration frequency was low ($1.1x10^{-7}$) but confirmed that TrwC is able to catalyse the integration reaction into the chromosome.

3.1.4. TrwC recruitment by T4SS

For its transfer to the recipient cell during R388 conjugation, TrwC must be recruited by its T4SS. Surprisingly, it has been reported that TrwC can also be efficiently translocated by the *B. henselae* VirB/D4 T4SS (see sections 1.1.1 and 4.1.1). For its

recruitment by any T4SS TrwC must possess the appropriate translocation signal, as it is the case of BID signal for *Bartonella* effector proteins translocated through its VirB/D4 system.

Recently, two putative translocation sequence motifs, TS1 (GDTIRIT at positions 796 to 802) and TS2 (GDRMKVV at positions 813 to 819), have been found in TrwC by BLAST alignment of TrwC with R1 TraI (**Figure 27**) (Alperi *et al.*, 2013).

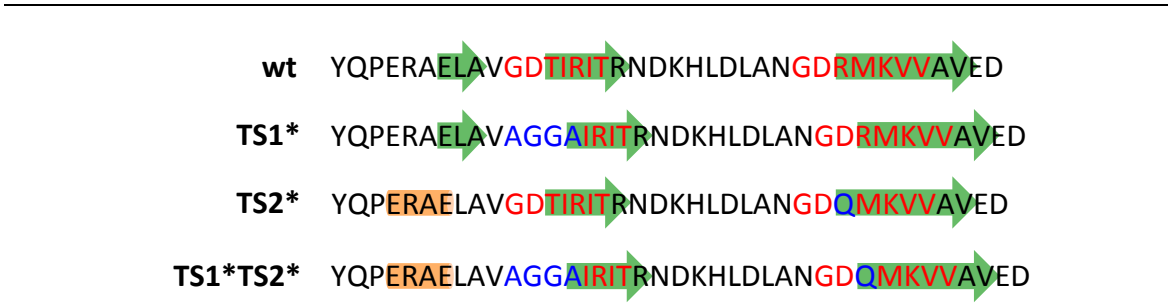

Figure 27. TrwC translocation signals. Protein secondary structure predictions are shown for TrwC wt as well as for the variants in the translocations signals (TS1* and TS2*). β-strands are indicated as green arrows while α-helices are shown as orange cylinders. The translocation signals are shown in red, and the changes introduced, in blue. Modified from (Alperi *et al.*, 2013).

Mutations in these sequence motifs were tested for bacterial conjugative transfer and translocation to human cells through the VirB/D4 T4SS of *B. henselae*. Mutations affecting TS1 drastically affected conjugation frequencies, while mutations affecting either motif had only a mild effect on DNA transfer rates through the VirB/D4 T4SS of *B. henselae*, indicating that TrwC can be recruited by two different T4SSs through different signals (Alperi *et al.*, 2013).

3.2. The relaxase F_TraI

The previously described *E. coli* DNA helicase I (Geider & Hoffmann-Berling, 1981) was found to correspond to F *traI* gene product (Abdel-Monem *et al.*, 1983). TraI is a 1756 aa protein with a molecular weight of 192 kDa (Traxler & Minkley, 1988). It belongs to the MOB_F group of relaxases (Garcillán-Barcia *et al.*, 2009) and it is biochemically and structurally related to TrwC.

3.2.1. TraI activities involved in conjugative DNA transfer

As R388_TrwC, TraI is a bifunctional enzyme with both relaxase and DNA helicase activities in the same molecule. These two activities must be physically linked to permit efficient conjugative DNA transfer (Matson *et al.*, 2001).

Relaxase activity

TraI is a site- and strand-specific transesterase that provides the nick required to initiate strand transfer. In contrast to TrwC, TraI is able to nick both supercoiled and linear dsDNA substrates carrying the *oriT*. Transesterification reaction is significantly stimulated by the addition of IHF and TraY on scDNA and is dependent on the addition of both accessory proteins on linear dsDNA molecules (Nelson *et al.*, 1995). TraM (through its *sbmC* binding site, Figure 12) was found to stimulate the relaxase reaction up to approximately 4-fold (Ragonese *et al.*, 2007). As a result of the reaction, TraI remains covalently bound to the 5´end of DNA. Addition of SDS or a proteinase is needed to obtain the DNA in a relaxed conformation.

A intrastrand three-DNA-base interaction of the DNA located in the active centre of the relaxase has been identified (Larkin *et al.*, 2005) and it was suggested to be a key specificity determinant of DNA recognition and cleavage since nearly all substitutions in those nucleotides reduce transfer efficiency (Hekman *et al.*, 2008).

TraI also has ssDNA endonuclease activity *in vitro*, being capable of binding and cleaving ssDNA oligonucleotides containing the TraI *oriT* binding site that includes the *nic* site in the only presence of a divalent metal ion (Matson *et al.*, 1993; Sherman & Matson, 1994). It was also shown that it is able to perform strand transfer to a second oligonucleotide also containing the *nic* site (Sherman & Matson, 1994).

The relaxase domain of TraI, TraI36, is located in the N-terminal 309 aa of the protein. It was found to be monomeric in solution under all conditions tested (Street *et al.*, 2003). Its three-dimensional structure with a 22 nt ssDNA oligonucleotide containing the *nic* site was resolved by Datta and collaborators in 2003 (Datta *et al.*, 2003). It shares structural features with other HUH proteins of known structure (Figure 17). The centre of the structure consists of a five-strand antiparallel β sheet (composed of strands β1, β2, β5, β8, and β9). The residues that constitute the catalytic active centre are located on one face of the β sheet. The opposite face of the β sheet packs against three α helices.

As in the case of R388_TrwC, this domain contains the three motifs described for relaxases (Ilyina & Koonin, 1992):

- Motif I: it comprises the four conserved tyrosines, Y16, Y17, Y23 and Y24. By mutagenesis in the catalytic residues, Y16 was found to be the only tyrosine catalysing ssDNA cleavage essential for F plasmid transfer (Larkin *et al.*, 2007). The structure of the TraI36 variant Y16F was obtained in 2005 (Larkin *et al.*, 2005). In contrast to what was described for R388_TrwC, TraI also catalyses the second cleavage reaction in the donor cell (Dostal *et al.*, 2011). Therefore, it was suggested that at least two TraI molecules are required for F conjugative transfer.
- Motif II: it presents an aspartic residue, D81, which may serve to position one of the histidine of motif III and may also participate in a charge relay that increases the affinity of the protein for the metal ion (Larkin *et al.*, 2005).
- Motif III: it contains the characteristic histidine triad (H146, H157 and H159) that coordinates the metal ion for the nicking reaction (Datta *et al.*, 2003).

DNA helicase activity

TraI DNA helicase activity was shown to be dependent on ssDNA, ATP and Mg^{+2} (Abdel-Monem *et al.*, 1976; Abdel-Monem & Hoffmann-Berling, 1976). Unwinding of duplex DNA was subsequently shown to occur with a 5´to 3´polarity (Dash *et al.*, 1992; Kuhn *et al.*, 1979). The length of DNA that is unwound is independent of its concentration, meaning that it is a highly processive reaction. In fact, it was found that it could unwind at least 850 bp of dsDNA under steady-state conditions (Lahue & Matson, 1988). TraI needs localized unwinding of at least 20 nucleotides at *nic* site to initiate unwinding of plasmid DNA strands (Csitkovits & Zechner, 2003). It can separate dsDNA at a rate of approximately 1100 bp s^{-1}, similar to the measured unwinding rate of the RecBCD helicase (Sikora *et al.*, 2006).

The DNA helicase domain is located between residues 310 and 1,476. Matson and co-workers showed that deletion of the region 309–349 disrupts helicase-associated ssDNA binding and the helicase activity (Byrd *et al.*, 2002). Dostal and Schildbach suggested that the helicase domain of TraI consists of two RecD-like subdomains (Fig. 1.22). The RecD-like domain II (residues 830-1473) contains helicase motifs and was proposed to be the motor domain, whereas the RecD-like domain I (residues 303-844) lacks critical helicase motifs and is likely specialized in ssDNA binding (Dostal & Schildbach, 2010). The region between the relaxase and the RecD-like II domains was

proposed to be responsible for TraI-mediated conjugative repression (Haft *et al.*, 2006) (**Figure 28**).

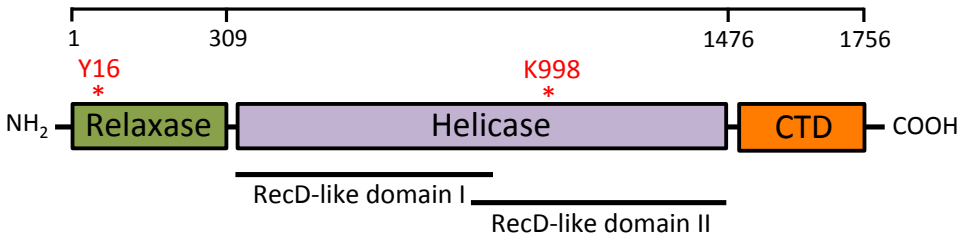

Figure 28. Functional domains of TraI. The asterisks indicate the catalytic residues Tyr-16 and Lys-998. RecD-like domains I (residues 303-844) and II (830-1473) are highlighted by black lines. Residue coordinates are given on the top. Modified from (Cheng *et al.*, 2011).

As seen in Figure 28, TraI presents an additional domain without known catalytic activity (CTD, C-terminal domain). It contains the C-terminal 280 aa of the protein. It is essential for conjugative DNA transfer. A TraI protein lacking the C-terminal 252 aa (TraIΔ252) was found to have wild-type transesterase and DNA helicase activity and to be able to interact with other components of the relaxosome. However, it was not able to support conjugative DNA transfer (Matson & Ragonese, 2005). It was then found that TraI interacts with TraM through this CTD domain (Ragonese *et al.*, 2007). The crystal structure of this CTD domain revealed a novel fold required for DNA transfer and single-stranded DNA binding (Guogas *et al.*, 2009).

3.2.2. TraI as a site-specific recombinase

Gao and collaborators originally showed that TraI was not capable of promoting recombination between two *oriT* copies in the absence of conjugation (Gao *et al.*, 1994). However, the recombining substrate plasmid used in that work was found to have the tandemly repeated *nic* sites in the leading strand (César *et al.*, 2006). Since Cesar and co-workers found that site-specific recombinase activity of R388_TrwC is influenced by host DNA replication, TraI-mediated recombination was tested using a newly constructed recombination cassette containing the *nic* sites located on the lagging strand. No recombination was observed in the absence of conjugation compared with 100% recombination observed with the R388 system in the same experiment (César *et al.*, 2006).

4. Genomic engineering in mammals

Gene therapy aims to treat disease through genomic modification in target tissues. For permanent modification, stable expression of exogenous DNA is required. Long-term expression is either achieved through extrachromosomal persistence or by integration of DNA into the human genome. Nowadays, there are two main challenges to get successful genomic modification. One is the requirement of an efficient DNA delivery method that must allow access to the specific cell type to be modified. The other is the ability to insert transgenes at a precise location in the genome.

4.1. Foreign DNA delivery

Any strategy for genome modification must include a way to deliver foreign DNA to the target cells. DNA can be introduced into human cells by a variety of methods, including naked DNA, synthetic vectors, viral vectors, or bactofection (**Figure 29**). A significant progress in the DNA delivery methods is the ability to target specific cells (tropism). Sophisticated vectors are being designed with modified surfaces to include specific ligands for attachment to target tissues.

Physical methods. These include electroporation, pressure-mediated DNA injection, ultrasound, and particle bombardment (Jafari *et al.*, 2012).

Chemical methods. They involve the use of chemical reagents such as calcium phosphate, cationic polymers or liposomes (Nishikawa & Huang, 2001).

Biological methods. They include the commonly used viral vectors like retroviruses, lentiviruses, or AAV. Bacterial ghosts, which are cell envelopes of gram-negative bacteria devoid of the cytoplasmic content, are another type of vehicle under study (Paukner *et al.*, 2005). In addition, certain bacteria have recently emerged as biological gene vectors capable of specifically delivering genes or gene products to tumour environments (bactofection) (Cummins *et al.*, 2014). Finally, T4SS of specific pathogens have been proposed as a new tool for targeted DNA delivery (Llosa *et al.*, 2012).

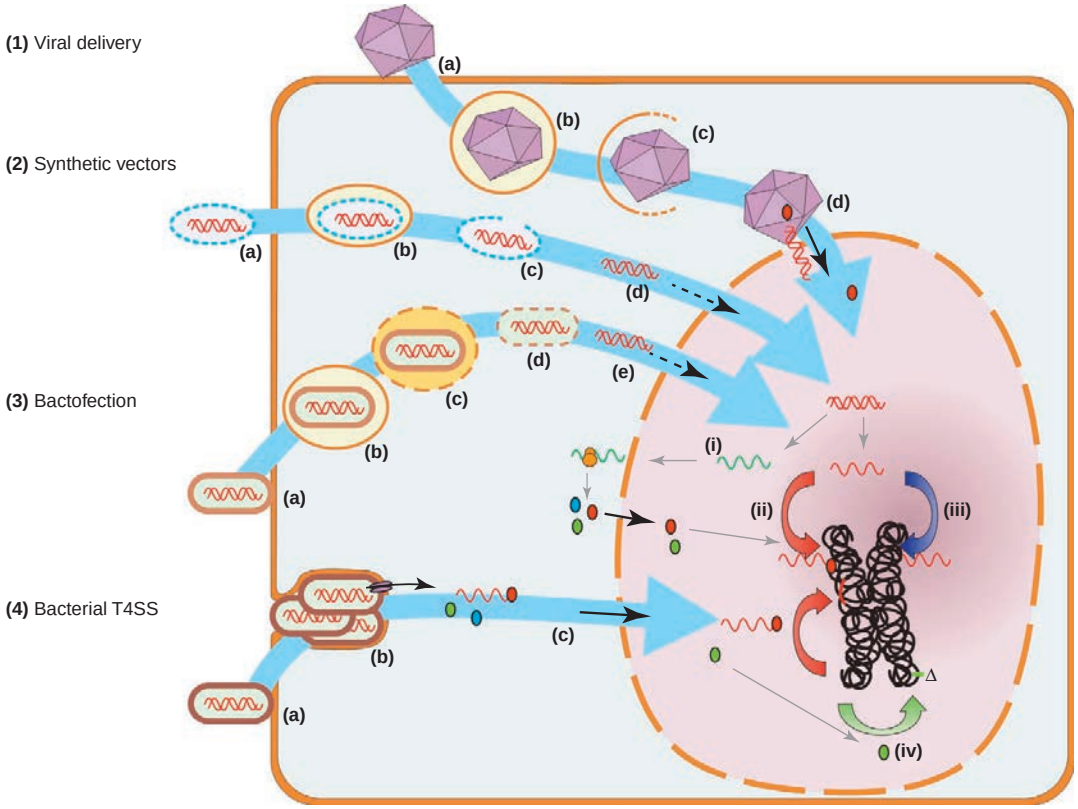

Figure 29. Schematic representation of main DNA delivery routes into mammalian cells. **(1)** Viral delivery involves the following steps: (a) binding to receptors, (b) endocytosis, (c) liberation and intracellular trafficking of the virus particle, and (d) entry of the viral DNA (and proteins, e.g., a viral integrase) into the nucleus. **(2)** Synthetic vectors follow the route: (a) receptor-mediated endocytosis, (b) liberation of the vector, (c) vector disassembly, and (d) entry into the nucleus. **(3)** Bactofection: (a,b) bacteria enter the cell via receptor binding and endocytosis, (c) the endosomal compartment is lysed, liberating the bacteria, (d) bacteria liberate DNA upon lysis and (e) DNA enters the nucleus. **(4)** Bacterial Type IV secretion (T4SS): (a) bacteria bind to receptors, (b) formation of the 'invasome' structure and secretion via the T4SS of a plasmid-derived single stranded DNA covalently bound to a relaxase (red sphere) and protein substrates (green and blue spheres) into the cytosol, and (c) entry of the bacterial DNA and proteins into the nucleus. Solid black arrows designate active nuclear targeting (1d and 4c); broken black arrows designate passive nuclear entry of DNA (2d and 3e). Taken from (Llosa *et al.*, 2012).

Physical and chemical methods are highly efficient. However, the transgene they introduce generally has a short-term expression. Another common problem is that the transgene integrates in a random manner and only in few cells (Colosimo *et al.*, 2000). Retroviruses insert their genome at non-specific sites into the host genome, potentially leading to insertional mutagenesis in the cell. Moreover, the expression of the transgene

is rapidly down-regulated due to epigenetic silencing phenomena (Pannell & Ellis, 2001). However, lentiviruses are thought to escape them and they have been successfully used to obtain transgenic animals (Hofmann *et al.*, 2003).

AAV shows a key advantage for its use as DNA delivery vector, which is that the virus naturally infects human cells posing no known health risks. In addition, the presence of AAV_Rep protein allows site-specific integration of the incoming DNA, as explained in section 4.2.2.2. Recombinant vectors have been widely used in preclinical as well as some clinical trials (Dalkara & Sahel, 2014; High *et al.*, 2014; Weinberg *et al.*, 2013). Moreover, the availability of capsids from different AAV serotypes allows for targeting specific tissues (Asokan *et al.*, 2012). The main hurdle of AAV as a targeting vector is the strict space limitation imposed by its capsid. In order to overcome this problem, attempts have been made to use dual infections (Ghosh *et al.*, 2008) or to incorporate Rep into other viruses, such as adenovirus or herpesvirus (Saydam *et al.*, 2012).

DNA delivery by bactofection has been accomplished mainly with invasive enterobacteria (Moreno *et al.*, 2010). Its efficiency may vary depending on the ability of the bacteria to replicate inside the host cell, or its location in intracellular vacuoles and the capacity to exit into the cytoplasm (Llosa *et al.*, 2012).

4.1.1. T4SS-mediated DNA delivery

Bartonella spp. infect vascular endothelial cells and erythrocytes with the help of two distinct T4SS, VirB/D4 and Trw, respectively (section 1.1.1). Two recent reports on *Bartonella*-mediated DNA transfer into human cells have demonstrated the versatility of bacterial secretion systems for macromolecular substrate transfer (Fernandez-Gonzalez *et al.*, 2011; Schroder *et al.*, 2011). In both works, the authors observed translocation of the substrate of two different bacterial conjugation systems (ssDNA covalently linked to the relaxase) mediated by the VirB/D4 T4SS of *B. henselae*.

Efficient DNA transfer into vascular endothelial cells was detected by expression of an enhanced green fluorescent protein (eGFP) cassette encoded on a bacterial plasmid. The plasmid further contained the *oriT* and conjugative genes from a cryptic *Bartonella* plasmid (pBGR1) (Schroder *et al.*, 2011) or from conjugative plasmid R388 (Fernandez-Gonzalez *et al.*, 2011). For pBGR1 system, 0.02 % of cells became GFP positive after infection with a wildt-type *B. henselae* strain carrying the mentioned plasmid (**Figure 30a**). For the R388 system, the efficiency of DNA transfer rendered 1-2 % of GFP-positive cells (Figure 30b) and was not affected by the presence of DNase I in

the culture medium. In addition to the relaxase domain, DNA helicase activity from the relaxase TrwC was found to be required for transfer of R388 derivatives, suggesting that the relaxase-DNA complex is translocated by a conjugation-like mechanism.

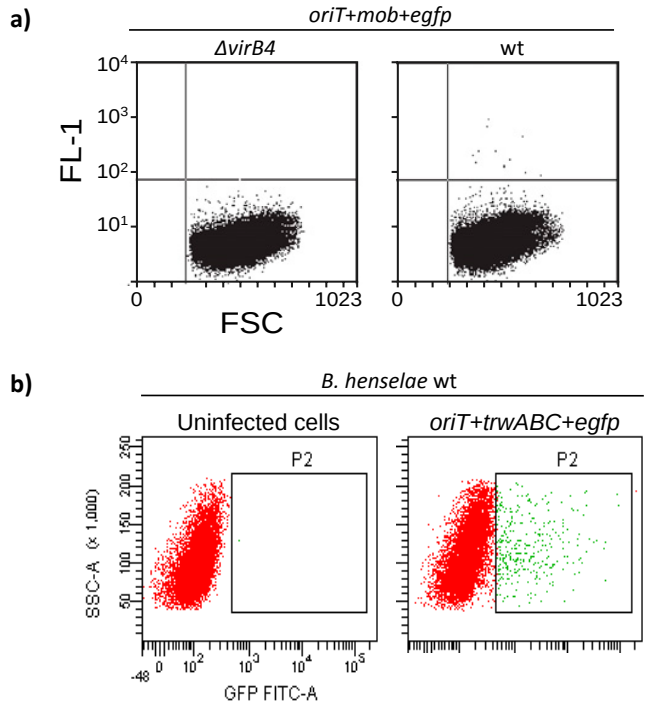

Figure 30. Transfer of DNA through *B. henselae* VirB/D4 T4SS. Fluorescence-activated cell sorting (FACS) graphs to detect GFP positive cells. **a)** DNA transfer of pBGR1-derivative carrying Mob relaxase gene. No DNA transfer was detected when using a T4SS-deficient Bartonella strain. When using the wild-type, 0.02 % og GFP positive cells were detected. FL-1, GFP fluorescence intensity. FSC, forward scatter. b) DNA transfer of R388-derivative containing TrwC coding gene. Uninfected cells were used to determine eGFP background. When using plasmid-carrying bacteria, 1-2 % of cells infected were GFP positive. SSC-A, side scatter. FITC-A, GFP fluorescence intensity. Modified from (Schroder *et al.*, 2011) and (Fernandez-Gonzalez *et al.*, 2011).

The addition of BID domain, the translocation signal for the *Bartonella* VirB/D4 T4SS (see section 1.1.1) to the relaxases increased DNA transfer. Only a 2-fold increase was obtained for R388 system (Alperi *et al.*, 2013). As noted before, wt TrwC is efficiently recruited by this T4SS (section 3.1.4). However, for pBRG1 system, a 100-fold increased transfer rate was observed (Schroder *et al.*, 2011). These data confirm that TrwC protein is a better substrate that Mob relaxase for the recruitment by the *Bartonella* VirB/D4 system. Moreover, they open they way for translocation of other proteins or protein-DNA complexes by addition of BID signal.

The potential advantages of DNA delivery by bacterial T4SS as a tool for genetic modification are outstanding (Llosa *et al.*, 2012). Different T4SS-carrying pathogens could be used to target specific cell types *in vivo*. Moreover, conjugative-like DNA transfer does no have a theoretical size limit and DNA could be translocated together with other proteins of interest.

4.2. Targeted integration of exogenous DNA

Random integration has been demonstrated to possess the risk of insertional mutagenesis, potentially leading to tumour growth (Hacein-Bey-Abina *et al.*, 2003). For that reason, the development of precise site-specific integration tools is essential for successful genomic modification. Different strategies for targeted integration are review in (Nieminen *et al.*, 2010) and (Cheng & Alper, 2014).

4.2.1. Homologous recombination approaches

Homologous recombination (HR)-based gene targeting strategies were the first approaches developed to modify site-specifically the mammalian genome (Capecchi, 1989; Smithies *et al.*, 1985). These strategies are based on the integration of an exogenous sequence that contains sequence homology to the target site. HR-mediated targeting has led to the generation of knock-in and knockout animal models via manipulation of stem cells with a high impact in many areas of research with the mouse being the typically modified organism (van der Weyden *et al.*, 2002). Despite its high site-specificity, the recombination events occur extremely infrequently (one event per 10^6-10^9 cells treated) (Capecchi, 1989).

Targeted double-strand breaks (DSBs) were found to greatly stimulate genome editing through HR-mediated recombination events (Choulika *et al.*, 1995; Rudin *et al.*, 1989). Since that observation, different strategies have been developed to introduce DSBs at a desired locus to induce HR-mediated integration (Gaj *et al.*, 2013). Different classes of customizable DNA-binding proteins have been engineered with that purpose, including zinc finger nucleases (ZFN) based on eukaryotic transcription factors, transcription activator-like effector nucleases (TALEN) from *Xanthomonas* bacteria, and most recently the RNA-guided DNA endonuclease Cas9 from the type II bacterial adaptive immune system CRISPR (Hsu *et al.*, 2014).

Zinc-finger nucleases

Chimeric nucleases that are hybrids between a nonspecific DNA cleavage domain and a zinc finger DNA recognition domain (**Figure 31**) (Kim *et al.*, 1996) were tested in the early 2000s by Bibikova and co-workers. They discovered that its use increased HR levels up to 100 % after injection of both engineered DNA substrates and nucleases into *Xenopus laevis* oocytes (Bibikova *et al.*, 2001). Meanwhile, Porteus and Baltimore found them to stimulate gene targeting in human cells (Porteus & Baltimore, 2003).

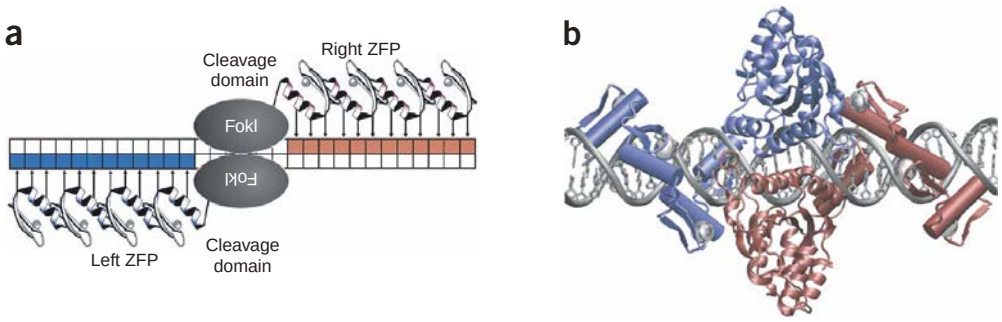

Figure 31. Sketch (a) and 3D model (b) of a ZFN dimer bound to a DNA target. Each ZFN consists of the cleavage domain of a nuclease (usually FokI Type IIS endonuclease) fused to a zinc-finger domain customized to specifically recognize either a 'left' or 'right' half-site (indicated in blue and red, respectively), with each finger recognizing a 3 nt sequence. Simultaneous binding by both ZFNs enables dimerization of the FokI nuclease domain and DNA cleavage in the region between the half-sites. Grey, DNA. Blue and red, ZFNs with reflective spheres denoting zinc ions. Modified from (Miller *et al.*, 2007).

ZFNs have been used in multiple applications ranging from precise sequence edits (Urnov *et al.*, 2005) to the targeted integration of entire genes (Coluccio *et al.*, 2013; Moehle *et al.*, 2007). In 2014 a clinical trial successfully used ZFNs to combat HIV infection by *ex vivo* knockout of the CCR5 receptor (Tebas *et al.*, 2014).

Despite the specificity provide by the ZF domain, off-target cleavage has been detected in many applications (Gabriel *et al.*, 2011; Pattanayak *et al.*, 2011). Moreover, ZFNs can be costly and time-consuming to produce, with mutation efficiencies only up to 18.8% in certain applications (Cheng & Alper, 2014; Sander *et al.*, 2011).

Transcription activator-like effector nucleases

They consist in the joining of a TALE protein to the catalytic domain of FokI nuclease (**Figure 32**). TALEs are naturally occurring DNA binding proteins found in the plant pathogenic bacteria *Xanthomonas* ssp.

TALENs have been successfully used for gene targeting in a number of different organisms, including yeast (Cermak *et al.*, 2011), plants (Mahfouz *et al.*, 2011), and human cells (Zhang *et al.*, 2011). They were found to mediate site-specific genome modification in human pluripotent cells with similar efficiency and precision as ZFNs do (Hockemeyer *et al.*, 2011). In other work, a specific TALEN showed less off-target activity compared to a ZFN, suggesting better specificity (Mussolino *et al.*, 2011). However, undesired off-target integration remains an important issue (Grau *et al.*, 2013).

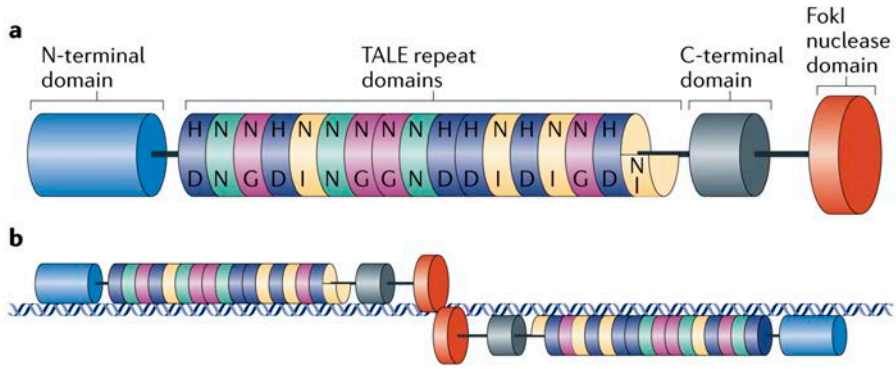

Figure 32. TALE nuclease. a) Schematic diagram of a TALEN. TALE repeats are shown as colored cylinders. Letters inside each repeat represent the hypervariable residues. The non-specific nuclease domain from the FokI endonuclease is shown as an orange cylinder. **b)** TALENs bind and cleave as dimers. Cleavage occurs in the "spacer" sequence that lies between the two regions of the DNA bound by the two TALEN monomers. Taken from (Joung & Sander, 2013).

The main advantage of TALENs over ZFNs is the mechanism by which these proteins bind to DNA. For TALENs, each module binds one individual nucleotide in the target sequence, whereas for ZFNs, each zinc finger binds 3 bp of DNA. Because zinc fingers that bind all possible 3-bp combinations are not available, a narrower range of DNA sequences can be specifically targeted by ZFNs than by TALENs (Scott *et al.*, 2014).

CRISPR/Cas9

In 2013, two different studies simultaneously showed successful engineering of type II CRISPR systems from *Streptococcus thermophilus* (Cong *et al.*, 2013) and *Streptococcus pyogenes* (Cong *et al.*, 2013; Mali *et al.*, 2013) to accomplish genome editing in mouse and human cells.

CRISPR arrays serve as a prokaryotic immune memory and defence mechanism. In type II CRISPR systems, small RNAs (crRNAs) direct Cas9 nuclease to exogenous target sequences that are then destroyed. Using design guide-RNAs, Cas9 activity can be directed to any genomic site of the form GN20GG (Mali *et al.*, 2013). This system does not require the engineering of a specific protein for each target, and the target recognition is based in DNA-RNA pairing rules (**Figure 33**). However, the need of a GN20GG sequence motif can limit the useful target sequences (Li *et al.*, 2014).

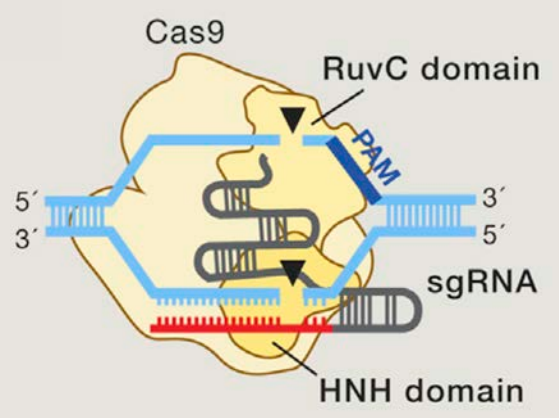

Figure 33. CRISPR/Cas9 target recognition and cleavage. The Cas9 nuclease is localized to specific DNA sequences via the guide sequence on its guide RNA (red), directed by base-pairing with the DNA target. Binding of a small recognition sequence (protospacer-adjacent motif or PAM, dark blue) downstream of the target locus helps to direct Cas9-mediated DSBs. HNH and RuvC, nuclease domains of Cas9. Modified from (Hsu *et al.*, 2014).

To date, the *S. pyogenes* Cas9 has been used broadly to achieve efficient genome editing in a variety of species and cell types, including human cell lines, bacteria, zebrafish, yeast, mouse, fruit fly, pig, and monkey (Hsu *et al.*, 2014; Sander & Joung, 2014). Homology between the guide RNA and target DNA, enzymatic concentration, and the duration of Cas9 expression are factors affecting the potential off-target activity (Hsu *et al.*, 2014). In fact, frequent off-target integration has ben found in some works (Wang *et al.*, 2015), while precise integration is shown in others (Osborn *et al.*, 2015).

4.2.2. Site-specific recombination approaches

In contrast to homologous recombination, site-specific recombination approaches do not need the design of synthetic nucleases, and the synapsis of the two DNA partners does not depend on a prior strand break. SSR-based systems have already proved their value for genetic engineering of plants and mammals (Turan & Bode, 2011). They have been mainly based on the used of classical site-specific recombinases and AAV_Rep.

4.2.2.1. Site-specifc recombinases

Site-specific recombinases (SSR) catalyse the recombination between specific target DNA sequences. They are characteristically sequence specific, very efficient, and often initiate recombination without cofactors. The recombination between both target sites is performed through four nicks and two strand-exchange reactions. They have been proved to be promising candidates as tools for genomic engineering.

Tyrosine recombinases: Cre and Flp

In this family of SSR, the residue responsible for the catalytic nucleophilic attack is a tyrosine. They mediate conservative recombination by sequential action of two pairs of monomers, each catalysing a strand-transfer reaction with the first crossover leading to the formation of a holiday junction (**Figure 34a**).

Cre, the recombinase from phage P1, mediates recombination between two copies of *loxP* target site (Abremski & Hoess, 1984). It is an example of SSR that catalyses recombination in both senses with equal efficiency. It has been widely used for the construction of transgenic animals, including the generation of conditional phenotypes (Sorrell & Kolb, 2005). More recently, it has been proved to be a highly robust and dependable gene modifier in human embryonic stem cells (hESCs) (Du *et al.*, 2009; Sakurai *et al.*, 2010). Different pseudo-*loxP* sites located in the human genome haven been found to be functional as targets for Cre-mediated integration and excision in a human cell environment at up to 100% of the efficiency of the native *loxP* site in bacterial assays (Thyagarajan *et al.*, 2000). The main problem of the use of Cre recombinase for gene therapy purposes is the reversibility of the reaction that would lead to a non-stable integration.

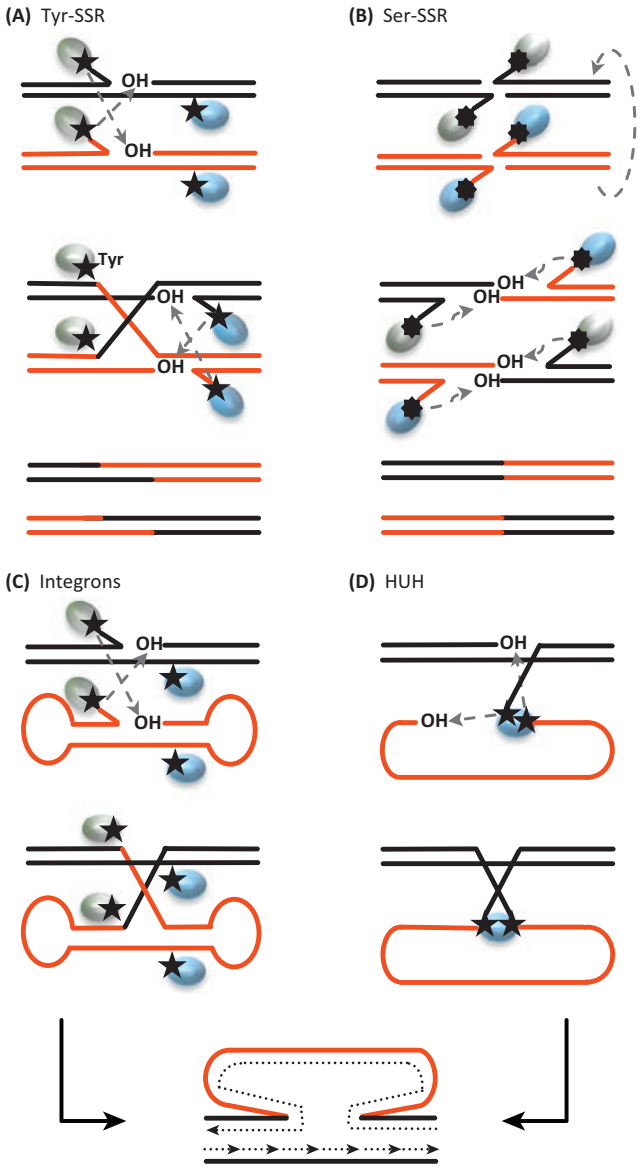

Figure 34. Comparison of recombination reactions catalysed by site-specific recombinases belonging to different families. (A) Tyr-SSRs. **(B)** Ser-SSRs. **(C)** Integrons. **(D)** HUH family proteins. The proteins are represented in blue and green while the DNA molecules are shown in red and black. The mechanism of the different SSRs is described in the text. Taken from (Gonzalez-Prieto *et al.*, 2013).

Flp is the recombinase of *S. cerevisiae* 2-micron plasmid and it recognizes FTR target site (Gronostajski & Sadowski, 1985). Flp has been shown to be able to mediate recombination in plant (Kilby *et al.*, 1995), insect (Chou & Perrimon, 1992), and mouse cells (Fiering *et al.*, 1995). It can effectively mediate site-specific excisional recombination in mouse embryonic stem cells and in transgenic mice (Dymecki, 1996)

as well as gene activation and integration in mammalian cells (O'Gorman *et al.*, 1991). However, a non-stable integration is also the main drawback of this system.

Serine recombinases: ΦC31

This is a very heterogeneous group of enzymes involved in resolution and inversion reactions. They all have in common that the nucleophilic attack is mediated by a serine residue. Ser-SSR mediate conservative recombination by concerted cleavage of four monomers, followed by a conformational switch of two of them, and both strand exchanges (Figure 34b).

The most outstanding member of this family is the integrase from *Streptomyces* phage ΦC31 (Int-ΦC31). It performs a unidirectional reaction between phage *attP* and host *attB* sites. Additional factors are needed to catalyse recombination in the newly created sites upon recombination (Khaleel *et al.*, 2011). The use of Int-ΦC31 for integration of exogenous DNA in mammals originated more than a decade ago. The system is very efficient, rendering a high percentage of viable transformed cells. It has been used for genetic correction in mice and also in cultured human cells, including stem cells. Notably, in the past few years this system has allowed phenotypic correction of hemophilia A and B in mice through the expression of human clotting factors (Chavez *et al.*, 2012; Keravala *et al.*, 2011). It has also been used for targeted integration in human muscle and cardiac progenitor cells (Lan *et al.*, 2012; Quenneville *et al.*, 2007), stem cell lines (Thyagarajan *et al.*, 2008), as well as in approaches to generate mouse iPS cells (Karow *et al.*, 2011).

The main drawback of using Int-ΦC31 is the presence of too many possible insertion sites (Chalberg *et al.*, 2006). This problem may be partially overcome by the use of mutant integrases obtained by directed evolution (Sclimenti *et al.*, 2001). Another hurdle is based on the observation that integration may lead to chromosomal rearrangements (Ehrhardt *et al.*, 2006).

Other Ser-SSR phage integrases have been characterized and used successfully in a mammalian environment. R4 and A118 integrases can integrate an incoming plasmid into endogenous pseudo *att* sites in the human genome, although aberrant chromosomal events have been associated with R4, and for A118, four out of 15 integration events at pseudo *attB* sites showed imperfect junctions (Keravala *et al.*, 2006; Olivares *et al.*, 2001).

4.2.2.2. HUH site-specific recombinases: AAV_Rep and TrwC

HUH proteins are defined by the presence of two conserved motifs: a His-hydrophobic-His motif required for metal ion binding, and a motif containing one or two catalytic tyrosines for nucleophilic attack of the DNA. HUH family members are strand transferases acting at target sites on a single DNA strand, and are involved preferentially in biological processes involving ssDNA intermediates, such as rolling-circle replication and transposition, or bacterial conjugation (Chandler *et al.*, 2013). In addition, some of them have been shown to catalyse site-specific recombination and integration reactions on dsDNA substrates. These include conjugative relaxase TrwC of plasmid R388 (César *et al.*, 2006); rolling-circle replicase Rep from AAV (Smith & Kotin, 2000); and transposase TnpA(REP) from *E. coli* K12 (Ton-Hoang *et al.*, 2012). To date, all proposed models involving ss-based SSR include a replication step, carried out by the host machinery, in order to complete the reaction (Figure 34d).

AAV_Rep

It is involved in all the important processes of human adeno-associated virus (AAV) life cycle, such as site-specific integration in the human genome. It recognizes a unique target sequence (AAVS1) located on chromosome 19, and several studies have demonstrated that insertion at this site poses no apparent health risks (Henckaerts *et al.*, 2009).

AAV_Rep-mediated site-specific integration of foreign DNA has been achieved, among others, in mouse (Henckaerts *et al.*, 2009) and human ESCs (Yang *et al.*, 2008), opening the possibility of using this system in replacement therapies in several human diseases. One of the main advantages of the use of AAV_Rep for genomic engineering is that the AAV itself is a vector to introduce the Rep protein into the target cell. To avoid cytotoxicity, inducible *rep* promoters could be engineered. Although some reports have challenged its site-specificity (Drew *et al.*, 2007), a recent, unbiased analysis of integration sites has provided further evidence for the specificity of Rep-mediated integration (Ward & Walsh, 2012).

R388_TrwC

TrwC is, apart from its role in conjugation, a site-specific recombinase (section 3.1.2) and a site-specific integrase (section 3.1.3). Although it has not been used with gene editing purposes yet, these activities confer to TrwC a high potential as a

biotechnological tool that could be delivered to the cell to be modified together with the DNA though a bacterial T4SS. Another fact supporting this potential is the similarity between TrwC and AAV_Rep (Figure 17) and the ability of the last one to catalyse integration into the human genome. This potential use of TrwC has been explored by Agúndez and co-workers (Agundez *et al.*, 2011).

As mentioned before (sections 3.1.1 and 3.1.2), an *oriT* core sequence of 17 bp is enough for TrwC to achieve conjugative mobilization and recombination. A search in the human genome database looking for putative TrwC target sequences was performed. No matches were found in the human genome for the (14+3) region. When allowing one mismatch, seven hits were found (**Figure 35**), all of them located in non-annotated or intronic regions. The top five matches shown in the Figure 35 present the mismatch out of the critical (6+2) region (Agundez *et al.*, 2011).

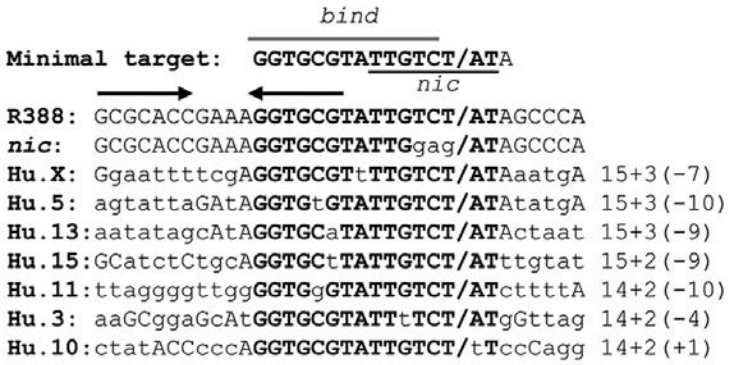

Figure 35. Possible TrwC target sequences in the human genome. The minimal region required for TrwC activity, with binding and *nic* regions, is shown in bold face. The inverted repeat is indicated by arrows. The minimal *oriT* sequence corresponds to nucleotides 173- 191 (Llosa *et al.*, 1991). The seven human sequences are named indicating the chromosome number and the homologous nucleotides 5'+3' with respect to the *nic* site (with the mismatch position in brackets). Nucleotides identical to the R388 sequence are in capital letters. Modified from (Agundez *et al.*, 2011).

In vitro, oligonucleotides containing both Hu.X and Hu.5 sequences, the most similar ones to the natural target, were bound by TrwC roughly with the same affinity as the R388 sequence and both acted efficiently as acceptors for TrwC-mediated strand transfer (Agundez *et al.*, 2011). *In vivo*, TrwC was found to catalyse site-specific recombination when the *oriT1* was replaced by either Hu.X or Hu.5 with the same efficiency as with its minimal wt target (Agundez *et al.*, 2011). Integration into HuX and

Hu5 human sequences in the presence of TrwA was also tested and TrwC was found to catalyse the reaction into both human target sequences. The efficiency of the reaction decreases by only 2–3 times compared to the wild-type minimal *oriT*. When DNA from the integrants was analyzed, 100% of the molecules were cointegrates, suggesting that the deviations from the consensus in the human sequences prevent resolution of the cointegrates formed (Agundez *et al.*, 2012).

TrwC was shown to localize in the cytoplasm of human cells, apparently due to the presence of a nuclear export signal in its C-terminal domain. However, a TrwC variant with nuclear localization has been obtained. It has a point A904T mutation and an extra peptide at its C terminus, and maintains its functionality in conjugation and recombination (Agundez *et al.*, 2011).

In summary, TrwC is active in the recipient cell, being able to catalyse site-specific integration of the incoming DNA. Moreover, a discrete number of putative targets have been found in the human genome and the possibility of nuclear targeting has already been proved. This together with the translocation of TrwC-DNA complexes through T4SS to specific human cells make TrwC a likely candidate as a tool for gene editing (Gonzalez-Prieto *et al.*, 2013).

Aims and scope

Conjugative relaxases are responsible for the processing of the DNA during bacterial conjugation. In addition, a few of them are also able to promote site-specific recombination between two tandemly repeated *oriT* copies on double-stranded DNA substrates in the absence of conjugation. Most remarkably, the MOB_F relaxase TrwC is able to promote site-specific integration of a conjugatively transferred DNA into a second copy of the *oriT* present in the recipient cell. At present, it is not known why some relaxases can perform these reactions while others, apparently more closely related, cannot. The biological role of such activities is also open to debate.

These abilities confer TrwC to a high potential as a biotechnological tool. TrwC and any covalently linked *oriT*-containing DNA could be conjugatively transferred to a recipient cell (which would not necessarily be restricted to prokaryotes) where the realxase could catalyse integration of the incoming DNA into a target sequence present in the recipient genome. The utility of this potential tool depends on our ability to manipulate the system.

The main objectives that we have addressed in this PhD thesis work are:

1. To understand the mechanism of site-specific recombination and integration reactions mediated by TrwC. The specific goals for that purpose were:
    a. To complete the characterization of TrwC-mediated integration reaction with respect to the target DNA sequence in both the donor and the recipient *oriT* copies.
    b. To select TrwC variants with enhanced site-specific recombinase and integrase activity, which would provide insight on the key TrwC residues involved in these processes.

2. To determine the extent of the *oriT*-specific recombination/integration abilities among conjugative relaxases. For this purpose, we wanted to test these activities in F_TraI and pKM101_TraI, two relaxases belonging to the same MOB_F family, previously reported as recombinase deficient.

3. To test the ability of TrwC to mediate site-specific integration in human cells, by selection of integration events after TrwC-mediated DNA transfer, followed by characterization of the integration pattern.

With this approach we intend to get insight into the molecular mechanism of theses intriguing reactions, and to evaluate their possible use as tool for genomic modification in eukaryotes that could have applications in biotechnology and biomedicine.

Experimental Procedures

1. Bacterial and yeast strains

Strain	Genotype	Reference
Bartonella henselae		
RSE247	SmR spontaneous mutant of ATCC 49882	(Schmid *et al.*, 2004)
Escherichia coli		
CMS1	Rf^R Km^R; HMS174::Km^R+*oriT*[(1)]	(Agundez *et al.*, 2012)
CMS2	Rf^R Km^R; HMS174::Km^R+*oriT*[(1)]	(Agundez *et al.*, 2012)
D1210	Sm^R*; recA hspR hsdM rpsI lacI*^{q}	(Sadler *et al.*, 1980)
DH5α	Nx^R; *F*$^-$ *endA1 hsdR17 supE44 thi-1 recA1 gyrA96 relA1 Δ(argF-lacZYA) U169 Φ80dlacZΔM15*	(Grant *et al.*, 1990)
HMS174	Rf^R; *F*$^-$ *recA hsdR rpoB*	(Campbell *et al.*, 1978)
S17.1 *λpir*	Sm^R; *pro res mod1 RP4-2 Tet::Mu Kan::Tn7*	(de Lorenzo & Timmis, 1994)
β2163	Km^R; Em^R; *F*$^-$ *RP4-2-Tc::Mu dapA::(erm-pir)*	(Demarre *et al.*, 2005)
Π1	Nx^R; Em^R; as DH5α, *Δthy::(erm-pir116)*	(Demarre *et al.*, 2005)
Saccharomyces cerevisiae		
Y13652	BY4742; *Mat a; his3D1; leu2D0; lys2D0; ura3D0; YDR293c::kanMX4*	EUROSCARF collection

Table 1. **Bacterial and yeast strains used in this work.**

[(1)] CMS1 and CMS2 have the *oriT* lying in the lagging and in the leading strand, respectively.

2. Plasmids

2.1. Published plasmids used in this work

Plasmid	Phenotype	Description	Reference
p99I^{+}	Ap^{R}	pTrc99A::*traI*(F)	(Haft *et al.*, 2006)
pCIG1000	Ap^{R}	pET3a::*trwAC*	(Draper *et al.*, 2005)
pCIG1028	Cm^{R} Km^{R}	R388 o*riT* recombination substrate	(César *et al.*, 2006)
pCIG1032	Cm^{R} Km^{R}	As pCIG1028, *oriT1* from pKM101	(César *et al.*, 2006)
pCIG1063	Cm^{R} Km^{R}	As pCIG1028, *oriT1* from F	(César *et al.*, 2006)
pCIG1064	Cm^{R} Km^{R}	F o*riT* recombination substrate	(César *et al.*, 2006)
pCIG1066	Cm^{R} Km^{R}	pKM101 *oriT* recombination substrate	(César *et al.*, 2006)
pCIG1077	Ap^{R}	pKK223-3::P_{ABC}*trwA-trwL*	(César & Llosa, 2007)
pCMS11	Cm^{R}	R6K(oriV)::$oriT_W$+$oriT_P$ **(1)**	(Draper *et al.*, 2005)
pDEL10	Cm^{R}	pMTX524::*trwB*(K275A)	(de Paz *et al.*, 2010)
pET3a	Ap^{R}	Expression vector	Novagen
pHP159	Gm^{R}	pBBR6::*oriT trwABC*+*egfp* **(2)**	(Fernandez-Gonzalez *et al.*, 2011)
pHP161	Gm^{R}	pBBR6::*oriT trwABC*+*egfp* **(2)**	(Fernandez-Gonzalez *et al.*, 2011)

Table 2. Published plasmids used in this work.

(1) $oriT_P$, RP4 *oriT*. $oriT_W$, R388 *oriT*. $oriT_F$, F oriT (in the whole Table).

(2) pHP159/161: pHP159 and pHP161 differ only in the orientation of the *egfp* cassette, which is in the same orientation as the Plac promoter in pHP161 and in the opposite orientation in pHP159.

Plasmid	Phenotype	Description	Reference
pHP181	Gm^R	pBBR6::*oriT trwAB+egfp*	(Fernandez-Gonzalez *et al.*, 2011)
pKK::*oriT-km*	Ap^R	pKK223-3::*oriT+ntp* gene	(Draper *et al.*, 2005)
pLA24	Gm^R	pHP159::BID	(Fernandez-Gonzalez *et al.*, 2011)
pLA31	Km^R	pSU36::$oriT_P$	(Agundez *et al.*, 2012)
pLA32	Km^R	pSU36::$oriT_W$	(Agundez *et al.*, 2012)
pLA33	Km^R	pSU36::$oriT_W$ (mutIR)	(Agundez *et al.*, 2012)
pLA34	Km^R	pSU36::$oriT_W$ (mut23-25)	(Agundez *et al.*, 2012)
pMTX326	Ap^R	pUC9::$oriT_p$	Matxalen Llosa
pMTX708	Ap^R Hyg^R	pTRE2hyg::Ptac-$oriT_W$ **(3)**	Matxalen Llosa
pMTX709	Ap^R Hyg^R	pTRE2hyg::$oriT_W$-Ptac **(3)**	Matxalen Llosa
pMTX712	Km^R	pET29c::*trwAC*	(Draper *et al.*, 2005)
pOX38	Km^R	Derivative of plasmid F	(Chandler & Galas, 1983)
pOX38Δ*traI*	Km^R, Tc^R	*traI::tetRA*	(Lang *et al.*, 2010)
pRS56	Ap^R	Cre-Stoplight::loxH+*neo*	(Schulein *et al.*, 2005)
pRS130	Km^R	pBGR::*mob:BID+gfp+neo*	(Schroder *et al.*, 2011)
pRS413IVb	Ap^R His^+	Yeast shuttle vector	(Kitazono, 2011)

Table 2. Published plasmids used in this work (continued).

(3) pMTX708/9: they differ in the orientation of the Ptac-*oriT* cassette. In pMTX708, the *oriT* is closer to the hygromycin resistance gene while in pMTX709 it is the Ptac what is closer to the resistance gene.

Plasmid	Phenotype	Description	Reference
pSU36	Km^R	Cloning vector	(Bartolomé *et al.*, 1991)
pSU1186	Ap^R	pUC8::$oriT_W$	(Llosa *et al.*, 1991)
pSU1445	Km^R, Tp^R	R388::*Tn5tac1* in *trwC*	(Grandoso *et al.*, 1994)
pSU1621	Ap^R	pET3a::*trwC*	(Guasch *et al.*, 2003)
pSU2007	Km^R Tp^R	R388 with Km^R cassette in Su^R gene	(Martinez & de la Cruz, 1988)
pSU4632	Cm^R	pSU24:: R388*trwAtrwBK136T*	(Moncalián *et al.*, 1999a)
pSU4633	Cm^R	pSU24::R388*trwAtrwB*	(Moncalián *et al.*, 1999a)
pSW23	Cm^R	Suicide cloning vector (R6K oriV)	(Demarre *et al.*, 2005)
R388	Su^R Tp^R	Wild-type conjugative plasmid	(Datta & Hedges, 1972)

Table 2. Published plasmids used in this work (continued).

2.2. Plasmids constructed for this work

				Construction[1]		
Plasmid	**Description**	**Phenotype**	**Size (kb)**	**Vector**	**Insert**	**Digestion / oligonucleotides**
pCOR16	pET3a::*trwC**[2]	Ap^R	7.6	pET3a	pMTX712	Mutagenic PCR[3] TAATACGACTCACTATAGGG GCTAGTTATTGCTCAGCGG NdeI-BamHI
pCOR17	pET3a::*trwC**	Ap^R	7.6	pET3a	pCOR16	NdeI-BamHI
pCOR18	pET3a::*trwAC**	Ap^R	8.1	pCOR17	pMTX712	XbaI-NdeI
pCOR26	pSU36::$oriT_F$[4]	Km^R	2.9	pSU36	pCIG1063	CAAGAATTCCGCACCGCTAGCAGCGC CAAAAGCTTGGCTCAACAGGTTGGTGG EcoRI-HindIII

Table 3. Plasmids constructed for this work.

[1] First column lists the vector plasmids; second column lists the plasmids from which the inserts were obtained; and third column indicates either the restriction enzymes used for cloning, or the oligonucleotides used for PCR amplification of the desired fragment (with the restriction sites underlined) and the enzymes subsequently used.

[2] *trwC** refers to *trwC* mutant #1 that showed enhanced recombination activity.

[3] Mutagenic PCR performed as described in section 3.4 of experimental procedures.

[4] $oriT_P$, RP4 *oriT*. $oriT_W$, R388 *oriT*. $oriT_F$, F oriT (in the whole Table).

				Construction		
Plasmid	**Description**	**Phenotype**	**Size (kb)**	**Vector**	**Insert**	**Digestion / oligonucleotides**
pCOR27	pCIG1064::*oriT*$_P$	Cm^R Km^R	7.9	pCIG1064	pMTX326	CAAAAGCTTGAATTCCGCTTGCCCTCATCTG CAAAAGCTTGCGCTTTTCCGCTGCATAA HindIII
pCOR28	R6K(oriV)::*oriT*$_F$-*oriT*$_P$	Cm^R	2.1	pSW23	pCOR27	EcoRI-XbaI
pCOR29	pRS413IVb::*oriT*$_W$	Ap^R His$^+$	5.2	pRS413IVb	pSU1186	AACGAATTCATGTCCTCTCCC AACCTCGAGCTCATTTTCTGCATC EcoRI-XhoI
pCOR30	pRS413IVb::*oriT*$_P$	Ap^R His$^+$	5.2	pRS413IVb	pLA31	CCAAGAATTCCGCTTGCCCTCATCTG CCAACTCGAGGCGCTTTTCCGCTGCATAA EcoRI-XhoI
pCOR31	pHP159::*neo*$^{(3)}$	Gm^R G418^R	12.8	pHP159	pRS56	CCAAATCGATCTACGTATTAGTCATCGCTATT CCAAATCGATGATATCCGGATATAGTTCC ClaI
pCOR33	pLA24::*neo*	Gm^R G418^R	13.3	pLA24	pRS56	CCAAATCGATCTACGTATTAGTCATCGCTATT CCAAATCGATGATATCCGGATATAGTTCC ClaI

Table 3. Plasmids constructed for this work (continued).

						Construction
Plasmid	**Description**	**Phenotype**	**Size (kb)**	**Vector**	**Insert**	**Digestion / oligonucleotides**
pCOR35	pHP181::*neo*	Gm^R $G418^R$	10.0	pHP181	pRS56	CCAAATCGATCTACGTATTAGTCATCGCTATT CCAAATCGATGATATCCGGATATAGTTCC ClaI
pCOR39	pDEL10::IS10	Cm^R	4.8	-	-	Obtained by IS10 insertion

Table 3. Plasmids constructed for this work (continued).

3. Molecular biology techniques

3.1. Standard cloning procedure

General procedures for molecular cloning were performed as described in (Sambrook & Russell, 2001). Specific procedures for each plasmid constructed are detailed in **Table 3**.

DNA extraction and purification. Different kits were used depending on the starting material and the applications of the purified product, following manufactures's recommendations.

For plasmid DNA, GenElute Plasmid Miniprep Kit (Sigma Aldrich) was used for general isolations. QIAprep Spin Midiprep Kit (Qiagen) was used for isolation of large amounts of DNA (~100 µg). ATP Gel/PCR DNA Fragment Purification Kit (ATP Biotech Inc) was used for purification of PCR products and DNA purification from agarose gels.

Instagene Matrix (BioRad) was used to extract total DNA from *B. henselae* and *E. coli* colonies.

High Pure PCR Template Preparation Kit (Roche) was used for purification of genomic DNA from human cells.

The concentration of DNA samples was determined with a Nano-Drop Spectrophotometer ND-1000 (Thermo Scientific).

PCR amplification. For the amplification of DNA fragments subsequently used for cloning, Kapa HiFi DNA polymerase (Kapa Biosystems) was used. For colony analysis, Kapa Taq Polymerase (Kapa Biosystems) was used. PCR reactions were set up to final volumes of 50 µl. An iCycler (BioRad) thermocycler was used with the following program: 10 min of denaturation at 95ºC; 25 cycles of amplification, including steps of denaturalization for 30 seconds at 95ºC, annealing for 30 seconds at appropriate annealing temperature (depending on the primers) and elongation for 60 seconds/kb to be amplified at 72ºC; and a last step of 10 min of final elongation at 72ºC. After completion of the reaction, the samples were maintained at 4ºC for short-term or at -20ºC for long-term conservation.

Restriction enzyme digestion. Restriction enzymes were purchased from Thermo Scientific. Reactions were usually performed in 25 µl for a minimum of two hours, following the manufacturer's indications. Inactivation was carried out during 10 min at 65 or 80ºC, as indicated for each enzyme.

DNA dephosphorylation. Dephosphorylation of vector DNA was performed to increase the cloning efficiency by avoiding vector religation. After restriction digestion (and enzyme inactivation), 1 U of Shrimp Alkaline Phosphatase (Thermo Scientific) was added and samples were incubated for 1 h at 37ºC. Inactivation was carried out by incubation at 65ºC during 10 min.

DNA ligation. Insert fragments were obtained by either restriction digestion or PCR amplification with primers incorporating the adequate restriction sites for ligation into the same sites of the vector. A 5:1 (insert/vector) molar ratio was used. Ligation reaction was performed with 0.5-10 ng of DNA and 2.5 U of T4 DNA ligase (Thermo Scientific) in a final volume of 35 µl. It was incubated overnight at 16ºC. For each ligation, the same reaction without insert DNA was used as negative control, adding water to reach the final volume. Inactivation was carried out by incubation at 65ºC during 10 min.

DNA sequencing. The DNA sequence of all cloned PCR fragments was determined. Samples were sent to Macrogen Inc. DNA Sequencing Service (Amsterdam, The Netherlands).

3.2. Electrophoresis

DNA electrophoresis

DNA was analysed by agarose gel eletrophoresis. Agarose was dissolved in TBE (Tri-HCl 45 mM, boric acid 45 mM, EDTA 0.5 mM, pH 8.2) to a final concentration of 1-2 % (w/v), depending on the size of the DNA fragments to be resolved. RedSafe (Intron Biotechnology) was used as staining solution at a final concentration of 1/50,000. Loading buffer [bromophenol blue 0.25 % (w/v), sucrose 40 % (w/v) in TBE] was added to DNA samples in a 1:6 ratio of blue dye:DNA sample. HyperLadder I (Bioline) was used as a molecular weight marker. A horizontal BioRad electrophoretic system was used (with constant voltage between 80-120 V). Agarose DNA gels were visualised with a Gel Doc 2000 UV system, and images were analysed with Quantity One software (BioRad).

Protein electrophoresis

SDS-polyacrylamide gels (acrylamide:bisacrylamide 29:1) were used for the visualisation of proteins (Sambrook & Russell, 2001). The concentration of polyacrylamide used for detection of TrwC was 10-12 %. Electrophoresis was carried out

using a Mini-PROTEAN II system (BioRad) in 6.1 cm x 0.75 mm gels. Samples were mixed with 5X loading buffer [Tris-HCl 250 mM pH 6.8, SDS 5 % (w/v), glycerol 50 % (v/v), bromophenol blue 0.05 % (w/v), DTT 250 mM] to a 1X final concentration. Samples were then incubated at 95ºC for 5 min before loading the gel. Precision Plus Protein Dual Color Standards (BioRad) was used as molecular weight marker. The electrophoretic run was performed at 200 V for 1 h in TGS buffer [Tris 25 mM, glycine 192 mM, SDS 0.1 % (w/v)]. After the run, gels were stained by incubation in staining solution [Coomassie blue R250 0.1 % (w/v), methanol 40% (v/v), glacial acetic acid 10% (v/v)] for 15 min at room temperature. Destaining was performed by incubation in destain solution [methanol 40 % (v/v), glacial acetic acid 10 % (v/v)] at room temperature. The images of the gels were taken using a ChemiDoc XRS System (BioRad).

3.3. Western blot

Total protein extracts were analysed as described in (Towbin *et al.*, 1979). *E. coli* DH5α cells containing the indicated plasmids were grown overnight. 1 ml of culture was then collected, centrifuged, and resuspended in 1/10 volume of 2X SDS-gel loading buffer [Tris HCl 250 mM pH 6.8, SDS 5 % (w/v), glycerol 50 % (w/v), bromophenol blue 0.05 % (w/v), DTT 250 mM] Samples were stored at -20ºC for at least one night.

Samples were boiled for 5 min and loaded in SDS-PAGE gels. After the run, samples were transferred to nitrocellulose filters in TGM buffer [Tris 20 mM, glycine 15 mM, methanol 20 % (v/v), pH 9.2] during 1 h at 180 mA and 4ºC. To confirm the transfer, gels were stained with staining solution as described in section 3.2 (protein electrophoresis). After the transfer, the filters were washed with TBST buffer [Tris HCl 50mM, NaCl 150mM, Tween-20 0.05 % (v/v), pH 7.5] during 5 min at room temperature. The filters were then incubated in blocking buffer [nonfat dry milk 3 % (w/v) in TBST] for 1 h at 37ºC. Then, filters were washed in TBST during 5 min at room temperature. Incubation with primary antibody was carried out during 1 h at room temperature. Anti-TrwC (Grandoso *et al.*, 1994) primary antibody was diluted 1:10,000 in blocking buffer [nonfat dry milk 1.5 % (w/v) in TBST]. After the incubation, filters were washed 3 times with TBST during 15 min at room temperature. Secondary antibody (IRDye 800CW anti-rabbit IgG, Li-Cor) was diluted as the anti-TrwC. Incubation was performed during 45 min at room temperature in the dark. Detection was performed with an Odyssey CLx Dual-Mode Imaging System (Li-Cor).

3.4. Random mutagenesis

Mutations in *trwC* were introduced by mutagenic PCR using the GeneMorph II Random Mutagenesis Kit (Agilent Technologies) following the manufacturer´s recommendations. We adjusted the reaction to obtain 1-4.5 mutations per kb. Template DNA was plasmid pMTX712 (Table 2). Oligonucleotides used were universal primers T7 (5'-TAATACGACTCACTATAGGG-3'), annealing in vector pET29c towards the 5´end of *trwA*, and T7 terminator (5'- GCTAGTTATTGCTCAGCGG -3'), annealing in pET29c towards the 3´end of *trwC*. Full-length *trwC* from the amplified PCR product was obtained by digestion with enzymes NdeI + BamHI and cloned into the corresponding sites of vector pET3a. Ligation containing the pool of mutants was then introduced in *E. coli* DH5α containing substrate plasmid pCIG1028 with two *oriT* copies. Transformations were incubated in liquid LB medium for 1 h at 37ºC, and then plated on antibiotics plus X-Gal containing plates to screen for blue colonies.

3.5. Characterization of integration sites by LAM PCR

Identification of genomic integration sites by LAM-PCR was performed as described in (Gabriel *et al.*, 2014). An outline of the protocol is shown in **Figure 36**.

a) Primers

HPLC-purified primers were ordered from Eurofins Genomics. They were designed using Primer3Plus online software selecting a primer length of 18-25 nt and a GC content between 40-60 %. Primers with more than three identical nucleotides at the 3' end were discarded. Primer Blast was used to confirm that primers do not align to host genome. All the primers are listed in **Table 4**.

The primers used to obtain the linker cassette (LC1 and LC2) contain the proper restriction enzymes overhang (RO) and a barcode sequence (N). The linker cassette used for each sample has a different barcode sequence, to label and allow the parallel sequencing of different samples.

Plasmid-specific primer for linear PCR (oriTI) was designed to align around 120 nt 5' to the *nic* site and is 5´biotinylated. Two additional plasmid-specific primers for first (oriTII; 5' biotinylated) and second (Mis-TrwC; unmodified) exponential PCRs were designed to align between the linear primer and the plasmid end. The plasmid-specific primer for second exponential PCR should hybridize at least 20 bp far from the plasmid end, leaving enough nucleotides to recognize true integration sites. For first and second exponential PCRs, two linker-specific primers were also used, LCI and Mis-LC,

respectively. Primers for second exponential PCR contain adaptor sequences (PE-PCR 2/1.0) for the specific sequencing technology used and plasmid-specific primer (Mis-TrwC) also contains an additional barcode sequence (N). A different Mis-TrwC primer (with a different barcode sequence) was used for amplification of each sample.

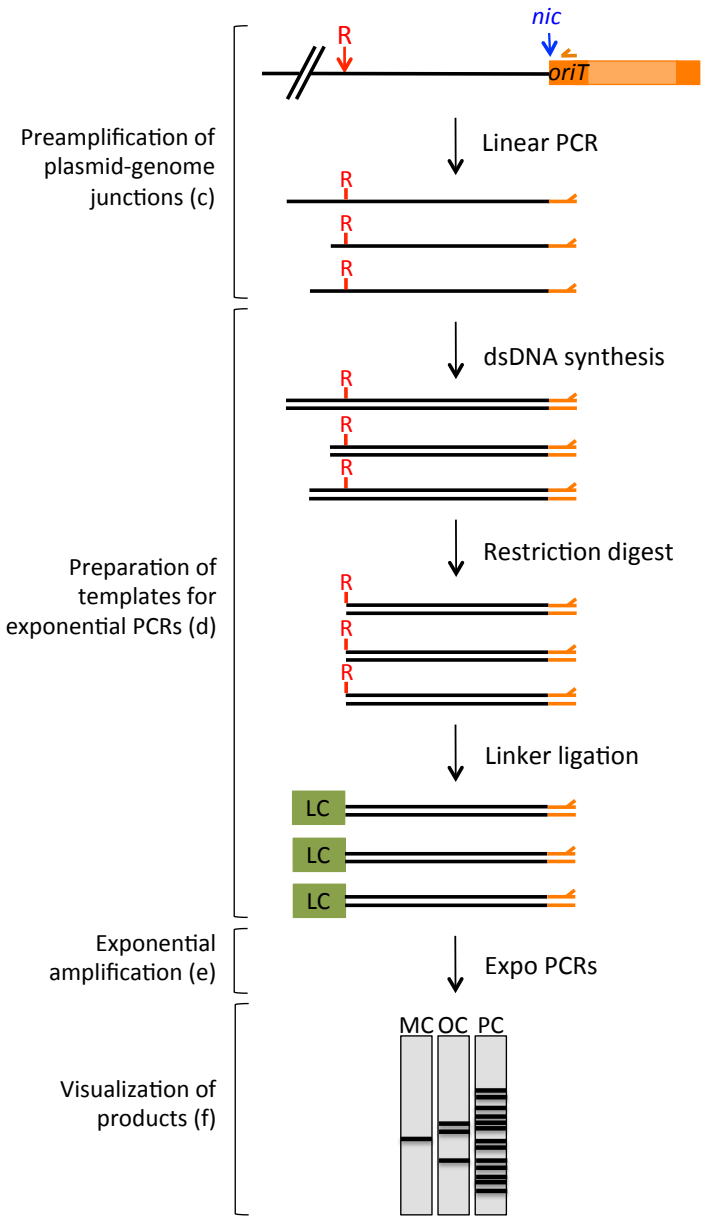

Figure 36. LAM-PCR. After linear amplification, dsDNA synthesis, digestion, and ligation were performed. The product is then amplified exponentially in two consecutive PCRs. After it, the samples were run on a 2 % agarose gel. The different steps are indicated as in the main text. Depending on the number of integration sites: MC, monoclonal sample; OC, oligoclonal sample; PC, polyclonal sample. *nic*, *oriT* nicking site. R, restriction site. LC, linker cassette. Modified from (Gabriel *et al.*, 2009).

Primer	Sequence	Specificity	Purpose
LC1	GACCCGGGAGATCTGAATTCAGTGGCAC AGCAGTTAGG(N)$_{12}$CTA(RO)	LC	LC
LCI	GACCCGGGAGATCTGAATTC		Expo. PCR I
LC2	(RO)TAG(N)$_{12}$CCTAACTGCTGTGCCACTG AATTCAGATC	LC	LC
Mis-LC	(PE-PCR 2.0)AGTGGCACAGCAGTTAGG	LC	Expo. PCR II
Mis-TrwC	(PE-PCR 1.0)(N)$_{10}$CGTCCTTAAAAGCCGG GTTG	Plasmid	Expo. PCR II
oriTI	CGATAACCCAATGCGCATAG	Plasmid	Preamplification
oriTII	TCTTTAGGGTCACGCTGGC	Plasmid	Expo. PCR I

Table 4. Oligonucleotides used for LAM-PCR. oriTI was used for linear amplification of plasmid-genome junctions. (N), barcode sequence of linker cassette. (RO), restriction enzyme overhang. (PE-PCR 2/1.0), adaptor sequences for high-throughput sequencing [see step h) in the main text for further details]. (N), barcode sequence introduced in second exponential PCR. oriTI and oriTII are 5'biotinylated. LC, linker cassette. Expo, exponential.

b) Preparation of linker cassette

40 µl of primer LC1 100 µM were mixed with 40 µl of primer LC2 µM (with appropriate restriction enzyme overhangs), 110 µl Tris-HCl (pH 7.5), and 10 µl $MgCl_2$ 250 mM. Mixture was incubated at 95ºC for 5 min and let to cool down slowly to room temperature. Then, 300 µl of H_2O were added and dsDNA was concentrated on a centrifugation filter. After that, 80 µl of H_2O were added to the eluate. Aliquots of 10 µl were prepared and stored at -20ºC.

As LAM-PCR protocol was carried out with both BfaI (C^TAG) and Tsp509I (^AATT) restriction enzymes in parallel, linker cassettes with the corresponding restriction overhangs were prepared for them to be ready to be ligated to the digested PCR product.

c) Preamplification of plasmid-genome junctions

Taq DNA polymerase (Genaxxon Bioscience) was used for all PCRs. A 50 µl PCR reaction was prepared for each sample. 200 ng of genomic DNA (in X µl) were used as template. DNA volume (X µl) was the same in all tubes and should be between 0.5 and 25 µl. Components of the reaction and PCR program are indicated in **Table 5**. After completion of PCR, 0.5 µl Taq Polymerase was added to each PCR tube and the whole PCR program was run again.

Reagent	Vol. (µl)	Concentration		Step	Temp. (ºC)	Time
H_20	43-X			Denaturation	95	5 min
Buffer	5	10X		Denaturation	95	45 sec
dNTPs	1	10 mM	X50	Annealing	58	45 sec
oriTI	0.5	0.17 µM		Elongation	72	60 sec
Taq Pol.	0.5	2.5 U/µl		Elongation	72	5 min

Table 5. Preamplification of plasmid-genome junctions. Reagents and concentrations are indicated on the left side of the table, while PCR program is shown on the right. The central steps of denaturation, annealing, and elongation were repeated 50 times. After one round of amplification, 0.5 µl of Taq were added and the whole program was repeated. Vol, volume. Temp, temperature. Taq pol, taq DNA polymerase.

d) Preparation of templates for exponential PCRs

Preparation of magnetic beads. 20 µl of the magnetic beads solution (containing 200 ng of magnetic beads) (Dynabeads M-280 streptavidin, Invitrogen) were exposed to a magnetic particle separator (MPS; MagnaSep magnetic particle separator, Dynal) at room temperature for 60 sec and supernatant was discarded. Beads were washed 2 times with 40 µl PBS + BSA 0.1 % (by resuspending, exposing to MPS, and discarding supernatant). An additional wash was performed with 20 µl LiCl 3M and the beads were then resuspended in 50 µl LiCl 6M. Both LiCl solutions must be prepared in advance by adding Tris-HCl (pH 7.5) and EDTA to a final concentration of 10 mM and 1 mM, respectively. LiCl solutions are used instead of the binding solution offered by the manufacturer of the magnetic beads.

Magnetic capture of PCR product. 50 µl of PCR product with 50 µl of bead solution were mixed and the mixture was incubated at room temperature on a horizontal shaker at 300 r.p.m. for 2-48 hours. After that time, DNA-beads complexes were exposed to MPS, supernatants were removed and complexes were washed 2 times with 100 µl of H_2O.

dsDNA synthesis. DNA-beads complexes were exposed to MPS, supernatants were removed, and each complex was resuspended in 8.5 µl of H_20, 1 µl of 10X Hexanucleotide Mix buffer (Roche), 0.25 µl dNTPs, and 0.5 µl of Klenow polymerase (Roche). Sample was incubated 1 h at 37ºC. After that time, 2 washes with 100 µl of H_2O each were done.

Restriction digest. dsDNA-beads complexes were exposed to MPS and supernatants were removed. 8.5 µl of H_2O, 1 µl 10X restriction enzyme buffer, and 0.5 µl restriction enzyme were added to each complex. DNA-beads complexes were resuspended and incubated for 1 h.

Liker ligation. FastLink DNA Ligation Kit (Epicentre) was used after digest. dsDNA-beads complexes were exposed to MPS and supernatants were removed. Each complex was resuspended in 5 µl of H_2O, 1 µl 10X buffer, 1 µl ATP (10mM), 2 µl linker cassette, and 1 µl DNA ligase. Reaction was performed at room temperature for 5 min.

Denaturation of dsDNA. dsDNA-beads complexes were then exposed to MPS and supernatants were removed. Each complex was resuspended in 5 µl of NaOH 0.1 M and incubated 10 min at room temperature on a horizontal shaker at 300 r.p.m. Freshly prepared NaOH was always used. After that time, complexes were exposed to MPS and DNA containing supernatants were collected and used for next PCR or stored at -20ºC.

e) Exponential amplification

First exponential amplification. 2 µl of denatured dsDNA were used as template. PCR mastermix and program used are shown in **Table 6**.

Magnetic capture of PCR products. Magnetic beads were prepared as described previously but resuspended in 25 µl LiCl 6 M. 25 µl of PCR product were mixed with 25 µl of beads and incubation was carried out at room temperature on a horizontal shaker (300 r.p.m.) for 2 h-overnight. After that time, DNA-beads complexes were exposed to MPS, supernatants were removed and complexes were washed 2 times with 100 µl of H_2O.

Reagent	Vol. (µl)	Concentration
H_20	19.25	
Buffer	2.5	10X
dNTPs	0.5	10 mM
oriTII	0.25	16.7 µM
LCI	0.25	16.7 µM
Taq Pol.	0.25	2.5 U/µl
DNA	2	

	Step	Temp. (ºC)	Time
	Denaturation	95	5 min
X35	Denaturation	95	45 sec
	Annealing	58	45 sec
	Elongation	72	60 sec
	Elongation	72	5 min

Table 6. First exponential amplification. Reagents and concentrations are indicated on the left side of the table, while PCR program is shown on the right. The central steps of denaturation, annealing, and elongation were repeated 35 times. Abbreviations as in Table 5.

Denaturation of dsDNA. DNA-beads complexes were exposed to MPS and supernatants were removed. Each complex was resuspended in 10 µl of NaOH 0.1 M and incubated as before. After that, complexes were exposed to MPS and DNA containing supernatants were collected and used for next PCR or stored at -20ºC.

Second exponential amplification (fusion-primer PCR). 2 µl of denatured dsDNA were used as template. In this second exponential PCR, sequencing specific adaptors were added to each LAM product. PCR mastermix was prepared and program was run as exemplified in **Table 7**.

f) Visualization of second exponential PCR products

8 µl of PCR product were visualized on a 2 % agarose gel. 100 bp ladder (Invitrogen) was used as molecular marker.

g) Purification of LAM-PCR products

40 µl of second exponential PCR reaction were mixed with 26 µl of room temperature AMPure XP Magnetic Beads (Life Technologies), incubated 5 min at room temperature and exposed to MPS for additional 2 min. Supernatant was discarded and

two washes with 200 µl EtOH 70 % were performed on the MPS. Supernantants was discarded again and DNA-beads complex was resuspended in 30 µl of water, incubated 1 min at room temperature, and supernatant was transferred to a fresh 0.2 ml tube. The concentration of the purified DNA was determined using a Qubit Fluorometer 2.0 (Life Technologies).

Reagent	Vol. (µl)	Concentration
H_20	40.5	
Buffer	5	10X
dNTPs	1	10 mM
Mis-TrwC	0.5	5 µM
Mis-LC	0.5	10 µM
Taq Pol.	0.5	2.5 U/µl
DNA	2	

	Step	Temp. (ºC)	Time
	Denaturation	95	5 min
X35	Denaturation	95	45 sec
	Annealing	58	45 sec
	Elongation	72	60 sec
	Elongation	72	5 min

Table 7. Second exponential amplification. Reagents and concentrations are indicated on the left side of the table, while PCR program is shown on the right. As for the first exponential PCR, the central steps of denaturation, annealing, and elongation were repeated 35 times. Abbreviations as in Table 5.

h) High-throughput sequencing

LAM samples were sequenced using MiSeq Benchtop next generation sequencing technology (Illumina). Once they are purified, the LAM-PCR products are ready for the sequencing reaction. The appropriate volumes of different purified samples were mixed together following the manufacturer´s recommendations. Primers used in the second exponential amplification already added the adaptor sequences needed for the sequencing reaction. These adaptor sequences are:

PE-PCR 1.0:

5´-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3´.

PE-PCR 2.0:

5´-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3´.

LAM-PCR products were sequenced in both directions. From PE-PCR 1.0 (adaptor present in primer Mis-TrwC) 400 nt were sequenced while only 50 nt were sequenced from PE-PCR 2.0 (adaptor present in primer Mis-LC). Information from PE-PCR 1.0 was used for sorting the sequences to the different samples and integration site detection, while information obtained from PE-PCR 2.0 was used only for sorting.

i) Bioinformatic analysis

It was performed as described in (Paruzynski *et al.*, 2010) for nrLAM-PCR. Briefly, sequences were trimmed by identification and removal of plasmid- and linker-specific sequence. Genomic sequences were aligned to the human genome using stand-alone BLAT (UCSC). Sequences with identities lower than 95 % were discarded. For each remaining sequence, the chromosome, the integration site, the nearest RefSeq protein-coding gene and other genomic features were recorded.

4. Microbiological techniques

4.1. Growth conditions and selection media

***Escherichia coli*.** Luria-Bertani broth (10 g tryptone, 5 g yeast extract, 5 g NaCl/ litre; Pronadisa) was used for bacterial growth, supplemented with agar 1.5 % (w/v) for solid culture. Cultures were incubated at 37ºC.

Selective media included antibiotics (Sigma Aldrich or Apollo Scientific) at the following concentrations: ampicillin (Ap), 100 µg/ml; chloramphenicol (Cm), 25 µg/ml; erythromycin (Em), 200 µg/ml; kanamycin monosulphate (Km), 50 µg/ml; nalidixic acid (Nx), 20 µg/ml; rifampicin (Rf), 100 µg/ml; streptomycin (Sm), 300 µg/ml; spectinomycin (Sp), 10 µg/ml; gentamicin sulphate (Gm), 10 µg/ml; trimethoprim (Tp), 20 µg/ml. When needed, media was supplemented with thymidine (Thy, Sigma Aldrich) at a final concentration of 0.3 mM. X-Gal (Apollo Scientific) was supplied at a concentration of 60 µg/ml, IPTG was used at a concentration of 0.5 mM, and DAP was added at a final concentration of 0.3 mM.

To preserve *E. coli* strains, a stationary phase culture was centrifuged and resuspended in peptone-glycerol [peptone 0.75 % (w/v), glycerol 50% (v/v)]. Duplicates of all the strains were kept at -20 and -80ºC.

***Bartonella henselae*.** *B. henselae* was grown in Columbia blood agar (CBA) plates with 5 % sheep blood (Difco) for 3-4 days at 37ºC in a 5 % CO_2 atmosphere. Antibiotics were added as follows: gentamicin sulphate (Gm), 10 µg/ml; streptomycin (Sm), 100 µg/ml; kanamycin monosulphate (Km), 50 µg/ml.

To preserve *B. henselae* strains, bacteria from a grown Columbia agar plate were recovered with a cotton swab and resuspended in LB-glycerol [LB 50 % (v/v), glycerol 50 % (v/v)]. Each strain was store at -20 and -80ºC.

***Saccharomyces cerevisiae*.** *S. cerevisiae* was grown in YP base medium (yeast extract 10 g/l, tryptone 20 g/l, Panreac AppliChem) or YP agar base medium (yeast extract 10 g/l, tryptone 20 g/l, bacteriological agar 15 g/l, Panreac AppliChem) when no selection was needed. Both media were supplemented with glucose (Panreac AppliChem) 20 % (w/v) for a final concentration of 2 %. For plasmid selection in histidine or tryptophan auxotrophic mutants synthetic defined (SD) dropout medium [yeast nitrogen base without amino acids 6.7 g/l (Sigma Aldrich), yeast synthetic dropout media supplement without histidine/tryptophan 1.92 g/l (Sigma Aldrich), glucose 40ml/l] was used. When solid medium was used, bacteriological agar 20 g/l (Pronadisa) was added. Yeast strains were incubated at 30ºC.

To preserve yeast strains, 500 µl of a saturated culture were mixed with 500 µl of glycerol 50 %. Yeast stocks were stored at -80ºC.

4.2. Transformation

E. coli transformation

Depending on the efficiency required, two different transformation methods were used. Chemically induced transformation was used for low efficiency transformation and electroporation was used when high efficiency of transformation was necessary.

Chemical transformation. Transformations were performed following the method described by (Chung & Miller, 1988). Bacterial cells were grown to early-mid log phase (OD_{600}= 0.3-0.8), and pelleted by centrifugation. Cells were resuspended in 1/10th volume of TSB [LB broth pH 6.1 containing PEG 10 % (w/v), DMSO 5 % (v/v), MgCl2 10 mM, and MgSO4 10 mM] at 4ºC, and incubated on ice for 10 min. 100 µl of cells were mixed with 1-3 µl (containing 50 ng) of plasmid DNA and incubated for another 30 min on ice. After that time, 900 µl of TSB plus glucose 20 mM were added

and cells were incubated 1 h at 37ºC to permit the expression of the antibiotic resistance gene. After incubation cells were plated on antibiotic containing media.

Electroporation. Transformation efficiency is a linear function of DNA and cell concentrations (Dower *et al.*, 1988). For preparation of electrocompetent cells, bacteria were grown to OD_{600} = 0.4-0.5, and pelleted by centrifugation at 4,000 r.p.m. and 4ºC. Cell were washed (by water addition, centrifugation, and supernatant removal steps) with 1 volume ice-cold milliQ water (Millipore Corporation), and an additional time with ½ volume. Cells were then washed with 1/50 volume ice-cold glycerol 10 % and resuspended in 1/400 volume glycerol 10 % and aliquoted in 50 µl samples. Aliquots were kept at -80ºC until usage. Aliquots were mixed with <10 ng of DNA in a 0.2 cm Gene Pulser cuvette (BioRad) and subjected to an electric pulse (2.5 kV/cm, capacitance 25 µF and 200 Ω) in a MicroPulser TM (BioRad). 1 ml LB was added to the electroporated cells, which were incubated at 37ºC to allow antibiotic-resistance gene expression. After incubation, cells were plated on antibiotic containing media.

B. henselae transformation

B. henselae strains were transformed by electroporation using a protocol based in the one described in (Grasseschi & Minnick, 1994). For preparing competent cells, the content of a 3-day-old CBA plate was harvested with a cotton swab into 950 µl ice-cold PBS. Cells were centrifuged at 4,000 r.p.m. for 5 min at 4ºC and the pellet was resuspended in 950 µl of ice-cold glycerol 10 %. Cells were centrifuged again in the same conditions. Two more washing steps (of resuspension, centrifugation, and pellet removal) were performed. After the final wash, pellet was resuspended in 100 µl of ice-cold glycerol 10 % (competent cells). For each electroporation, 40 µl of competent cells were transferred to a new tube and 3 µl of DNA 300 ng/µl were added. Mixture was incubated on ice for 15 min and pipetted into a cooled BioRad 0.2 cm Gene Pulser cuvette and subjected to an electric pulse (2.5 kV/cm, capacitance 25 µF and 200 Ω) in a MicroPulser TM. Directly after electroporation, 1 ml of SB broth [RPMI 1641 + L-glutamine 74.8 % (v/v) (Lonza), Hepes 42mM (Sigma Aldrich), sodium pyruvate 11 mM (Sigma Aldrich), FBS 5 % (v/v), defibrinated sheep blood 5 % (v/v) (Oxoid)] was added at room temperature. Mixture was incubated for 3.5 h at 37ºC under 5 % CO_2 with slow shaking. After that time, cells were centrifuged at 4,000 r.p.m. for 4 min and pellet was resuspended in 40 µl SB broth and plated on a CBA plate with appropriate antibiotics.

S. cerevisiae transformation

Chemical transformation. Yeast strains were transformed by the LiAc-PEG method (Gietz & Woods, 2001). For preparation of competent cells, yeast were grown to OD_{600} = 0.8. Two series of washes and centrifugations with 1 volume of milliQ water at room temperature were performed. Cells were resuspended in 1/4 volume of LiSORB (Lithium acetate 100 mM, sorbitol 1 M in TE) and centrifuged again. After removing supernatant, 2 ml of LiSORB supplemented with DMSO 5 % were added to the pellet. Aliquots of 50 µl were prepared and kept at -80ºC until usage.

For transformation, 2 µl of DNA and 2.5 µl of carrier (salmon sperm ssDNA 10 mg/ml, Sigma Aldrich) were added to a competent yeast aliquot. 300 µl of LiPEG (LiAc 100 mM, PEG_{3350} 40 % in TE) were added. The mixture was inverted to mix and incubated at 30ºC for 30 minutes without shaking, followed by incubation at 42ºC for 10 minutes without shaking. The cells were then spined to get a pellet that was resuspended in 100 µl of TE 1X and plated on selective medium.

Electroporation. Electrocompetent cells were prepared as described in (Becker & Guarente, 1991). Yeast culture was grown to OD_{600} (1:10 dilution)= 0.3. It was centrifuged at 4,000 g for 5 min at 4ºC. The cell pellet was kept on ice and resuspended in 1/6 volume of cold milliQ water. 1/50 volume of TE 10X buffer and LiAc 1M were added and cells were incubated 45 min at 30ºC shaking at 85 r.p.m. After that time 1/200 volume of DTT 1M was added and cells were incubated 15 min in the same conditions. The volume was brought to 50 ml with water, and cells were pelleted at 4,000 r.p.m. for 5 min at 4ºC. This washing step was repeated once with water and then with ice-cold sorbitol 1M. Cells were then resuspended in 500 µl sorbitol 1M and aliquots of 40 µl were prepared. 5 µl (5-100 ng) of DNA were added to an aliquot and the mixture was transferred to a 0.2 cm cuvette and subjected to an electric pulse (1.5 kV, capacitance 25 µF and 200 Ω) in a Gene Pulser Xcell main unit (BioRad). 1 ml of sorbitol 1M was added. Cells were incubated for 3 h at 30ºC and then plated on selective plates.

4.3. Generation of yeast *petite* mutants

To obtain Y13652 derivatives with defects in the respiratory chain, Y13652 was streaked in a YPD (YP supplemented with glucose) agar plate containing 150 µg/ml G418 and incubated at 28ºC for 1.5 to 2 days. Grown yeast colonies were then picked up with a toothpick, resuspended in 100 µl of YPD and diluted to 10^4 cells/ml. 100 µl of the dilution were spread on another YPD plate and incubated at 28ºC for 1.5 to 2 days. After that time, colonies that were smaller comparing with the size of surrounding colonies

were streaked on both YPD and YPG (YP medium supplemented with glycerol 2 %). The colonies that were able to grow on YPD but not on YPG were selected.

4.4. Conjugation

All matings were performed on solid media by mixing donor and recipient strains. After washing both strains and mixing them, the mix was centrifuged and transferred to a conjugation filter (0.2 µm cellulose acetate filter, Sartorius) on a pre-warmed plate. After incubation, appropriate dilutions were plated on selective media for both donors and transconjugants. The frequency of conjugation is expressed as the number of tranconjugants per donor cells. The growth of donor and recipient strains, the incubation time, and the selection conditions depend on the species used in the mating and are explained below.

From *E. coli* to *E. coli*

Standard mating assays were performed as described in (Grandoso *et al.*, 2000). DH5α and D1210 were generally used as donor or recipient strains. Donor and recipient cells were grown to stationary phase. 100 µl of each culture were used. They were centrifuged and resuspended in fresh LB in order to remove antibiotics from the media. Then donor and recipient strains were mixed, centrifuged, resuspended in 20 µl of LB, and then transferred to the cellulose acetate filter. Mating plate was incubated 1 h at 37ºC. After incubation, the filter was introduced into 2 ml LB, and vortexed for complete cell resuspension. Different dilutions were then plated.

From *E. coli* to *B. henselae*

β2163 donor strain was grown to stationary phase. 200 µl were mixed with 800 µl of PBS. Recipient *B. henselae* was grown in CBA plates for 3-4 days. After that time, bacteria from half of the plate were collected with a cotton swab and resuspended in 1 ml of PBS. Both donor and recipient aliquots were centrifuged and the supernatants discarded. Donor cells were resuspended in 20 µl of PBS while recipient cells were resuspended in 80 µl, and they were both mixed together. The mixture was placed in a cellulose acetate filter on a CBA plate supplemented with DAP. The mating plate was incubated at 37ºC in a 5 % CO_2 atmosphere during 6 h.

As conjugation was performed to transform *B. henselae* with a specific plasmid and not to calculate conjugation frequencies, transconjugants were selected by recovering the mating mixture with a loop and streaking it on a CBA plate with appropriate antibiotics. The plate was incubated for 6-9 days at 37ºC in a 5 % CO_2 atmosphere.

From *E. coli* to *S. cerevisiae* (TKC reaction)

Conjugation protocol to Y13652 and the different *petite* mutants derivatives was adapted from (Moriguchi *et al.*, 2013). Bacterial donor strain was grown overnight, centrifuged and resuspended in TNB (Tris-HCl 40 mM pH 7.5, NaCl 0.5 %) at a concentration of $1.5x10^8$ cfu/ml. Yeast recipient strain was streaked in YPD medium 48 h before conjugating and incubated at 28ºC. The day of the conjugation, one colony (around $1x10^6$ cfu) was picked up with a toothpick and resuspended in 25 µl of donors-containing TNB solution to obtain a donor:recipient ratio of approximately 3.8:1. The mixed suspension was then placed on a cellulose acetate filter in a TNB agar plate and incubated 1 h at 28ºC. After that time, the filter was transferred to 1 ml of TNB to resuspend the cells and the suspension was plated in YPD and selective medium (both supplemented with Cm to kill donor bacteria) to select for recipients and transconjugants, respectively. TKC efficiency was expressed as the number of colonies on selection plate divided by the number of colonies on YPD plate.

4.5. Site-specific recombination assay

Relaxase-mediated site-specific recombination activity was assayed by the loss of a DNA segment between two *oriT* copies, leading to the expression of *lacZ*α gene, as previously described (César *et al.*, 2006) (**Figure 37**).

Upon recombination and subsequent loss of the intervening DNA, a lactose promoter localised to one side of the intervening cassette comes into close proximity with the lacZα gene, located on the other side. Thus, in the appropriate genetic background and in a X-gal containing medium, recombination can be monitored by the appearance of blue bacterial colonies.

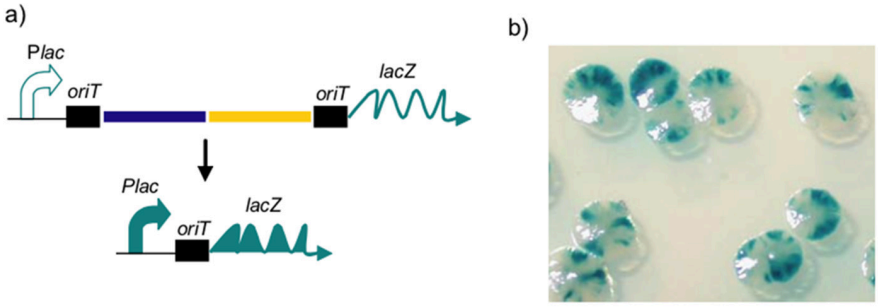

Figure 37. Site-specific recombination assay. a) Structure of the recombination cassette and the resulting product after recombination between two *oriT* copies. Genes between both *oriTs*: blue, *lacIq*; yellow, *ntpII*. On the left, *oriT1*; on the right, *oriT2*. *Plac*, lactose promoter. **b)** DH5α colonies harbouring substrate plasmid depicted in a) and transformed with a helper plasmid expressing *trwAC* plated on X-Gal plates. Each blue sector is originated from a single recombination event. Modified from (César *et al.*, 2006).

Strain DH5α (*lacZΔM15*) was used as host for β-galactosidase complementation. Intramolecular recombination was tested by transformation of a helper plasmid (coding for the relaxase and other Tra proteins) into DH5α strain containing the substrate plasmid (with the two *oriT* copies repeated *in tandem*). 100 µl of the transformation (T0) were diluted 1/50 in LB broth with antibiotics and grown overnight (T20). A 1/5000 dilution of T20 was grown overnight again in selective media (T40) and, finally, a 1/5000 dilution of T40 was grown overnight again in selective media (T60). Appropriate dilutions of T0, T20, T40, and T60 were plated on selective media containing X-Gal. Colonies showing at least a blue sector were considered recombinants.

For testing pKM101_TraI, recombination was measured by the loss of Km resistance. For each generation, 100 colonies were replica-plated in +/- Km media.

Recombination was then confirmed by restriction analysis of plasmid DNA extracted from the different generations analysed. Recombination frequency is expressed as the number of blue (or Km-sensitive) colonies divided by the total number of colonies x 100.

4.6. Site-specific integration assay

Site-specific integration activity was measured using the assay described by (Agundez *et al.*, 2012) that is depicted in **Figure 38**.

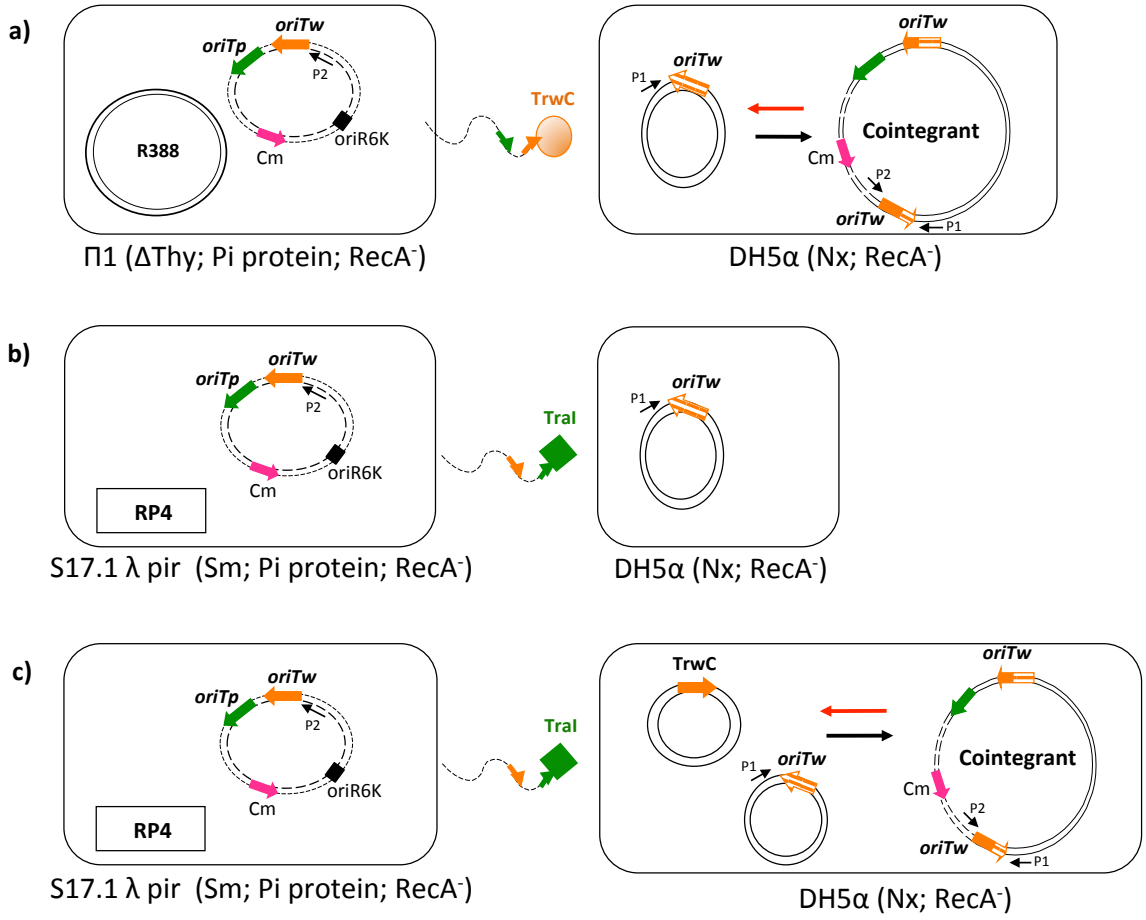

Figure 38. TrwC-mediated site-specific integration assay. a) The suicide plasmid is mobilized by TrwC (orange sphere) from a Pi^+ strain into a Pi^- recipient harbouring an *oriT*$_W$-containing plasmid where site-specific integration takes place. For testing F_TraI integrase activity, the donor strain contained F plasmid instead R388, and the suicide plasmid had an *oriT*$_F$ instead of the *oriT*$_W$. **b)** Negative control of the integration assay, in which the suicide plasmid is mobilized by RP4_TraI. **c)** Integration assay expressing the relaxase in the recipient cell. In this case, the suicide plasmid is mobilized by RP4_TraI (green square) to a recipient strain hosting the *oriT*-containing plasmid and another one coding for the relaxase.

Matings were carried out using Π1 as donor strain. This donor strain contained pCMS11, a plasmid harbouring the cognate *oriT* of the relaxase under study (*oriT*$_W$ for R388_TrwC; *oriT*$_F$ for F_TraI), a RP4 *oriT* (*oriT*$_P$) and a R6K replicon (that only replicates in strains expressing *pir*), used as a suicide plasmid for mobilization into strains lacking the *pir*-encoded protein (Pi protein). The suicide plasmid was mobilized from Π1 by a helper plasmid coding for the appropriate transfer region except for the *oriT*. DH5α harbouring a plasmid carrying the same cognate *oriT* (or the *oriT*$_P$ as a negative control) was used as recipient. The recipient strain does not code for Pi protein, so that the

suicide plasmid cannot be maintained. However, if the relaxase pilots the suicide plasmid to the recipient cell and is able to integrate it site-specifically into the *oriT*-containing recipient plasmid, the non-replicative plasmid is kept as a cointegrate with the host plasmid (Figure 38a). As a negative control, the suicide plasmid was mobilized from S17.1 λpir strain by RP4_TraI to the same recipient as in the test (Figure 38b).

To study the integrase activity of TrwC*, the integration assay was carried out expressing the relaxase from the recipient cell as described (Agundez *et al.*, 2012) (Figure 38c). In this case, the suicide plasmid was mobilized by RP4_TraI and the recipient cell contained two plasmids: one with the recipient *oriT* and another one producing the relaxase.

To study chromosomal integration mediated by TrwC, we used CMS1 and CMS2 strains (Agundez *et al.*, 2012) as recipients. They contain a copy of the *oriT* in place of the chromosomal *lacZ* gene, with the *nic* site lying in the lagging and leading strands, respectively. The isogenic strain without *oriT*, HMS174, was used as negative control.

For each mating, 100 μl of both donor and recipient strains were washed with 900 μl of fresh LB. They were mixed and the mixture was centrifuged again and resuspended in 20 μl of LB, that were transfer to a cellulose acetate filter on a LB plate supplemented with Thy when Π1 was used. The plate was incubated at 30ºC for 4 hours. Integrants were detected by selecting in Cm, as it is the resistance provided by the suicide plasmid upon integration. Integration events were also analysed at the molecular level. DNA was checked by PCR with primer P1 (5´-AGCGGATAACAATTTCACACAGGA-3´), annealing to the recipient plasmid, and P2 (5´-GCAGGATCCGCTAAGCTTTGTCGGTCATTTCGA-3´), annealing to the suicide plasmid (Figure 38). An amplicon is only expected for the cointegrate molecules. DNA was also analysed by restriction digestion. After confirming the number of positive integrants, the frequency is reported as the number of integrants per donor cell.

5. Cellular biology techniques

5.1. Cell culture

The cell lines used in this work are listed in **Table 8**.

Cell line	Description	ATCC number
EA.hy926	Fusion of HUVEC cells and adenocarcinomic human alveolar basal epitelial A549 cells	CRL-2922
HeLa	Epithelial human cells of cervix adenocarcinoma	CCL-2

Table 8. Human cell lines used in this work.

All cell lines were routinely grown in DMEM medium (glucose 4.5 g/l, with ultraglutamine, Lonza) supplemented with FBS 10 % (Lonza) at 37ºC under 5 % CO_2. To prevent bacterial contamination penicillin-streptomycin 1 % (Lonza) was added to the medium. When needed, antibiotics were use at the following concentrations: G418 disulfate salt (Sigma Aldrich), 500 μg/ml; hygromycin B (Invitrogen), 80 μg/ml. An Axiovert 25 inverted microscope (Zeiss) was used to visualize the cells in culture.

To preserve cell lines, cells of a confluent T75 flask were trypsinized and centrifuged 10 min at 1,000 r.p.m. After discarding supernatant, cells were resuspended in 2 ml of cold freezing medium (DMEM 71.5 %, FBS 21.5 %, DMSO 7 %) and two aliquots of 1 ml each were prepared. Aliquots were stored at -80ºC for a week and then transferred to -140ºC.

5.2. Transfection

For routine transfection of HeLa cells, the cationic JetPei transfection reagent (Polyplus Transfection) was used. The amount of DNA and JetPei reagent were adjusted depending on the cell culture format used, following the manufacturer´s instructions. To generate stably transfected cell lines, HeLa cells were allowed to grow and to express the drug resistance gene under non-selective conditions for 24-48 hours after transfection. Then, cells were cultivated in standard medium supplemented with the appropriate drug during 4-5 weeks, until outgrowth of resistant cells. Medium was changed every 2-3 days to avoid loss of selection pressure.

With the aim of obtaining stable cell lines, we attempted to transfect EA.hy926 (defined as a hard-to-transfect cell line) using different transfection reagents including JetPei, Lipofectamine 2000, Lipofectamine 3000, Lipofectamine LTX & Plus Reagent (Life Technologies), TurboFect (Thermo Scientific), and GeneJammer (Agilent Technologies), according to the manufacturer´s instructions. The transfection efficiency obtained with any of them was not high enough to obtain stable transformants.

5.3. Electroporation

Stable transfection of EA.hy926 cells was also tried by electroporation in a Gene Pulser Xcell main unit (BioRad). 1-10x10^5 cells of a 50-70 % confluent flask were used for each electroporation. The cells were washed with PBS, trypsinized, collected by centrifugation at 400 g for 6 min at room temperature, and resuspended in Optimem medium (Life Technologies). Cells and plasmid DNA were mixed in a cuvette and the pulse was applied in a Gene pulser Xcell main unit. Immediately after, cells were transferred to a plate with non-selective medium and the plate was incubated for 24-48 h before starting the selection.

All the pre-set conditions of the Gene Pulser Xcell device (mammalian 1 to mammalian 12) were tested. In all the cases cells were electroporated in a 100 μl volume and pulse was performed in a 0.2 cm cuvette.

5.4. Infection of human cells with *B. henselae*

Human EA.hy926 and HeLa cells were infected with *B. henselae*. *B. henselae* containing the indicated plasmids was grown on a CBA plate with appropriate antibiotics for 3-4 days (this was considered the first passage). A quarter of plate was recovered with a cotton swab and streaked in a new plate to obtain the next passage. Infection assays were done using passages 2-4.

Routine infections were done in a 6-well plate. The day before the infection, 80,000 cells per well were seeded and incubated in DMEM + FBS 10 %. When the purpose of the infection was to select human cells that had stably acquired the plasmid transferred from *B. henselae*, infections were performed in 150 mm tissue culture dishes. In that case, 1.2x10^6 cells were seeded the day before the infection.

The day of infection, the medium was changed to M199 medium (Gibco) supplemented with FBS 10 % and appropriate antibiotics to select for the *B. henselae* strains added right after. The bacteria were recovered from the CBA plate and resuspended in 1 ml of PBS. The number of bacteria was calculated considering that an OD_{600}=1 corresponds to 10^9 bacteria/ml (Kirby & Nekorchuk, 2002). Bacteria were added to the human cells to get a multiplicity of infection (MOI) of 400. The mixture of human cells and bacteria was incubated for 72 hours at 37ºC under 5 % CO_2. After that time, cell were washed three times with PBS and analysed by flow cytometry (section 5.6). When needed, antibiotic treatment was initiated right after the flow cytometry.

5.5. Fluorescence microscopy

This technique was used to check the morphology and eGFP expression of transfected or infected cells. A Nikon Eclipse E400 microscopy was used. Filters used for excitation and emission spectra were 488 and 530 nm, respectively.

5.6. Flow cytometry

Flow cytometry was used to detect and quantify human cells expressing eGFP after transfection or infection with *B. henselae*. After the transfection/infection incubation time, the medium was removed and cells were washed with PBS three times and then trypsinized. After centrifugation at 1,000 r.p.m. for 10 min, the pellet was resuspended in 600 µl of PBS in special FACS tubes. Samples were analysed in a Cytomics FC500 flow cytometer (Beckman Coulter). For each sample, 20,000 events were analysed. The viable population was established eliminating attaches, dead cells, and bacteria by means of size and complexity. An untransfected / uninfected control was used to adjust the GFP background level, delimiting the startpoint of the population of GPF positive cells.

6. Bioinformatic analysis

6.1. Statistical analysis

The different results obtained in this work were compared to the corresponding controls for statistically significance performing analysis of variance (ANOVA). In the graphs through the text, the significance is indicated by one or more asterisks and the corresponding p-value is indicated. Excel and GraphPad Prism programs were used with this purpose.

6.2. Software

Vector NTI Advance. It is a sequence analysis and design tool that can be used to view, analyse, transform, create, annotate, and share nucleotide and protein sequences. There is no free version available but a trial version can be downloaded from the website. http://www.lifetechnologies.com/es/en/home/life-science/cloning/vector-nti-software.html.

BLAST. The Basic Local Alignment Tool finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can also be used to infer functional and evolutionary relationships between sequences as well as to help in the identification of members of gene families. http://blast.ncbi.nlm.nih.gov/Blast.cgi.

BLAT. The BLAST-like Alignment Tool is a sequence alignment tool (UCSC Genome Bioinformatics) similar to BLAST but structured differently. BLAT finds similarity in DNA and protein but it needs an exact or nearly-exact match to find a hit. Therefore, BLAT is not as flexible as BLAST (Kent, 2002). https://genome.ucsc.edu/cgi-bin/hgBlat?command=start.

Primer3Plus. It is a program for designing PCR and sequencing primers, as well as hybridization probes. It offers the possibility to set several thermodynamic properties of the primers as desired (Untergasser *et al.*, 2012). http://primer3plus.com.

ImageJ. It is a public domain, java-base image-processing program. It can be used for multiple applications such as to display, annotate, edit, calibrate, measure, process or save image data. http://imagej.nih.gov/ij/.

GraphPad Prism. It is scientific 2D graphing and statistics software. It is useful for performing different statistical analysis and displaying experimental results in a graphical way. A trial version can be downloaded from the website. http://www.graphpad.com/scientific-software/prism/.

Results and Discussion

Chapter 1. Site-specific recombination mediated by conjugative relaxases

1. A mutagenesis approach to search for TrwC derivatives with improved recombinase activity

TrwC can catalyse site-specific recombination between two *oriT* copies repeated *in tandem* (see Introduction, section 3.1.2). The reaction is dependent on TrwC, since a R388 TrwC mutant is recombination deficient, and it is markedly favoured by TrwA, the accessory protein that assists TrwC nicking. The N-terminal 84 residues of TrwA were sufficient to enhance recombination and it was found to exert its influence independently from its putative *oriT* binding sites, which were dispensable at both recombination loci, suggesting that the effect is mediated by protein-protein interactions (César *et al.*, 2006).

Mutagenesis has been previously found to be useful as a method to obtain derivatives with increased efficiency for the ΦC31 integrase, a serine recombinase (Keravala *et al.*, 2009). In an attempt to increase the recombinase activity of TrwC, we carried out random mutagenesis of *trwC* by error-prone PCR and subsequent mutant screening in a TrwA deficient background. Under these conditions, the recombination efficiency of TrwC should be low enough to allow detection of any increase in the efficiency of the reaction, either due to a catalytically improved TrwC recombinase, or to a TrwC variant that has become independent of the TrwA requirement to accomplish efficient recombination.

The characterization of a TrwC variant with enhanced recombinase activity could contribute to the understanding of the catalytic basis of the reaction and it could be used with higher efficiency than TrwC for genetic engineering purposes.

1.1. Screening of *trwC* mutants

Before setting up the mutagenesis protocol, a screening method was needed. To this end, we used the recombination assay described by (César *et al.*, 2006) in which recombination leads to *lacZ* expression, allowing *in vivo* screening of colonies on selective plates. We used the expression vector pET3a as helper plasmid to carry *trwC* or its mutants. The helper plasmid is introduced in *lacZ*$^-$ strain DH5α containing pCIG1028, the substrate plasmid with both *oriT* copies. After chemical transformation of the helper plasmid, host cells are plated on X-Gal containing plates with the appropriate antibiotics to select for both the helper and the substrate plasmids. After overnight growth at 37ºC, the percentage of blue colonies is used to estimate recombination efficiency.

To test the screening efficiency of this system, we assayed helper plasmids pCIG1000 (pET3a::*trwAC*) and pSU1621 (pET3a::*trwC*), identical except for the presence or absence of *trwA* upstream of wild-type *trwC*, respectively. The results obtained with these two plasmids are shown in **Figure 39**. Although higher levels of recombination in the absence of TrwA were described in (César *et al.*, 2006), under our assay conditions the recombination efficiency of TrwC dropped to almost undetectable levels, while almost 100 % of the colonies underwent recombination in the presence of TrwA. We concluded that pET3a was an excellent vector for our purposes and that the incubation with pCIG1028 was a powerful tool for the screening. pCIG1000 and pSU1621 were assayed in parallel with all the mutants as internal controls.

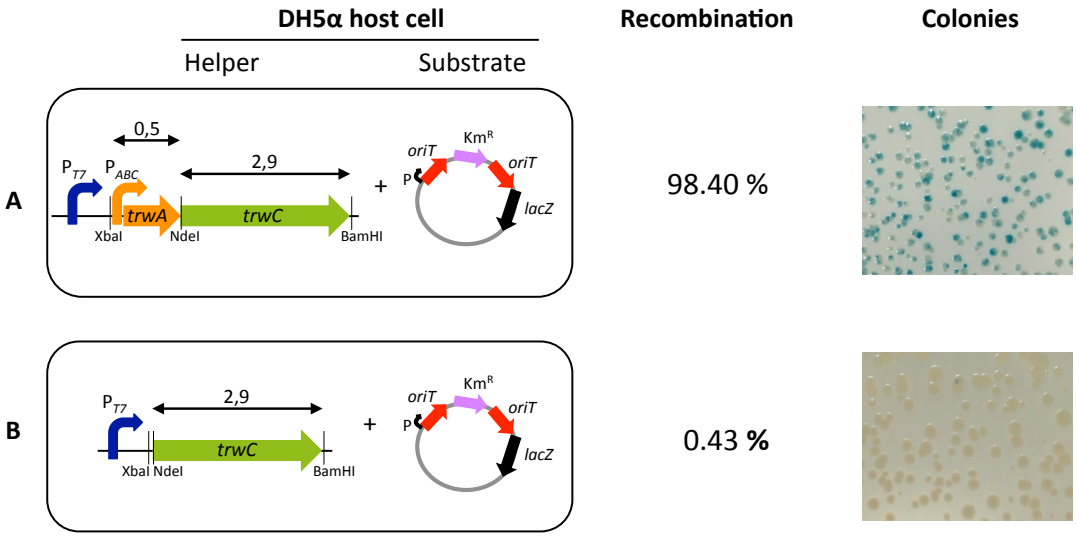

Figure 39. Effect of TrwA in TrwC-mediated recombination. Scheme of the DNA fragments cloned in pET3a in pCIG1000 (**A**) and pSU1621 (**B**), used as helper plasmids coding for the relaxase. The host cell used for the assay also contained a substrate plasmid that allowed the appearance of blue colour in the colonies where recombination had occurred. The percentage of recombinant colonies is shown in the middle column. The appearance of colonies obtained after plating on X-gal contained medium is shown on the right. P*T7*, T7 promoter. P*ABC*, *trwABC* promoter. P, P*lac* promoter.

1.2. Optimization of the mutagenesis procedure

Random mutagenesis was performed using GeneMorph II Random Mutagenesis Kit. Plasmid pMTX712 (pET29c::*trwAC*) was used as template for the mutagenic PCR. Using T7 primers (see section 3.4 of experimental procedures) a band of 3.5 kb

corresponding to *trwA* plus *trwC* was amplified. Adjusting the amount of template DNA and the number of cycles, the mutagenesis kit allows the selection of the mutation rate between 1 and 16 mutations per kilobase. Following the manufacturer´s recommendations, we set the reaction to obtain 1-4.5 mutations per kilobase.

With the aim of optimizing the PCR protocol to increase the product yield we tested different conditions. We observed no significant differences when using different possible combinations of amount of template DNA and number of cycles to obtain a specific mutation rate (e.g. 100 ng template and 25 cycles, versus 500 ng template and 30 cycles). Changes in the amount of polymerase (1-2 µl), annealing temperature (45-50ºC) and elongation time (3-6 min) rendered similar amounts of amplified product. We decided to use 100 ng of template DNA and 1 µl of Mutazime II DNA polymerase. The program selected was: 10 min of initial denaturation at 94ºC; 25 cycles of 30 seconds of denaturation at 94ºC, 30 seconds of annealing at 45ºC, and 3 min of elongation at 72ºC; and a final elongation step of 10 min at 72ºC.

To determine the actual mutation rate, we sequence the *trwC** coding gene from several colonies picked up randomly with T7 primers (section 3.4, Experimental Procedures). An average frequency of 4 mutations per kilobase was found.

1.3. Selection of TrwC variants with increased site-specific recombinase activity

The different steps carried out are outlined in **Figure 40**. After mutagenic PCR was performed, the PCR product was digested to obtain the 3 kb band corresponding to the pool of full-length *trwC* mutants (*trwC**), which was cloned into pET3a vector and introduced into the host strain already containing the substrate plasmid. pET3a coding for wild-type TrwC and TrwA+TrwC were assayed in parallel to have reference levels for low and efficient recombination. After overnight growth on X-gal containing plates, the colonies that showed blue sectors were selected for further analysis.

After around 40,000 colonies screened, one blue colony and a few more showing light blue sectors were found. To confirm that the blue colour was due to a higher recombination rate, plasmid DNA from those colonies was purified and used to transform DH5α selecting for the helper plasmid only. The colonies carrying a helper plasmid (Ap^R) but not a substrate plasmid (Cm^R) were selected by colony replica-plating in ampicillin and chloramphenicol plates. Ampicillin-resistant and chloramphenicol-sensitive colonies were selected and DNA was extracted. The different helper plasmids obtained were assayed individually in a recombination assay during 60 generations (see section 4.5 of

experimental procedures). Cells harbouring the substrate plasmid were transformed with the different helper plasmids carrying *trwC* mutants. pET3a::*trwC* and pET3a::*trwAC* were used again as reference. The different generations were plated on X-gal containing plates to calculate the percentage of recombination. If the blue colonies selected from the mutagenesis experiment had become blue due to a TrwC variant-mediated recombination event, most of the colonies generated in the second transformation should be blue. However, if the initial blue colonies had turned blue due to a background recombination event, in a TrwC variant-independent way, the colonies generated in the second transformation would be mostly white. **Table 9** shows the results in percentage of recombination for the different generations (T0-T60).

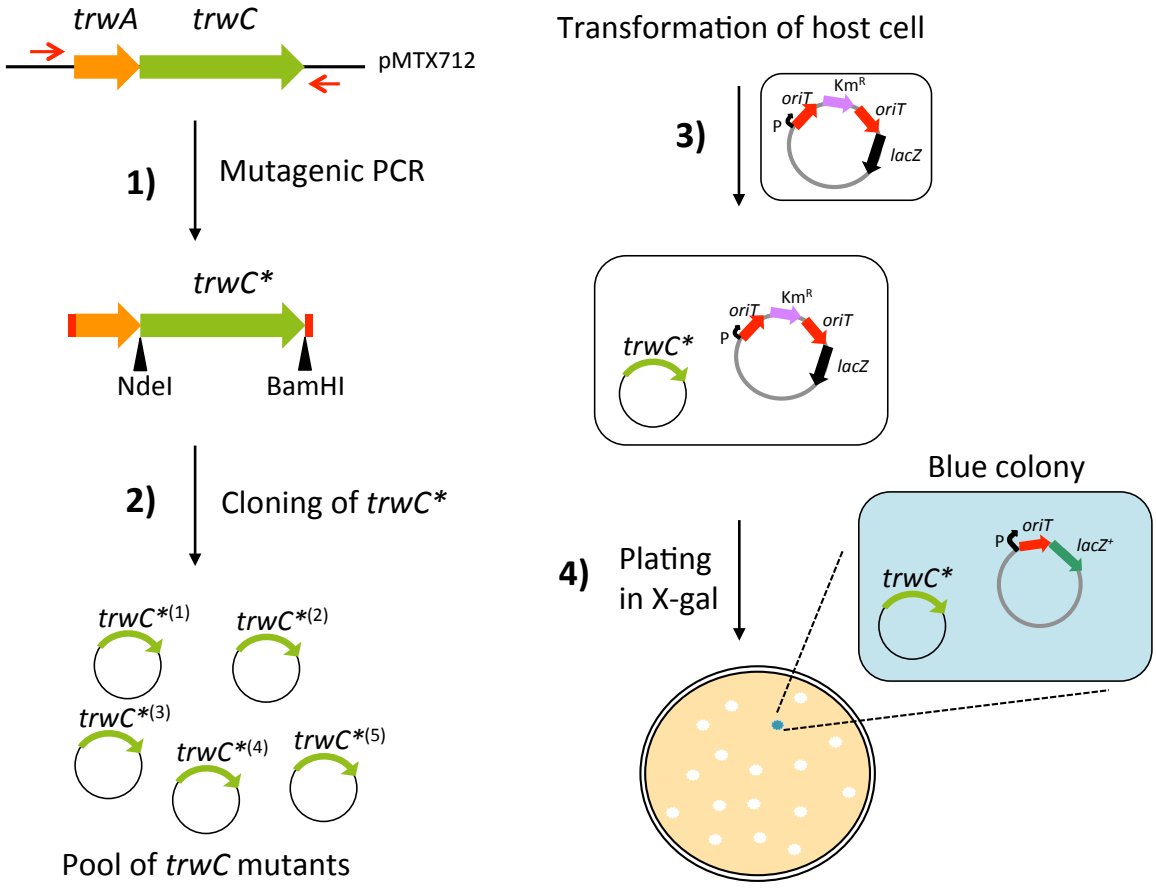

Figure 40. Outline of the mutagenic PCR and mutant screening. The pool of *trwC* mutants obtained by PCR was cloned into pET3a and each molecule was analysed by incubation with the substrate plasmid. Blue colonies were selected as indicative of recombination events. Green arrows, *trwC*(*). Red arrows, *oriT(s)*. P, P*lac*.

Helper plasmid	% Recombination			
	T0	**T20**	**T40**	**T60**
pET3a::*trwAC*	98.10	99.66	100	100
pET3a::*trwC*	0.05	0.38	0.83	0.86
pET3a::*trwC** (1)	0.72	0.84	1.52	2.20
pET3a::*trwC** (2)	0.05	0	0	0.54
pET3a::*trwC** (3)	0	0	0	0.11
pET3a::*trwC** (4)	0	0	0.09	0
pET3a::*trwC** (5)	0.15	0.28	0.62	0.83
pET3a::*trwC** (6)	0	0	0.12	0

Table 9. Recombination assay with *trwC* mutants. Percentage of blue colonies in different generations for the different helper plasmids analysed is shown. Numbers in brackets indicate the different helper plasmids carrying *trwC* mutants. trwC* (1) initially produced a blue colony, while *trwC**(2) to *trwC**(6) produced colonies with blue sectors.

As expected, with pET3a::*trwAC* the vast majority of colonies were recombinant, while with pET3a::*trwC* less than 1 % of the colonies had the recombined plasmid. It was found that TrwC* (1) variant showed enhanced recombination levels compared to TrwC under the same assay conditions.

DNA of the different helper plasmids was analysed with restriction enzymes NdeI and BamHI (**Figure 41**). Apart from the 4.6 kb-band corresponding to pET3a, a band of 3 kb corresponding to *trwC* or *trwC** should be visible.

Mutants that did not show the expected bands or that mediated site-specific recombination with the same or even less efficiency than TrwC wild-type were discarded. After this, mutant #1 seemed worth to be analysed further. This TrwC variant will be henceforth referred to as TrwC*.

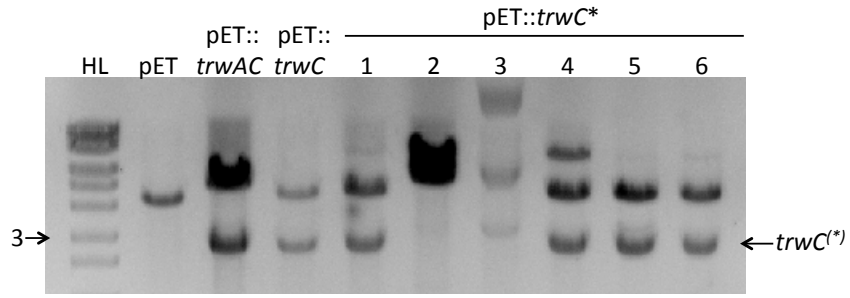

Figure 41. Restriction analysis of helper plasmids. Two bands were expected, one of 4.6 kb corresponding to pET3a vector (or 5.1 kb for pET3a::*trwA*), and the 3 kb band of *trwC* (wild-type or mutant). HL, Hyperladder I.

As we had obtained a TrwC variant with increased recombinase activity, we decided to repeat the mutagenesis approach but using *trwC** as well as *trwC* as templates to try to increase even more the efficiency of the mutant. The strategy carried out was the same. pMTX712 and pCOR16 (pET3a::*trwC**) were used as templates for the mutagenic PCR using the same conditions as before. The pool of *trwC* mutants was cloned into pET3a and analysed by incubation with pCIG1028 in DH5α. As previously done, around 40,000 colonies were screened looking for recombination events. The helper plasmids of the colonies showing at least blue sectors were obtained and assayed again. Some of the helper plasmids found in the new screening were able to catalyse the reaction with higher efficiency than the wild-type protein, but none of them reached the level of the mutant previously isolated (**Table 10**). For that reason, we decided to continue with the characterization of TrwC* only.

Helper plasmid		**% Rec.**	**Helper plasmid**		**% Rec.**
pET3a::*trwC*		0.11	pET3a::*trwAC*		99.62
pET3a::*trwC**		0.87	pET3a::*trwAC**		99.85
Template		**% Rec.**	**Template**		**% Rec.**
trwC	(1)	0.19	*trwC**	(1)	0.24
	(2)	0.04		(2)	0.36
	(3)	0.18		(3)	0.43
	(4)	0.02		(4)	0.09
	(5)	0.05		(5)	0
	(6)	0.26		(6)	0.02
	(7)	0.11		(7)	0
	(8)	0.02		(8)	0.19
	(9)	0		(9)	0.20
	(10)	0.02		(10)	0.38
	(11)	0		(11)	0.45
	(12)	0.04		(12)	0.20
	(13)	0.04		(13)	0.03
	(14)	0.28		(14)	0.63
	(15)	0.32		(15)	0
	(16)	0		(16)	0.08
	(17)	0.11		(17)	0.68
	(18)	0.63		(18)	0.07
				(19)	0.20

Table 10. Recombination efficiency of helper plasmids isolated in the second round of mutagenesis. pET3a::*trwC* and pET3a::*trwAC* with *trwC* wild-type and mutant were used as reference. Numbers in brackets indicate the different helper plasmids that produced a blue colony in the screening. Data represent the mean of at least 3 independent experiments. % Rec., percentage of recombination.

1.4. Characterization of TrwC*

We firstly tested if TrwC* was efficient in conjugation. We assayed its ability to complement *in trans* plasmid pSU1445, a *trwC*-deficient R388 derivative. The results (**Table 11**) showed no significant difference when compared with the wild-type protein so we concluded that TrwC* is completely functional in conjugation.

Donor	**TrwC protein**	**Transfer frequency**
DH5α (R388 *trwC*$^-$ + pET3a::*trwC*)	TrwC	0.62
DH5α (R388 *trwC*$^-$ + pET3a::*trwC**)	TrwC*	0.58
DH5α (R388 *trwC*$^-$*)*	none	$<1.94x10^{-7}$

Table 11. Complementation of a TrwC-deficient R388 in conjugative transfer. Transfer frequencies are shown as transconjugants per donor (mean of 3 different experiments).

To be sure that the increase in the activity was due to the relaxase and not to a side effect produced by the vector, we obtained *trwC** gene by digestion with NdeI and BamHI and we cloned it again in pET3a vector (to obtain pCOR17; See Experimental Procedures, Table 3). We then characterized its site-specific recombinase activity during different generations by incubation with pCIG1028. Different generations were plated in the presence of X-gal to monitor the colour of the colonies. The results (**Figure 42a**) confirmed that there is a statistically significant difference in the percentage of recombination mediated by the mutant and the wilt-type protein. The assays were also performed in the presence/absence of TrwA using four different helper plasmids: *trwC*, *trwAC*, *trwC**, and *trwAC**. The results (Figure 42b) show that an increase in the level of recombination was observed in the absence as well as in the presence of the accessory protein.

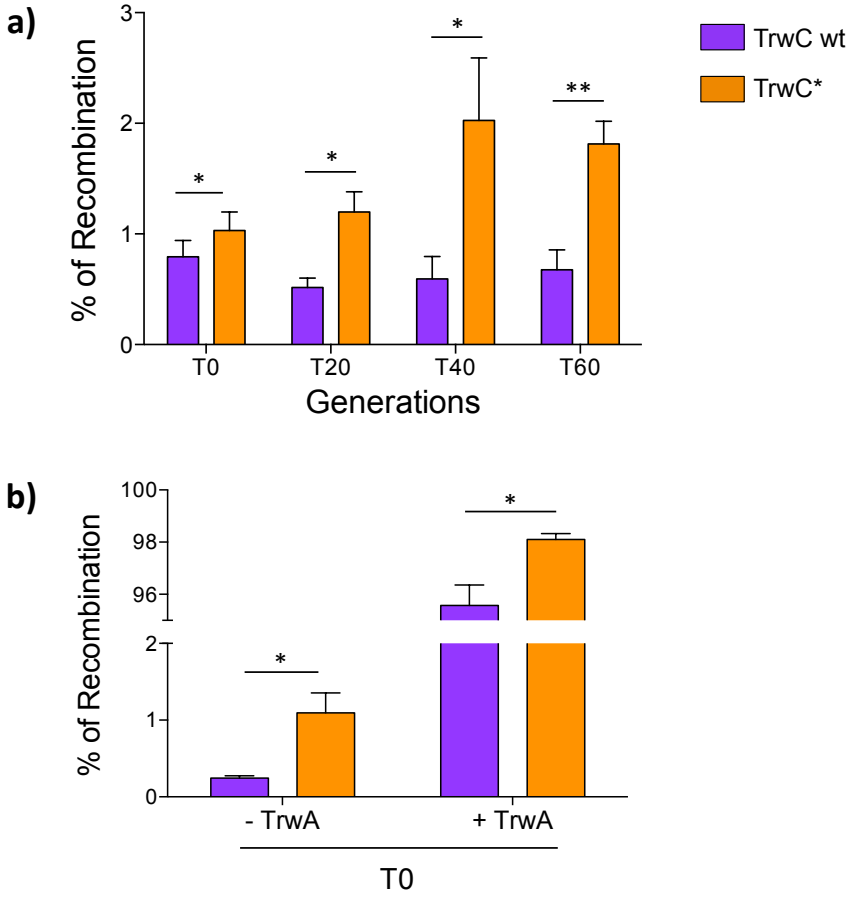

Figure 42. Site-specific recombination mediated by TrwC*. a) Percentage of recombinant colonies was calculated for different generations compared to TrwC wt in the absence of TrwA. **b)** Effect of TrwA in TrwC*-mediated recombination. When needed, TrwA was expressed from the same helper plasmid as TrwC$^{(*)}$. Data represent the mean of at least 5 independent experiments. * $p<0.05$. **$p<0.01$.

We determined the *trwC** DNA sequence and found three changes from the *trwC* sequence publically available at NCBI GenBank database (accession number X63150.2): C1572G, G1573C, and T1986G. They were found to produce a unique change in the amino acid sequence of the protein: Ala525Pro. However, after sequencing other *trwC*-containing DNA molecules we found the same changes, so we assume that what we had found were mistakes in the GenBank sequence submitted in 2002. We have already requested the update to include the mistakes detected in this work.

This means that *trwC* and *trwC** do not have any difference at the nucleotide sequence level; still we observe a significant difference in recombination efficiency. Trying to find an explanation, we measured the amount of plasmid DNA (by gel

electrophoresis) and TrwC protein (by western blot with anti-TrwC antibody) on *E. coli* cells coding for *trwC* or *trwC**, under the same assay conditions. pET3a::*trwC* and pET3a::*trwC** plasmids were introduced into DH5α. The cultures were grown overnight, diluted to an optical density (OD_{600}) of 0.6, grown for 2 hours and then diluted again to have both cultures at the same OD_{600} to be sure that both samples contained the same amount of cells. Immediately after, DNA was extracted and digested with NdeI and BamHI. Equivalent samples were also analysed by western blot. Results are shown in **Figure 43**.

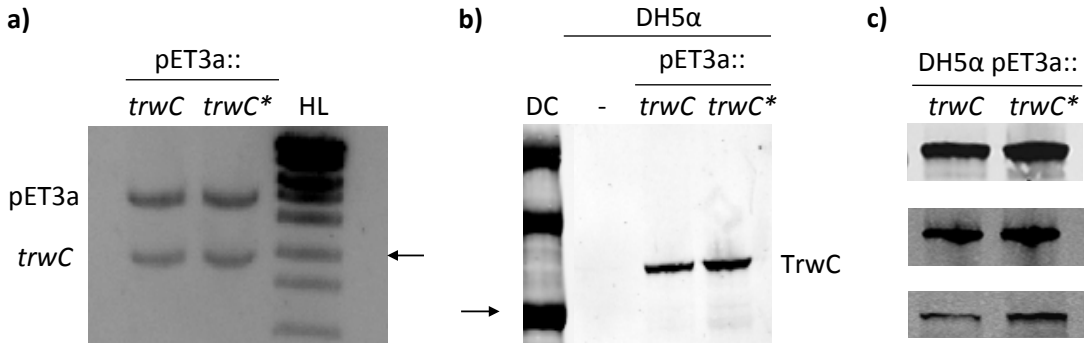

Figure 43. DNA and protein amount of TrwC and TrwC*. a) Gel electrophoresis of pSU1621 (pET::*trwC*) and pCOR17 (pET::*trwC**). The bands of *trwC(*)* and pET3a are indicated on the left side. The band of 3 kb of the molecular marker is indicated with an arrow. **b)** Western blot using cell extracts of DH5α carrying pSU1621 or pCOR17. The wild-type strain without any plasmid was used as negative control (-). The arrow indicates the band of 100 kDa of the marker. TrwC(*) band is indicated on the right. **c)** TrwC and TrwC* western-blot results of other experiments. HL, Hyperladder. DC, Protein dual colour standard.

Figure 43 (a and b) shows the results of one representative experiment out of at least four independent experiments with the same results. In all of them, no differences in the amount of DNA were found, while an appreciable difference in the amount of protein was observed. Western-blot results of other experiments are shown in Figure 43c.

This indicates that a higher amount of TrwC* protein is produced, and this higher amount of protein presumably causes the observed increase in recombination efficiency. To determine if this increase in the protein abundance could be due to a difference in the promoter region of both pET3a vectors, we sequenced them. No differences were found in the kilobase located upstream of *trwC(*)* ORF.

Despite not having an explanation for what has been found, TrwC* is not what we were looking for when we planned the mutagenesis of TrwC. It seems that the increase in the recombinase activity is due to a higher abundance of protein and not to an increased catalytic activity. Apart from TrwC*, we have not found any mutant with increased recombinase activity. As we were mutating a 966 aa protein and each position in the protein sequence could be mutated to 19 different amino acids, a total number of 18,354 single mutants could be generated. Taking into account that we have performed a screening of approximately 80,000 colonies, the results suggest that mutants with increased recombinase activity cannot be obtained by altering one or two residues of the protein, in contrast to what was found for the ΦC31 integrase (Keravala *et al.*, 2009). For that reason, we decided to discontinue this research approach.

2. Site-specific recombination mediated by F_TraI

Site-specific recombination has been described as not a prototypical reaction of conjugative relaxases (César *et al.*, 2006). In fact, it has been barely described for other relaxases than TrwC. These include relaxases as different as the MOB_P R64_NikB (Furuya & Komano, 2003), the MOB_V relaxases of *Enterococcus faecalis* plasmids pAD1 (Francia & Clewell, 2002a) and pAMα1 (Francia & Clewell, 2002b), and the MOB_Q R1162_MobA (Meyer, 1989). However, other relaxases closely related to TrwC were reported to be recombination-deficient. It was published that F_TraI was not capable of promoting recombination between two *oriT* copies in the absence of conjugation (Gao *et al.*, 1994). Cesar and co-workers tested it again using an optimized recombination cassette containing both *nic* sites in the lagging strand and the same recombination assay described here. Under these conditions, no blue colonies were detected, so it was inferred that TraI-mediated recombination was not taking place (César *et al.*, 2006). It was an unexpected result, since both R388_TrwC and F_TraI belong to the same MOB_F family of relaxases and are highly related, even at structural level (see introduction, Figure 17).

2.1. Revisiting F_TraI-mediated recombination assay

We wanted to determine if the lack of F_TraI recombinase activity could be due to any problem in the setting of the assay. We firstly focused in the substrate plasmid, pCIG1064. As the one for testing TrwC-mediated recombination, it has two *oriT* copies repeated in tandem separated by a long DNA segment. As pCIG1028, it should lead to

lacZ expression upon recombination (see Experimental Procedures, section 4.5). When using DH5α (*lacZΔM15*, *recA*) as host cell, recombination should produce blue colonies in X-gal containing medium. To check that this actually happened when using pCIG1064, conjugation-dependent recombination (**Figure 44**) was tested.

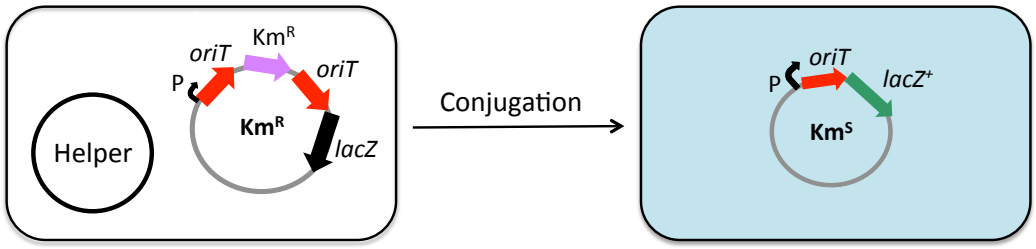

Figure 44. Conjugation-dependent recombination. In conjugative processing of the recombination substrate, the intervening sequence between the two *oriTs* (containing a Km resistance gene) is lost. Thus, transconjugants mimic the expected recombinants. Red arrows indicate the *oriTs*. P, P*lac*.

In this assay, recombination substrates were used as mobilizable plasmids. They were mobilized to the recipient cell by the corresponding relaxase expressed in a helper plasmid. In conjugation, the relaxase should recognize one of the *oriTs* of the recombination substrate and process it, ending the reaction when finding the second *oriT* copy. Consequently, the substrate plasmid is transferred to the recipient without the DNA segment between the *oriTs*. Thus, transconjugants mimic the expected recombinants. D1210 was used as donor strain, while DH5α was used as recipient. pCIG1028 and helper R388 were used as positive control. For testing pCIG1064, pOX38 was used as helper plasmid. pOX38 is a derivative of plasmid F that contains the whole conjugative Tra region. As a negative control, a TraI-deficient pOX38 was used as helper. **Table 12** shows the results obtained.

As expected, no transconjugants were found when there was not a functional relaxase. Blue transconjugants were found after TrwC-mediated mobilization. For TraI-mediated mobilization, transconjugants were found to be white. Plates were stored at 4ºC for a few days and eventually all the colonies had turned blue. This observation suggested that maybe the X-gal containing plates where TraI-mediated recombination was previously tested (César *et al.*, 2006) could have shown blue colonies if incubated for longer time.

As both R388 and F recombined substrates have a different *oriT* between the promoter and the *lacZ* ORF, this difference in *lacZ* expression between both plasmids could be due to the presence or absence of a promoter or a terminator in one (or both) *oriTs*.

Plasmids in donor	Relaxase	Recombination substrate	Transfer frequency	Color of transconjugants
R388+pCIG1028	TrwC	R388_*oriTs*	$9.8x10^{-2}$	Blue
pOX38+pCIG1064	TraI	F_*oriTs*	$1.2x10^{-2}$	Blue*
pOX38Δ*traI*+pCIG1064	none	F_*oriTs*	$<4.7x10^{-6}$	-

Table 12. F conjugation-mediated recombination assay. Transfer frequency is shown as transconjugants per donor. Data represents the mean of at least 3 independent experiments. * Blue after 1-3 days.

2.2. F_TraI-mediated site-specific recombination

We decided to test again F_TraI-mediated recombination, monitoring the colour of the colonies during longer time. pOX38 was used as helper plasmid and it was incubated with its cognate substrate plasmid containing two tandemly repeated copies of F *oriT*. To be sure that the recombination (if detected) was dependent on TraI, the same substrate plasmid was also incubated with helper pOX38Δ*traI*. R388 with its cognate substrate plasmid was used as positive control for recombination (**Figure 45**).

Aliquots of bacterial cultures were plated on X-gal containing media during different generations. Plates were incubated overnight at 37ºC and then stored at 4ºC during 2-4 days.

After overnight incubation at 37ºC, colonies in the positive control were mostly blue. In both F-derived systems, the colonies were white. However, after incubation at 4ºC, some colonies of pOX38 turned blue while all the colonies obtained after transformation with pOX38Δ*traI* remained white. We considered these blue colonies as recombinants and used them to calculate recombination frequencies, which are shown in **Figure 46**. Blue colonies in the F_TraI assay reached 11 % in T60.

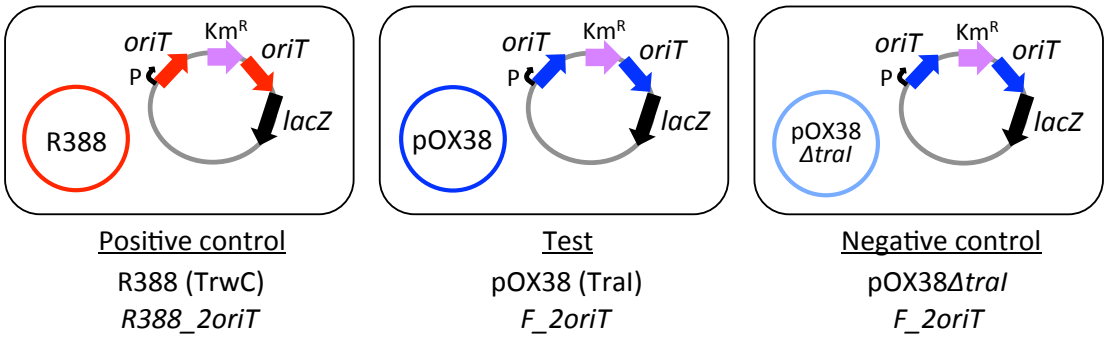

Figure 45. Experimental design for TraI-mediated recombination. Cells harbouring the substrate plasmid (with R388 or F *oriTs*, shown as red and blue arrows, respectively) were transformed with the appropriate helper plasmid. Assays were performed as explained in experimental procedures (section 4.5). P, P*lac*.

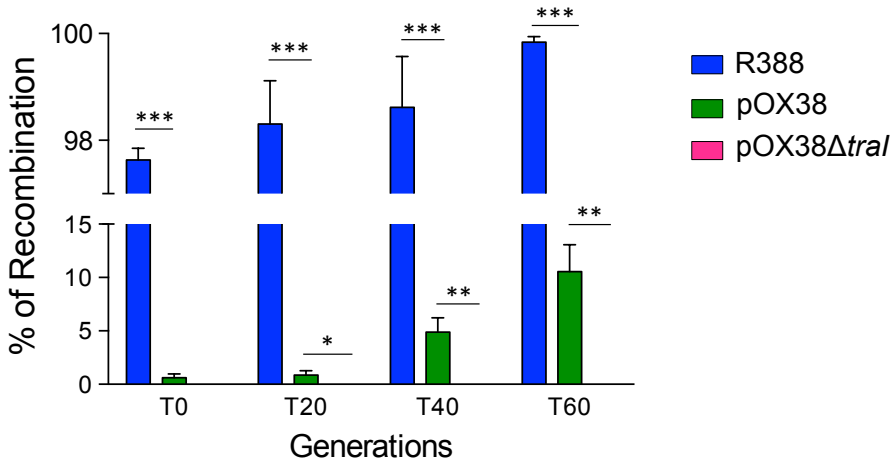

Figure 46. Site-specific recombination mediated by F_TraI. Blue and white colonies were counted to obtain recombination frequencies. For simplicity, only the helper plasmids are indicated. Data represent the mean of at least 6 independent experiments. *, p<0.05. **p<0.01. ***p<0.001.

Recombination assays were then performed including samples with no helper plasmid, and pOX38 was also tested with the R388 substrate plasmid. Recombination was confirmed by visualization of DNA in agarose gels to check the size of the helper plasmids (**Figure 47**), since recombination implies the loss of the 3.2 kb cassette in between both *oriT* copies.

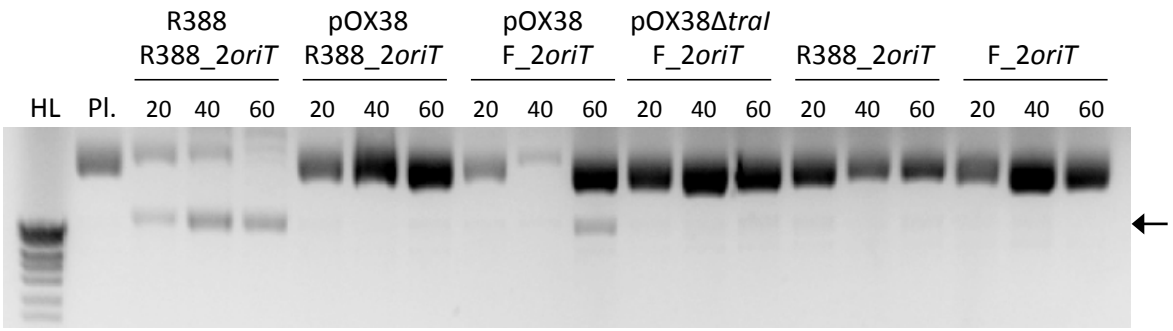

Figure 47. Analysis of the DNA obtained from different generations of a recombination assay. The helper (if present) and substrate plasmids are indicated on the top. Recombinant plasmid size is indicated with an arrow. 20, 40, and 60 indicate the generations analysed. Pl., F substrate plasmid DNA. HL, Hyperladder I.

As expected, in the positive control (R388 system) a smaller plasmid is detected as an evidence of recombination. The same band is also present in the last generation of pOX38 system (with its cognate substrate). When pOX38 was incubated with a substrate plasmids containing R388 *oriTs* no recombinant plasmid was detected. The same was found when using pOX38Δ*traI*, confirming that the recombination detected with pOX38 is mediated by its relaxase TraI.

In contrast to what was previously reported, these results clearly indicate that F_TraI is, as TrwC, a conjugative relaxase with site-specific recombinase activity. However, the efficiency of F_TraI is very low compared to that of TrwC under similar assay conditions.

2.3. Complementation of pOX38Δ*traI*

To test the effect of different amounts of F_TraI in the recombination efficiency, plasmid p99I^{+} was used. It has the *traI* gene cloned after *trc* promoter. As it also has the *lacI*^{q} gene, which represses Plac promoter, detection of recombination by *lacZ* expression was not possible. Recombination assays were performed with helper plasmids R388, pOX38, pOX38Δ*traI*, and pOX38Δ*traI* complemented with p99I^{+} in the presence of their respective R388/F *oriT* substrate plasmids. *traI* expression was induced by incubation of the liquid cultures on IPTG-containing media during 3 hours before plating and extracting DNA of each generation. DNA samples were analysed by restriction digestion (**Figure 48**).

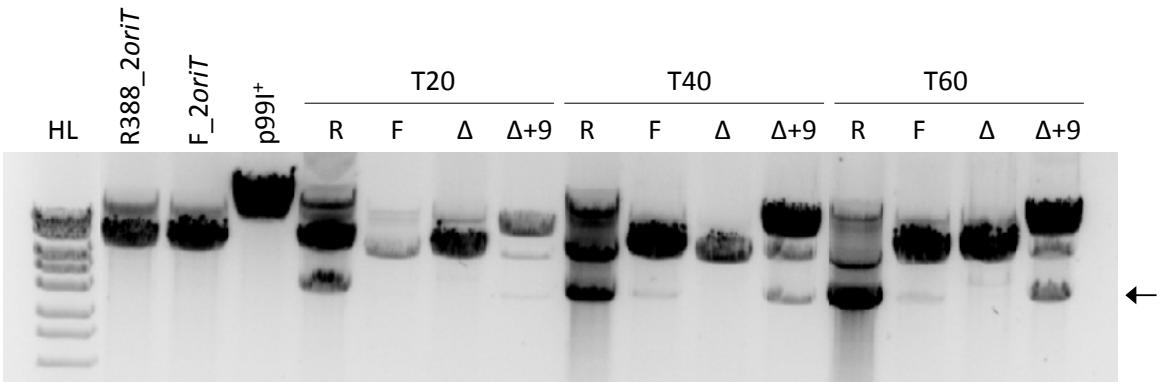

Figure 48. EcoRI digestion of DNA samples from recombination assays. EcoRI enzyme cuts once in the recombination substrate plasmids, resulting in a band of approximately 7 kb when there is no recombination and in a band of 4 kb when the plasmids are recombinant (indicated by an arrow). HL, Hyperladder. R, R388+R388_2*oriT*; F, pOX38+F_2*oriT*; Δ, pOX38ΔtraI+F_2*oriT*; Δ+9, pOX38ΔtraI+F_2*oriT*+p99I$^+$.

A band of recombinant substrate is present in all the generations of the R388 system and in T40 and T60 generations of pOX38 system, while it is not detected with pOX38Δ*traI.* However, when p99I$^+$ is also present with pOX38Δ*traI*, a recombinant band is visible even in the first generation analysed.

These results show that the amount of recombinant product obtained is higher when TraI is produced under a strong promoter that when it is produced from its natural promoter. Together with the results obtained with the TrwC variant selected for its increased recombinase activity and found to be wild-type TrwC but produced in higher amount (see section 1.4), this is another evidence indicating that an increase in the amount of relaxase protein present in the cell produces an increase in the amount of recombinant plasmid. This result also indicates that efficient site-specific recombination can be obtained using the F_TraI relaxase.

3. Site-specific recombination mediated by pKM101_TraI

Apart from R388_TrwC and F_TraI, the other best characterized member of MOB_F family of conjugative relaxases is TraI from IncN conjugative plasmid pKM101. It is also a bifunctional enzyme with both relaxase and helicase domains (Paterson *et al.*, 1999). Its site-specific recombinase activity was also tested by Cesar and co-workers and no

recombination was detected (César *et al.*, 2006). However, a previous work showing intermolecular recombination suggested the opposite (Paterson *et al.*, 1999).

The substrate plasmid used in the study carried out by César and co-workers was similar to the ones used with TrwC and F_TraI. *lacZ* is expected to be expressed upon recombination, making recombinant colonies become blue in the appropriate X-gal containing medium. However, we have shown that the appearance of blue colour when using F substrate is slower than when using R388 substrate. For that reason, we decided to test pKM101 substrate in a conjugation-mediated recombination assay (Figure 44) as previously done for F substrate plasmid. Results are shown in **Table 13**.

In this case, no blue colour was found in the transconjugants obtained with the pKM101 substrate plasmid, not even after up to 6 days of storage at 4ºC. Although white, 100 % of the transconjugants were found to be Km sensitive, indicating the loss of *ntpII* gene, as expected after conjugative transfer from *oriT2* to *oriT1*.

Although the three recombination substrates (R388/F/pKM101) were constructed in the same way (César *et al.*, 2006), and they are all based in the same original pRec2*oriT*-Cm plasmid (Draper *et al.*, 2005), they differ, as mentioned in the previous section, in the sequence of the *oriT*, where the presence of transcription promoters or terminators could make a difference in *lacZ* expression levels.

Plasmids in donor	**Relaxase**	**Recombination substrate *oriTs***	**Transfer frequency**	**Color of transconjugants**
R388+pCIG1028	TrwC	R388	$4.1x10^{-1}$	Blue
pOX38+pCIG1064	F_TraI	F	$9.6x10^{-2}$	Blue*
pKM101+pCIG1066	pKM_TraI	pKM101	$7.3x10^{-1}$	White
pCIG1066	none	pKM101	$<5.1x10^{-7}$	-

Table 13. Results of the conjugation-mediated recombination assay. Transfer frequency is shown as transconjugants per donor. Data represent the mean of at least 3 independent experiments. * Blue after 1-3 days.

Site-specific recombination mediated by pKM101_TraI was then analysed by the loss of Km resistance and DNA restriction analysis (Experimental Procedures, section 4.5). The results obtained are shown in **Table 14**.

	% Recombination			
	T0	**T20**	**T40**	**T60**
R388 + R388_2*oriT*	0	39.35	76.59	89.58
pKM101 + R388_2*oriT*	0	0	0	0
pKM101 + pKM101/R388_2*oriT*	0	0	0	0
pKM101+ pKM101_2*oriT*	0	0	2.08	6.25

Table 14. pKM101_TraI-mediated site-specific recombination. Recombination was calculated as number of Km sensitive colonies per total number of colonies x 100. Data represent the mean of 3 independent experiments.

R388-based system showed lower levels of recombination than when it is tested by *lacZ* expression. This is probably due to the fact that when recombination is tested by *lacZ* expression, a single recombination event produces a blue sector in the colony and it is considered as recombinant. However, when it is tested by Km resistance, mostly all the plasmid copies in a bacterium must have undergone recombination for the cell to become Km sensitive and so, considered as recombinant. Despite that, pKM101_TraI-mediated recombination was detected in T40 and T60 generations when the substrate plasmid had two pKM101 *oriT* copies (pCIG1066), while no recombination was detected when the substrate plasmid had two R388 *oriT* copies (pCIG1028) or one *oriT* from each conjugative plasmid (pIG1032).

DNAs obtained from the last generation (T60) were analysed by restriction digestion with EcoRI, that cuts once in the recombination substrates (**Figure 49**). The recombinant band was observed in R388 positive control and in pKM101 system. As for F_TraI, this confirms that pKM101_TraI is also able to catalyse recombination on its cognate substrate, in contrast to what was previously reported.

It can be observed in Figure 49 that the recombination substrate showed two different sizes (asterisk). This puzzling phenomenom was investigated and we found out it was due to the presence/absence of a copy of IS10. This is explained in detail in the Appendix of this PhD thesis work.

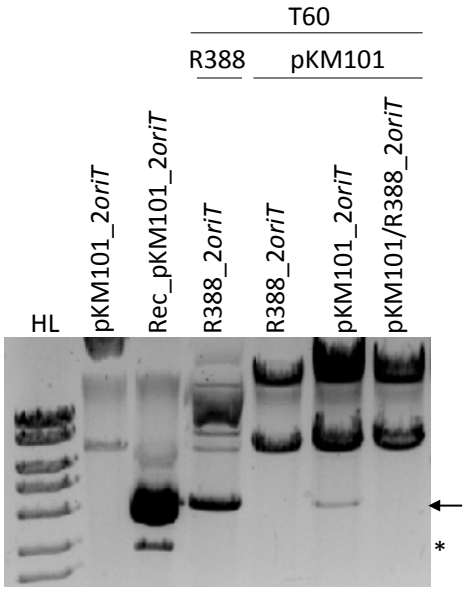

Figure 49. EcoRI digestion of T60 DNAs. The 4 kb band of recombinant substrates is indicated with an arrow. * indicates recombinant plasmid that had lost IS10 insertion sequence (changing from 4 to 3 kb in size; see text for further explanation). Rec_pKM101_2*oriT*, recombinant substrate plasmid control DNA obtained by conjugation-mediated recombination.

All together, these results indicate that F_TraI and pKM101_TraI are, as TrwC, conjugative relaxases with site-specific recombinase activity. Both relaxases had been previously tested and no recombinase activity was found (César *et al.*, 2006). We decided to revisit these recombination assays since both TraI proteins are highly related to TrwC and the three of them share many different features. The previous negative results could be due to the fact that the expression of the *lacZ* reporter gene varies a lot between the similarly constructed recombination substrate plasmids, leading to the misinterpretation of the results.

Although the three relaxases belong to the same family (MOB_F) (Garcillán-Barcia *et al.*, 2009), this ability is not exclusive for relaxases belonging to it. As it has been previously mentioned (see section 2), other known relaxases catalysing this activity are remarkably different to MOB_F relaxases (Garcillán-Barcia *et al.*, 2009).

In the case of R64_NikB, *oriT*-specific recombination on dsDNA require the presence of the *oriT* DNA binding protein NikA, whereas only NikB is necessary for recombination of ssDNA substrates (Furuya & Komano, 2003). However, site-specific recombination mediated by R1162_MobA in the absence of conjugation has only been

detected when two directly repeated copies of its *oriT* were cloned into a ssDNA bacteriophage (Barlett *et al.*, 1990; Meyer, 1989).

The efficiency of TrwC and both TraI proteins in recombination is very different under our assay conditions, with the first one being much more efficient. In the case of F_TraI, 11 % of colonies underwent recombination in T60, while TrwC-mediated recombination reaches the 99.8 %. For pKM101_TraI, 6.3 % of the colonies of the last generation analysed had become Km-sensitive, compared to the 89.6 % of Km-sensitive colonies obtained with TrwC.

This could be due to different properties of each relaxase, such as *oriT* binding affinity, the mechanism by which they cut the DNA at the *oriT,* or their interaction with the replication machinery.

The different efficiency of these relaxases as recombinases could be also due, at least in part, to the number of relaxase molecules in the cell, since we have observed that an increase in the amount of protein produces an increase in the recombination efficiency. It is know that a high number of TrwC molecules are present in the cell (Grandoso, 1994) whereas TraI proteins could be expressed in lower levels as happens in the case of TraI of plasmid RP4 (Grahn *et al.*, 2000). To elucidate this, the recombination assays should be complemented with the quantification of protein amounts (by western blot, for example). Until that, we cannot compare efficiencies, although everything seems to indicate that TrwC is more active as recombinase than the others.

Chapter 2. Site-specific integration mediated by conjugative relaxases

4. Site-specific integration mediated by TrwC

4.1. Previous results on DNA requirements of the site-specific integration reaction

TrwC was firstly described as a site-specific integrase by Draper and collaborators (Draper *et al.*, 2005). In this work, the authors showed that TrwC was capable of integrating a suicide plasmid (unable to replicate in the recipient cell) containing the R388 *oriT*$_W$ and the RP4 *oriT*$_P$ into an *oriT*$_W$-containing plasmid present in the recipient after *in vivo* mobilization during conjugation. A frequency of integration of 5.4x10^{-6} was reported when DH5α was used as recipient. The reaction was later optimized and further characterized by Agúndez and collaborators (Agundez, 2011; Agundez *et al.*, 2012) (see Introduction, section 3.1.3).

Two different assays were used to characterize the reaction (illustrated in **Figure 50**): the standard assay in which TrwC is provided from the donor cell covalently linked to the transferred DNA, and the assay in which the T-DNA is introduced by another relaxase (RP4_TraI), and *trwC* is expressed in the recipient.

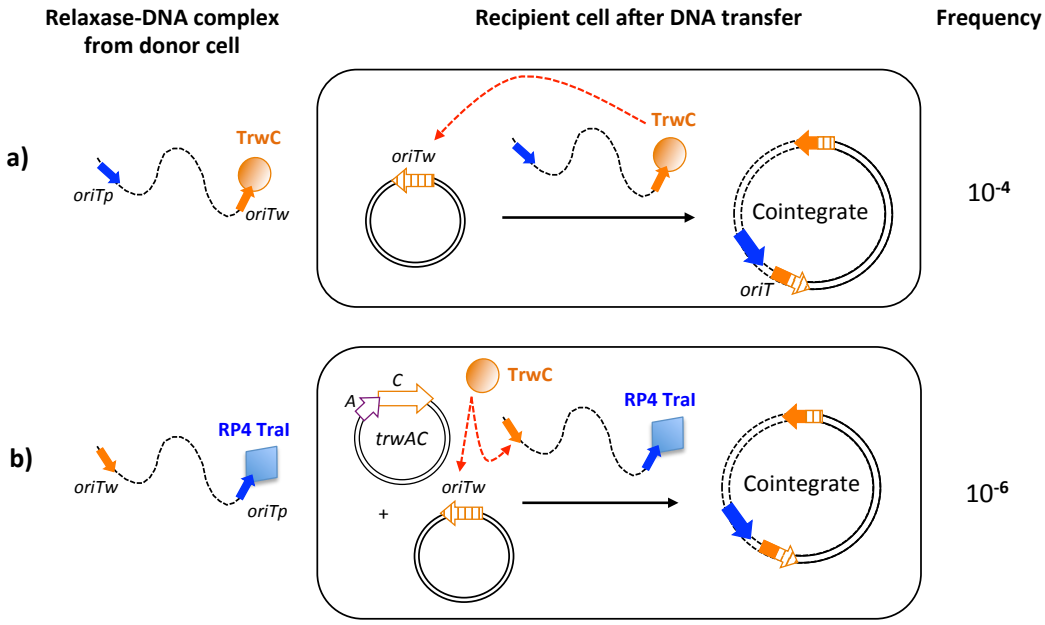

Figure 50. TrwC-mediated site-specific integration assays. a) TrwC-DNA complexes are transferred from the donor. **b)** The suicide plasmid is mobilized by RP4_TraI and *trwC* is expressed in the recipient, where it must localize and process both *oriTs*. Red arrows indicate the action of TrwC in its cognate *oriTs*. *A, trwA. C, trwC*. *oriTw*, R388 *oriT*. *oriTp*, RP4 *oriT*.

The latter was designed in order to assess the effect of *trwC* mutations on integration. The two assays require different actions to complete integration, since in the first one (Figure 50a), the TrwC-DNA complex is only required to catalyse the strand-transfer reaction into the acceptor *oriT* in the recipient, while in the second (Figure 50b), free TrwC must act on both intact *oriT* copies, located on the supercoiled DNA of the recipient plasmid, and on the transferred single strand coming from the donor. Thus, the first assay mimics the end of conjugative DNA transfer, in which a transferred TrwC-DNA complex enters a recipient cell and requires a strand-transfer reaction to complete the process; while in the second assay, TrwC is required to act on intact *oriT* copies as it would at the beginning of the conjugative DNA process.

The minimal *oriT* required for TrwC-mediated relaxation and recombination is 17 bp (Introduction, sections 3.1.1 and 3.1.2). TrwC integrase activity was assayed using a target plasmid containing this minimal *oriT* of 17 bp. Upon TrwC-DNA transfer to the recipient, integration was detected in this minimal target. However, the frequency of the reaction decreases by around 2 logs. When *trwA* was also expressed in the recipient cell, a modest increase in the integration frequency was obtained (Agundez *et al.*, 2012). It was also found that DNA sequences of human origin deviating slightly from the consensus 17 bp *oriT* could also work as acceptor sites (see Introduction, section 4.2.2.2); the decrease in efficiency was less than 1 log.

The authors also addressed the requirements of *oriT* sequences essential for TrwC binding and nicking. Using the assay which expresses *trwC*+*trwA* in the recipient cell (Fig 50b), no integration was detected on two mutant *oriTs* affecting the IR_2 where TrwC binds and the three nucleotides 5' to the *nic* site, confirming that TrwC requires its binding and nicking sites to catalyse the complete integration process (Agundez *et al.*, 2012).

As these last results were obtained in the integration assay expressing *trwC* in the recipient, it must act on intact *oriT* copies. As discussed before, the use of the integration assay in which TrwC-DNA complexes are already formed could distinguish between DNA sequence requirements at the first or the last step of TrwC-mediated integration. In addition, in order to use TrwC as a genomic modification tool, TrwC-DNA complexes will enter the recipient cell, where it will catalyse integration of the incoming DNA if it can act on existing genomic sequences. Thus, we reasoned that in order to test the specificity of TrwC for its target, it would be necessary to test integration into different acceptor sites once a TrwC-DNA complex is delivered into the recipient cell.

4.2. Site-specific integration of TrwC-DNA complexes into mutant *oriTs*

We decided to use the integration assay depicted in Figure 50a, i.e. with TrwC-DNA complexes formed in the donor cell, to test TrwC-mediated integration into two different *oriT* mutants. Those mutants were described in (Lucas *et al.*, 2010) and are shown in **Figure 51**:

- Mutation in TrwC binding site (mutIR): the sequence of both arms of the inverted repeat IR_2, to which TrwC binds, was changed by another sequence maintaining the secondary structure. In spite of that maintenance, this mutation was described to cause a drastic 5-log reduction of the mobilization frequency, indicating that TrwC recognizes the specific nucleotide sequence. *In vitro* TrwC binding to oligonucleotides containing this mutation is completely abolished.

- Mutation in the *nic* site (mut23-25): in this mutant, the three nucleotides 5' to the nic site are changed (TCT to GAG). This mutation almost abolishes mobilization, as this region is critical for the initial nicking reaction. *In vitro*, TrwC was reported to be unable to nick oligonucleotides containing this mutation.

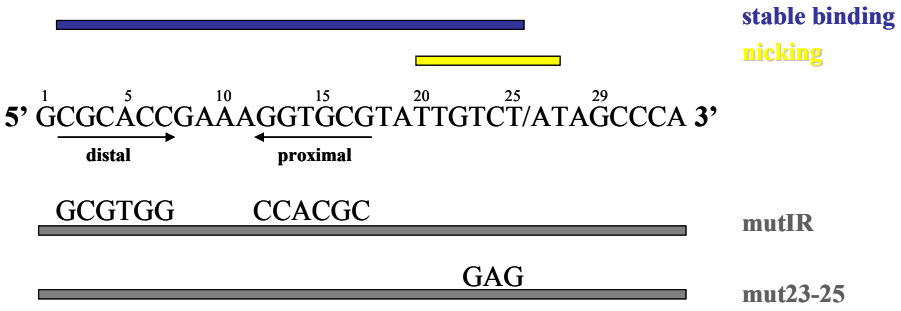

Figure 51. *oriT* mutations tested for TrwC-mediated integration, affecting IR_2 binding site and the *nic* site. The nucleotide sequence shown corresponds to coordinates 201 to 173 from (Llosa *et al.*, 1991). The arrows indicate the arms of IR_2. The *nic* site is indicated by a slash. Blue and yellow bars indicate minimal sequence requirements for TrwC stable binding and nicking (Lucas *et al.*, 2010).

The suicide plasmid pCMS11 containing R6K replication origin and both R388 and RP4 *oriTs* was mobilized by TrwC from Π1 donor strain to DH5α recipient strain harbouring the *oriT*-containing recipient plasmid where integration was tested. Different

recipient plasmids were analysed in parallel: pLA31 (carrying RP4 *oriT*), pLA32 (with R388 wild-type *oriT*), pLA33 (R388 *oriT* with mutIR mutation), and pLA34 (R388 *oriT* with mut23-25 mutation). RP4_TraI-mediated integration into the same recipient plasmids was assay in parallel as negative control. As explained in Experimental Procedures (section 4.6), integration events were selected in chloramphenicol and integration was checked by PCR and restriction analysis (**Figure 52**). For PCR analysis, primers P1, annealing to the *oriT*-containing recipient plasmid, and P2, annealing to the suicide plasmid (see Experimental Procedures, section 4.6) were used. They amplify a specific region of the cointegrate molecule with a size of 1.2 kb. As positive control, plasmid pKK::*oriT-Km* was used. It contains 826 bp of the suicide plasmid mimicking a cointegrate molecule. For restriction analysis, we used XcmI, which cuts once in the recipient plasmid.

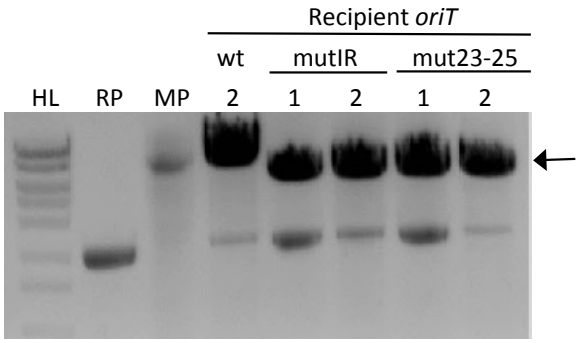

Figure 52. Restriction analysis of integrants. After checking it by PCR, the presence of the cointegrate was confirmed by XcmI restriction analysis. This restriction enzyme cuts once in the *oriT*-containing recipient plasmid. The cointegrate is indicated by an arrow. RP, recipient plasmid. MP, suicide mobilizable plasmid. HL, Hyperladder I.

Once integration was confirmed by PCR and restriction analysis, integration frequencies were calculated. **Table 15** shows the results obtained. As expected, no integration events were detected when assaying RP4_TraI. For TrwC, we found integration into all *oriTs*, in contrast to what was obtained when the suicide plasmid was mobilized by RP4_TraI and *trwC* was expressed in the recipient cell (previous section). When TrwC entered the recipient cell covalently bound to the transferred DNA strand, it was capable to catalyse integration into both mutant recipient *oriTs* and only a modest decrease in the integration frequency was observed when compared to the integration into wild-type *oriT*.

	Plasmid in the recipient cell			
		R388_*oriT*		
Relaxase	**RP4_*oriT***	wt	mutIR	mut23-25
R388_TrwC	$<3.3x10^{-8}$	$5.5x10^{-5}$	$1.3x10^{-5}$	$7.3x10^{-6}$
RP4_TraI	$<1.4x10^{-7}$	$<1.4x10^{-7}$	$<1.4x10^{-7}$	$<1.4x10^{-7}$

Table 15. Site-specific integration into mutant *oriTs* when expressing TrwC from the donor. No relaxase was expressed in the recipient cells. Frequencies are shown as integrants per donor. Data represent the mean of at least 3 independent experiments.

These surprising results suggest the existence of different DNA requirements for the initiation and termination steps of the integration reaction mediated by TrwC. **Figure 53** illustrates the different effect of *oriT* mutations when TrwC-DNA complex enters the recipient (a) or TrwC is produced in the recipient (b).

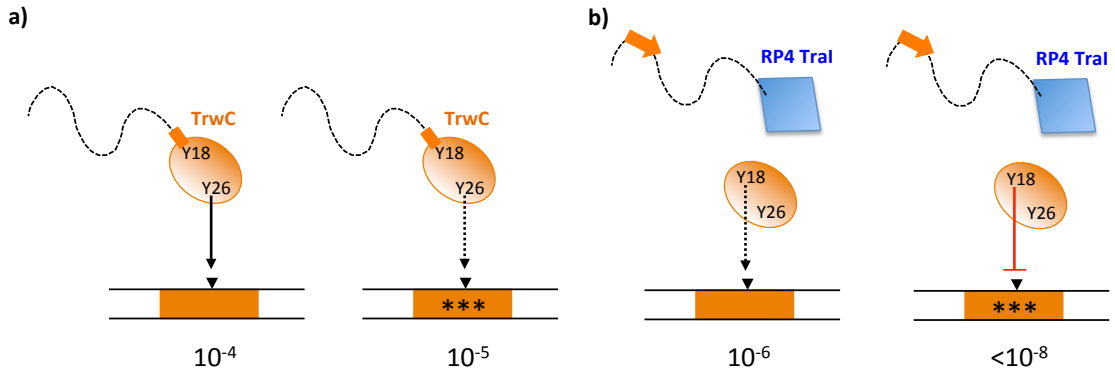

Figure 53. Different DNA requirements for initiation and termination of TrwC-mediated integration. a) When TrwC arrives covalently bound to the DNA, it is able to integrate the incoming DNA into wt (orange rectangle) and mutant (asterisks) *oriTs*. **b)** However, TrwC catalyses integration into wt *oriT* when produced in the recipient cell but it is not able to perform the initial nicking in the mutant *oriTs*. The integration frequencies are shown in the bottom. Y18, Y26, catalytic tyrosines.

In the first case, the complex is made through Y18 (the catalytic tyrosine responsible for the initial cleavage of supercoiled *oriT*-containing DNA; see Introduction, sections 3.1.1 and 3.1.3), and the free Y26 can act on the acceptor *oriT* with lower

sequence specificity. In the second situation, TrwC will find two types of *oriTs* in the cell, but the preferred substrate to perform the initial nicking reaction by Y18 will be the supercoiled substrate (Agundez *et al.*, 2012). This reaction would have higher demands in terms of sequence specificity, thus not allowing *oriT* mutations in the *oriT* present in the recipient.

Another interesting observation from these results is that they highlight the low specificity of TrwC for its target in the recipient genome, which will have implications for the use of TrwC as a tool for genomic modification.

In summary, the reactions mediated by TrwC to accomplish integration mimic those for initiation and termination of conjugation (see Introduction, sections 3.1.1 and 3.1.3), while showing substantial differences with TrwC-mediated recombination in terms of protein and DNA requirements (Introduction, section 3.1.2).

This difference probably reflects the fact that recombination is boosted by two initial nicking reactions catalysed by Y18, while integration in the recipient requires a final strand-transfer reaction mediated by Y26, which is less sequence-specific.

5. Site-specific integration mediated by TrwC*

TrwC*, which was isolated because of its higher recombinase activity and it was characterized in the first part of this work, was also tested in a site-specific integration assay in bacteria.

As it has been mentioned before, TrwC is able to catalyse the integration reaction when it is produced in the recipient cell and the suicide plasmid is mobilized by RP4_TraI. This allowed us to study the integrase activity of TrwC* expressing it in the recipient. S17.1λ*pir* carrying the mobilizable plasmid (pCMS11) was used as donor strain while DH5α carrying the *oriT*-containing plasmid (pLA31 for RP4 *oriT* and pLA32 for R388 *oriT*) plus a plasmid expressing *trwC** (pCOR17) was used as recipient strain. To test the effect of the presence of TrwA in the recipient cell, the assay was also carried out using a plasmid expressing *trwAC** (pCOR18) in the recipient cell. TrwC wild-type was analysed in parallel for comparison (using pSU1621 and pCIG1000, for –TrwA and +TrwA respectively) (**Figure 54**).

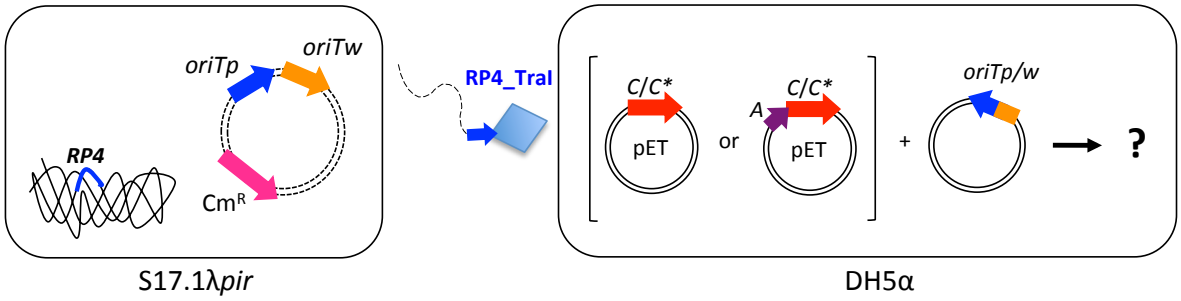

Figure 54. Site-specific integration mediated by TrwC*. TrwC$^{(*)}$ was produced in the recipient cell, in the absence or in the presence of TrwA, together with the plasmid containing the recipient *oriT*. R388 and RP4 *oriTs* were both tested. As negative control, no helper plasmid was used in the recipient. As explained before, integration events were selected in chloramphenicol. *oriTw*, R388 *oriT*. *oriTp*, RP4 *oriT*. C, TrwC. A, TrwA.

Results are shown in **Table 16**. No significant differences were found between TrwC and TrwC* under either of the conditions tested (+/- TrwA). When adding TrwA, integration frequencies increased in more than 2 times for both TrwC and TrwC*.

The increase in recombination mediated by TrwC* was found to be caused by the presence of a higher amount of protein in the cell (section 1.4). These results indicate that the amount of TrwC is not a limiting factor for the efficiency of the integration reaction, in contrast to what happens in the recombination reaction.

Recipient *oriT*	**Proteins in the recipient**				
		- TrwA		**+ TrwA**	
	None	TrwC	TrwC*	TrwC	TrwC*
R388 *oriT*	$<1.45x10^{-8}$	$4.03x10^{-7}$	$3.27x10^{-7}$	$9.21x10^{-7}$	$8.19x10^{-7}$
RP4 *oriT*	$<1.45x10^{-8}$	$<1.45x10^{-8}$	$<1.45x10^{-8}$	$<1.45x10^{-8}$	$<1.45x10^{-8}$

Table 16. Integration when expressing TrwC* in the recipient cell. TrwC* was compared with the wild-type protein with and without TrwA. Integration frequencies are shown as integrants per donor. Data represent the mean of at least 5 independent experiments.

6. Site-specific integration mediated by F_TraI

The ability to catalyse the formation of cointegrates in a site-specific way has been reported for a few relaxases (Miyazaki & van der Meer, 2011; Wang *et al.*, 2013). TrwC is, together with Mob02281 (Wang *et al.*, 2013), the only relaxase whose site-specific integrase activity in the recipient has been reported, being able to integrate an *oriT*-containing DNA into a second copy of the *oriT* present in the recipient cell after conjugative transfer (Agundez *et al.*, 2012; Draper *et al.*, 2005).

Since its site-specific recombinase activity was not detected under the assay conditions previously tested, F_TraI site-specific integrase activity had not been tested before. Our results showing the site-specific recombinase activity of F_TraI prompted us to test its site-specific integrase activity, adapting the integration assay used for TrwC. A new R6K(oriV)::$oriT_F$-$oriT_P$ suicide plasmid was constructed (pCOR28), as pCMS11 but with a copy of the F *oriT* replacing the *oriT* from R388. A recipient plasmid was also constructed (pCOR26), using the same cloning vector as for TrwC recipient plasmid (pLA32) but carrying an *oriT* from F. The assay was carried out as exemplified in **Figure 55**.

Π1 containing the suicide plasmid and pOX38 as helper was used as donor strain. The suicide plasmid was mobilized to two different recipients. Both were DH5α and only differed in the *oriT* present in the recipient plasmid they carried. F_TraI-mediated integration was tested in RP4 *oriT* at the same time as in its cognate *oriT*. R388 and RP4 systems were used as positive and negative control of integration into their respective *oriTs*. Integrants were selected in chloramphenicol, and analysed at the molecular level as described for TrwC. To detect the presence of the cointegrate, PCR was performed using primers P1 (Experimental procedures, section 4.6), annealing in the recipient plasmid, and Hind_oriTp_280 (5´-CAAAAGCTTGCGCTTTTCCGCTGCATAA-3´), annealing in the suicide plasmid.

The results obtained are shown in **Table 17**. No integration mediated by F_TraI into its cognate *oriT* was detected under the same assay conditions in which TrwC-mediated integration was detected.

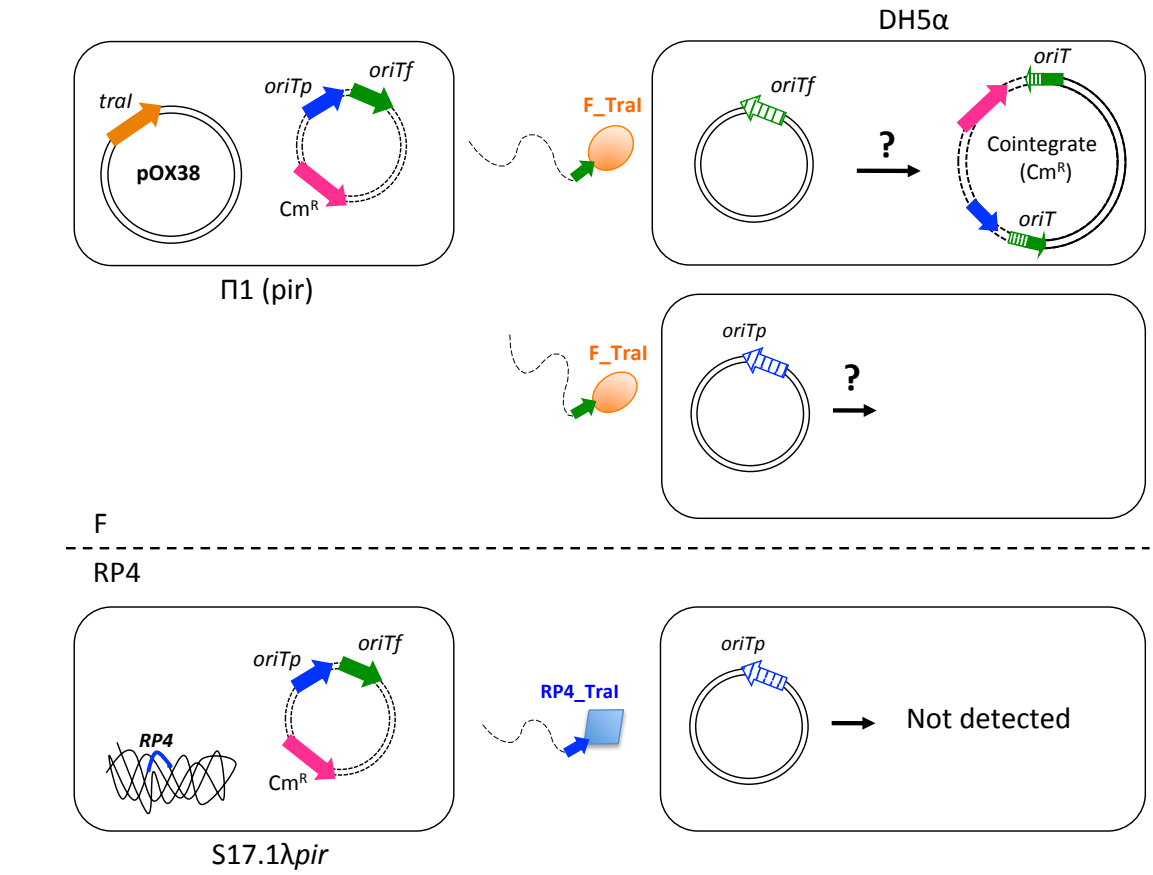

Figure 55. Scheme of TraI-mediated site-specific integration assay. The suicide plasmid was mobilized by F_TraI to DH5α carrying a recipient plasmid with either F or RP4 *oriT*. RP4-mediated integration on its cognate *oriT* was assayed in parallel as negative control. The dotted line separates the two different donors used, providing F or RP4 conjugative machinery. $oriT_P$, RP4 *oriT*. $oriT_F$, F *oriT*.

Donor	Relaxase	Recipient plasmid		
		F *oriT*	RP4 *oriT*	R388 *oriT*
Π1 (pOX38+pCOR28)	F_TraI	< $5.5x10^{-9}$	< $5.5x10^{-9}$	nt
S17.1λ*pir* (pCOR28)	RP4_TraI	nt	< $6.3x10^{-9}$	nt
Π1 (R388+pCMS11)	TrwC	nt	<$4.9x10^{-8}$	$8.1x10^{-5}$

Table 17. F_TraI mediated integration assay. Frequencies are shown as integrants per donor. Data represent the mean of at least 5 independent experiments. nt, not tested.

These results indicate that although both TrwC and F_TraI relaxases are able to catalyse site-specific recombination of two tandemly repeated *oriT* copies, there is a significant difference in their site-specific integrase activity in the recipient cell, with F_TraI lacking this activity under the assay conditions tested. The reason of this difference could reside in the differences in the catalytic mechanism of both relaxases. As mentioned before, TrwC needs two catalytic tyrosines for its role during conjugative DNA processing, whereas in the case of TraI, only one tyrosine was found to be essential for plasmid F transfer (see Introduction, sections 3.1.1 and 3.2.1, respectively).

It has to be noted that under optimal assay conditions, TrwC has an efficiency of nearly 100 % as recombinase while its higher efficiency as integrase corresponds to a frequency of $2.3x10^{-4}$ integrants per donor, obtained when expressing TrwA in the recipient cell (Agundez *et al.*, 2012). In the case of TraI, its efficiency as recombinase is 1 log lower than that of TrwC. Considering this, it could be possible that TraI had also integrase activity but we were not able to detect it under the assay conditions used. However, we do not favour this hypothesis, since the recombinase activity of TraI is roughly 1 log lower than that of TrwC under our assay conditions (Figure 46), while the integration activity, if exist, should be at least 4 logs lower than that of TrwC according to the results in Table 17.

Future experiments with the aim of trying to detect its integrase activity (if exists), could include TraY and/or TraM accessory proteins (see Introduction, section 2.3.2) in the recipient cell in an attempt to increase integration efficiency, as osbserved for TrwA in TrwC-mediated integration.

Chapter 3. Site-specific integration mediated by TrwC in eukaryotes

7. Testing R388 conjugation to yeast

The site-specific integration activity of TrwC in recipient cells could be used as a tool for genomic modification. Its use would have wider applications if TrwC were active in eukaryotic cells. With the aim of testing TrwC integrase activity in an eukaryotic cellular environment, we focused our attention in *Saccharomyces cerevisiae* as model organism. Our idea was to set up a site-specific integration assay similar to the one used for bacteria (see Experimental Procedures, section 4.6) but using the baking yeast as recipient.

A prerequisite to set up such system would be that plasmid R388 could be transferred by conjugation from *E.coli* to *S. cerevisiae*. Conjugative transfer of plasmid RP4 from bacteria to yeast was firstly described in 1989 (Heinemann & Sprague, 1989). R388 conjgation to yeast had been previously tested (Larrea, 2013), using *S. cerevisiae* L40 strain as recipient and the assay conditions originally described by Heinemann and Sprague. RP4 transfer was detected, with similar rates to the previously published ones, but no R388 conjugative transfer was detected. However, publication of a mutant *S. cerevisiae* strain especially receptive for conjugation prompted us to test again R388 transfer to yeast.

7.1. Conjugation to *ssd1* yeast mutant

Moriguchi and co-workers recently tested strains from a collection of 4,823 knock-out mutants of *S. cerevisiae* MAT-α haploids for their individual conjugative receptivity using *E. coli* as donor, in what was named a trans-kingdom conjugation (TKC) assay (Moriguchi *et al.*, 2013). Two types of mutants, an *ssd1* mutant and respiratory mutants, were identified as highly receptive mutants. Ssd1 is a RNA-binding protein with impact in many different cellular functions, including minichromosome stability, stress tolerance, and maintenance of cell wall integrity (Li *et al.*, 2009). Although that work was carried out using IncP- and IncQ-derived plasmids, we decided to test R388 transfer using these highly receptive yeast mutants as recipients.

Using an *E. coli*/yeast shuttle vector, we constructed two mobilizable plasmids, one carrying the R388 *oriT* (pCOR29) and the other one carrying the RP4 *oriT* (pCOR30). The plasmids have selectable markers for both bacteria (ampicillin resistance gene) and yeast (His3 gene). Firstly, mobilization of the plasmids between bacteria was tested using DH5α as recipient. pCOR29 was mobilized from D1210 carrying a Km-resistant R388 derivative (pSU2007), and a conjugation frequency of $3.6x10^{-1}$ transconjugants

per donor was obtained. pCOR30 was mobilized from S17.1λ*pir* and a conjugation frequency of 1.4 transconjugants per donor was obtained (frequencies are the mean of at least 3 independent experiments).

Once both *oriTs* were proved to be functional, conjugation to yeast was assayed. As donors, the same two strains as in the bacteria-to-bacteria conjugation were used. As recipient, we used *S. cerevisiae* Y13652 strain, the *ssd1* mutant described by (Moriguchi *et al.*, 2013) as highly receptive in conjugation. Conjugation frequencies (**Table 18**) were calculated as number of transconjugants per recipient and they are shown as a TKC efficiency value, also converted to Log10, to be able to compare the results with what was previously published.

Donor	**Conjugative machinery**	**TKC efficiency**	**Log(TKC_{effi})**
D1210 (pCOR29)	none	$<5.7x10^{-6}$	-
D1210 (pSU2007+pCOR29)	R388	$<5.3x10^{-6}$	-
D1210 (pCOR30)	none	$<4.3x10^{-6}$	-
S17.1λ*pir* (pCOR30)	RP4	$4.2x10^{-4}$	-3.4

Table 18. R388 and RP4 conjugation to *S. cerevisiae* Y13652. Conjugations in the absence of conjugative machinery were negative controls. TKC efficiencies are shown as number of transconjugants per recipient. Log(TKC_{effi}) represents the value of TKC efficiency converted to Log_{10}. Data represent the mean of at least 3 independent experiments.

Mobilization of pCOR30 (the RP4 *oriT*-containing plasmid) was detected, with a TCK value similar to the one previously found by Moriguchi *et al*. Under the same assay conditions, no conjugation of pCOR29 (carrying the R388 *oriT*) was detected.

7.2. Conjugation to *ssd1 petite* mutants

Yeast *petite* mutants are respiratory-deficient mutants unable to grow in respiratory proficient media containing glycerol as sole carbon source because of the absence of functional mitochondria (Chen & Clark-Walker, 2000). Moriguchi and co-workers found that *ssd1 petite* double mutants are even more receptive in conjugation (Moriguchi *et al.*, 2013), so we decided to test R388 conjugation using those double

mutants as recipients. Different spontaneous *petite* mutants were obtained as explained in Experimental Procedures (section 4.3) and assayed in conjugation (**Table 19**).

Recipient	**Donor**	**TKC efficiency**	**Log(TKC_{effi})**
Y13652	pSU2007+pCOR29	$<1.4x10^{-6}$	-
	pCOR30	$<7.8x10^{-6}$	-
	S17.1λ*pir* (pCOR30)	$5.1x10^{-4}$	-3.3
Y13652 pm clone 1	pSU2007+pCOR29	$<5.2x10^{-6}$	-
	pCOR30	$<9.3x10^{-6}$	-
	S17.1λ*pir* (pCOR30)	$1.6x10^{-3}$	-2.8
Y13652 pm clone 2	pSU2007+pCOR29	$<1.2x10^{-5}$	-
	pCOR30	$<1.1x10^{-5}$	-
	S17.1λ*pir* (pCOR30)	$4.4x10^{-4}$	-3.5
Y13652 pm clone 3	pSU2007+pCOR29	$<1.2x10^{-5}$	-
	pCOR30	$<3.5x10^{-5}$	-
	S17.1λ*pir* (pCOR30)	$9.7x10^{-4}$	-3.0
Y13652 pm clone 4	pSU2007+pCOR29	$<1.6x10^{-5}$	-
	pCOR30	$<1.2x10^{-5}$	-
	S17.1λ*pir* (pCOR30)	$1.2x10^{-3}$	-2.9

Table 19. R388 and RP4 conjugation to Y12652 *petite* mutants. Conjugation frequencies are calculated as in Table 18. When not indicated, donor strain was D1210. pm, *petite* mutant.

As expected, higher rates of RP4 conjugation were obtained with the majority of the *petite* mutants tested. However, no R388 conjugation was detected in any case.

This result suggest that plasmid R388 does not have the plasticity described for RP4 Mpf system, which presumably allows for effective transkingdom gene transfer (Bates *et al.*, 1998). In view of this result, we discarded conjugation to yeast as a strategy to test TrwC activity in eukaryotes.

8. Site-specific integration mediated by TrwC in human cells

We decided to test the TrwC-mediated integration reaction directly in human cells. Previous results on TrwC-mediated chromosomal integration (Introduction, section 3.1.3) and integration into human sequences resembling the *oriT* (Introduction, section 4.2.2.2) validated this approach.

TrwC-DNA complexes cannot be delivered to human cells by conjugation. Although there was a report on conjugative DNA transfer to mammalian cells (Waters, 2001), those results have been never reproduced (Agundez, 2011). However, a previous work in our laboratory demonstrated that DNA can be transferred to human cells through *B. henselae* T4SS using a combination of the R388 conjugative machinery and the pathogenic T4SS (Fernandez-Gonzalez *et al.*, 2011) (Introduction, section 4.1.1). In this work, a plasmid (pHP161) carrying the R388 Dtr region (Introduction, section 2.2.2) and an *egfp* cassette was mobilized. A plasmid without *trwC*, pHP181, was used as a negative control. In the absence of TrwC, GFP expression was not detected, indicating that TrwC was responsible for the DNA mobilization.

Our strategy was to add a selectable marker to that mobilizable plasmid, and carry out an antibiotic treatment after the DNA transfer experiment to select for the cells where a stable integration event in the chromosomal DNA had taken place (**Figure 56**). Those events would be later characterized to determine the integration junctions and find out if they were due to TrwC-mediated site-specific integration events.

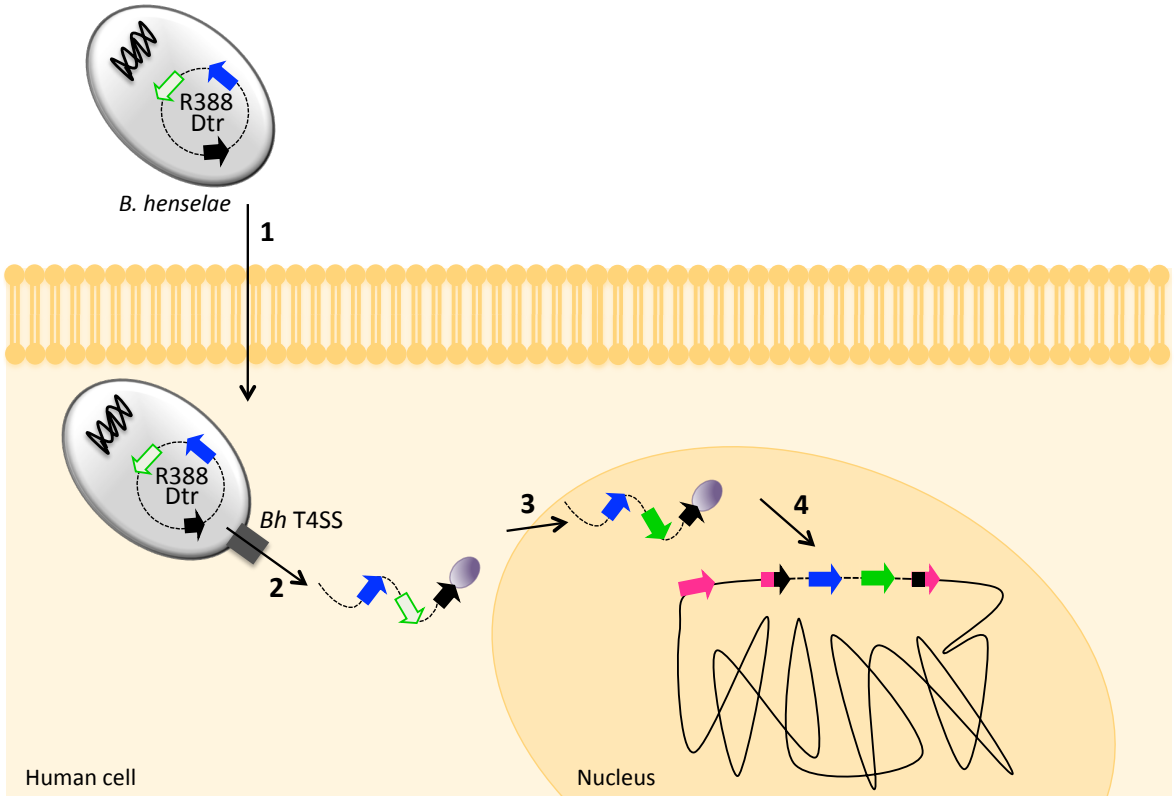

Figure 56. Analysis of TrwC integrase activity in human cells. *B. henselae* carrying a plasmid with the R388 Dtr region (with the *oriT* shown as a black arrow), a eukaryotic *egfp* cassette (in pale green), and a selectable marker (in blue) was used to infect human cells (1). Once inside, TrwC (purple sphere)-DNA complexes are translocated through the *Bartonella* VirB/D4 T4SS, represented as a grey rectangle (2). These complexes are able to reach the nucleus, since *egpf* expression (in brilliant green) is detected three days post-infection (3). TrwC-mediated integration of the plasmid DNA into potential targets of the human genome (pink arrow) would be selected by antibiotic treatment. *Bh, B. henselae*.

8.1. Construction of mobilizable plasmids

The mobilizable plasmid used by (Fernandez-Gonzalez *et al.*, 2011), pHP161, contains the R388 Dtr region, but not the genes of the T4SS, and a eukaryotic *egfp* cassette. With the aim of analysing the integration pattern of the transferred DNA after *Bartonella* infection of human cells, we added a neomycin phosphotransferase expression cassette to be able to select for stable chromosomal integration events. This *neo* cassette was added to mobilizable plasmids coding for TrwC and TrwC:BID fusion relaxases and to the negative control lacking *trwC*. **Figure 57** shows a schematic representation of the plasmids constructed.

DNA transfer through *B. henselae* VirB/D4 T4SS was also described by Schröder and co-workers (Introduction, section 4.1.1). In their work, the mobilizable plasmid was a derivative of the *Bartonella*-specific cryptic plasmid pBGR1, which encodes a relaxase (Mob) and the respective *oriT*, plus the *gfp* cassette. Upon addition of a BID domain to the relaxase (to obtain plasmid pRS130), transfer rate reached similar levels to TrwC-driven DNA transfer. After selection of a few stable integrants, mapping of chromosomal junctions indicated that the Mob relaxase could be protecting the 5′ end of the transferred DNA during the transformation process (Schroder *et al.*, 2011). Plasmid pRS130 was used in our experiments as an additional control.

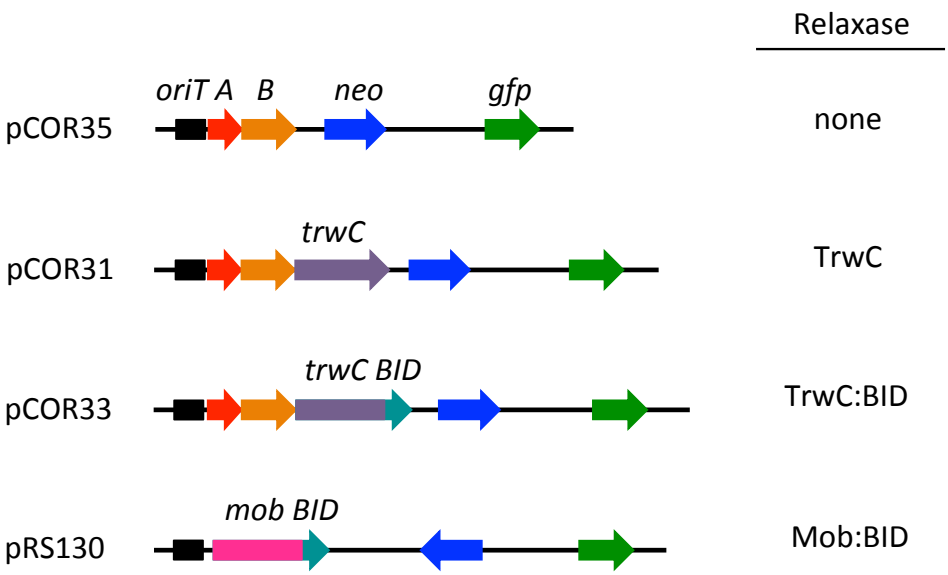

Figure 57. Mobilizable plasmids for selection of stable integration events. pCOR31/33/35 were constructed by adding a neomycin resistance gene to the plasmids previously used for DNA transfer assays. pRS130 was kindly provided by Christoph Dehio (University of Basel, Switzerland).

Prior to carrying out the selection, the functionality of the neomycin resistance cassette was tested by transfection of HeLa cells with plasmid DNA. After 3-4 weeks of selection, G418 resistant colonies were perfectly visible for cells transfected with pCOR31, pCOR33, pCOR35 and pRS130, but no surviving cells were found after transfection with pHP159, pLA24 and pHP181, the mobilizable plasmids to which the *neo* cassette was added (see Experimental Procedures, Table 2). This confirmed that the plasmids we constructed could be used for selecting integration events with G418 treatment.

8.2. Selection of target cell lines

Some human sequences resembling the minimal *oriT* of 17 bp have been proved to be functional as acceptors for site-specific integration mediated by TrwC in bacteria (see Introduction, section 4.2.2.2), so our purpose was to detect TrwC-mediated integration into these natural targets. However, the full-length *oriT* is likely a more effective target for TrwC-mediated integration. For this reason, we decided to integrate the R388 *oriT* in a human cell line and collect the pool of integrants. As we planned to characterize the integration pattern mediated by TrwC in the human genome, we thought that it would be interesting to compare both integration patterns in the presence/absence of the *oriT*, and see if the presence of the natural target produces any significant changes in it.

As vascular endothelial cells are natural host cells for *B. henselae*, EA.hy926 cell line has always been used for the infection experiments previously done in our laboratory. For that reason, we wanted to modify EA.hy926 cells to carry a copy of the *oriT* in the chromosome. Plasmids pTRE2hyg::*oriT* (pMTX708 and pMTX709; Table 2) were used with that purpose. They only differ in the orientation of the Ptac-*oriT* cassette cloned into a vector coding for hygromycin resistance for selection in human cells. Ea.hy926 cells are defined as difficult-to-transfect (Hunt *et al.*, 2010; Truttmann *et al.*, 2011) and a very low transient transfection efficiency is obtained when transfecting them using JetPei, the transfection reagent used in the laboratory for routine transfections. We were unable to obtain stable transfected EA.hy926 when selecting with hygromycin. Other different transfection reagents as well as electroporation were tried (see Experimental Procedures, sections 5.2 and 5.3) but all cells died after several days of antibiotic treatment.

Since we were unable to obtain an EA.hy926::*oriT* cell line, we decided to use a cell line easier to transfect, such as HeLa cells.

A previous requirement was to assay *Bartonella*-mediated DNA transfer to this cell line. Infection of HeLa cell by *B. henselae* had been previously reported (Truttmann *et al.*, 2011). In that work, a 50 % of invasome-positive cells was found after infection. However, no DNA transfer has been tested to this cells line, and it could happen that infection efficiency was not high enough to detect eGFP positive cells. EA.hy926 and HeLa cells were infected in parallel, both with plasmids pHP161 (*oriT*+*trwABC*+*egfp*) and pHP181 (*oriT*+*trwAB*+*egfp*). After three days, DNA transfer was tested by flow cytometry (**Figure 58**).

a)

Cell line	Mobilizable plasmid	
	oriT+trwABC	*oriT+trwAB*
EA.hy926	1.54	0
HeLa	0.43	0

b)

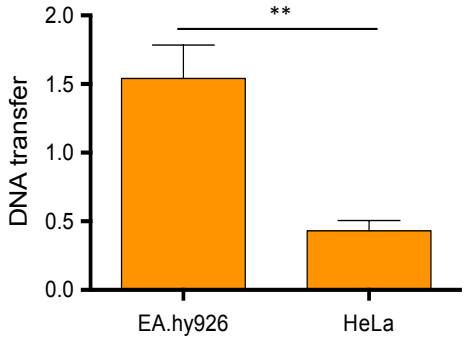

Figure 58. DNA transfer to EA.hy926 and HeLa cells. a) DNA transfer values indicated are the percentage of GFP positive cells detected after 3 days of infection. **b)** DNA transfer of pHP161 (*oriT+trwABC*) to both cell lines. Data are the mean of at least 5 independent experiments. **, $p<0.01$.

Since HeLa cell line is not a natural host for *B. henselae*, a lower infection rate and so, a lower DNA transfer rate were expected. In fact, the DNA transfer was found to be lower than when using EA.hy926. However, DNA transfer to HeLa cells still occurred efficiently.

HeLa cells were transfected with plasmids pTRE2hyg::*oriT* and stable transfectants were selected by hygromycin treatment. Both plasmids were used to avoid any possible influence of the orientation of the *oriT* cloned downstream the P_{minCMV} promoter, since transcription through the *oriT* has been shown to affect recombination (César & Llosa, 2007), and so it could affect integration. We transfected both plasmids separately, and hygromycin-resistant colonies appeared equally for the two of them. The resistant colonies obtained in both transfections were pooled together to establish a HeLa::*oriT* cell line, in which the *oriT* is supposed to be located in many different chromosomal locations and in the two possible orientation with respect to the promoter. In this way, we do not risk selecting a particular clone in which the *oriT* copy may lie in a region which may be particularly inaccessible for TrwC or in an orientation that could affect integration of the mobilizable plasmid.

The presence of the *oriT* was tested by PCR on genomic DNA samples from HeLa and HeLa::*oriT* cells using primers oriT1 (5'-CTCATTTTCTGCATCATCA-3') and oriT330 (5´-CCTCTCCCGTAGTGTTA-3') that amplify the 330 bp band of the *oriT* (**Figure 59**). Water and pMTX708 plasmid DNA were used as negative and positive controls, respectively. As seen on the gel, only one band corresponding to the *oriT* was present in the sample obtained from HeLa::*oriT* cells, while no amplification was detected in the sample obtained from unmodified cells.

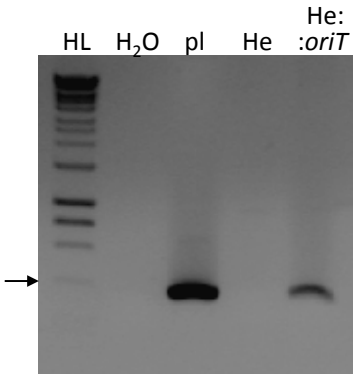

Figure 59. PCR amplification of R388 *oriT*. Genomic DNA was extracted from unmodified HeLa and from Hyg^R HeLa::*oriT* cells. The arrow indicates the 400 bp band of the molecular weight marker. HL, Hyperladder. pl, pMTX708 plasmid DNA. He, unmodified HeLa cells. He::*oriT*, pool of HeLa cells transfected with pTRE2hyg::*oriT*.

8.3. Selection of stable integrants

EA.hy926, HeLa and HeLa::*oriT* cell lines were used for infection with *B. henselae* carrying pCOR31 (*trwC*), pCOR33 (*trwC:BID*), pCOR35 (*ΔtrwC*) or pRS130 (*mob:BID*) plasmids (see Figure 57). Three days post infection DNA transfer was measured by flow cytometry of the cell populations. The results (**Figure 60**) indicate that DNA transfer occurred to the three different cell lines when a relaxase (TrwC, TrwC:BID, or Mob:BID) was coded in the plasmid, while no DNA transfer was detected when there was not relaxase. This confirms that in all cases the DNA transfer took place by a conjugation-related mechanism, as previously described.

As shown in the previous section, the transfer efficiency is higher when using EA.hy926 as host cell, but HeLa and HeLa::*oriT* were sufficiently good recipients for *Bartonella* infection so that gene transfer and expression of the eukaryote-specific *egfp* reporter cassette could be detected.

As usual, TrwC:BID-mediated DNA transfer rate was higher than that of TrwC. Mob:BID-mediated DNA transfer to EA.hy926 cells was previously reported to be 10 times higher than what we obtained (Schroder *et al.*, 2011). As the plasmid we used was the same used in that study, we attribute this difference to the assay conditions in which the experiments were carried out.

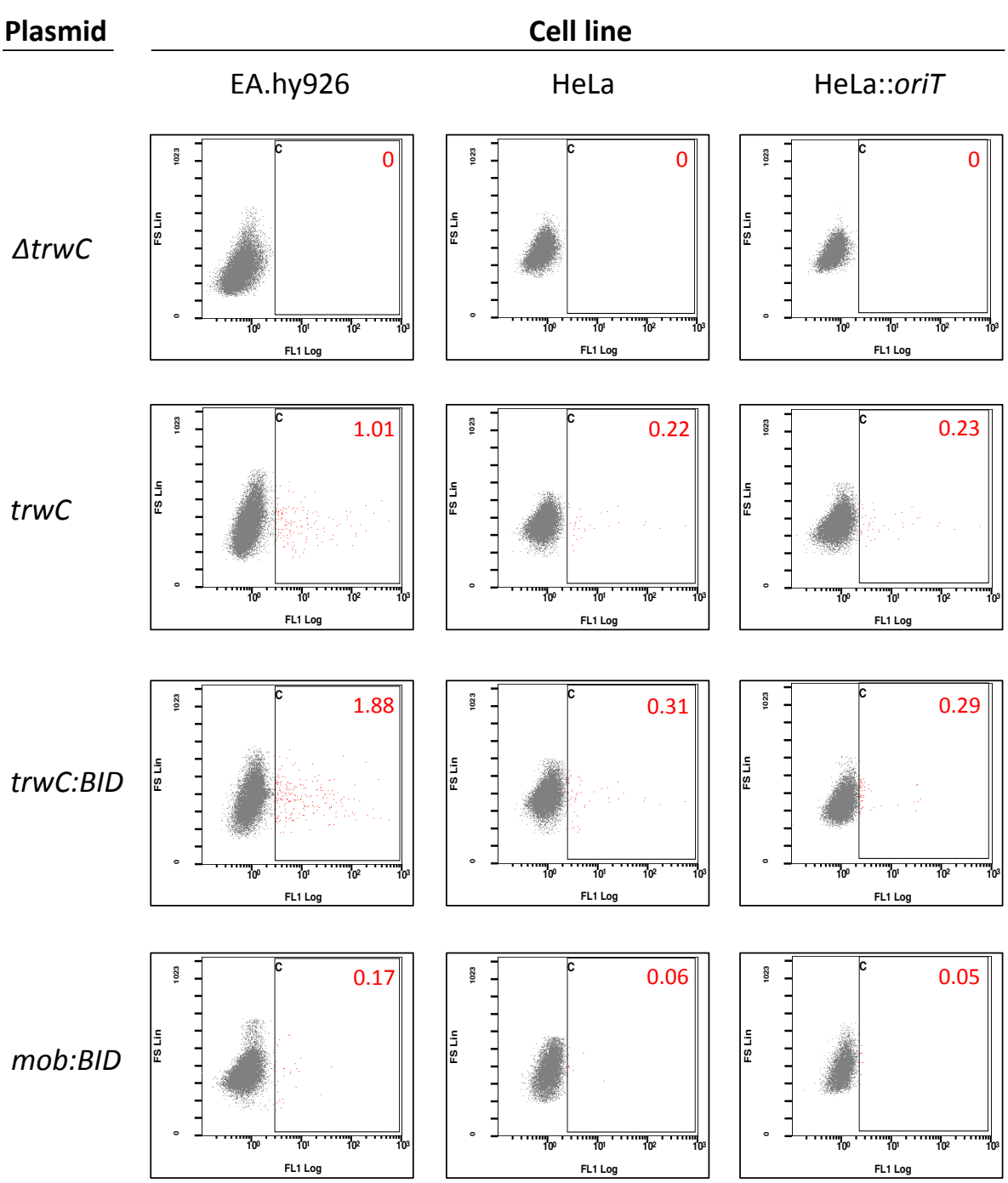

Figure 60. *Bartonella* mediated DNA transfer to different cell lines. Each row shows the DNA transfer of one plasmid (which is indicated by the relaxase gene) to the three different cell lines used. The percentage of GFP positive cells is indicated in red in each plot. These are the results of one experiment, representative of at least 5 independent experiments.

For the three cell lines, DNA transfer rate was 0.5-1 log lower when using *mob:BID* plasmid compared to *trwC* or *trwC:BID*. This could be indicating that TrwC, even in the absence of a BID domain, is a better substrate for the *B. henselae* VirB/D4 T4SS than Mob. Another possibility is that, once in the human cell, TrwC somehow contributes to the entrance into the nucleus or the expression of the exogenous DNA. Regarding the entrance into the nucleus, TrwC has been reported to have cytoplasmic localization (Agundez *et al.*, 2011) while a passive entry of the Mob-guided DNA has been suggested (Schroder *et al.*, 2011), so none of the relaxases is expected to have an active role in nuclear import. It would be necessary to set up an assay to detect translocation of the relaxase (rather than detection of DNA transfer and expression) in order to distinguish between these possibilities (higher relaxase recruitment versus better expression of transferred DNA). Fluorescence-based β-lactamase, calmodulin-dependent adenylate cyclase (Chen *et al.*, 2010), or Cre recombinase assays are commonly used with that purpose. This last assay has been used to study the T4SS-mediated translocation of, for example, *Bartonella* effector proteins (Schulein *et al.*, 2005).

Another conclusion we can infer from the results is that no significant difference in TrwC-mediated DNA transfer was found when using HeLa and HeLa::*oriT*. This is not surprising, since we are detecting transient expression of transferred DNA.

After confirming that DNA transfer and expression of the mobilizable plasmid had taken place, antibiotic treatment with G418 was carried out during 4-5 weeks to select for stable integration of the transferred plasmids into the human genome. After selection, the number of G418-resistant cell colonies was calculated by counting using an optical microscope. Approximately 10^3 resistant colonies were obtained after mobilization of *trwC*- and *trwC:BID*-carrying plasmids. Surprisingly, no resistant colonies were obtained after mobilization of *mob:BID* plasmid. This will be discussed later.

In order to check if G418-resistant colonies represented the expected integration events, resistant cells from each experiment were pooled together and genomic DNA (gDNA) was extracted from the different pools. gDNA preparations were later analysed by PCR for the presence of *trwA* and *trwC* genes. For *trwA* amplification, we used primers Hind3_TrwA_F (5'-AACAAAGCTTATGGCACTAGGCGACCCC-3') and BamHI_TrwA_R (5´-AACAGGATCCTCAATCCTCCTTCCCCTCCC-3´), which amplify a 400 bp band. For *trwC*, primers Hind3_TrwC_F (5´-AACAAAGCTTATGCTCAGTCACATGGTATT-3´) and 670_TrwC (5´-TGTGTGCTAGGTCGAA-3´), that amplify a 680 bp band, were used. As positive control, pHP161 plasmid DNA was included. **Figure 61** shows the results of PCR reactions with the expected amplification products in each case.

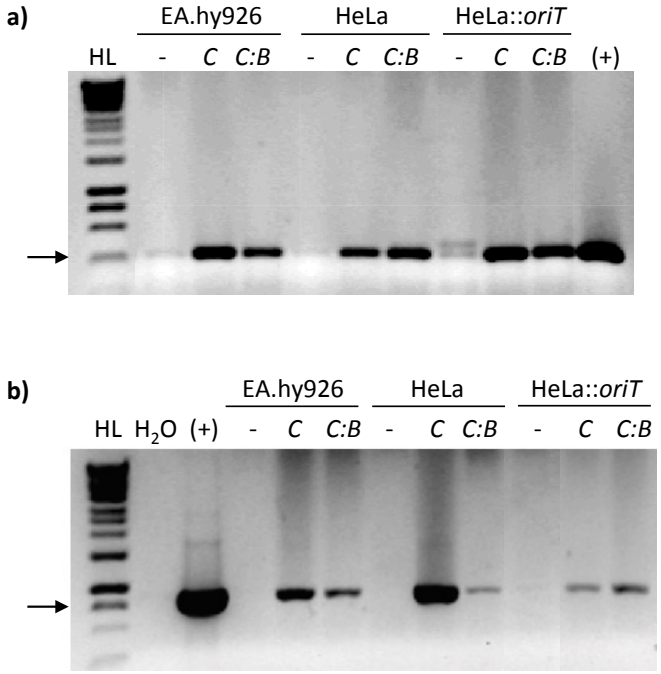

Figure 61. PCR amplification of TrwA (a) and TrwC (b) genes. gDNA was extracted from the G418-resistant cell pools. The arrows indicate the band of 400 bp in a) and the one of 800 bp in b). C, cell pools obtained after mobilization of *trwC* plasmid. C:B, pools obtained after mobilization of *trwC:BID* plasmid. -, unmodified cells. HL, Hyperladder I. (+), pHP161 plasmid DNA.

The pools were also analysed by flow cytometry to detect eGFP expression as another evidence of integration of the mobilizable plasmids. As negative controls, the three unmodified cell lines were used. Results are shown in **Figure 62**. Although eGFP expression is clearly detected, the percentage of GFP positive cells (10-15 %) was lower than expected. This could be due to integration events into silenced regions of the genome, and/or to integration of plasmid fragments (containing the neomycin resistance gene but not the *egfp* cassette).

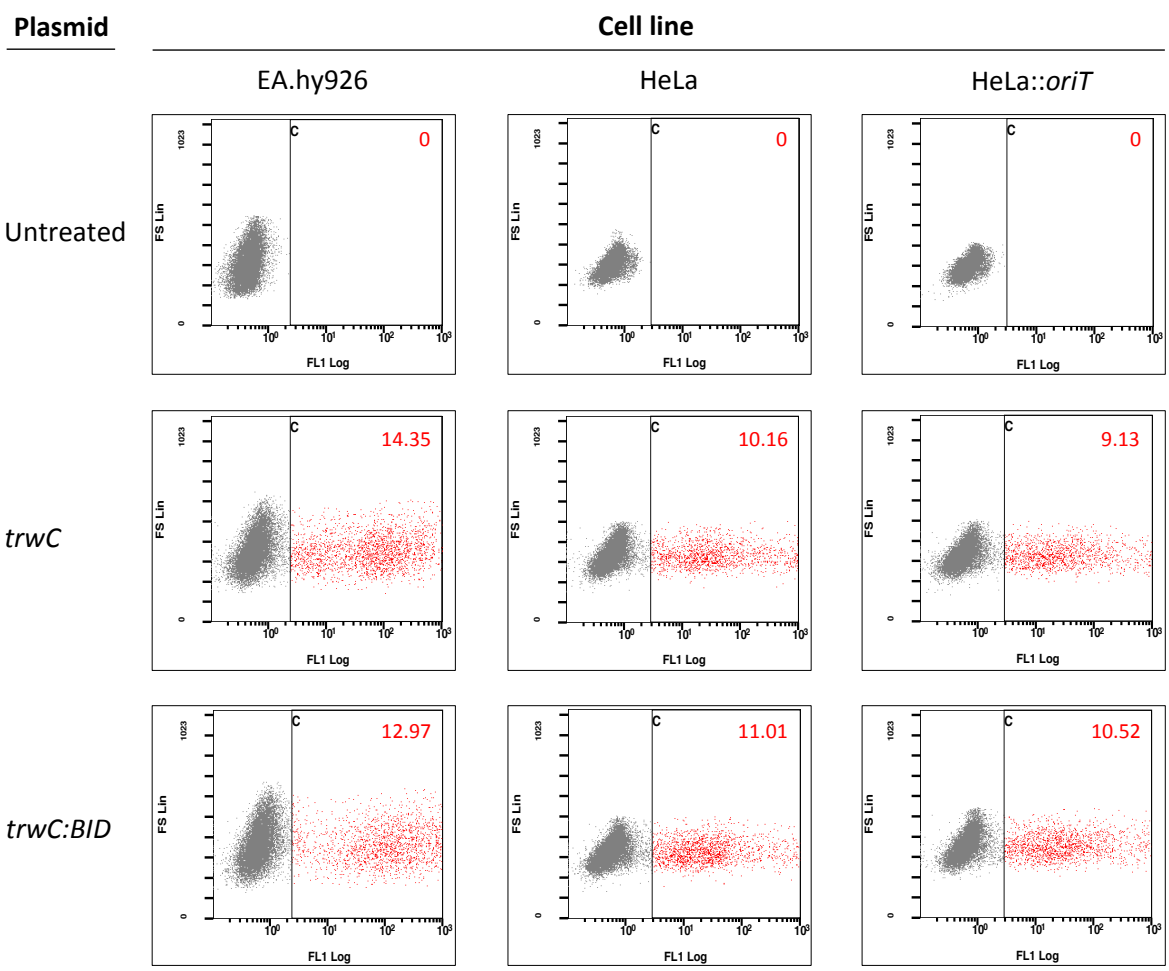

Figure 62. GFP expression of G-418-resistant cell pools. Pools obtained after integration of both *trwC*- and *trwC:BID*-coding plasmids were analysed by flow cytometry using uninfected cells as control. The percentage of GFP positive cells is indicated in red in each plot. Data represent the results of one experiment but similar results were obtained in another independent experiment.

Table 20 shows the number of eGFP positive cells detected three days post infection, as well as the number of surviving cell colonies after the antibiotic treatment. As the dishes where infections were done were confluent at the moment of the flow cytometry, a theoretical number of cells was estimated (see the following web page for reference: http://www.lifetechnologies.com/es/en/home/references/gibco-cell-culture-basics/cell-culture-protocols/cell-culture-useful-numbers.html). The most relevant results of Table 20 are graphically shown in **Figure 63** to facilitate interpretation.

Cell line	Relaxase in mobilizable plasmid			
	None	TrwC	TrwC:BID	Mob:BID
EAhy926				
Cells	$2.0x10^{7}$	$2.0x10^{7}$	$2.0x10^{7}$	$2.0x10^{7}$
% GFP^{+}	0	1.01	1.88	0.17
GFP^{+}	0	$2.0x10^{5}$	$3.8x10^{5}$	$3.4x10^{4}$
Neo^{R}	0	$1.1x10^{3}$	$1.6x10^{3}$	0
Neo^{R}/cells	$<5x10^{-8}$	$5.5x10^{-5}$	$8.0x10^{-5}$	$<5x10^{-8}$
Neo^{R}/GFP^{+}	-	$5.5x10^{-3}$	$4.2x10^{-3}$	$<2.9x10^{-5}$
HeLa				
Cells	$2.0x10^{7}$	$2.0x10^{7}$	$2.0x10^{7}$	$2.0x10^{7}$
% GFP^{+}	0	0.22	0.31	0.06
GFP^{+}	0	$4.4x10^{4}$	$6.2x10^{4}$	$8.0x10^{3}$
Neo^{R}	0	$2.5x10^{3}$	$3.7x10^{3}$	0
Neo^{R}/cells	$<5x10^{-8}$	$1.3x10^{-4}$	$1.9x10^{-4}$	$<5x10^{-8}$
Neo^{R}/GFP^{+}	-	$5.7x10^{-2}$	$5.9x10^{-2}$	$<1.3x10^{-4}$
HeLa::*oriT*				
Cells	$2.0x10^{7}$	$2.0x10^{7}$	$2.0x10^{7}$	$2.0x10^{7}$
% GFP^{+}	0	0.23	0.29	0.05
GFP^{+}	0	$4.6x10^{4}$	$5.8x10^{4}$	$1.0x10^{4}$
Neo^{R}	0	$2.8x10^{3}$	$4.2x10^{3}$	0
Neo^{R}/cells	$<5x10^{-8}$	$1.4x10^{-4}$	$2.1x10^{-4}$	$<5x10^{-8}$
Neo^{R}/GFP^{+}	-	$6.1x10^{-2}$	$7.2x10^{-2}$	$<1.0x10^{-4}$

Table 20. GFP positive and G418-resistant cells. Cells: approximate total number of cells in a 150 mm dish in confluent growth conditions. % GFP: percentage of eGFP-positive cells, obtained by flow cytometry. GFP^{+}: number of eGFP-posivie cells, calculated in base of the theoretical total number of cells. Neo^{R}: approximate number of G418-resistant colonies (counted by microscopy). Data represents the results of one experiment but similar results were obtained in another independent experiment.

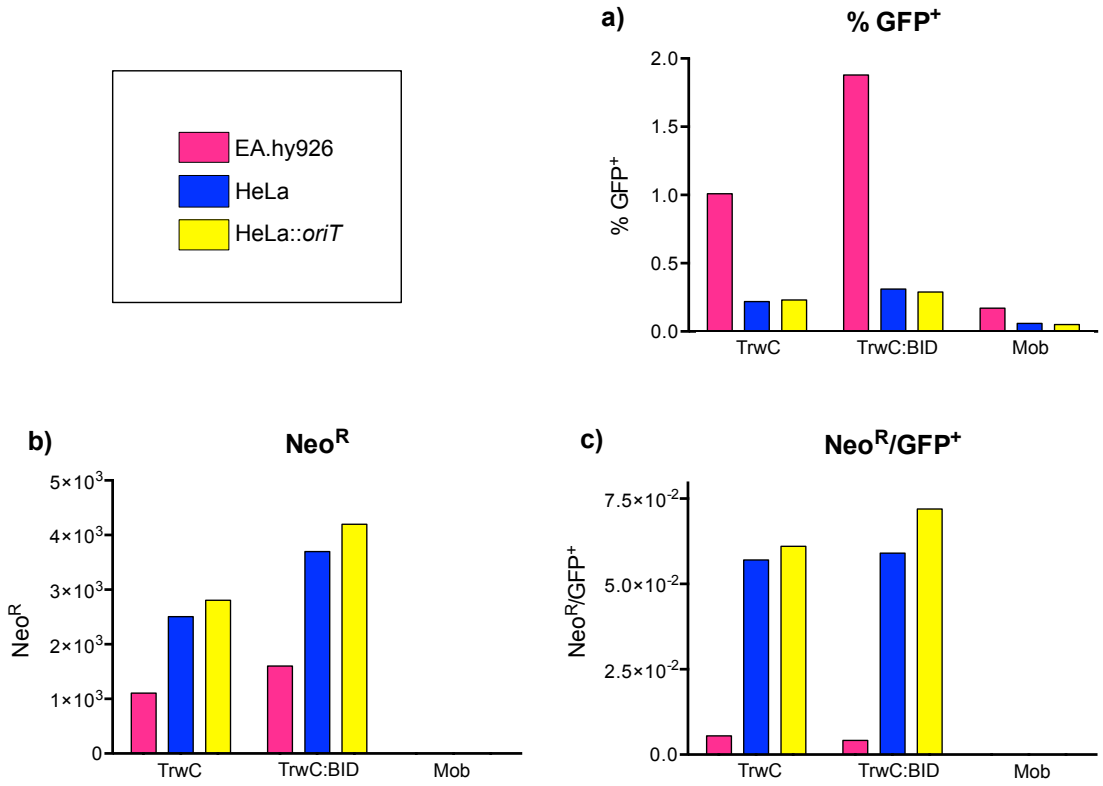

Figure 63. GFP-positive and G418-resistant cells. Graphical representation of the percentage of GFP-positive cells and the number of G418-resistant colonies, as well as the Neo^R/cells and Neo^R/GFP^+ ratios.

When comparing the results in the different cell lines, it was found that a similar number of integrants was obtained in all cases (Figure 63b) despite the fact that DNA transfer rate was at least 10 times lower in HeLa or HeLa::*oriT* cells (Figure 63a). A possible explanation for this observation is that the viability of EA.hy926 cells in lower than that of HeLa cells, so some EA.hy926 integrants could have died without developing a resistant colony while almost each HeLa integrant would have formed a G418-resistant colony. Another possibility is that the higher cell-division rate of HeLa cells could facilitate the integration, since *gfp* expression was found to depend on the disassembly of the nuclear envelope during cell division (Schroder *et al.*, 2011), presumably by granting access of the T-DNA to the nuclear genome.

After the antibiotic treatment, cell colonies were visible in the three cell lines infected with *Bartonella* carrying the plasmids coding for TrwC or TrwC:BID, while no single cell survived the selection when using *ΔtrwC* plasmid. This is in concordance with the flow cytometry results, which showed no DNA transfer in the absence of relaxase. No significant differences were found in the integration rate of transferred DNA (Neo^R/GFP^+

ratio) obtained for plasmids coding for TrwC or TrwC:BID (Figure 63c). Surprisingly, no resistant colonies were found in infections with plasmids coding for Mob:BID in any of the cell lines. It was an unexpected finding since formation of G418-resistant colonies after mobilization of pRS130 (*mob:BID* plasmid) had been already reported (Schroder *et al.*, 2011). However, although they had a 10 times higher DNA transfer rate, only a few cell lines were established. Our transfer rate could be too low to be able to obtain stable integrants.

This difference in the number of stable integrants between TrwC/TrwC:BID and Mob:BID is significant. If we look at the Neo^R/GFP^+ rate in Table 20, which represents the rate of integration of the transferred DNA, it can be observed in HeLa cells that about 6 % of the DNA molecules transferred by TrwC or TrwC-BID are integrated, while this integration is less that 1 in 10,000 in the case of Mob:BID-driven DNA. Based on these results, we conclude that TrwC facilitates the integration of the mobilizable plasmids, while Mob does not.

8.4. Characterization of integration sites

In order to determine if the integrants obtained were site-specific, we characterized the integration pattern in the cell pools obtained with TrwC and TrwC:BID.

With the purpose of having a control of a random integration pattern, HeLa and HeLa::*oriT* cells were transfected with plasmids pCOR31 and pCOR33, and random integration events were selected with G418 as previously done. This was not done using EA.hy926 due to the impossibility to obtain stable transfection using those cells. After several weeks of selection, antibiotic resistant colonies were pool together. These pools were analysed in parallel with the ones obtained after TrwC-mediated mobilization of plasmid DNA, to compare both plasmid integration patterns. This control was used to check if integration at the *nic* site could occur in the absence of TrwC.

For the characterization of the integration sites in the cell pools, linear amplification mediated PCR (LAM-PCR) technique was chosen. It is a technique that allows the characterization of unknown flanking sequences adjacent to known DNA. It was initially developed to identify viral vector integration sites (IS) within the host genome (Schmidt *et al.*, 2007) and, since then, it has been used in many different integrome analysis projects. We used LAM-PCR for the characterization of the integration sites of events that occurred at the *nic* site in the pools of G418-resistant cells. This characterization was performed during a short-term stay done in the laboratory of Dr. Manfred Schmidt, in the German Cancer Research Center (Heidelberg, Germany).

Since LAM-PCR was carried out to characterize genomic DNA-*nic* junctions, it would not permit the identification of insertions happening in the full-length *oriT* copy of HeLa::*oriT*, as integration there would result in two copies of the *oriT*. However, it would permit the identification of integration events in other chromosomal locations and its comparison with the integration pattern obtained when the *oriT* was not present in the genome to be modified.

LAM-PCR protocol is described in detail in Experimental Procedures (section 3.5). Briefly, it consists on an initial linear amplification of genome-plasmid junctions with a plasmid-specific primer. After synthesis of dsDNA, the PCR product is digested and a linker cassette of known sequence is ligated. Exponential PCR amplification is then performed with plasmid- and linker-specific primers. PCR-obtained bands are then analysed by gel electrophoresis and high-throughput sequencing. Genome-plasmid junctions are sequenced and bioinformatic analysis is carried out to obtain the integration sites.

Primers were designed as described in Experimental procedures. To select the plasmid region for annealing of plasmid-specific primers we focused in the 5' region to the *nic* site, as this would be the plasmid region lying nearest to the human DNA after TrwC-mediated integration (**Figure 64**).

LAM-PCR protocol was carried out using two different enzymes for the restriction digest step, BfaI and Tsp509I, none of them cutting in the plasmid region between the sequence where oriTI primer anneals and the junction with the genomic DNA (**Figure 65**). It was done in this way with the aim of eliminating the restriction bias resulting of using just one enzyme. In that case, if the nearest restriction site were too close to the junction, the resulting fragment would not be mappable. On the other hand, if the restriction site were too far from the junction, the amplification would be inefficient.

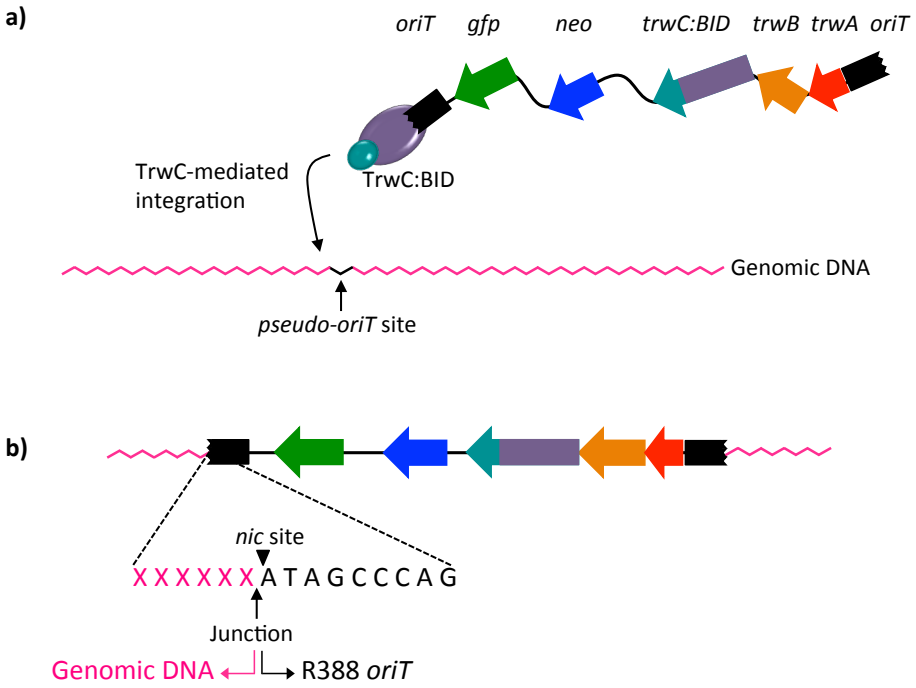

Figure 64. Scheme of the expected integration events of pCOR33. a) TrwC is mobilized through the *Bartonella* T4SS covalently bound to the 5' end of the *nic* site; this region would be integrated when TrwC found a *pseudo-oriT* target in the human genome (in pink). **b)** After that integration event, the 5' region to the *nic* site is expected to be part of the junction with the human genomic DNA. For pCOR31, the cointegrate molecule would be the same but without BID sequence.

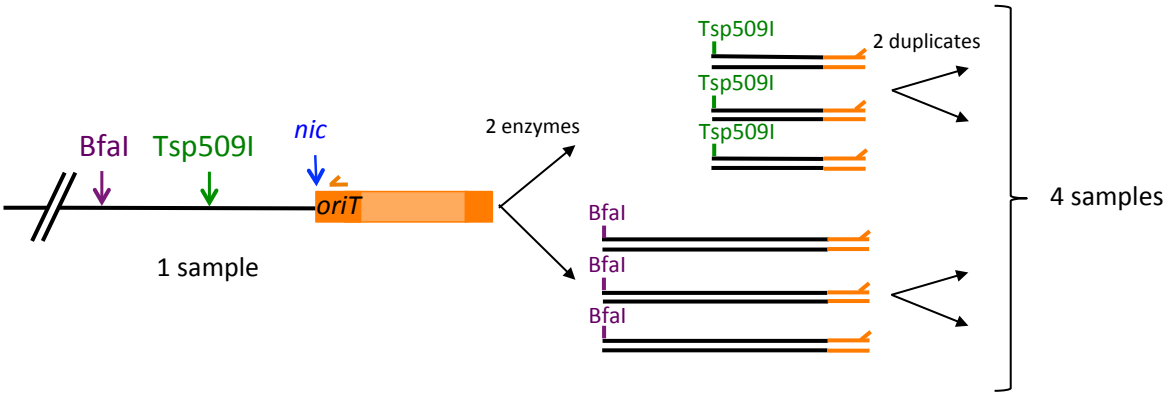

Figure 65. Restriction enzymes used during LAM-PCR protocol. Both enzymes recognize a commonly found 4 bp target site. Each sample was digested in parallel with both enzymes and each digestion was done in duplicate.

Once digestion, linker ligation and exponential amplification were carried out, LAM-PCR reactions were first checked by electrophoresis in agarose gels. Two of these gels are shown in **Figure 66**. The results obtained in these two gels are representative of the ones obtained for all the samples analysed, including the controls obtained after transfection of plasmid DNA. As explained in Experimental Procedures (section 3.5, Figure 36) this gives a first idea of the integration pattern. In these gels we expect to see only one band if integration occurs in a single target in the genome, or several bands if there are different integration sites.

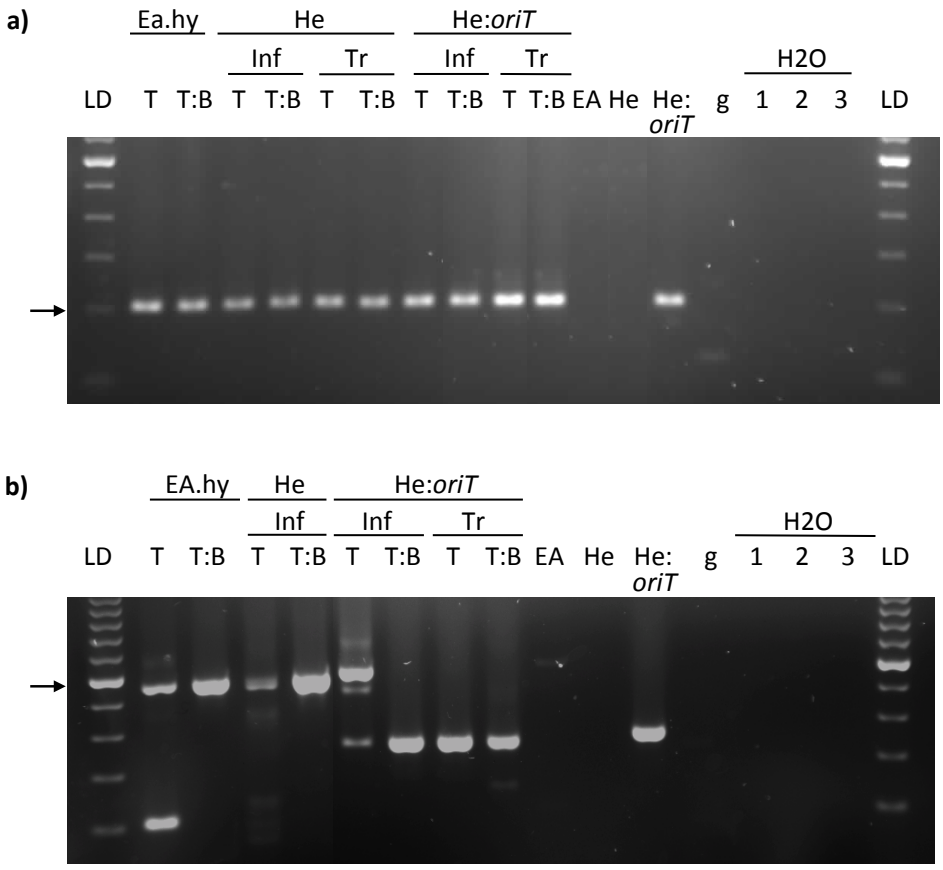

Figure 66. Gel electrophoresis of LAM-PCR products. LAM-PCR products obtained when using **a)** BfaI and **b)** Tsp509I restriction enzymes. Inf, samples obtained after *Bartonella* infection. Tr, samples obtained by transfection of plasmid DNA. T, *trwC*-coding plasmid (pCOR31). T:B, *trwC:BID*-coding plasmid (pCOR33). EA, He, and He::*oriT*, samples from uninfected cell lines (Ea.hy926, HeLa, and HeLa::oriT, respectively). g, Human genomic DNA (Roche), used as negative control. H2O (1,2,3) water controls of linear, first, and second exponential PCRs, respectively. LD, 100 bp ladder. The arrows indicate the band of 200 bp in a) and the band of 600 bp in b).

The first observation is that the controls are correct: water samples lines were always empty, as were the genomic DNA samples from untreated EA.hy926 and HeLa cells. In contrast, a band was always visible in untreated HeLa::*oriT* cells due to the presence of the R388 *oriT* in these cells. For constructing HeLa::*oriT* cell line, pTRE2hyg::*oriT* plasmids DNA was transfected. As their integration for generating the cell line occurred randomly, each molecule will have become integrated through a different plasmid site. For that reason, the vast majority of integrated plasmid molecules will have an intact full length *oriT*, and so a predominant band is obtained after LAM-PCR. The size of that band is determined by the distance between the sequence where primer oriTI anneals and the nearest recognition sequence for the specific enzyme in pCOR31/33 sequence. That band was expected to be of 220 bp when using BfaI (that cuts in the *oriT*) and of 345 bp when using Tsp509I (that cuts in pTRE2hyg::*oriT* but downstream the *oriT*). As seen on the gels, that was exactly what we found.

As for the pattern observed in the rest of the samples, obtained either after plasmid transfection or after translocation of TrwC-DNA complexes through *B. henselae* T4SS, to our surprise they all looked monoclonal, with a predominant band in all of them. This was foreseeable for the controls obtained after transfection of pCOR31 and pCOR33, for the same reason explained for HeLa::*oriT* cell line. However, different bands were expected in the samples obtained after *Bartonella* infections, each one resulting from an integration event in any potential target site. It could happen that a single target site was being recognized by TrwC in the human genome.

However, the band that appeared in those samples was found to have the same size as the *oriT*-derived band present in the untreated HeLa::*oriT* sample when using BfaI, indicating that a non-nicked *oriT* was part of the cointegrate. For Tsp509I samples, the size of the band obtained is also in agreement with that hypothesis, since the band expected in that situation would be 580 bp in size (Tsp509I cuts in *trwA*) (**Figure 67**).

These results strongly suggested that the transferred DNA had not become integrated at the *nic* site at the *oriT* (where TrwC remains covalently bound once the DNA is cut). Rather, this band presumably reflected random integration events. We could not discard however the presence of a main target site lying at that distance. Also, low-frequence site-specific integration events could be masked by this main band, and it would be interesting to analyse those. So, despite these findings, LAM-PCR resulting bands were analysed by high-throughput sequencing with primers Mis-TrwC and Mis-LC as described in Experimental Procedures (section 3.5).

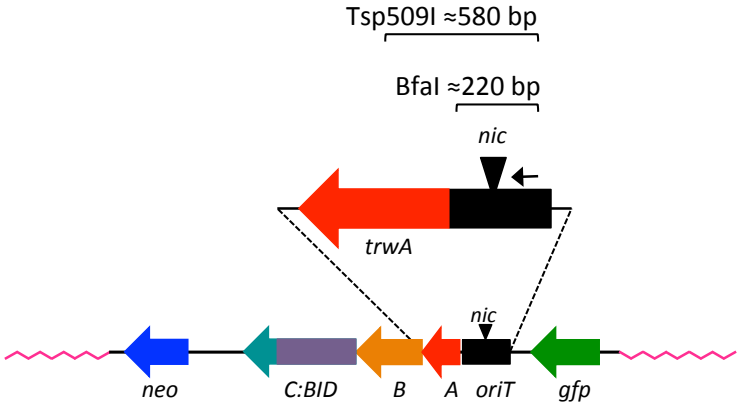

Figure 67. Structure of a cointegrate molecule with intact full *oriT* + *trwA*. In this example, integration would have occurred at a nucleotide between *neo* and *gfp* genes. Primer oriTI annealing sequence is shown by an arrow. *trw* has been omitted from *trwA/B/C* names for clarity.

Sequencing results were analysed as described in (Paruzynski *et al.*, 2010). Most of the 2,000,000 reads obtained were found to be plasmid DNA, as expected after the visualization of the samples in agarose gels, suggesting that most of the integrants obtained represent random integration events.

The remaining 11.317 reads could be mapped to the human genome, resulting in 42 potential integration events. To discard false positives, identity to the human genome threshold was raised to 98%. **Table 21** shows the resulting potential integration events.

IE	Cell line	Relaxase	Sequence count	Genomic length	Span	Identity	Chromosome	Integration locus	Missing bp
1	EA.hy926	TrwC	19	6-399(400)	395	99.75	11	35225119	0
2	EA.hy926	TrwC	402	1-400(401)	401	100	11	35225119	0
3	EA.hy926	TrwC:BID	11	187	187	99.47	3	126767392	15
4	HeLa	TrwC:BID	21	3-23(23)	21	100	13	83099211	0
5	HeLa	TrwC:BID	1	5-24(24)	20	100	2	77468902	0
6	HeLa	TrwC:BID	8	6-91(91)	86	98.84	1	178168973	13
7	HeLa	TrwC:BID	15	98	98	100	19	18303607	14
8	HeLa	TrwC:BID	15	114	114	98,25	2	37383046	15
9	HeLa	TrwC:BID	84	14-207(207)	194	99.48	12	28128063	16
10	HeLa	TrwC:BID	95	63-417(417)	356	99.72	16	68832248	16
11	HeLa	TrwC:BID	114	155	155	100	2	111118923	16
12	HeLa	TrwC:BID	3	151	151	99.34	6	170575027	16
13	HeLa	TrwC:BID	2	327	327	99.39	14	106208154	18
14	HeLa	TrwC:BID	15	15-55(55)	41	100	6	9173423	18
15	EA.hy926	-	1	226	226	99.56	13	61187811	14
16	EA.hy926	-	1	1-196(197)	196	99.49	5	1363474	16
17	EA.hy926	-	4	17-37(38)	21	100	8	94488444	16

Table 21. Integration events detected after LAM-PCR and DNA sequencing. The information collected for each integration event (see Experimental Procedures, section 3.5) is shown. For detailed information, refer to the main text. IE, integration event.

In Table 21, the cell line and the relaxase of the mobilizable plasmid are indicated for each integration event. Sequence count is the number of times the sequence read was found. Genomic length indicates the size of the alignment. In brackets is shown the length of the sequence (determined by the nearest restriction site) and the numbers before indicate which part of the whole sequence aligns to the genome, while span indicates the size of the alignment in the genomic DNA. The identity of the aligned sequence to the reference genome is shown in the next row. Regarding the chromosomal context of the integration site, the chromosome and the integration locus are also shown. Missing base pairs indicates the number of base pairs that are missed in the read with respect to the plasmid sequence until the *nic* site.

Most integration events showed more than 12 missing base pairs of a total of 41 bp amplified from primer oriTI (Table 4) binding site to the *nic* site, so they were considered as false positives (or random integration events that by chance occurred near the *nic* site) and discarded. Integration event IE5 was also discarded since it was found just once. IE1 and IE2 integration events were found to occur at the same site, the differences in the sequencing reads were assumed to be sequencing errors. From here on, they will be considered together as one integration event and we will refer to it as IE2. After those filters, two potential site-specific integration events were detected, IE2 and IE4. **Table 22** shown the chromosomal information of the corresponding integration sites (IS).

	IS2	**IS4**
Chromosome	11	13
Integration locus	35225119	83099211
In gene, distance to TSS	64,702	-
Intron/exon	Intron 9	-
Distance to gene (bp)	-	1,352,133
RefSeq gene	CD44	SLITRK1
Gene length (bp)	93,530	5,185
Gene orientation	+	-
RefSeq ID	NM_000610	NM_052910

Table 22. IS2 and IS4 integration sites. A more extensive explanation is given in the main text. TSS, transcription start site.

In Table 22, the chromosome, the orientation, and the locus where integration took place are shown. For IS2, integration took place in a gene, so the gene, the distance to the corresponding transcription start site and the intron/exon in which integration occurred are shown. For IS4, since integration took place in a non-coding region, the nearest RefSeq (NCBI Reference Sequence Database) gene was identified. As this SLITRK1 gene was located upstream the integration site, distance to gene indicates the number of bp from the end of the gene to the integration site.

We analysed these two integration events in more detail. **Figure 68a** shows the genomic-plasmid integration junctions (IJ), with the plasmid DNA shown in green, starting at the *nic* site. In Figure 68b the corresponding insertion sites in the genomic DNA are shown, aligned to the TrwC target sequence. These genomic sites were obtained after performing the BLAT alignment of the sequencing reads against the genome (see Experimental Procedures, section 3.5).

When we analysed the integration sites, we found that a thymidine that was present in the sequencing read of IJ2 was absent in the corresponding genomic sequence available in the database.

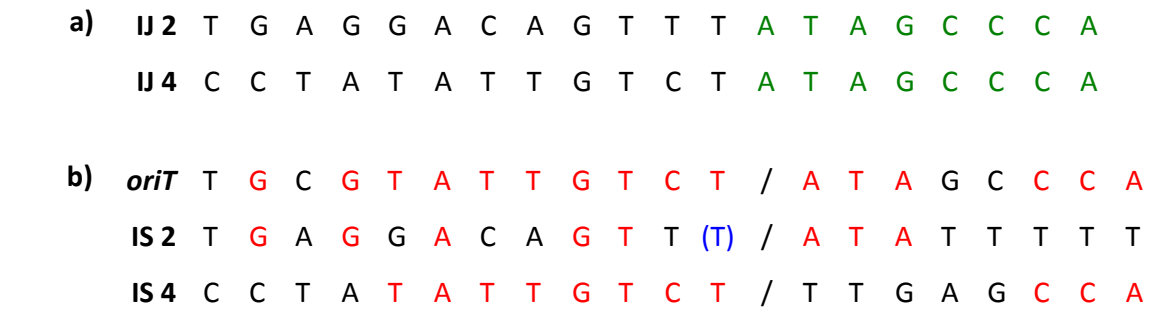

Figure 68. Analysis of selected integration events. a) Genome-plasmid integration junctions (IJ). *oriT* DNA starting at the *nic* site is shown in green. **b)** Chromosomal integration sites (IS) aligned with TrwC *nic* region. Nucleotides that are the same as in R388 *oriT* are shown in red. The thymidine observed in IJ2 that was absent in the genomic sequence available in the database is shown in blue.

Integration site IS2 was characterized in a genomic DNA sample from EA.hy926 cells obtained after translocation and integration of plasmid pCOR31, while the alignment of the sequencing reads against the genome was done using the human

genomic sequence downloaded from the UCSC Downloads web page. Due to this, it could be that this thymidine was part of the genomic DNA of EA.hy926 cells.

To elucidate it, we amplified the genomic region containing IS2 of untreated EA.hy926 cells, using primers IS6_Hu11 (5'-AAGAAAGTCAACCTTCATCTTTTAC-3') and Xba_IS6_Hu11 (5´-CAACTCTAGAGGAAAAGTCAGAAAGACACCAAC-3'). Sequencing results confirmed that the genomic sequence of EA.hy926 cells was the same as the one downloaded from the UCSC database.

As the nucleotide 3´to the *nic* site in R388 *oriT* is a thymidine, the thymidine present in IJ2 must be of plasmid origin, implying that this integration event had occurred at *nic*+1 site instead of exactly at the *nic* site. Considering the high specificity of conjugative relaxases for nicking exactly at their *nic* site, this result suggests that this event was not mediated by the site-specific integrase activity of TrwC, in spite of the considerable homology that IS2 shares with the *nic* region (Figure 68b). Thus, IE2 may represent one more random integration event, which just by chance occurred close to the *nic* site.

In contrast, IE4 represents a bona-fide TrwC-mediated site-specific integration event. As it can be seen in Figure 68, the eight nucleotides 5' to the integration site are identical to the eight nucleotides 5' to the *nic* site in R388 *oriT*. Since the probability of integration at any position of the human genome is approximately $1/3x10^9$, it is very unlikely that this was the result of a random event.

Taken all together, LAM-PCR results show the presence of non-nicked *oriT* in almost all the sequencing reads. One possible explanation for this could be the integration of plasmids concatemers, as it happens in the integration of T-DNA mediated by *A. tumefaciens* (Sullivan *et al.*, 2002). However, *oriT*-host genome junctions should have been detected at the end of the concatemer. As this did not happen, this hypothesis seems very unlikely.

The other, most likely, possibility is that the vast majority of reads obtained by next-generation sequencing were random integration events of the recircularized plasmid. Considering that the DNA enters the human cell cut at the *nic* site (where TrwC is covalently bound), recircularization implies that TrwC is acting on the human recipient cell, mimicking in the human cell its activity in the bacterial recipient cell during conjugation, where it recircularizes the transferred plasmid.

The detection of IE4 integration event proves that TrwC can act as a site-specific integrase in human cells. However, it seems that random integration is much more efficient in human cells, so after TrwC-mediated recircularization of the DNA (as inferred

from the presence of full-length *oriTs*) most molecules would undergo random integration events.

So, TrwC is active in human cells, but it does not work as an efficient site-specific integrase on natural target sites in the human genome. This indicates that TrwC is not useful as a site-specific integrase on natural targets for gene therapy purposes.

The presence of the full-length *oriT* was also detected when analysing integration after Mob-mediated transfer (Schroder *et al.*, 2011). This result implies that the Mob relaxase is also active in recircularization of the transferred DNA in the human cell. Analysis of eight integration events showed in one case preservation of the 5´end of the transferred DNA molecules, integrated in a DNA sequence with no homology to the *nic* site. The authors inferred that Mob could be preserving the 5´end of the DNA strand but did not play any role in the integration process (Schroder *et al.*, 2011). In our case, the detection of IE4, in which integration takes place in a human sequence with 100 % homology to the 8 bp 5´ to the *nic* site, argues for an active role of TrwC in this particular integration event.

In spite of the negligible proportion of integration events attributable to an integrase activity of TrwC, as shown on Table 20, we obtained at least 3 logs more G418-resistant colonies when DNA was introduced with TrwC than with Mob relaxase piloting the DNA into the human cell under the same assay conditions. If not due to its site-specific integration activity as we expected, this may suggest that TrwC protects the DNA ends or even maintains the transferred DNA strand in a close circular form, allowing long-term presence of incoming DNA molecules, which would favour its subsequent random integration.

Chapter 4. General Discussion

Conjugative relaxases are the proteins in charge of initiating and terminating DNA processing during bacterial conjugation. TrwC is the relaxase of conjugative plasmid R388. Apart from its role in conjugation, TrwC is known to be able to catalyse site-specific recombination between two *oriT* copies repeated in tandem in the absence of conjugation, and site-specific integration of the mobilizable DNA once transferred to the recipient cell into a recipient *oriT* copy (Draper *et al.*, 2005). The recombinase activity has been reported in only a few relaxases, while the ability to promote site-specific integration in the recipient cell after conjugative transfer has been reported only for TrwC and the relaxase Mob02281 of the *Enterococcus faecalis* pBMB0228 cointegrate (Wang *et al.*, 2013).

This situation poses many relevant questions: How are these reactions performed? Why some relaxases can catalyse them while other related ones cannot? What could be the biological role of this ability of conjugative relaxases, which is not involved in conjugative DNA transfer itself? Could we use this ability as a tool for genomic engineering? The purpose of this work was to address these questions, using the following approaches:

- A mutagenesis protocol to obtain improved recombinases, in order to determine the key residues involved in these reactions.
- An analysis of two relaxases related to TrwC that were previously reported not to be able to catalyse these reactions.
- Testing TrwC-mediated integration of conjugatively transferred DNA into the genome of non-permissive hosts.
- Analysing TrwC-mediated integration events in the human genome, in order to evaluate its possible use for human genomic modification.

Molecular characterization of site-specific recombination and integration reactions mediated by relaxases

In order to get insight into the molecular mechanism of relaxase-mediated recombination, we reasoned that selection of proteins with improved activities could give us valuable information on the protein requirements. Unfortunately, such approach was not successful; after extensive *trwC* mutagenesis and screening of about 80,000 colonies, the only variant that showed increased recombinase activity turned out to be caused by a higher amount of protein production. It seems that the improvement of TrwC by altering one or two residues is not possible. Future work could include the increase of the mutation rate or the use of approaches for high-throughput screening in

order to analysed a higher amount of variants generated, in a similar way as done for zinc-finger nucleases (Durai *et al.*, 2005).

We also reasoned that the study of related relaxases being able/unable to catalyse these reactions could shed light into the requirements for a relaxase to act as a recombinase/integrase, so we extended our study to F_TraI and pKM101_TraI relaxases, highly related to TrwC, previously reported to lack the recombinase activity that TrwC has (César *et al.*, 2006). However, in this PhD thesis work we have shown that both of them are actually able to catalyse site-specific recombination between two cognate *oriT* copies repeated *in tandem*, although the results suggest that the efficiency of both TraI proteins is much lower than that of TrwC under the same assay conditions. As discussed in section 3, the different efficiency could be due to biochemical properties of each relaxase, such as *oriT* binding affinity or a differential interaction with the replication machinery. But the relevant finding is that the three proteins, belonging to the same family of relaxases and sharing extensive structural similarity, are catalytically able to catalyse the same reaction. In contrast, we did not detect F_TraI-mediated site-specific integration of the conjugatively mobilized DNA in the recipient cell under similar assay conditions to those used for TrwC.

A significant step forward in the molecular characterization of these reactions was accomplished by analysis of DNA requirements in the TrwC target sequences involved in the site-specific integration process. We have found that the *oriT* sequences required for initiation and termination have very different DNA requirements, which mimic those for the start and end of conjugative DNA transfer. Changes in the critical area affecting either the *nic* or the binding sites abolished TrwC integration activity when TrwC was only expressed in the recipient cell and thus, it is required to act on the supercoiled mutant *oriT* (Agundez, 2011). In contrast, incoming TrwC-DNA complexes could transfer DNA into acceptor sites with mismatches in the core *oriT* sequence (Agundez *et al.*, 2012).

Possible biological role of site-specific recombinase and integrase activities

Why would a conjugative relaxase have site-specific recombinase and integrase activities? Neither is required for conjugative DNA transfer; in fact recombination happens in the absence of conjugation.

The first evidences of relaxases with this ability were found in small staphylococcal plasmids whose relaxases were shown to promote site-specific

recombination at *oriT* rendering plasmid cointegrates during conjugative mobilization (Gennaro *et al.*, 1987; Novick *et al.*, 1984; Projan & Novick, 1988). Priebe and Lack suggested that the nicks produced at those sites were the first step of plasmid processing for conjugative transfer, being the recombination-mediated formation of cointegrates an incidental result of the system (Priebe & Lacks, 1989).

However, the number of relaxases shown to have recombinase activity continues to grow, and taking into account that the newly defined SSR F_TraI and pKM101_TraI were previously considered as recombination deficient, the possibility exists that any conjugative relaxase might have this activity when assayed under proper conditions. If this were the case, it would support the idea of a functional role for such activity.

One possibility is that *oriT*-specific recombination could be used for resolution of plasmid multimers. Several works have reported the importance of this resolution for stable plasmid maintenance. Hakkaart and co-workers showed in 1984 that copy-control mutants of plasmid CloDF13 were not stably inherited in *E. coli*. They identified specific regions (*parB*) in which a *recA*-independent site-specific resolution process occurred when multimers were formed (Hakkaart *et al.*, 1984). Similar systems to maintain a plasmid monomeric state were found in different plasmids, such as ColE1 and ColK (Summers *et al.*, 1985), or RP4 (Eberl *et al.*, 1994). Although in the majority of plasmids in which a multimer resolution system has been identified, they are not *oriT*- and relaxase-related, it makes sense to use the relaxase to resolve dimers when two *oriT* copies are detected in the same molecule, as an additional control against dimer maintenance.

Regarding the biological role of TrwC integrase activity in the recipient, our hypothesis is that it could provide R388 the possibility to disseminate in nature when it is mobilized to a recipient cell in which it cannot be replicated. In this case, if a DNA sequence resembling the TrwC *nic* site is found in the recipient genome, TrwC could catalyse the integration of R388 into the recipient genome. In support of this possibility, it has been shown that TrwC can catalyse integration into the chromosome of a recipient bacterium, and it can also integrate into sequences that do not match exactly the R388 *nic* site (Agundez *et al.*, 2012). This hypothesis is strongly reinforced by our results, which showed that TrwC-mediated integration is possible into *oriT* copies with several nucleotides deviating from the consensus, which affect TrwC functions essential for the formation of the TrwC-DNA complex. This flexibility in the acceptor site would contribute to the dissemination of plasmids beyond their host range and to the genetic information exchange between distantly related bacteria.

If that were the functional role, this would be in agreement with F_TraI being unable to catalyse integration, since F is a conjugative plasmid of narrow host range that only conjugates within *E. coli* and other related Enterobacteriaceae strains in which it is able to replicate, while R388 is a conjugative plasmid of broader host range. We have tried to test the limits of this promiscuity by assaying R388 transfer from bacteria to yeast, but we found that, under conditions where plasmid RP4 is readily transferred, no R388 transfer to *S. cerevisiae* was detected.

In summary, a possible role for the site-specific recombination/integration activity of relaxases would be to form/resolve cointegrates through two *oriT* copies, allowing different plasmids to share their conjugation and/or replication machineries, when appropriate. This ability could be extended, with lower efficiency, to chromosomal integration into DNA sequences resembling the *oriT nic* region.

This putative role of the *oriT*-specific recombination reaction in the resolution and regeneration of cointegrate plasmids could be indicative of a relaxase-mediated mechanism of plasmid evolution. Wang and collaborators have recently reported the resolution and regeneration of a cointegrate plasmid via *oriT*-specific recombination mediated by a Mob protein. They showed that the cointegrate contained two functional replication regions and two functional mobilization regions and that the two constituent plasmids were mobilizable, being able to fuse back via *oriT* site-specific recombination after entering into the same cell by conjugation (Wang *et al.*, 2013).

In the discussion, Wang and collaborators suggested that something similar could occur for pAM1α cointegrate (Francia & Clewell, 2002b). They proposed that when tetracycline resistance plasmid pAM1$\alpha\Delta$1, which is not able to replicate in *Enterococcus faecalis*, was mobilized into an *E. faecalis* strain carrying the plasmid pAM1$\alpha\Delta$2, the latter plasmid captured pAMα1 via MobB/MobE-mediated recombination at the two similar *oriT* sites, acquiring its tetracycline resistance gene for host adaptation (Wang *et al.*, 2013). However, the pAM1α cointegrate formation has not been directly observed experimentally yet.

The presence of two *oriT* sequences and two relaxases together with two or even more origins of replication have also been shown for plasmids R6K (Avila *et al.*, 1996), pAD1 (Francia & Clewell, 2002a) and pAMα1 (Francia & Clewell, 2002b). Although resolution has not been demonstrated *in vivo*, the presence of such duplicated mobilization regions supports the idea that cointegration/resolution could be an adaptation strategy also for these plasmids.

Conjugative relaxases belong to the HUH family of proteins. HUH proteins are defined by the presence of a series of motifs involved in their catalytic activity (see Introduction, sections 3 and 4.2.2.2) and its members are sequence-specific endonucleases that perform strand-transfer reactions on ssDNA, required in processes such as bacterial conjugation, rolling-circle replication, or transposition (see (Chandler *et al.*, 2013) for a recent review). Some of them have been shown to catalyse site-specific recombination and integration reactions on dsDNA substrates, and thus they can also be considered as SSRs. Our results add two new members to the HUH family of SSRs. HUH recombinases may represent a family of moonlighting proteins evolutionarily selected to perform site-specific recombination/integration, in addition to their role in viral replication, bacterial conjugation, or transposition. There are evidences that this molecular strategy has been preserved from bacteria, to plants, to mammalian viruses (Gonzalez-Prieto *et al.*, 2013), supporting the notion that this activity adds evolutionary advantage to the plasmids encoding a SSR-relaxase.

Potential use of relaxases for genomic modification of human cells

Site-specific recombinase and integrase activity confers to TrwC a high potential as a biotechnological tool. TrwC and any covalently linked *oriT*-containing DNA could be conjugatively transferred to the recipient cell, which is not restricted to prokaryotes. In fact, it has been shown that TrwC-DNA complexes can be efficiently sent to human cells through the T4SS of *B. henselae* (Fernandez-Gonzalez *et al.*, 2011). In addition, we could accomplish site-specific integration into *oriT*-like sequences of the recipient chromosome. Agúndez and collaborators showed in a bacterial system that TrwC could catalyse integration into DNA sequences of human origin. However, it remained to be determined if TrwC was active in human cells (Agundez *et al.*, 2012).

Using human cells, we have analysed TrwC-mediated site-specific integration into genomic DNA after the mobilization of TrwC-DNA complexes from *B. henselae*. We found that TrwC is active in the human cell, since the *oriT* sequence was recircularized in most integration events. Moreover, we have detected at least one integration event that reflects TrwC site-specific integration activity in human cells. However, it does not work as an efficient integrase, as most molecules were found to undergo random integration, thus precluding its use as a SSR for human genomic modification.

Although not useful as a site-specific integrase, TrwC may be useful in combination with other proteins to deliver DNA molecules to human cells. We obtained at least a 3-log increase in the number of integration events when DNA molecules were

led by TrwC compared to other relaxase. These results suggest that TrwC is protecting the transferred DNA molecules in the human cell, favouring their long-term presence and subsequent genomic integration.

This effect of TrwC could also be useful as a tool for genomic engineering, as outlined in **Figure 69**.

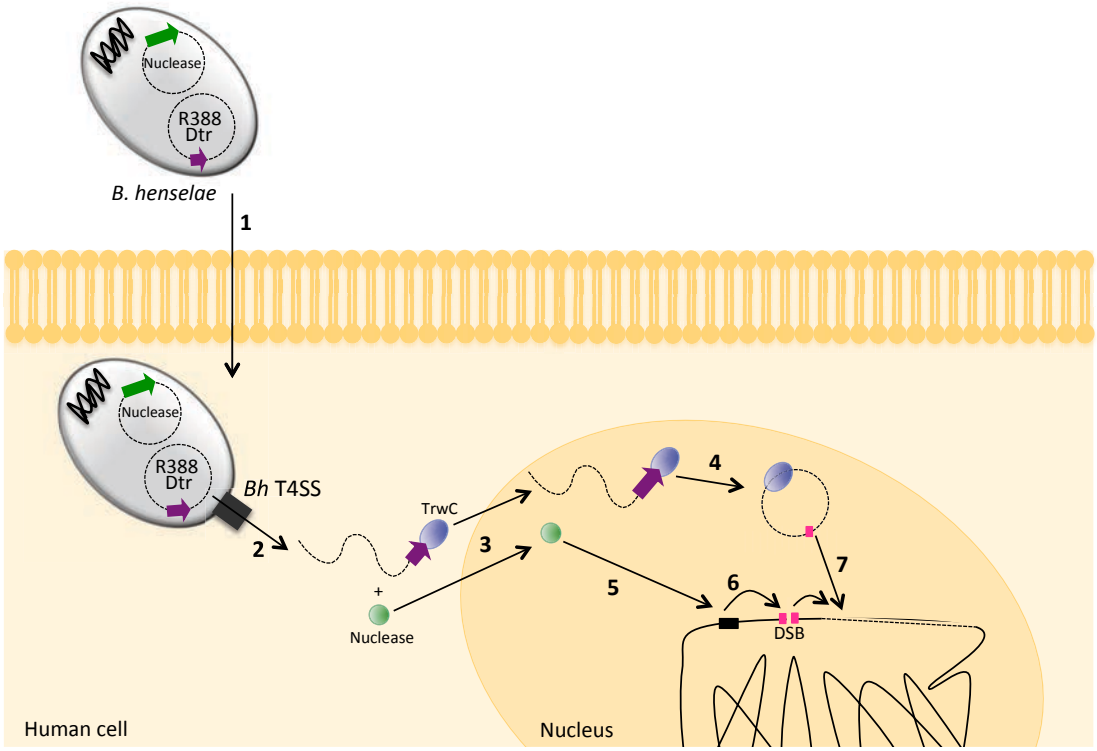

Figure 69. Potential use of TrwC in cooperation with a nuclease. 1) Bacteria containing T4SS infect their specific human cell type. **2)** T4SS-mediated secretion of TrwC-DNA complexes and a site-specific nuclease. **3)** Nuclear entry, promoted by the addition of a NLS if needed. **4)** TrwC stabilizes the transferred DNA strand, protecting one or both ends. **5)** The nuclease recognizes its specific target in the genome, **6)** producing a double strand break (DSB). **7)** If the exogenous DNA has homologous regions (in pink) to the flanking regions of the DSB, homologous recombination-mediated integration can then take place. Blue oval, TrwC. Green sphere, nuclease. Purple arrow, *oriT*.

TrwC could be used to transfer the DNA of interest through the T4SS of different bacteria that infect specific human cell types. By adding the appropriate secretion signal (BID signal for *Bartonella* VirB/D4 T4SS), a site-specific nuclease such as Zinc-finger or TALEN nucleases could be also translocated. Once in the nucleus of the human cell, the

nuclease could produce a break in the chromosome while TrwC protects and stabilizes the exogenous DNA, both activities jointly promoting integration of the incoming DNA molecule by homologous recombination.

In support of this approach, it has been recently reported that concomitant translocation of I-SceI homing site-specific endonuclease together with VirD2 relaxase-T-DNA complexes through *A. tumefaciens* T4SS enhanced T-DNA site-specific integration into yeast chromosome when I-SceI target site was present (Rolloos *et al.*, 2015).

This strategy would contribute with several advantages over existing methods for genome modification:

- DNA delivery based on bacterial T4SS systems: T4SS machineries permit *in vivo* delivery of the DNA of interest directly to the cell to be modified. Moreover, using different T4SS-containing pathogens (conveniently attenuated), different cell types could be targeted.
- The use of a conjugation-derived system would permit the delivery of DNA of any origin and sequence, since conjugative-like transfer has no theoretical limit for transferred DNA length and the only DNA requirement is the *oriT*.
- Bacterial DNA has been shown to be potentially toxic in the human cell. If the DNA of interest is flanked by two copies of the *oriT,* no bacterial DNA would be transferred to the human cell, since the relaxase would cut at the two *nic* sites in the donor molecule, as shown in the conjugation-mediated recombination assay described in this work.
- TrwC-DNA complexes as well as other proteins of interest (such as the nuclease) would be sent to the target cell, eliminating the need for expression of potentially toxic proteins in the human cell. Virtually any protein could be sent through a T4SS by the addition of T4S and nuclear localization signals.

With this work, we have provided a step forward by showing that TrwC is active in human cells and it promotes integration into the recipient genome. Future experiments could involve the evaluation of TrwC in combination with other nucleases, followed by transfer of a wild-type human gene into a defective cell line, to evaluate the efficiency of recovery of the gene function.

Appendix:

Insertional hotspot for IS10 in the *cat* promoter region of widely used cloning vectors

Universal cloning vectors offer a series of advantages to facilitate their manipulation, which include their small size, presence of multiple cloning sites, selection markers, and high copy number. The most commonly used cloning vectors are associated with the pMB1 replicon. They carry most often the beta-lactamase gene from Tn*3*, originally present in the historic vector pBR322 [GeneBank Accession number J01749; (Bolivar *et al.*, 1977)] allowing plasmid selection by resistance to ampicillin. Cloning vectors based on the p15A replicon are the common alternative to pMB1-based vectors. They have a moderate copy number and therefore they are a better option when potentially toxic genes are to be cloned. Most of these vectors encode the *cat* gene from *Tn9*, as in original vector pACYC184 [GeneBank Accession number X06403; (Chang & Cohen, 1978)], conferring chloramphenicol resistance. The most widely used family of p15A-derived cloning vectors is the pSU8 family (Bartolomé *et al.*, 1991; Martinez *et al.*, 1988), which was designed to allow coexistence with a pMB1-Ap^R vector in the same host. pSW family of vectors are conditionally replicating plasmids based on the IncX *oriV* origin of replication which are dependent on the *pir*-encoded protein (Demarre *et al.*, 2005). These vectors are also Cm^R and during their construction the *cat* gene was amplified from pSU18. In spite of their successful use as cloning vectors for more than 20 years, during this thesis work we have detected an apparent genetic instability in several constructs based on pSU and pSW vectors.

IS10R, the right module of the bacterial transposon Tn10, can act as an individual insertion sequence, since it codifies a transposase protein (Foster *et al.*, 1981), with a transposition frequency of 10^{-4} per cell per bacterial generation (Shen *et al.*, 1987).

Although the originally sequenced genomes of both *Salmonella* and *E. coli* strains (Blattner *et al.*, 1997; Parkhill *et al.*, 2001) did not show any copy of IS10, there have been several reports of the presence of IS10 elements in different *Salmonella* and *E. coli* strains, some as widely used as *E. coli* JM109, DH5α, DH10B or XL-2 Blue (Kovarik *et al.*, 2001; Matsutani, 1991), and IS10 transposition events are frequently detected in *E. coli* (Rodriguez *et al.*, 1992). Thus, we can assume that IS10 copies are present in the genomes of many commonly used *E. coli* K-12 laboratory strains. In fact, there have been previous reports on cloning artefacts due to IS10 transposition from *E. coli* genome to a plasmid. IS10 insertion has been reported to take place frequently during cloning in pUC19 vector (Kobori *et al.*, 2009). Kovarik et al (2001) performed a database search that concluded that IS10 was inserted into numerous eukaryotic clones.

1. A new insertional hotspot for IS10 in the pSU8 family of vectors

We routinely use the pSU8 family of vectors in *E. coli*, and we have observed several times that the restriction pattern of the different plasmid constructions based on these vectors was not the expected one. An increase in size of about 1 kb was observed. This phenomenon was observed only in particular constructs: recombination substrate plasmids (encoding two *oriTs*) pCIG1028 (R388), pCIG1032 (pKM101/R388), pCIG1066 (pKM101) and pCIG1064 (F) (César *et al.*, 2006); plasmids coding for *trwB* mutations pSU4632 (Moncalián *et al.*, 1999a), pDEL10 (de Paz *et al.*, 2010); and the suicide plasmid pCMS11 (Draper *et al.*, 2005), carrying the *oriTs* of R388 and RP4, once integrated into the recipient *oriT* copy in the chromosomal integration assay described in (Agundez *et al.*, 2012).

Figure 70a shows a restriction analysis of the recombination substrate plasmids used in this work where this phenomenon can be observed. The band of 3.2 kb corresponding to the recombination cassette was present in all of them. However, the band corresponding to pSU19 vector that should be 2.3 kb is size was not visible. Instead, a band of approximately 3.5 kb was detected. Figure 70b shows restriction analysis of pCIG1028 DNA in the four different forms found during this PhD thesis work.

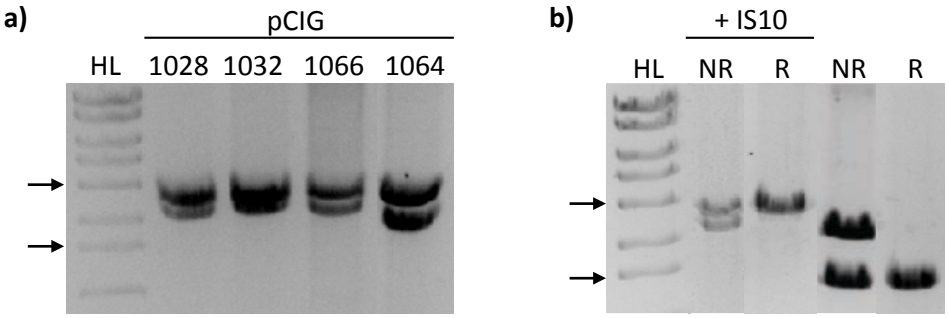

Figure 70. Restriction analysis of recombination substrate plasmids with EcoRI and HindIII restriction enzymes. **a)** In not recombined substrate plasmids, we expected a band of 2.3 kb for vector pSU19 and a band of 3.2 kb (3.0 for pCIG1064) for the recombination cassette containing both *oriTs*. The band of the vector is not detected, instead a band of about 3.5 kb is visible. **b)** The expected bands are seen in pCIG1028 when IS10 was not present (2.3+3.2 kb bands for non-recombined and 2.3 kb for recombined plasmid, as the band of 0.4 kb corresponding to the unique *oriT* is not visible). When IS10 is present, the vector band increases up to around 3.5 kb. The arrows indicate the bands of 4 and 2.5 kb. HL, Hyperladder. NR, not recombined. R, recombinant.

Similar results were found when analysing the rest of the plasmids mentioned above. Although they are independent constructs, all of them had acquired the same increase in size.

We delimited the region containing the extra DNA by restriction analysis and found that it was always the same region of the vector backbone. We determined the DNA sequence from the vector DNA until the junction with the foreign DNA, in four plasmid constructs that had gained the extra DNA independently during their manipulation. The four of them carry different DNA segments cloned in vectors pSU19, pSU24, and pSW23. **Figure 71** shows the results obtained.

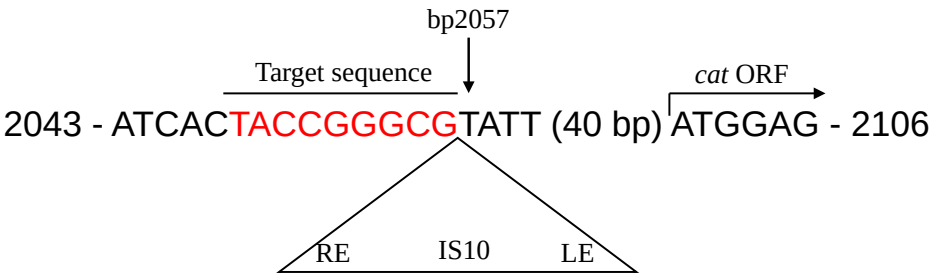

Figure 71. New IS10 target site located upstream of the *cat* ORF. The target site is shown in red and was found 44 bp upstream the ATG of *cat* ORF. Coordinates correspond to the DNA sequence of pSU8 vector (accession number X53939). IS10 insertion sequence begins in nucleotide 1329 of its DNA sequence (accession number J01829).

The inserted DNA was identical to the published sequence of insertion sequence IS10 (GenBank Accession number J01829). The junction with the vector DNA occurred exactly at the IS right end, indicating that the observed sequences were the result of IS10 transposition events. Insertion is obviously not random, since all independently obtained integration events occur exactly at the same nucleotide, so this DNA sequence could represent a previously unknown hotspot for IS10 transposition.

The site of insertion was located 44 bp upstream the *cat* ORF. The target sequence was 5'-TACCGGGCG-3', which does not correspond to the preferred target site for IS10 originally described, 5'-NGCTNAGCN-3' (Halling & Kleckner, 1982). No such consensus target site is present in the pSU vectors. However, (Kovarik *et al.*, 2001) defined a new consensus sequence based on analysis of a large number of IS10 insertions: 5'-NPuCNN-NGPyN-3'. Our target sequence matches this consensus.

2. Insertion of IS10 is selected at high Cm concentrations.

As explained above, IS10 insertions were detected only in a subset of pSU- and pSW-based constructs. A plausible explanation for this result is that the insertion events were selected because of the Cm selection applied. It has been reported previously that IS10 carries outward-facing promoters at both ends (Martinez-Garcia *et al.*, 2003; Simons *et al.*, 1983) and IS10 insertion events that confer resistance to fluoroquinolones by transcriptional activation of adjacent genes in *Salmonella enterica* have been previously reported (Olliver *et al.*, 2005). So, transcription of *cat* from the IS10 promoter facing outwards could contribute to Cm resistance under conditions in which the intrinsic Cat levels would be insufficient to cope with the Cm selection applied.

In order to test this hypothesis, we have grown *E. coli* strain DH5α containing pSU19-derived constructs for up to 100 generations in the presence of different concentrations of Cm. Plasmids used were pDEL10, pCOR39, and pSU4633, three pSU19::*trwB* constructs. pSU4633 encodes wild-type TrwB; pDEL10 codes for TrwB variant K275A; as pDEL10 had been observed to change in size before, pCOR39 was obtained from incubation of plasmid pDEL10 until the increase size appeared, and it was used as a control of the size after IS10 insertion. Plasmid pSU4633, not previously shown to change in size after extensive manipulation, was used as a control of size stability.

At the end of the experiment, plasmid DNA was extracted and aliquots were run on an agarose gel to compare sizes (**Figure 72**).

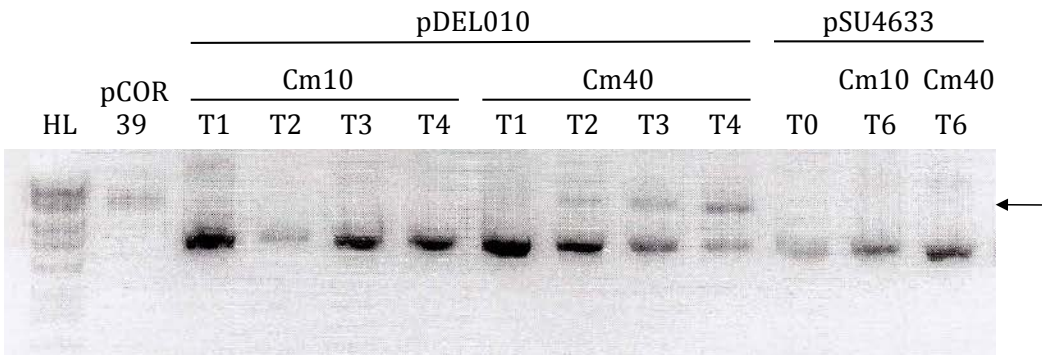

Figure 72. IS10 insertion at different Cm concentrations. The arrow indicates the molecular size after IS10 insertion. T0, DNA used to transform cells at the beginning of the assay; T1, T2, T3, T4, and T6 refer to DNA obtained from cultures of different consecutive days. HL, Hyperladder I.

Plasmid pSU4633 remained in its original size at the end of the assay, independently of the Cm concentration applied. In contrast, plasmid pDEL10 remained stable growing the host strain in LB supplemented with 10 µg/ml of Cm, but acquired IS10 insertion upon growth under 40 µg/ml of Cm.

3. Insertion of IS10 permits selection of chromosomal integration events at higher Cm concentrations

TrwC-mediated chromosomal integration events described in (Agundez *et al.*, 2012) took place after mobilization of suicide plasmid pCMS11, which was constructed using pSW23 backbone, into an *E. coli* strain carrying a chromosomal copy of R388 *oriT* in place of the *lacZ* gene. Integrants were selected with Cm 25 µg/ml and integration was firstly checked by PCR analysis with primers amplifying a region of 1 kb of the cointegrate molecule. However, an amplicon of 2 kb was sometimes found, suggesting again IS10 integration into a pSU-derived vector. This was confirmed by DNA sequencing of the PCR product. We found IS10 insertion in the same target sequence located upstream of the *cat* gene.

To elucidate if the pressure exerted by the Cm selection was responsible for IS10 insertion, we tested chromosomal integration (**Figure 73**) selecting integrants in both Cm 10 and 40 µg/ml.

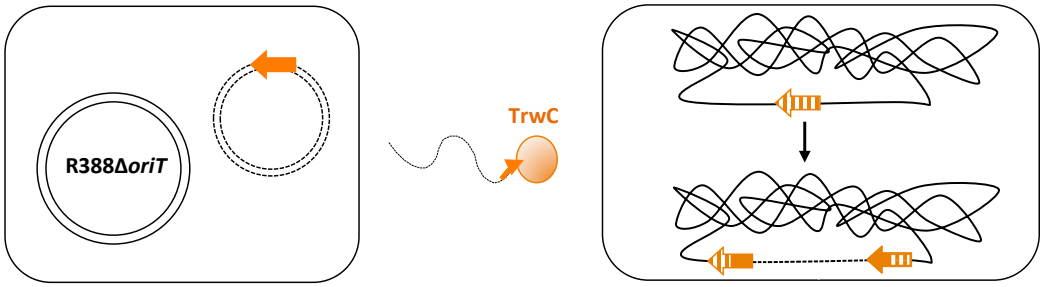

Figure 73. TrwC-mediated site-specific integration into a chromosomal *oriT* copy. TrwC-DNA complexes arrive to the recipient cell, where TrwC is able to act on a copy of its cognate *oriT* located in the chromosome. R388 *oriTs* are shown in orange.

Chromosomal integration events were analysed by PCR with primers 5´lacZ (5'-ATGACCATGATTACGGATTCA-3´) and 3'72Cm (5'-GCCTCAAAATGTTCTTTACGA-3´) (**Figure 74**). 100 % of the integrants obtained with Cm40 selection were found to have IS10 insertion, while only 66.6 % of the ones selected under lower antibiotic pressure had acquired it (mean of at least 3 independent experiments, with approximately 10 colonies analysed in each experiment), suggesting that IS10 insertion is necessary for the integrants to survive under the higher Cm selection applied.

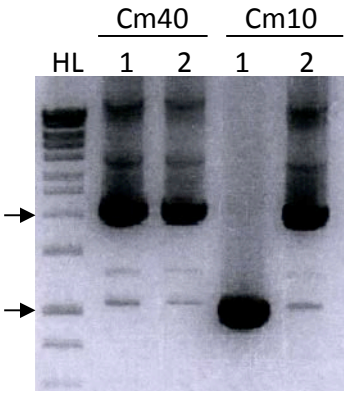

Figure 74. PCR to detect chromosomal integration events. 1 and 2 indicate two different colonies obtained under each condition tested. The arrows indicate the bands of 2 and 1 kb. Cm40/10 refers to the concentration of Cm (in µg/ml) used for selecting integration events. HL, Hyperladder.

Taking these results together, it is clear that transposition events are selected upon high Cm selection. The most surprising result is the occurrence of this phenomenon only in certain plasmid constructs, and not in others. pSU4633 (not undergoing any change in size) is identical to pDEL10 and pSU4632, only the two latter code for different point mutations in *trwB*, whose phenotype is described in (de Paz *et al.*, 2010) and (Moncalián *et al.*, 1999a), respectively. The other plasmids found to acquire IS10 are recombination substrate plasmids and the cointegrate obtained after pCMS11 integration. In all cases, they have two *oriT* copies repeated *in tandem*.

The most plausible explanation is that the DNA segment cloned in these plasmids could either confer instability to the cell or affect plasmid stability. Insertions lying in the promoter of the chloramphenicol resistant gene would lead to higher levels of Cat, which could compensate the slower metabolism or lower number of plasmid molecules coding for the gene. This idea is supported by the observation that other plasmids based on the

same family of vectors do not suffer this phenomenon under the same laboratory conditions.

The reason why certain TrwB variants and the presence of two *oriTs* in the same molecule could be deleterious to the plasmid or to the cell, remains to be determined. TrwB has DNA binding and ATPase activities, both of which could be toxic if deregulated. Both K136T and K275A variants have been analysed in detail and they are affected in their ATPase activity (Larrea, 2013; Moncalián *et al.*, 1999a). As for the two *oriT* copies, which are in all cases either identical or highly related, *oriTs* are DNA sequences where many DNA binding proteins are recruited, and two copies *in tandem* may undergo recombination. In summary, they could act as protein recruiting factors that could interfere with normal plasmid replication or transcription processes.

From a practical point of view, our suggestion for the ample scientific community using the pSU/pSW families of cloning vectors would be to use low Cm selection conditions, to avoid selection of IS10 transposition events into their constructs. In fact, transposition activity has been reported to be increased under certain stress conditions, such as long incubation of bacterial cultures in the stationary phase of growth (Skaliter *et al.*, 1992) or UV light induction (Eichenbaum & Livneh, 1998). In a similar way, incubation with high concentrations of chloramphenicol could be a stressful situation that increases IS10 transposition.

Conclusions

1. We have set up a mutagenesis protocol followed by *in vivo* detection of recombination to screen for *trwC* mutants with enhanced *oriT*-specific recombination activity.

2. A variant (TrwC*) with enhanced recombinase activity was selected. No mutations were found in the nucleotide sequence of *trwC**. The cause of the higher recombination levels observed is probably a higher abundance of TrwC* in the cells. This higher abundance does not have any effect in the integration frequency mediated by TrwC*, compared to TrwC.

3. We have identified three mistakes in the *trwC* sequence publically available in the GenBank database. The corrected sequence has been submitted to the database.

4. The conjugative relaxases TraI of plasmids F and pKM101 can catalyse recombination between two copies of their cognate *oriT*s, in contrast to what was previously reported. Their recombinase activity is very low compared to that of TrwC under similar assay conditions. There is an increase in the frequency of *oriT*-specific recombination mediated by F_TraI when the relaxase is produced in higher amounts.

5. TrwC covalently linked to DNA can catalyse strand-transfer reactions into mutant *oriTs* affected in nucleotides relevant for TrwC binding and nicking, with only a moderate decrease in the integration frequency compared to that of wild-type *oriT*.

6. No *oriT*-specific integration mediated by F_TraI has been detected under the same assay conditions used for TrwC.

7. We have not detected R388 conjugative transfer to *ssd1* and *ssd1* "*petite*" yeast mutants under assay conditions in which transfer of plasmid RP4 is detected.

8. We have detected DNA transfer through the VirB/D4 T4SS of *B. henselae* to HeLa cells. However, there is significant decrease in the transfer rate when compared to EA.hy926 cells. DNA transfer is one-log more efficient when DNA is led by TrwC or TrwC:BID than by Mob:BID.

9. We have obtained pools of genomic integration events of TrwC- and TrwC:BID-driven plasmids transferred from *B. henselae* into HeLa, HeLa::*oriT* and EA.hy926

cells. No stable integration of the Mob:BID-driven plasmid was obtained under the same assay conditions.

10. The rate of integration of TrwC-driven DNA into the human genome is at least 3 logs higher than for Mob:BID-driven DNA.

11. Analysis of the genomic integration pattern of TrwC- and TrwC:BID-driven DNA by LAM-PCR suggests that integration is mainly not site-specific.

12. Most TrwC-driven plasmids integrated into the human genome showed a recircularized *oriT*, confirming that TrwC is active in human cells.

13. At least one site-specific integration event in the human genome mediated by TrwC has been identified. The eight nucleotides 5´ to the integration site are identical to the eight nucleotides located 5´ to the *nic* site in wild-type *oriT*.

14. We have identified a new insertional hotspot for IS10 in the *cat* promoter of pSU/pSW families of vectors. Insertion into this site is selected at high chloramphenicol concentrations.

Resumen en español

Introducción y objetivos

Los sistemas de secreción tipo IV (T4SS) constituyen una familia de transportadores moleculares capaces de enviar sustratos desde el citoplasma bacteriano directamente al medio extracelular o al citoplasma de una célula receptora. Tienen una gran versatilidad, tanto en la naturaleza del sustrato (ADN, proteínas o complejos nucleoproteicos) como de la célula receptora (otras bacterias o células eucariotas, de origen vegetal o animal). Esta versatilidad hace que estén implicados es procesos muy diversos, como la conjugación bacteriana o la translocación de efectores a células humanas durante la infección de diversos patógenos (Alvarez-Martinez & Christie, 2009).

La conjugación bacteriana es un mecanismo de transferencia genética horizontal que requiere el contacto celular a través de un T4SS. Bajo condiciones de laboratorio, se ha descrito conjugación de bacterias a levaduras, plantas e incluso células de mamífero. Las relaxasas conjugativas son las proteínas encargadas del procesamiento del ADN que va a ser transferido durante la conjugación a través de su acción sobre el origen de transferencia u *oriT*.

TrwC es la relaxasa del plásmido R388. Aparta de su papel en la conjugación, sabemos que TrwC es capaz de catalizar la recombinación sitio-específica entre dos copias del *oriT* repetidas en tándem así como la integración sitio-específica del ADN que es movilizado si en la célula receptora encuentra una copia del *oriT* (Draper *et al.*, 2005). La actividad recombinasa de otras relaxasas conjugativas relacionadas a TrwC había sido también analizada. TraI, la relaxasa del plásmido F y TraI, la relaxasa del plásmido pKM101 no mostraron dicha actividad bajo las condiciones ensayadas (César *et al.*, 2006).

Recientemente se ha descrito la translocación de dos complejos relaxasa-ADN a células humanas a través del T4SS VirB/D4 del patógeno *Bartonella henselae*. Estas relaxasas son Mob (Schroder *et al.*, 2011), la relaxasa de un plásmido natural de *B. henselae* (pBRG1) y TrwC, la relaxasa de R388 (Fernandez-Gonzalez *et al.*, 2011). Los plásmidos movilizados contienen el *oriT* correspondiente y el gen de la proteína verde fluorescente (GFP) bajo un promotor eucariota, para detectar la llegada del ADN al núcleo de la célula humana. Los autores encontraron que cuando *B. henselae* portaba el plásmido derivado de pBRG1, un 0,02 % de células eran GFP-positivas. En cambio, hasta un 3 % de las células fueron positivas cuando se analizó el plásmido derivado de R388.

Con todo esto, TrwC muestra un gran potencial como herramienta biotecnológica para la modificación genética de, incluso, células humanas. Trabajo previo en el laboratorio demostró que en el genoma humano existe un numero concreto de

potenciales dianas para la integración mediada por TrwC y que dos de ellas funcionan como dianas para la integración en bacterias (Agundez *et al.*, 2011; Agundez *et al.*, 2012).

Los objetivos que nos plateamos para este trabajo fueron:

1. Aumentar nuestro conocimiento sobre el mecanismo de las reacciones de recombinación e integración sitio-específica mediadas por TrwC. Los objetivos concretos fueron:
    a. Completar la caracterización de la reacción de integración mediada por TrwC con respecto a los requerimientos de ADN en ambos *oriTs* diana de la reacción.
    b. Seleccionar variantes de TrwC con actividad recombinasa e integrasa aumentada, lo que podría proveer valiosa información sobre los residuos clave de TrwC que participan en dichas reacciones.

2. Determinar la presencia de las actividades recombinasa e integrasa específicas de *oriT* en otras relaxasas conjugativas. Para ello, nos planteamos analizar dichas actividades en F_TraI y pKM101_TraI, dos relaxasas que pertenecen, como TrwC, a la familia MOB_F de relaxasas conjugativas, y que habían sido previamente descritas como deficientes en estas actividades (César *et al.*, 2006).

3. Analizar la habilidad de TrwC de mediar integración sitio-específica en células humanas mediante la selección de eventos de integración tras la transferencia de ADN mediada por la propia TrwC, seguida de la caracterización de patrón de integración.

Resultados y discusión

1. Reacción de recombinación sitio-específica mediada por relaxasas conjugativas.

Con el fin del obtener mutantes de TrwC que catalicen la reacción de recombinación con mayor eficiencia llevamos a cabo una mutagénesis al azar mediante PCR. A la hora de poner a punto el ensayo, lo primero que necesitamos fue un buen sistema para realizar un análisis rápido y sencillo de los mutantes generados. El sistema elegido fue el clonaje de los mutantes en el vector pET3a y su incubación en DH5α junto a un plásmido sustrato de la recombinación portador de dos copias del *oriT*; tras la recombinación, se expresa el gen *lacZ*, lo que permite la selección por el color de las

colonias en el medio selectivo adecuado (César *et al.*, 2006). *trwC* y *trwA+trwC* clonados en el mismo pET3a se usaron para comprobar los niveles de recombinación obtenidos con la proteína salvaje. Comprobamos que cuando TrwA está presente, el nivel de recombinación mediada por TrwC llega casi al 100 %, mientras que baja a niveles prácticamente indetectables cuando no está dicha proteína accesoria, lo que supone una situación ideal para la búsqueda de mutantes.

Para optimizar la PCR mutagénica, probamos distintas condiciones cambiando la cantidad de ADN molde, el número de ciclos, la temperatura de annealing, la cantidad de enzima y el tiempo de elongación. Tras elegir las condiciones óptimas, confirmamos que obteníamos la tasa de mutación esperada (1-4,5 mutaciones por kilobase) y llevamos a cabo la mutagénesis utilizando *trwC* salvaje como molde. Tras clonar los pools de mutantes generados y transformar la células portadoras del plásmido sustrato, la recombinación fue detectada sembrando en placas suplementadas con X-gal, ya que el plásmido sustrato contiene una copia del gen *lacZ*, que se expresa solo cuando es recombinado. Por ello, sectores de color azul en las colonias representaban eventos de recombinación.

Tras obtener unas 40.000 colonias, aquellas de color azul fueron seleccionadas. A partir de ellas obtuvimos los distintos plásmidos portadores de los mutantes correspondientes, que fueron analizados de nuevo en un ensayo de recombinación durante distintas generaciones. Encontramos que uno de los mutantes (al que nos referiremos como TrwC*) mostraba mayores niveles de recombinación que la proteína salvaje.

Antes de continuar con la caracterización de TrwC*, repetimos la estrategia de mutagénesis, igual que anteriormente pero en este caso utilizando tanto *trwC* como *trwC** como moldes para la PCR mutagénica. Como previamente, analizamos unas 40.000 colonias y los plásmidos portadores de los mutantes que produjeron colonias azules fueron obtenidos y analizados de nuevo. Aunque alguno de los mutantes mostró niveles de recombinación más altos que TrwC, ninguno de ellos llegó a los niveles obtenido con TrwC*, por lo que decidimos continuar adelante centrándonos en este primer mutante seleccionado.

TrwC* fue analizado primero en conjugación, por complementación de un R388 deficiente en el gen *trwC*. La frecuencia de conjugación obtenida con TrwC* fue similar a la obtenida con la proteína salvaje, indicando que TrwC* es totalmente funcional en conjugación. Para descartar la posibilidad de que el efecto observado en recombinación se debiera a alguna mutación en el vector pET3a, el gen *trwC** fue re-clonado y analizado de nuevo en un ensayo de recombinación. Los resultados mostraron mayores

niveles de recombinación obtenidos con TrwC*, estadísticamente significativos cuando se compararon con los valores obtenido con TrwC (Figura 42). Este aumento fue observado tanto en presencia como en ausencia de TrwA, sugiriendo que es debido a un aumento intrínseco de la actividad catalítica y no a la adquisición de cierta independencia de TrwA.

Al secuenciar *trwC**, inicialmente encontramos 3 mutaciones a nivel de secuencia nucleotídica, que producían un único cambio a nivel de proteína: A525P. Sin embargo, tras secuenciar el gen salvaje clonado en diversos plásmidos nos percatamos de que esas mutaciones estaban también presentes en él. Atribuimos entonces las supuestas mutaciones a fallos en la secuencia de *trwC* depositada en la base de datos GenBank (NCBI). La secuencia corregida ha sido enviada para su sustitución.

Para ver si la causa de las diferencias catalíticas observadas era resultado de una diferencia en la secuencia de la región promotora responsable de la expresión de $trwC^{(*)}$, secuenciamos dicha región en ambos plásmidos. En la kilobase localizada "aguas arriba" de los genes no encontramos ninguna diferencia entre ambos.

A continuación analizamos la cantidad de ADN (mediante electroforesis en gel de agarosa) y la cantidad de proteína (mediante western blot anti-TrwC) tanto para TrwC como para TrwC*. Encontramos que, aunque la cantidad de ADN era la misma en ambos casos, una mayor cantidad de proteína estaba presente en los lisados de células expresando TrwC* (Figura 43). Esto parece indicar que obtenemos mayores niveles de recombinación porque hay más cantidad de proteína. Es decir, TrwC* no es lo que estábamos buscando, no es un mutante con actividad recombinasa aumentada. Por esa razón, en este punto decidimos no continuar con su caracterización.

Aparte de TrwC*, tras analizar unas 80.000 colonias, no encontramos ningún otro mutante que mostrara un aumento significativo de la actividad recombinasa. Esto puede estar indicando que no es posible obtener dichos mutantes introduciendo solamente uno o dos cambios.

Dado que otro de los objetivos de este trabajo era la búsqueda de otras relaxasas conjugativas con actividad recombinasa, centramos nuestra atención en la anteriormente estudiada TraI del plásmido F. Encontramos que las colonias portadoras del sustrato de recombinación de TraI tardan más en mostrar color azul que aquéllas que tienen el sustrato de TrwC. Sabiendo esto, analizamos su actividad recombinasa de manera similar a como se había analizado TrwC. Encontramos que, en contra de lo que se había publicado previamente, TraI tiene actividad recombinasa sitio-específica. Sin

embargo, la eficiencia con la que cataliza la reacción es mucho más baja que la que muestra TrwC bajo las mismas condiciones de ensayo (unas 10 veces menor).

Comprobamos también si TraI era capaz de catalizar la recombinación cuando se expresaba *en trans* el resto de la maquinaria conjugativa de F por complementación de un plásmido carente del gen *traI*. Vimos que si que es capaz de catalizar la recombinación en dichas condiciones y que, además, cuando *traI* se expresa bajo un promotor fuerte (produciéndose mayor cantidad de proteína) la cantidad de plásmido recombinado es mayor. Este resultado respalda lo observado con TrwC*: si hay más proteína, el nivel de recombinación es mayor.

Aunque TraI del plásmido pKM101 también había sido descrita como deficiente en recombinación, analizamos su plásmido sustrato para ver si ocurría lo mismo que con el de TraI de F. En este caso, encontramos que aunque recombinado, las colonias portadoras no se volvían azules. Por ello, analizamos su actividad recombinasa pero por la pérdida de resistencia a la kanamicina (lo que también ocurre tras la recombinación). De nuevo encontramos que TraI de pKM101 es también una recombinasa sitio-específica, aunque mucho menos eficiente que TrwC (aproximadamente un 6 % de colonias recombinantes con TraI frente a casi el 100 % con TrwC).

Por tanto, TraI de los plásmidos F y pKM101 pueden, como TrwC, catalizar la recombinación entre dos oriTs. Sin embargo, la eficiencia en ambos casos es mucho más baja. Esto puede deberse, como hemos visto, a diferencias en la cantidad de proteína (se sabe que TrwC es muy abundante en la célula) o a diferencias bioquímicas de las relaxasas, como puede ser la afinidad por el *oriT* o la interacción con la maquinaria de replicación.

Estos últimos resultados indican que, dentro de las relaxasas conjugativas, existe una sección de ellas que muestra actividad recombinasa. La razón para que esta actividad innecesaria para la conjugación esté presente en las relaxasas no está clara. Puede ser que esté implicada en la resolución de multímeros, que generan inestabilidad plasmídica. Podría ser también el reflejo de un papel de las relaxasas conjugativas en la evolución de los plásmidos, la formación temporal de cointegrados puede ser un mecanismo de adaptación de algunos plásmidos a determinadas condiciones del ambiente (por ejemplo, la presencia de un antibiótico).

2. Reacción de integración sitio-específica mediada por relaxasas conjugativas.

La reacción de integración es analizada mediante la movilización de un plásmido suicida resistente a cloranfenicol, que contiene el *oriT* correspondiente de la relaxasa bajo estudio y que no es capaz de replicar en la célula receptora. Además de dicho plásmido suicida, la célula donadora contiene un segundo plásmido portador de toda la maquinaria conjugativa necesaria (incluida la relaxasa), mientras que la célula receptora porta un plásmido con la segunda copia del *oriT*, que va a actuar como receptora de la integración. Los integrantes son seleccionados en cloranfenicol, ya que si la célula receptora se vuelve resistente es porque el plásmido suicida se ha integrado tras ser movilizado desde la donadora (Figura 38a). La integración se confirma después mediante análisis molecular del cointegrado.

La integración mediada por TrwC puede ser también analizada movilizando el plásmido suicida por otra relaxasa (TraI del plásmido RP4, ya que el plásmido suicida contiene también una copia del *oriT* de dicho plásmido) y produciendo TrwC en la célula receptora. En esta situación, TrwC es capaz de localizar ambas copias de su *oriT* en la célula receptora y catalizar la integración (Figura 38c).

Trabajo previo en el laboratorio había analizado la integración mediada por TrwC en *oriTs* mutantes cuando la relaxasa era expresada en la receptora (Agundez, 2011). Los mutantes tenía afectada las regiones necesarias para la unión y el corte de TrwC. En esa situación, no se detectó integración en ninguno de los mutantes, indicando que dichas regiones son necesarias para completar la integración.

El ensayo de integración movilizando TrwC junto al DNA desde la donadora nos permitiría analizar los requerimientos de ADN del inicio y el final de la reacción de integración. Además, de cara a un posible uso de TrwC como herramienta de modificación genética, la situación sería ésta última: TrwC llegaría unida a ADN e integraría en *pseudo-oriTs*. Por ello, nos propusimos analizar la integración en *oriTs* mutantes cuando TrwC llegaba en complejo covalente con el ADN desde la donadora. Los mutantes utilizados fueron los descritos previamente en (Lucas *et al.*, 2010): mutIR, que tiene alterada la secuencia de la repetición invertida 2 (IR_2), necesaria para la unión de TrwC; y mut(23-25), en el que se cambiaron los tres nucleótidos localizados en 5´del sitio *nic*, donde TrwC corta el *oriT*.

Se realizaron ensayos en los que la donadora utilizada portaba un plásmido que codifica la maquinaria conjugativa de R388 (sin *oriT*) y el plásmido suicida que contiene el *oriT*. De este modo, TrwC es movilizada desde la donadora como parte de un complejo

nucleoproteico con el plásmido suicida, covalentemente unido a TrwC por el sitio *nic*. La receptora portaba el plásmido que contiene el *oriT* receptor, bien el salvaje o los mutantes mutIR o mut23-25.

En este caso, sí que detectamos integración en ambos mutantes (Tabla 15). De hecho, solo observamos un moderado descenso en la frecuencia de integración al compararlas con la obtenida en el *oriT* salvaje. Este resultado sugiere que los requerimientos de ADN que muestra TrwC para el inicio y la terminación de la reacción de integración son distintos, siendo menos estrictos en el final de la reacción, lo que queda reflejado por el hecho de que TrwC pueda integrar en los *oriTs* mutantes cuando el primer *oriT* ha sido ya procesado por ella en la célula donadora, pero no pueda realizar el corte inicial en ADN superenrollado cuando el *oriT* contiene mutaciones en nucleótidos críticos para la unión y el corte.

Estos resultados indican, desde un punto de vista biotecnológico, que TrwC sería capaz de integrar en secuencias que se desvíen notablemente de su secuencia diana cuando llega a la célula a modificar ya unida a ADN.

Por otro lado, como parte de la caracterización de TrwC*, analizamos su actividad integrasa. Para ello, utilizamos un ensayo de integración en el que movilizamos el plásmido suicida con la maquinaria del plásmido RP4 a una célula receptora con dos plásmidos: el portador del *oriT* que actúa como receptor de la integración y otro para la expresión de TrwC$^{(*)}$ (+/- TrwA). A diferencia de lo que ocurre en recombinación, en integración no encontramos diferencias significativas entre TrwC y TrwC*, lo que indica que la cantidad de TrwC no es limitante para la eficiencia de la reacción de integración.

Como durante este trabajo hemos descubierto que TraI del plásmido F tiene actividad recombinasa sobre el *oriT*, decidimos probar su actividad como integrasa en un ensayo similar al diseñado para TrwC. En este caso, la maquinaria conjugativa presente en la célula donadora es la de F y tanto el plásmido suicida como el presente en la célula receptora tienen un copia del *oriT* de F. Bajo las mismas condiciones experimentales utilizadas para TrwC, no hemos detectado integración mediada por TraI de F (Tabla 17).

Hay que tener en cuenta que TrwC tiene una eficiencia del 100 % como recombinasa, mientras TraI es aproximadamente 10 veces menos eficiente como recombinasa. Se podría pensar que TraI tuviera actividad integrasa pero que no fuéramos capaces de detectarla bajo nuestras condiciones de ensayo. Sin embargo, considerando que la frecuencia de integración de TrwC es del orden de 10^{-5}, y que el nivel de detección del ensayo es del orden de 10^{-9} eventos de integración por donador,

aunque TraI fuese una integrasa 10.000 veces menos eficiente que TrwC, seríamos capaces de detectarlo.

Esto indicaría que aunque tanto TrwC como TraI son capaces de recombinar dos copias en tándem de su respectivo *oriT*, TraI no catalizaría la integración del ADN en la célula receptora, bajo las condiciones experimentales en la que TrwC sí lo hace. Esta diferencia podría ser debida a diferencias intrínsecas entre ambas relaxasas, como puede ser el distinto número de residuos Tyr catalíticos (Grandoso *et al.*, 2000; Larkin *et al.*, 2007).

Nuestra hipótesis es que esta actividad integrasa de TrwC podría proveer a R388 de la posibilidad de diseminarse en la naturaleza, incluso cuando es movilizado a una célula receptora en la que no puede replicar, situación en la que TrwC podría integrar el plásmido en el genoma de la célula receptora si existiese alguna secuencia diana potencial. Si ese fuera el caso, tendría sentido que TraI careciera de actividad integrasa ya que F es un plásmido conjugativo de estrecho rango de hospedador que solo conjuga entre cepas de *Escherichia coli* y otras enterobacterias relacionadas donde es capaz de replicar, mientras que R388 es un plásmido de mayor rango de hospedadores.

3. Caracterización de la reacción de integración mediada por TrwC en eucariotas.

Con el objetivo de analizar la reacción de integración mediada por TrwC en un ambiente celular eucariota, centramos nuestra atención en la levadura *Saccharomyces cerevisiae* como organismo modelo. Nuestra idea era poner a punto un ensayo similar al utilizado en bacterias pero utilizando dicha levadura como célula receptora. Para ello, el primer requisito necesario era la conjugación de derivados de R388 a levaduras.

Trabajo previo en el laboratorio (Larrea, 2013) no había detectado transferencia de R388 por conjugación de *E. coli* a la cepa de levaduras L40 en las condiciones en las que RP4 sí conjuga. Sin embargo, el descubrimiento reciente de una cepa de levadura altamente receptiva en conjugación nos animó a volver a intentarlo.

Moriguchi y colaboradores publicaron recientemente un trabajo en el que analizaron una colección de 4.823 mutantes de levaduras como receptores en conjugación (Moriguchi *et al.*, 2013). Encontraron dos tipos de mutantes altamente receptivos: el mutante *ssd1* (Ssd1 es una proteína que une RNA y está implicada en diversos procesos celulares) y mutantes en la cadena respiratoria, conocidos como mutantes "*petite*" por su característico menor tamaño.

Para probar la conjugación de R388 a estos mutantes, construimos dos plásmidos lanzadera capaces de replicar y ser seleccionados tanto en bacterias como en levaduras, uno de ellos con el *oriT* de R388 y el otro con el de RP4. Estos plásmidos fueron inicialmente conjugados desde *E. coli* al mutante *ssd1*. Para el caso del plásmido movilizado por RP4 obtuvimos una frecuencia de $4{,}2x10^{-4}$ transconjugantes por receptor, similar a la obtenida por Moriguchi *et al*. Sin embargo, no obtuvimos tranconjugantes para el plásmido movilizado por R388.

A partir del mutante en *ssd1* obtuvimos mutantes "*petite*" ya que los dobles mutantes habían sido descritos como aún más receptivos. Obtuvimos varios mutantes que fueron ensayados como receptores en la conjugación de los plásmidos mencionados anteriormente. Una vez más, fuimos capaces de reproducir los resultados publicados para RP4, obteniendo frecuencias entre $1{,}6x10^{-3}$ y $4{,}4x10^{-4}$. Sin embargo, no se detectó conjugación en el sistema de R388. Esto nos hizo descartar a la levadura como modelo de estudio de las reacciones que TrwC realiza en la célula receptora eucariota.

Decidimos entonces analizar la reacción de integración directamente en células humanas. Como se ha mencionado en la introducción, trabajo previo en el laboratorio había demostrado que complejos TrwC-ADN pueden ser transferidos a células humanas desde *Bartonella henselae* a través de su T4SS VirB/D4 (Fernandez-Gonzalez *et al.*, 2011). Además, se había visto que TrwC tenía la capacidad de catalizar la integración usando como dianas secuencias presentes en el genoma humano (Agundez *et al.*, 2012).

Para ensayar la transferencia e integración de ADN a células humanas, se utilizaron plásmidos movilizables que codifican para las siguientes relaxasas:

- TrwC.
- La proteína de fusión TrwC:BID. BID es la señal de translocación reconocida por el T4SS VirB/D4, presente en los efectores de *Bartonella* que son translocados de manera natural a través de él.
- Mob:BID. La relaxasa Mob procede de un plásmido natural críptico de *Bartonella*, y también se había demostrado que puede llevar ADN a células humanas a través del T4SS de *B. henselae*, especialmente cuando se fusionaba a la señal de translocación BID (Schroder *et al.*, 2011).
- Un plásmido sin relaxasa para utilizar como control negativo.

Con el fin de seleccionar eventos de integración en las células humanas, añadimos a todos ellos el gen de resistencia al antibiótico G418.

Como líneas celulares receptoras se utilizaron tanto Ea.hy926 (células diana naturales para *B. henselae*) como HeLa, previamente descrita en modelos de infección

como huésped menos efectivo (Truttmann *et al.*, 2011). Para comparar el patrón de integración mediada por TrwC cuando su diana natural está presente o no en el genoma de las células a modificar, transfectamos células HeLa de manera estable con un plásmido que contiene una copia salvaje del *oriT* de R388. Las colonias resistentes obtenidas fueron crecidas juntas para obtener una mezcla celular, con el *oriT* integrado al azar en distintas localizaciones genómicas.

Bartonella portando los distintos plásmidos fue utilizada para infectar las distintas células humanas, con o sin *oriT*. Tres días tras la infección, la transferencia de ADN fue analizada detectando expresión de GFP por citometría de flujo, como se había descrito previamente. Como era esperable, no se detectó transferencia de ADN con el plásmido que carece de relaxasa. Con los demás plásmidos (*trwC*, *trwC:BID*, *mob:BID*) sí se detectó transferencia, unas 6 y 11 veces mayor con TrwC y TrwC:BID, respectivamente, que con Mob:BID (Figura 60).

Tras confirmar la transferencia de ADN, llevamos a cabo una selección de aquellas células que habían integrado de manera estable el ADN movilizado, por tratamiento con G418 durante 4-5 semanas. Sorprendentemente, tras dicha selección, todas las células que habían recibido el plásmido con Mob:BID habían muerto, mientras que sí que obtuvimos células resistentes a G418 en los casos de TrwC y TrwC:BID (Tabla 20 y Figura 63). Este resultado indica que, bajo nuestras condiciones experimentales, hay una diferencia en la actividad de Mob y TrwC una vez ambas han llegado a la célula humana unidas a ADN. La eficiencia de integración de TrwC es, al menos, 10^3 veces mayor que la de Mob.

Tanto para TrwC como para TrwC:BID, las células resistentes fueron crecidas juntas para obtener poblaciones heterogéneas de células en las que caracterizar los sitios de integración. La presencia del plásmido movilizado en el pool celular resultante fue detectada por PCR y citometría de flujo. El porcentaje de células positivas para la expresión de GFP fue de aproximadamente el 11 %, lo que puede indicar que en muchos casos se habían integrado fragmentos de plásmido que contenían el gen de resistencia a G418 pero no el gen de la GFP, o que el ADN se había integrado en regiones silenciadas del genoma.

Los sitios de integración fueron caracterizados por LAM-PCR (“linear amplification mediated PCR”). En resumen, consiste en la amplificación por PCR de la uniones plásmido-ADN genómico que han ocurrido por el sitio *nic,* basada en una amplificación lineal inicial utilizando un oligonucleótido específico del plásmido, cuya secuencia es conocida. Tras el protocolo se obtienen una serie de fragmentos de ADN, cada uno

representativo de un sitio de integración concreto, que son secuenciados utilizando técnicas de secuenciación masiva.

Si la integración del plásmido ocurrió mediada por TrwC, lo esperado sería que ocurriera por el sitio *nic*, donde TrwC corta el *oriT*. Sin embargo, tras llevar a cabo el protocolo, encontramos que los fragmentos obtenidos se correspondían, en tamaño, con lo esperado para DNA plasmídico y no para uniones plásmido-ADN genómico ocurridas por el sitio *nic*, lo que podría indicar que el plásmido se había integrado por otro sitio diferente al *nic*.

Tras la secuenciación masiva, en la gran mayoría de las 2.000.000 de lecturas obtenidas se secuenció el sitio *nic* y, a continuación, en vez de ADN genómico, el final del *oriT* y más secuencia plasmídica. Esto confirma que, en contra de lo esperado para una integración mediada por TrwC, el plásmido no se integró por el sitio *nic* en la mayoría de los eventos producidos. Así mismo, este resultado demuestra que TrwC es activa en células humanas, puesto que ha recircularizado el *oriT* que entró en la célula humana unido a TrwC por el sitio *nic*.

Las 11.317 secuencias restantes sí que se pudieron mapear en el genoma humano, resultando en 42 posibles sitios de integración. Debido al alto número de falsos positivos encontrados, decidimos descartar todas aquellas secuencias que tuvieran menos del 98 % de homología con el genoma humano y aquellas que no ocurrieron exactamente por el sitio *nic*. Tras esto, quedaron dos posibles eventos de integración sitio-específica. El análisis de uno de ellos demostró que la unión con el ADN genómico había ocurrido en el nucleótido +1 del sitio *nic*, lo que probablemente indica que no es un evento de integración mediado por TrwC. El evento restante ha sido caracterizado y sí refleja una integración sitio-específica mediada por TrwC. Los ocho nucleótidos localizados en 5´ del sitio de integración son idénticos a los ocho nucleótidos presentes en 5´ del sitio *nic* en el *oriT* de R388. El sitio de integración está localizado en el gel 13, en una región intergénica y a 1.352 kb del gen más cercano.

Los resultados obtenidos con LAM-PCR indican que la gran mayoría de secuencias corresponden al ADN no cortado en el sitio *nic*. Como el ADN sí que llega a la célula humana cortado por este punto, esto indica que, una vez en la célula humana, TrwC recircularia el plásmido, mimetizando su función en la célula receptora durante la conjugación bacteriana, lo que nos confirma que TrwC es activa en células humanas. Sin embargo, una vez recircularizado, el plásmido podría integrase por cualquier punto al azar.

La detección de un evento de integración sitio-específica indica que TrwC sí es capaz de catalizar dicha integración en células humanas. Sin embargo, la eficiencia con la que lo hace es tan baja que esos eventos quedan diluidos entre todos los inespecíficos. Por tanto, TrwC no es útil como integrasa sitio-específica para la modificación genética de células humanas.

Sin embargo, el hecho de que, bajo nuestras condiciones experimentales, obtuviéramos células resistentes a G418 con TrwC y TrwC:BID pero no con Mob:BID parece indicar que TrwC desempeña alguna función en la célula humana que facilita luego la integración inespecífica del ADN. TrwC podría estar protegiendo los extremos del ADN introducido, de manera que éste puede persistir periodos más largos de tiempo en la célula humana, lo que facilitaría su integración.

Perspectivas

Con este trabajo hemos aumentado nuestro conocimiento sobre las relaxasas conjugativas como recombinasas e integrasas sitio-específicas, y hemos estudiado su potencial como herramientas de modificación genética en células humanas.

La gran mayoría de relaxasas conjugativas pertenecen a la familia de proteínas HUH. Dicha familia de proteínas está definida por la presencia de una serie de motivos catalíticos y sus miembros son endonucleasas que llevan a cabo reacciones de transferencia de cadena en sustratos de ADN de cadena sencilla requeridas en procesos biológicos como la conjugación bacteriana, la replicación de circulo rodante o la transposición (Chandler *et al.*, 2013).

Se ha demostrado que algunas de estas proteínas HUH catalizan, además, reacciones de recombinación e integración sitio-específica en sustratos de ADN de doble cadena y pueden ser consideradas, por tanto, como recombinasas sitio-específicas. Nuestros resultados han añadido dos nuevos miembros a la familia HUH de recombinasas sitio-específicas. Las proteínas HUH pueden ser, por tanto, una familia de proteínas seleccionadas a lo largo de la evolución para catalizar recombinación o integración sitio-específica aparte de su papel en replicación viral, conjugación bacteriana o transposición. Existen evidencias de que esta estrategia se ha llevado a cabo tanto en bacterias como en plantas y virus de mamíferos (Gonzalez-Prieto *et al.*, 2013), lo que apoya la idea de que está actividad recombinase puede aportar una ventaja evolutiva a los plásmidos que contiene una relaxasa con actividad recombinase.

Esta actividad recombinasa/integrasa y el hecho de que lleguen a una célula receptora en forma de complejo covalente con el ADN de interés hace que estas relaxasas tengan un alto potencial como herramienta biotecnológica (Gonzalez-Prieto *et al.*, 2013).

Sin embargo, los resultados obtenidos con TrwC en células humanas indican que, en ese contexto, no es capaz de actuar como integrasa sitio-específica de manera eficiente. A pesar de ello, todo parece indicar que TrwC estabiliza el ADN foráneo, potenciando su integración al azar. Esta habilidad de TrwC puede ser también muy interesante si se coordina con la acción de una nucleasa sitio-específica que fomente, mediante la introducción de cortes de doble cadena, la integración del ADN plasmídico por recombinación homóloga, como hacen las nucleasas “zinc-finger”.

Esta estrategia seguiría basada en la introducción de complejos TrwC-ADN a través de T4SS, en este caso acompañados de una nucleasa (a la que se puede añadir la señal BID para que sea también reconocida por el T4SS). Podríamos entonces conseguir transferencia e integración sitio-específica del ADN de interés (Figura 69), con un sistema que aportara las siguiente ventajas sobre otros utilizados hoy en día:

- Como la transferencia está basada en un sistema conjugativo, se podrían transferir e integrar moléculas de ADN de cualquier longitud y secuencia, con el único requerimiento del *oriT*.
- Tanto en el caso de TrwC como de la nucleasa, las propias proteínas pueden ser translocadas a través del T4SS, lo que evitaría la expresión de genes potencialmente tóxicos en la célula humana.
- Si a la hora de construir el ADN plasmídico de interés, los genes a integrar se clonan flanqueados por dos copias del *oriT*, todo lo que quede fuera de esa región no será transferido a la célula humana, evitando la presencia en ella de ADN bacteriano.
- La utilización de distintos patógenos (con distintos T4SS) que infectan tipos celulares concretos permitiría el acceso *in vivo* al tejido de interés de manera específica.

Los resultados obtenidos en la última parte de este trabajo indican que, aunque no de la manera esperada, las relaxasas conjugativas siguen teniendo un gran potencial para su uso aplicado en la modificación genética. Su complementación con otras proteínas podría resultar en el establecimiento de un sistema con grandes ventajas para su uso en el campo de la biotecnología.

Anexo 1. Inserción frecuente de IS10 en la región promotora del gen *cat* de vectores ampliamente utilizados.

Durante este trabajo hemos encontrado que distintos plásmidos, todos ellos codificando el gen *cat* y derivados de los vectores de las familias pSU/pSW, habían sufrido inserción de IS10, todos ellos de manera independiente y en el mismo punto, a 44 pares de bases del inicio del gen *cat*. Todos los plásmidos en los que hemos encontrado dicho fenómeno presentan algún tipo de potencial inestabilidad, como puede ser la presencia de dos copias del *oriT* en los plásmidos sustrato de la recombinación. Tras caracterizar el sitio de integración, comprobamos que no se corresponde con la secuencia diana consenso de IS10, por lo que hemos identificado una nueva diana.

Diferentes plásmidos fueron crecidos a distintas concentraciones de Cm y comprobamos que aquellos plásmidos “inestables” adquirían IS10 solo al aplicar altas concentraciones de Cm, mientras que los plásmidos estables no lo hacían ni a la concentración más alta.

La presencia de promotores en IS10 nos lleva a sugerir que dicha integración produce mayores niveles de Cat, lo que podría compensar, de alguna manera, el metabolismo más lento de esos plásmidos inestables. Por ello, y para evitar la integración de IS10 no deseada, recomendamos que, siempre que sea posible, se utilicen concentraciones de Cm bajas cuando se trabaje con plásmidos derivados de pSU/pSW.

Conclusiones

1. Hemos puesto a punto un protocolo de mutagénesis por PCR y un sistema para la detección de eventos de recombinación *in vivo* para la búsqueda de variantes de TrwC con actividad recombinasa *oriT*-específica aumentada.

2. Una variante con actividad recombinasa aumentada (TrwC*) fue seleccionada. Sin embargo, el gen no tiene mutaciones respecto al salvaje. Encontramos que el aumento en la recombinación se debe, probablemente, a un aumento en la cantidad de proteína en la célula. Esta mayor cantidad de proteína no produce ningún efecto en la integración mediada por TrwC.

3. Hemos identificado tres errores en la secuencia de *trwC* disponible en la base de datos GenBank. Hemos solicitado su corrección.

4. Las relaxasas TraI de los plásmidos F y pKM101 son capaces de catalizar la recombinación entres dos copias de sus respectivos *oriTs*, en contra de lo que se había publicado anteriormente. Su eficiencia como recombinasas es muy baja comparada con la de TrwC bajo las mismas condiciones de ensayo. Cuando F_traI es producida en mayor cantidad, se produce un aumento de la frecuencia de recombinación.

5. TrwC covalentemente unida al ADN puede catalizar reacciones de transferencia de cadena a *oriTs* con mutaciones en nucleótidos relevantes para la unión y el corte por TrwC, con solo un moderado descenso en la frecuencia de integración comparada con la frecuencia en el *oriT* salvaje.

6. No hemos detectado integración mediada por F_TraI bajo las mismas condiciones utilizadas para TrwC.

7. No hemos detectado transferencia conjugativa de R388 a mutantes de levadura *ssd1* y *ssd1 "petite"* bajo las condiciones de ensayo en las que sí detectamos transferencia de RP4.

8. Hemos detectado transferencia de ADN mediada por TrwC a través del T4SS VirB/D4 de *B. henselae* a células HeLa. Sin embargo, hay una disminución significativa de la tasa de transferencia comparada con la obtenida a células EA.hy926. La transferencia de ADN es 10 veces más eficiente cuando el ADN es movilizado por TrwC o TrwC:BID que por Mob:BID.

9. Hemos obtenido pools celulares portadores de eventos de integración de plásmidos movilizados por TrwC y TrwC:BID desde *B. henselae* a células HeLa, HeLa::*oriT* y EA.hy926. No obtuvimos integración estable del plásmido movilizado por Mob:BID bajo las mismas condiciones de ensayo.

10. La tasa de integración de ADN mediada por TrwC es, al menos, 10^3 veces mayor que la mediada por Mob:BID.

11. El análisis del patrón de integración mediada por TrwC y TrwC:BID mediante LAM-PCR sugiere que la integración es mayoritariamente no sitio-específica.

12. La gran mayoría de las copias de plásmido movilizado por TrwC integradas en el genoma humano muestran recircularización previa del plásmido, lo que confirma que TrwC es activa en las células humanas.

13. Al menos un evento de integración en el genoma humano mediado por TrwC ha sido detectado. Los 8 nucleótidos localizados 5´ al sitio de integración son iguales a aquellos localizados en 5´ del sitio *nic* en el *oriT* de R388.

14. Hemos identificado un sitio recurrente de integración de IS10 en la región del promotor del gen *cat* en los vectores de las familias pSU/pSW. La integración es seleccionada a altas concentraciones de cloranfenicol.

Bibliography

Abdel-Monem, M., Durwald, H. & Hoffmann-Berling, H. (1976). Enzymic unwinding of DNA. 2. Chain separation by an ATP-dependent DNA unwinding enzyme. *Eur J Biochem* **65**, 441-449.

Abdel-Monem, M. & Hoffmann-Berling, H. (1976). Enzymic unwinding of DNA. 1. Purification and characterization of a DNA-dependent ATPase from Escherichia coli. *Eur J Biochem* **65**, 431-440.

Abdel-Monem, M., Taucher-Scholz, G. & Klinkert, M. Q. (1983). Identification of Escherichia coli DNA helicase I as the traI gene product of the F sex factor. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 4659-4663.

Abremski, K. & Hoess, R. (1984). Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *The Journal of biological chemistry* **259**, 1509-1514.

Achtman, M., Manning, P. A., Kusecek, B., Schwuchow, S. & Willetts, N. (1980). A genetic analysis of F sex factor cistrons needed for surface exclusion in Escherichia coli. *J Mol Biol* **138**, 779-795.

Agundez, L. (2011).TrwC, a conjugative relaxase with site-specific integrase activity: Characterization of the reaction and biotechnological potential. In Molecular Biology: University of Cantabria.

Agundez, L., Machon, C., Cesar, C. E., Rosa-Garrido, M., Delgado, M. D. & Llosa, M. (2011). Nuclear targeting of a bacterial integrase that mediates site-specific recombination between bacterial and human target sequences. *Applied and environmental microbiology* **77**, 201-210.

Agundez, L., Gonzalez-Prieto, C., Machon, C. & Llosa, M. (2012). Site-specific integration of foreign DNA into minimal bacterial and human target sequences mediated by a conjugative relaxase. *PloS one* **7**, e31047.

Alperi, A., Larrea, D., Fernandez-Gonzalez, E., Dehio, C., Zechner, E. L. & Llosa, M. (2013). A translocation motif in relaxase TrwC specifically affects recruitment by its conjugative type IV secretion system. *Journal of bacteriology* **195**, 4999-5006.

Alvarez-Martinez, C. E. & Christie, P. J. (2009). Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* **73**, 775-808.

Arutyunov, D. & Frost, L. S. (2013). F conjugation: back to the beginning. *Plasmid* **70**, 18-32.

Asokan, A., Schaffer, D. V. & Samulski, R. J. (2012). The AAV vector toolkit: poised at the clinical crossroads. *Molecular therapy : the journal of the American Society of Gene Therapy* **20**, 699-708.

Avila, P., Nunez, B. & de la Cruz, F. (1996). Plasmid R6K contains two functional oriTs which can assemble simultaneously in relaxosomes in vivo. *J Mol Biol* **261**, 135-143.

Backert, S. & Meyer, T. F. (2006). Type IV secretion systems and their effectors in bacterial pathogenesis. *Current opinion in microbiology* **9**, 207-217.

Barlett, M. M., Erickson, M. J. & Meyer, R. J. (1990). Recombination between directly repeated origins of conjugative transfer cloned in M13 bacteriophage DNA models ligation of the transferred plasmid strand. *Nucleic acids research* **18**, 3579-3586.

Bartolomé, B., Jubete, Y., Martínez, E. & de la Cruz, F. (1991). Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. *Gene* **102**, 75-78.

Bates, S., Cashmore, A. M. & Wilkins, B. M. (1998). IncP plasmids are unusually effective in mediating conjugation of Escherichia coli and Saccharomyces cerevisiae: involvement of the tra2 mating system. *Journal of bacteriology* **180**, 6538-6543.

Becker, D. M. & Guarente, L. (1991). High-efficiency transformation of yeast by electroporation. *Methods in enzymology* **194**, 182-187.

Beranek, A., Zettl, M., Lorenzoni, K., Schauer, A., Manhart, M. & Koraimann, G. (2004). Thirty-eight C-terminal amino acids of the coupling protein TraD of the F-like conjugative resistance plasmid R1 are required and sufficient to confer binding to the substrate selector protein TraM. *Journal of bacteriology* **186**, 6999-7006.

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G. & Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Molecular and cellular biology* **21**, 289-297.

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A. & other authors (1997). The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453-1462.

Boer, R., Russi, S., Guasch, A., Lucas, M., Blanco, A. G., Perez-Luque, R., Coll, M. & de la Cruz, F. (2006). Unveiling the Molecular Mechanism of a Conjugative Relaxase: The Structure of TrwC Complexed with a 27-mer DNA Comprising the Recognition Hairpin and the Cleavage Site. *J Mol Biol*.

Bolivar, F., Betlach, M. C., Heyneker, H. L., Shine, J., Rodriguez, R. L. & Boyer, H. W. (1977). Origin of replication of pBR345 plasmid DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5265-5269.

Bolland, S., Llosa, M., Avila, P. & de la Cruz, F. (1990). General organization of the conjugal transfer genes of the IncW plasmid R388 and interactions between R388 and IncN and IncP plasmids. *Journal of bacteriology* **172**, 5795-5802.

Bowie, J. U. & Sauer, R. T. (1990). TraY proteins of F and related episomes are members of the Arc and Mnt repressor family. *J Mol Biol* **211**, 5-6.

Bradley, D. E. & Cohen, D. R. (1976). Basic characterization of W-pili. *J Gen Microbiol* **97**, 91-103.

Bradley, D. E. (1980). Morphological and serological relationships of conjugative pili. *Plasmid* **4**, 155-169.

Bradley, D. E., Taylor, D. E. & Cohen, D. R. (1980). Specification of surface mating systems among conjugative drug resistance plasmids in Escherichia coli K-12. *Journal of bacteriology* **143**, 1466-1470.

Breg, J. N., van Opheusden, J. H., Burgering, M. J., Boelens, R. & Kaptein, R. (1990). Structure of Arc repressor in solution: evidence for a family of beta-sheet DNA-binding proteins. *Nature* **346**, 586-589.

Buchanan-Wollaston, V., Passiatore, J. E. & Cannon, F. (1987). The mob and oriT mobilization functions of a bacterial plasmid promote its transfer to plants. *Nature* **328**, 172-175.

Bundock, P., den Dulk-Ras, A., Beijersbergen, A. & Hooykaas, P. J. (1995). Trans-kingdom T-DNA transfer from Agrobacterium tumefaciens to Saccharomyces cerevisiae. *The EMBO journal* **14**, 3206-3214.

Byrd, D. R. & Matson, S. W. (1997). Nicking by transesterification: the reaction catalysed by a relaxase. *Molecular microbiology* **25**, 1011-1022.

Byrd, D. R., Sampson, J. K., Ragonese, H. M. & Matson, S. W. (2002). Structure-function analysis of Escherichia coli DNA helicase I reveals non-overlapping transesterase and helicase domains. *The Journal of biological chemistry* **277**, 42645-42653.

Cabezón, E. & de la Cruz, F. (2006). TrwB: an F(1)-ATPase-like molecular motor involved in DNA transport during bacterial conjugation. *Res Microbiol* **157**, 299-305.

Cambronne, E. D. & Roy, C. R. (2006). Recognition and delivery of effector proteins into eukaryotic cells by bacterial secretion systems. *Traffic* **7**, 929-939.

Campbell, J. L., Richardson, C. C. & Studier, F. W. (1978). Genetic recombination and complementation between bacteriophage T7 and cloned fragments of T7 DNA. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 2276-2280.

Campos-Olivas, R., Louis, J. M., Clerot, D., Gronenborn, B. & Gronenborn, A. M. (2002). The structure of a replication initiator unites diverse aspects of nucleic acid metabolism. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10310-10315.

Capecchi, M. R. (1989). Altering the genome by homologous recombination. *Science* **244**, 1288-1292.

Carballeira, J. D., Gonzalez-Perez, B., Moncalian, G. & de la Cruz, F. (2014). A high security double lock and key mechanism in HUH relaxases controls oriT-processing for plasmid conjugation. *Nucleic acids research* **42**, 10632-10643.

Cavalli, L. L., Lederberg, J. & Lederberg, E. M. (1953). An infective factor controlling sex compatibility in Bacterium coli. *J Gen Microbiol* **8**, 89-103.

Cermak, T., Doyle, E. L., Christian, M. & other authors (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research* **39**, e82.

César, C. E., Machón, C., de la Cruz, F. & Llosa, M. (2006). A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences. *Molecular microbiology* **62**, 984-996.

César, C. E. & Llosa, M. (2007). TrwC-mediated site-specific recombination is controlled by host factors altering local DNA topology. *Journal of bacteriology* **189**, 9037-9043.

Chalberg, T. W., Portlock, J. L., Olivares, E. C., Thyagarajan, B., Kirby, P. J., Hillman, R. T., Hoelters, J. & Calos, M. P. (2006). Integration specificity of phage phiC31 integrase in the human genome. *J Mol Biol* **357**, 28-48.

Chandler, M. & Galas, D. J. (1983). Cointegrate formation mediated by Tn9. II. Activity of IS1 is modulated by external DNA sequences. *J Mol Biol* **170**, 61-91.

Chandler, M., de la Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G. & Ton-Hoang, B. (2013). Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nature reviews Microbiology* **11**, 525-538.

Chang, A. C. & Cohen, S. N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *Journal of bacteriology* **134**, 1141-1156.

Chavez, C. L., Keravala, A., Chu, J. N., Farruggio, A. P., Cuellar, V. E., Voorberg, J. & Calos, M. P. (2012). Long-term expression of human coagulation factor VIII in a tolerant mouse model using the phiC31 integrase system. *Human gene therapy* **23**, 390-398.

Cheah, K. C. & Skurray, R. (1986). The F plasmid carries an IS3 insertion within finO. *J Gen Microbiol* **132**, 3269-3275.

Chen, C., Banga, S., Mertens, K., Weber, M. M., Gorbaslieva, I., Tan, Y., Luo, Z. Q. & Samuel, J. E. (2010). Large-scale identification and translocation of type IV secretion substrates by Coxiella burnetii. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21755-21760.

Chen, X. J. & Clark-Walker, G. D. (2000). The petite mutation in yeasts: 50 years on. *International review of cytology* **194**, 197-238.

Cheng, J. K. & Alper, H. S. (2014). The genome editing toolbox: a spectrum of approaches for targeted modification. *Current opinion in biotechnology* **30C**, 87-94.

Cheng, Y., McNamara, D. E., Miley, M. J., Nash, R. P. & Redinbo, M. R. (2011). Functional characterization of the multidomain F plasmid TraI relaxase-helicase. *The Journal of biological chemistry* **286**, 12670-12682.

Chou, T. B. & Perrimon, N. (1992). Use of a yeast site-specific recombinase to produce female germline chimeras in Drosophila. *Genetics* **131**, 643-653.

Choulika, A., Perrin, A., Dujon, B. & Nicolas, J. F. (1995). Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Molecular and cellular biology* **15**, 1968-1973.

Christie, P. J. & Vogel, J. P. (2000). Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends in microbiology* **8**, 354-360.

Christie, P. J. (2001). Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Molecular microbiology* **40**, 294-305.

Christie, P. J. & Cascales, E. (2005). Structural and dynamic properties of bacterial type IV secretion systems (review). *Mol Membr Biol* **22**, 51-61.

Chung, C. T. & Miller, R. H. (1988). A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucleic acids research* **16**, 3580.

Clewell, D. B. & Helinski, D. R. (1969). Supercoiled circular DNA-protein complex in Escherichia coli: purification and induced conversion to an open circular DNA form. *Proceedings of the National Academy of Sciences of the United States of America* **62**, 1159-1166.

Colosimo, A., Goncz, K. K., Holmes, A. R., Kunzelmann, K., Novelli, G., Malone, R. W., Bennett, M. J. & Gruenert, D. C. (2000). Transfer and expression of foreign genes in mammalian cells. *Biotechniques* **29**, 314-318, 320-312, 324 passim.

Coluccio, A., Miselli, F., Lombardo, A. & other authors (2013). Targeted gene addition in human epithelial stem cells by zinc-finger nuclease-mediated homologous recombination. *Molecular therapy : the journal of the American Society of Gene Therapy* **21**, 1695-1704.

Cong, L., Ran, F. A., Cox, D. & other authors (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823.

Coulthurst, S. J. (2013). The Type VI secretion system - a widespread and versatile cell targeting system. *Res Microbiol* **164**, 640-654.

Couturier, M., Bex, F., Bergquist, P. L. & Maas, W. K. (1988). Identification and classification of bacterial plasmids. *Microbiological reviews* **52**, 375-395.

Craig, N. L. & Nash, H. A. (1984). E. coli integration host factor binds to specific sites in DNA. *Cell* **39**, 707-716.

Crawford, E. M. & Gesteland, R. F. (1964). The adsorption of bacteriophage R-17. *Virology* **22**, 165-167.

Csitkovits, V. C. & Zechner, E. L. (2003). Extent of single-stranded DNA required for efficient TraI helicase activity in vitro. *The Journal of biological chemistry* **278**, 48696-48703.

Cummins, J., Cronin, M., van Pijkeren, J. P., Gahan, C. G. & Tangney, M. (2014). Bacterial systems for gene delivery to systemic tumors. *Methods in molecular biology* **1141**, 201-209.

Dalkara, D. & Sahel, J. A. (2014). Gene therapy for inherited retinal degenerations. *Comptes rendus biologies* **337**, 185-192.

Dash, P. K., Traxler, B. A., Panicker, M. M., Hackney, D. D. & Minkley, E. G., Jr. (1992). Biochemical characterization of Escherichia coli DNA helicase I. *Molecular microbiology* **6**, 1163-1172.

Datta, N. & Hedges, R. W. (1971). Compatibility groups among fi - R factors. *Nature* **234**, 222-223.

Datta, N. & Hedges, R. W. (1972). Trimethoprim resistance conferred by W plasmids in Enterobacteriaceae. *J Gen Microbiol* **72**, 349-355.

Datta, S., Larkin, C. & Schildbach, J. F. (2003). Structural insights into single-stranded DNA binding and cleavage by F factor TraI. *Structure* **11**, 1369-1379.

Davies, J. & Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **74**, 417-433.

de la Cruz, F. & Davies, J. (2000). Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends in microbiology* **8**, 128-133.

de Lorenzo, V. & Timmis, K. N. (1994). Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons. *Methods in enzymology* **235**, 386-405.

de Paz, H. D., Sangari, F. J., Bolland, S., Garcia-Lobo, J. M., Dehio, C., de la Cruz, F. & Llosa, M. (2005). Functional interactions between type IV secretion systems involved in DNA transfer and virulence. *Microbiology* **151**, 3505-3516.

de Paz, H. D., Larrea, D., Zunzunegui, S., Dehio, C., de la Cruz, F. & Llosa, M. (2010). Functional dissection of the conjugative coupling protein TrwB. *Journal of bacteriology* **192**, 2655-2669.

Dehio, C. (2005). Bartonella-host-cell interactions and vascular tumour formation. *Nature reviews Microbiology* **3**, 621-631.

Dehio, C. (2008). Infection-associated type IV secretion systems of Bartonella and their diverse roles in host cell interaction. *Cellular microbiology* **10**, 1591-1598.

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

Dempsey, W. B. (1987). Integration host factor and conjugative transfer of the antibiotic resistance plasmid R100. *Journal of bacteriology* **169**, 4391-4392.

Deonier, R. C. & Mirels, L. (1977). Excision of F plasmid sequences by recombination at directly repeated insertion sequence 2 elements: involvement of recA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 3965-3969.

Deonier, R. C., Hadley, R. G. & Hu, M. (1979). Enumeration and identification of IS3 elements in Escherichia coli strains. *Journal of bacteriology* **137**, 1421-1424.

Di Laurenzio, L., Frost, L. S. & Paranchych, W. (1992). The TraM protein of the conjugative plasmid F binds to the origin of transfer of the F and ColE1 plasmids. *Molecular microbiology* **6**, 2951-2959.

Disque-Kochem, C. & Dreiseikelmann, B. (1997). The cytoplasmic DNA-binding protein TraM binds to the inner membrane protein TraD in vitro. *Journal of bacteriology* **179**, 6133-6137.

Dostal, L. & Schildbach, J. F. (2010). Single-stranded DNA binding by F TraI relaxase and helicase domains is coordinately regulated. *Journal of bacteriology* **192**, 3620-3628.

Dostal, L., Shao, S. & Schildbach, J. F. (2011). Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer. *Nucleic acids research* **39**, 2658-2670.

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

Draper, O., Cesar, C. E., Machon, C., de la Cruz, F. & Llosa, M. (2005). Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 16385-16390.

Drew, H. R., Lockett, L. J. & Both, G. W. (2007). Increased complexity of wild-type adeno-associated virus-chromosomal junctions as determined by analysis of unselected cellular genomes. *The Journal of general virology* **88**, 1722-1732.

Du, Z. W., Hu, B. Y., Ayala, M., Sauer, B. & Zhang, S. C. (2009). Cre recombination-mediated cassette exchange for building versatile transgenic human embryonic stem cells lines. *Stem cells* **27**, 1032-1041.

Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M. H. & Chandrasegaran, S. (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic acids research* **33**, 5978-5990.

Dymecki, S. M. (1996). Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6191-6196.

Eberl, L., Kristensen, C. S., Givskov, M., Grohmann, E., Gerlitz, M. & Schwab, H. (1994). Analysis of the multimer resolution system encoded by the parCBA operon of broad-host-range plasmid RP4. *Molecular microbiology* **12**, 131-141.

Ehrhardt, A., Engler, J. A., Xu, H., Cherry, A. M. & Kay, M. A. (2006). Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. *Human gene therapy* **17**, 1077-1094.

Eichenbaum, Z. & Livneh, Z. (1998). UV light induces IS10 transposition in Escherichia coli. *Genetics* **149**, 1173-1181.

Fekete, R. A. & Frost, L. S. (2002). Characterizing the DNA contacts and cooperative binding of F plasmid TraM to its cognate sites at oriT. *The Journal of biological chemistry* **277**, 16705-16711.

Fernandez-Gonzalez, E., de Paz, H. D., Alperi, A., Agundez, L., Faustmann, M., Sangari, F. J., Dehio, C. & Llosa, M. (2011). Transfer of R388 derivatives by a pathogenesis-associated type IV secretion system into both bacteria and human cells. *Journal of bacteriology* **193**, 6257-6265.

Fernandez-Lopez, R., Garcillan-Barcia, M. P., Revilla, C., Lazaro, M., Vielva, L. & de la Cruz, F. (2006). Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS microbiology reviews* **30**, 942-966.

Fiering, S., Epner, E., Robinson, K. & other authors (1995). Targeted deletion of 5'HS2 of the murine beta-globin LCR reveals that it is not essential for proper regulation of the beta-globin locus. *Genes & development* **9**, 2203-2213.

Firth, N., Ippen-Ihler, K. & Skurray, R. (1996). Structure and function of the F factor and mechanism of conjugation. In Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology, pp. 2377-2401: American Society for Microbiology.

Forterre, P. & Gadelle, D. (2009). Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic acids research* **37**, 679-692.

Foster, T. J., Davis, M. A., Roberts, D. E., Takeshita, K. & Kleckner, N. (1981). Genetic organization of transposon Tn10. *Cell* **23**, 201-213.

Francia, M. V. & Clewell, D. B. (2002a). Transfer origins in the conjugative Enterococcus faecalis plasmids pAD1 and pAM373: identification of the pAD1 nic site, a specific relaxase and a possible TraG-like protein. *Molecular microbiology* **45**, 375-395.

Francia, M. V. & Clewell, D. B. (2002b). Amplification of the tetracycline resistance determinant of pAMalpha1 in Enterococcus faecalis requires a site-specific recombination event involving relaxase. *Journal of bacteriology* **184**, 5187-5193.

Francia, M. V., Varsaki, A., Garcillan-Barcia, M. P., Latorre, A., Drainas, C. & de la Cruz, F. (2004). A classification scheme for mobilization regions of bacterial plasmids. *FEMS microbiology reviews* **28**, 79-100.

Francia, M. V., Clewell, D. B., de la Cruz, F. & Moncalian, G. (2013). Catalytic domain of plasmid pAD1 relaxase TraX defines a group of relaxases related to restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 13606-13611.

Frost, L. S., Ippen-Ihler, K. & Skurray, R. A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiological reviews* **58**, 162-210.

Furuya, N. & Komano, T. (2003). NikAB- or NikB-dependent intracellular recombination between tandemly repeated oriT sequences of plasmid R64 in plasmid or single-stranded phage vectors. *Journal of bacteriology* **185**, 3871-3877.

Gabriel, R., Eckenberg, R., Paruzynski, A. & other authors (2009). Comprehensive genomic access to vector integration in clinical gene therapy. *Nature medicine* **15**, 1431-1436.

Gabriel, R., Lombardo, A., Arens, A. & other authors (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* **29**, 816-823.

Gabriel, R., Kutschera, I., Bartholomae, C. C., von Kalle, C. & Schmidt, M. (2014). Linear amplification mediated PCR--localization of genetic elements and characterization of unknown flanking DNA. *Journal of visualized experiments : JoVE*, e51543.

Gaj, T., Gersbach, C. A. & Barbas, C. F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology* **31**, 397-405.

Gao, Q., Luo, Y. & Deonier, R. C. (1994). Initiation and termination of DNA transfer at F plasmid oriT. *Molecular microbiology* **11**, 449-458.

Garcillan-Barcia, M. P., Jurado, P., Gonzalez-Perez, B., Moncalian, G., Fernandez, L. A. & de la Cruz, F. (2007). Conjugative transfer can be inhibited by blocking relaxase activity within recipient cells with intrabodies. *Molecular microbiology* **63**, 404-416.

Garcillán-Barcia, M. P., Francia, M. V. & de la Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS microbiology reviews* **33**, 657-687.

Geider, K. & Hoffmann-Berling, H. (1981). Proteins controlling the helical structure of DNA. *Annu Rev Biochem* **50**, 233-260.

Gennaro, M. L., Kornblum, J. & Novick, R. P. (1987). A site-specific recombination function in Staphylococcus aureus plasmids. *Journal of bacteriology* **169**, 2601-2610.

Ghosh, A., Yue, Y., Lai, Y. & Duan, D. (2008). A hybrid vector system expands adeno-associated viral vector packaging capacity in a transgene-independent manner. *Molecular therapy : the journal of the American Society of Gene Therapy* **16**, 124-130.

Gietz, R. D. & Woods, R. A. (2001). Genetic transformation of yeast. *Biotechniques* **30**, 816-820, 822-816, 828 passim.

Gomis-Rüth, F. X., Moncalián, G., Pérez-Luque, R., González, A., Cabezón, E., de la Cruz, F. & Coll, M. (2001). The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. *Nature* **409**, 637-641.

Gonzalez-Perez, B., Lucas, M., Cooke, L. A., Vyle, J. S., de la Cruz, F. & Moncalian, G. (2007). Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. *The EMBO journal* **26**, 3847-3857.

Gonzalez-Prieto, C., Agundez, L., Linden, R. M. & Llosa, M. (2013). HUH site-specific recombinases for targeted modification of the human genome. *Trends in biotechnology* **31**, 305-312.

Grahn, A. M., Haase, J., Bamford, D. H. & Lanka, E. (2000). Components of the RP4 conjugative transfer apparatus form an envelope structure bridging inner and outer membranes of donor cells: implications for related macromolecule transport systems. *Journal of bacteriology* **182**, 1564-1574.

Grandoso, G. (1994).Funciones de las DNA helicasas en la conjugación bacteriana: purificación y caracterización de TrwC. In Department of Molecular Biology: University of Cantabria.

Grandoso, G., Llosa, M., Zabala, J. C. & de la Cruz, F. (1994). Purification and biochemical characterization of TrwC, the helicase involved in plasmid R388 conjugal DNA transfer. *Eur J Biochem* **226**, 403-412.

Grandoso, G., Avila, P., Cayón, A., Hernando, M. A., Llosa, M. & de la Cruz, F. (2000). Two active-site tyrosyl residues of protein TrwC act sequentially at the origin of transfer during plasmid R388 conjugation. *J Mol Biol* **295**, 1163-1172.

Grant, S. G., Jessee, J., Bloom, F. R. & Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 4645-4649.

Grasseschi, H. A. & Minnick, M. F. (1994). Transformation of Bartonella bacilliformis by electroporation. *Canadian journal of microbiology* **40**, 782-786.

Grau, J., Boch, J. & Posch, S. (2013). TALENoffer: genome-wide TALEN off-target prediction. *Bioinformatics* **29**, 2931-2932.

Grohmann, E., Muth, G. & Espinosa, M. (2003). Conjugative plasmid transfer in gram-positive bacteria. *Microbiol Mol Biol Rev* **67**, 277-301.

Gronostajski, R. M. & Sadowski, P. D. (1985). The FLP protein of the 2-micron plasmid of yeast. Inter- and intramolecular reactions. *The Journal of biological chemistry* **260**, 12328-12335.

Guasch, A., Lucas, M., Moncalián, G., Cabezas, M., Pérez-Luque, R., Gomis-Rüth, F. X., De La Cruz, F. & Coll, M. (2003). Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. *Nat Struct Biol* **10**, 1002-1010.

Guogas, L. M., Kennedy, S. A., Lee, J. H. & Redinbo, M. R. (2009). A novel fold in the TraI relaxase-helicase c-terminal domain is essential for conjugative DNA transfer. *J Mol Biol* **386**, 554-568.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M. & other authors (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415-419.

Haft, R. J., Palacios, G., Nguyen, T., Mally, M., Gachelet, E. G., Zechner, E. L. & Traxler, B. (2006). General Mutagenesis of F Plasmid TraI Reveals Its Role in Conjugative Regulation. *Journal of bacteriology* **188**, 6346-6353.

Haft, R. J., Gachelet, E. G., Nguyen, T., Toussaint, L., Chivian, D. & Traxler, B. (2007). In vivo oligomerization of the F conjugative coupling protein TraD. *Journal of bacteriology* **189**, 6626-6634.

Hakkaart, M. J., van den Elzen, P. J., Veltkamp, E. & Nijkamp, H. J. (1984). Maintenance of multicopy plasmid Clo DF13 in E. coli cells: evidence for site-specific recombination at parB. *Cell* **36**, 203-209.

Halling, S. M. & Kleckner, N. (1982). A symmetrical six-base-pair target site sequence determines Tn10 insertion specificity. *Cell* **28**, 155-163.

Hanai, R. & Wang, J. C. (1993). The mechanism of sequence-specific DNA cleavage and strand transfer by phi X174 gene A* protein. *The Journal of biological chemistry* **268**, 23830-23836.

Heinemann, J. A. & Sprague, G. F., Jr. (1989). Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* **340**, 205-209.

Hekman, K., Guja, K., Larkin, C. & Schildbach, J. F. (2008). An intrastrand three-DNA-base interaction is a key specificity determinant of F transfer initiation and of F TraI relaxase DNA recognition and cleavage. *Nucleic acids research* **36**, 4565-4572.

Henckaerts, E., Dutheil, N., Zeltner, N. & other authors (2009). Site-specific integration of adeno-associated virus involves partial duplication of the target locus. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7571-7576.

Hernando, M. A. (2000). " Bioquímica de las reacciones de transferencia de cadenas de DNA catalizadas por la proteína TrwC". *PhD thesis Santander: Universidad de Cantabria*.

Hickman, A. B., Ronning, D. R., Kotin, R. M. & Dyda, F. (2002). Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. *Mol Cell* **10**, 327-337.

High, K. H., Nathwani, A., Spencer, T. & Lillicrap, D. (2014). Current status of haemophilia gene therapy. *Haemophilia : the official journal of the World Federation of Hemophilia* **20 Suppl 4**, 43-49.

Hockemeyer, D., Wang, H., Kiani, S. & other authors (2011). Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29**, 731-734.

Hofmann, A., Kessler, B., Ewerling, S. & other authors (2003). Efficient transgenesis in farm animals by lentiviral vectors. *EMBO reports* **4**, 1054-1060.

Howard, M. T., Nelson, W. C. & Matson, S. W. (1995). Stepwise assembly of a relaxosome at the F plasmid origin of transfer. *The Journal of biological chemistry* **270**, 28381-28386.

Hsu, P. D., Lander, E. S. & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278.

Hu, S., Otsubo, E., Davidson, N. & Saedler, H. (1975). Electron microscope heteroduplex studies of sequence relations among bacterial plasmids: identification and mapping of the insertion sequences IS1 and IS2 in F and R plasmids. *Journal of bacteriology* **122**, 764-775.

Hunt, M. A., Currie, M. J., Robinson, B. A. & Dachs, G. U. (2010). Optimizing transfection of primary human umbilical vein endothelial cells using commercially available chemical transfection reagents. *Journal of biomolecular techniques : JBT* **21**, 66-72.

Ilyina, T. V. & Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic acids research* **20**, 3279-3285.

Jafari, M., Soltani, M., Naahidi, S., Karunaratne, D. N. & Chen, P. (2012). Nonviral approach for targeted nucleic acid delivery. *Current medicinal chemistry* **19**, 197-208.

Joung, J. K. & Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nature reviews Molecular cell biology* **14**, 49-55.

Karl, W., Bamberger, M. & Zechner, E. L. (2001). Transfer protein TraY of plasmid R1 stimulates TraI-catalyzed oriT cleavage in vivo. *Journal of bacteriology* **183**, 909-914.

Karow, M., Chavez, C. L., Farruggio, A. P. & other authors (2011). Site-specific recombinase strategy to create induced pluripotent stem cells efficiently with plasmid DNA. *Stem cells* **29**, 1696-1704.

Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome research* **12**, 656-664.

Keravala, A., Groth, A. C., Jarrahian, S., Thyagarajan, B., Hoyt, J. J., Kirby, P. J. & Calos, M. P. (2006). A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. *Molecular genetics and genomics : MGG* **276**, 135-146.

Keravala, A., Lee, S., Thyagarajan, B., Olivares, E. C., Gabrovsky, V. E., Woodard, L. E. & Calos, M. P. (2009). Mutational derivatives of PhiC31 integrase with increased efficiency and specificity. *Molecular therapy : the journal of the American Society of Gene Therapy* **17**, 112-120.

Keravala, A., Chavez, C. L., Hu, G., Woodard, L. E., Monahan, P. E. & Calos, M. P. (2011). Long-term phenotypic correction in factor IX knockout mice by using PhiC31 integrase-mediated gene therapy. *Gene therapy* **18**, 842-848.

Khaleel, T., Younger, E., McEwan, A. R., Varghese, A. S. & Smith, M. C. (2011). A phage protein that binds phiC31 integrase to switch its directionality. *Molecular microbiology* **80**, 1450-1463.

Kilby, N. J., Davies, G. J. & Snaith, M. R. (1995). FLP recombinase in transgenic plants: constitutive activity in stably transformed tobacco and generation of marked cell clones in Arabidopsis. *The Plant journal : for cell and molecular biology* **8**, 637-652.

Kim, Y. G., Cha, J. & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 1156-1160.

Kirby, J. E. & Nekorchuk, D. M. (2002). Bartonella-associated endothelial proliferation depends on inhibition of apoptosis. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 4656-4661.

Kitazono, A. A. (2011). Optimized protocols and plasmids for in vivo cloning in yeast. *Gene* **484**, 86-89.

Kobori, S., Ko, Y. & Kato, M. (2009). A target site for spontaneous insertion of IS10 element in pUC19 DNA located within intrinsically bent DNA. *The open microbiology journal* **3**, 146-150.

Kovarik, A., Matzke, M. A., Matzke, A. J. & Koulakova, B. (2001). Transposition of IS10 from the host Escherichia coli genome to a plasmid may lead to cloning artefacts. *Molecular genetics and genomics : MGG* **266**, 216-222.

Kuhn, B., Abdel-Monem, M., Krell, H. & Hoffmann-Berling, H. (1979). Evidence for two mechanisms for DNA unwinding catalyzed by DNA helicases. *The Journal of biological chemistry* **254**, 11343-11350.

Lahue, E. E. & Matson, S. W. (1988). Escherichia coli DNA helicase I catalyzes a unidirectional and highly processive unwinding reaction. *The Journal of biological chemistry* **263**, 3208-3215.

Lan, F., Liu, J., Narsinh, K. H. & other authors (2012). Safe genetic modification of cardiac stem cells using a site-specific integration technique. *Circulation* **126**, S20-28.

Lane, H. E. (1981). Replication and incompatibility of F and plasmids in the IncFI Group. *Plasmid* **5**, 100-126.

Lang, S., Gruber, K., Mihajlovic, S. & other authors (2010). Molecular recognition determinants for type IV secretion of diverse families of conjugative relaxases. *Molecular microbiology* **78**, 1539-1555.

Lanka, E. & Wilkins, B. M. (1995). DNA processing reactions in bacterial conjugation. *Annu Rev Biochem* **64**, 141-169.

Larkin, C., Datta, S., Harley, M. J., Anderson, B. J., Ebie, A., Hargreaves, V. & Schildbach, J. F. (2005). Inter- and intramolecular determinants of the specificity of single-stranded DNA binding and cleavage by the F factor relaxase. *Structure* **13**, 1533-1544.

Larkin, C., Haft, R. J., Harley, M. J., Traxler, B. & Schildbach, J. F. (2007). Roles of active site residues and the HUH motif of the F plasmid TraI relaxase. *The Journal of biological chemistry* **282**, 33707-33713.

Larrea, D. (2013).La interfase molecular entre los sistemas de secreción tipo IV y sus sustratos. In Molecular Biology: University of Cantabria.

Larrea, D., de Paz, H. D., Arechaga, I., de la Cruz, F. & Llosa, M. (2013). Structural independence of conjugative coupling protein TrwB from its Type IV secretion machinery. *Plasmid* **70**, 146-153.

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

Lessl, M., Pansegrau, W. & Lanka, E. (1992). Relationship of DNA-transfer-systems: essential transfer factors of plasmids RP4, Ti and F share common sequences. *Nucleic acids research* **20**, 6099-6100.

Li, L., Lu, Y., Qin, L. X., Bar-Joseph, Z., Werner-Washburne, M. & Breeden, L. L. (2009). Budding yeast SSD1-V regulates transcript levels of many longevity genes and extends chronological life span in purified quiescent cells. *Molecular biology of the cell* **20**, 3851-3864.

Li, M., Suzuki, K., Kim, N. Y., Liu, G. H. & Izpisua Belmonte, J. C. (2014). A cut above the rest: targeted genome editing technologies in human pluripotent stem cells. *The Journal of biological chemistry* **289**, 4594-4599.

Llosa, M., Bolland, S. & de la Cruz, F. (1991). Structural and functional analysis of the origin of conjugal transfer of the broad-host-range IncW plasmid R388 and comparison with the related IncN plasmid R46. *Mol Gen Genet* **226**, 473-483.

Llosa, M., Bolland, S. & de la Cruz, F. (1994a). Genetic organization of the conjugal DNA processing region of the IncW plasmid R388. *J Mol Biol* **235**, 448-464.

Llosa, M., Bolland, S., Grandoso, G. & de la Cruz, F. (1994b). Conjugation-independent, site-specific recombination at the oriT of the IncW plasmid R388 mediated by TrwC [published erratum appears in J Bacteriol 1994 Oct;176(20):6414]. *Journal of bacteriology* **176**, 3210-3217.

Llosa, M., Grandoso, G. & de la Cruz, F. (1995). Nicking activity of TrwC directed against the origin of transfer of the IncW plasmid R388. *J Mol Biol* **246**, 54-62.

Llosa, M., Grandoso, G., Hernando, M. A. & de la Cruz, F. (1996). Functional domains in protein TrwC of plasmid R388: dissected DNA strand transferase and DNA helicase activities reconstitute protein function. *J Mol Biol* **264**, 56-67.

Llosa, M., Gomis-Rüth, F.-X., Coll, M. & de la Cruz, F. (2002). Bacterial conjugation: a two-step mechanism for DNA transport. *Molecular microbiology* **45**, 1-8.

Llosa, M., Zunzunegui, S. & de la Cruz, F. (2003). Conjugative coupling proteins interact with cognate and heterologous VirB10-like proteins while exhibiting specificity for cognate relaxosomes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10465-10470.

Llosa, M. & de la Cruz, F. (2005). Bacterial conjugation: a potential tool for genomic engineering. *Res Microbiol* **156**, 1-6.

Llosa, M., Roy, C. & Dehio, C. (2009). Bacterial type IV secretion systems in human disease. *Molecular microbiology* **73**, 141-151.

Llosa, M., Schroder, G. & Dehio, C. (2012). New perspectives into bacterial DNA transfer to human cells. *Trends in microbiology* **20**, 355-359.

Loper, J. E. & Kado, C. I. (1979). Host range conferred by the virulence-specifying plasmid of Agrobacterium tumefaciens. *Journal of bacteriology* **139**, 591-596.

Low, H. H., Gubellini, F., Rivera-Calzada, A. & other authors (2014). Structure of a type IV secretion system. *Nature* **508**, 550-553.

Lu, J. & Frost, L. S. (2005). Mutations in the C-terminal region of TraM provide evidence for in vivo TraM-TraD interactions during F-plasmid conjugation. *Journal of bacteriology* **187**, 4767-4773.

Lu, J., Wong, J. J., Edwards, R. A., Manchak, J., Frost, L. S. & Glover, J. N. (2008). Structural basis of specific TraD-TraM recognition during F plasmid-mediated bacterial conjugation. *Molecular microbiology* **70**, 89-99.

Lucas, M., Gonzalez-Perez, B., Cabezas, M., Moncalian, G., Rivas, G. & de la Cruz, F. (2010). Relaxase DNA binding and cleavage are two distinguishable steps in conjugative DNA processing that involve different sequence elements of the nic site. *The Journal of biological chemistry* **285**, 8918-8926.

Lum, P. L. & Schildbach, J. F. (1999). Specific DNA recognition by F Factor TraY involves beta-sheet residues. *The Journal of biological chemistry* **274**, 19644-19648.

Luo, Y., Gao, Q. & Deonier, R. C. (1994). Mutational and physical analysis of F plasmid traY protein binding to oriT. *Molecular microbiology* **11**, 459-469.

Mahfouz, M. M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X. & Zhu, J. K. (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 2623-2628.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826.

Maneewannakul, S., Maneewannakul, K. & Ippen-Ihler, K. (1992). Characterization, localization, and sequence of F transfer region products: the pilus assembly gene product TraW and a new product, TrbI. *Journal of bacteriology* **174**, 5567-5574.

Manning, P. A., Morelli, G. & Achtman, M. (1981). traG protein of the F sex factor of Escherichia coli K-12 and its role in conjugation. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 7487-7491.

Martinez, E., Bartolome, B. & de la Cruz, F. (1988). pACYC184-derived cloning vectors containing the multiple cloning site and lacZ alpha reporter gene of pUC8/9 and pUC18/19 plasmids. *Gene* **68**, 159-162.

Martinez, E. & de la Cruz, F. (1988). Transposon Tn21 encodes a RecA-independent site-specific integration system. *Mol Gen Genet* **211**, 320-325.

Martinez-Garcia, E., Navarro-Llorens, J. M. & Tormo, A. (2003). Identification of an unknown promoter, OUTIIp, within the IS10R element. *Journal of bacteriology* **185**, 2046-2050.

Matson, S. W. & Morton, B. S. (1991). Escherichia coli DNA helicase I catalyzes a site- and strand-specific nicking reaction at the F plasmid oriT. *The Journal of biological chemistry* **266**, 16232-16237.

Matson, S. W., Nelson, W. C. & Morton, B. S. (1993). Characterization of the reaction product of the oriT nicking reaction catalyzed by Escherichia coli DNA helicase I. *Journal of bacteriology* **175**, 2599-2606.

Matson, S. W., Sampson, J. K. & Byrd, D. R. (2001). F plasmid conjugative DNA transfer: the TraI helicase activity is essential for DNA strand transfer. *The Journal of biological chemistry* **276**, 2372-2379.

Matson, S. W. & Ragonese, H. (2005). The F-plasmid TraI protein contains three functional domains required for conjugative DNA strand transfer. *Journal of bacteriology* **187**, 697-706.

Matsutani, S. (1991). Multiple copies of IS10 in the Enterobacter cloacae MD36 chromosome. *Journal of bacteriology* **173**, 7802-7809.

Mazel, D. & Davies, J. (1999). Antibiotic resistance in microbes. *Cell Mol Life Sci* **56**, 742-754.

Meyer, R. (1989). Site-specific recombination at oriT of plasmid R1162 in the absence of conjugative transfer. *Journal of bacteriology* **171**, 799-806.

Mihajlovic, S., Lang, S., Sut, M. V. & other authors (2009). Plasmid r1 conjugative DNA processing is regulated at the coupling protein interface. *Journal of bacteriology* **191**, 6877-6887.

Miller, J. C., Holmes, M. C., Wang, J. & other authors (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**, 778-785.

Miyazaki, R. & van der Meer, J. R. (2011). A dual functional origin of transfer in the ICEclc genomic island of Pseudomonas knackmussii B13. *Molecular microbiology* **79**, 743-758.

Moehle, E. A., Rock, J. M., Lee, Y. L., Jouvenot, Y., DeKelver, R. C., Gregory, P. D., Urnov, F. D. & Holmes, M. C. (2007). Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 3055-3060.

Moncalián, G., Grandoso, G., Llosa, M. & de la Cruz, F. (1997). oriT-processing and regulatory roles of TrwA protein in plasmid R388 conjugation. *J Mol Biol* **270**, 188-200.

Moncalián, G., Cabezón, E., Alkorta, I., Valle, M., Moro, F., Valpuesta, J. M., Goñi, F. M. & de la Cruz, F. (1999a). Characterization of ATP and DNA binding activities of TrwB, the coupling protein essential in plasmid R388 conjugation. *The Journal of biological chemistry* **274**, 36117-36124.

Moncalián, G., Valle, M., Valpuesta, J. M. & de la Cruz, F. (1999b). IHF protein inhibits cleavage but not assembly of plasmid R388 relaxosomes. *Molecular microbiology* **31**, 1643-1652.

Moncalián, G. (2000). Arquitectura y ensamblaje del relaxosoma en el plásmido conjugativo R388. *PhD thesis Santander: Universidad de Cantabria*.

Moncalián, G. & de la Cruz, F. (2004). DNA binding properties of protein TrwA, a possible structural variant of the Arc repressor superfamily. *Biochim Biophys Acta* **1701**, 15-23.

Moreno, M., Kramer, M. G., Yim, L. & Chabalgoity, J. A. (2010). Salmonella as live trojan horse for vaccine development and cancer gene therapy. *Current gene therapy* **10**, 56-76.

Moriguchi, K., Yamamoto, S., Tanaka, K., Kurata, N. & Suzuki, K. (2013). Trans-kingdom horizontal DNA transfer from bacteria to yeast is highly plastic due to natural polymorphisms in auxiliary nonessential recipient genes. *PloS one* **8**, e74590.

Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T. & Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic acids research* **39**, 9283-9293.

Nelson, W. C., Howard, M. T., Sherman, J. A. & Matson, S. W. (1995). The traY gene product and integration host factor stimulate Escherichia coli DNA helicase I-catalyzed nicking at the F plasmid oriT. *The Journal of biological chemistry* **270**, 28374-28380.

Nelson, W. C. & Matson, S. W. (1996). The F plasmid traY gene product binds DNA as a monomer or a dimer: structural and functional implications. *Molecular microbiology* **20**, 1179-1187.

Nieminen, M., Tuuri, T. & Savilahti, H. (2010). Genetic recombination pathways and their application for genome modification of human embryonic stem cells. *Experimental cell research* **316**, 2578-2586.

Nishikawa, M. & Huang, L. (2001). Nonviral vectors in the new millennium: delivery barriers in gene transfer. *Human gene therapy* **12**, 861-870.

Novick, R. P., Projan, S. J., Rosenblum, W. & Edelman, I. (1984). Staphylococcal plasmid cointegrates are formed by host- and phage-mediated general rec systems that act on short regions of homology. *Mol Gen Genet* **195**, 374-377.

Núñez, B. & de la Cruz, F. (2001). Two atypical mobilization proteins are involved in plasmid CloDF13 relaxation. *Molecular microbiology* **39**, 1088-1099.

O'Gorman, S., Fox, D. T. & Wahl, G. M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**, 1351-1355.

Olivares, E. C., Hollis, R. P. & Calos, M. P. (2001). Phage R4 integrase mediates site-specific integration in human cells. *Gene* **278**, 167-176.

Olliver, A., Valle, M., Chaslus-Dancla, E. & Cloeckaert, A. (2005). Overexpression of the multidrug efflux operon acrEF by insertional activation with IS1 or IS10 elements in Salmonella enterica serovar typhimurium DT204 acrB mutants selected with fluoroquinolones. *Antimicrobial agents and chemotherapy* **49**, 289-301.

Olsen, R. H., Siak, J. S. & Gray, R. H. (1974). Characteristics of PRD1, a plasmid-dependent broad host range DNA bacteriophage. *Journal of virology* **14**, 689-699.

Osborn, M. J., Gabriel, R., Webber, B. R. & other authors (2015). Fanconi anemia gene editing by the CRISPR/Cas9 system. *Human gene therapy* **26**, 114-126.

Padmalayam, I., Karem, K., Baumstark, B. & Massung, R. (2000). The gene encoding the 17-kDa antigen of Bartonella henselae is located within a cluster of genes homologous to the virB virulence operon. *DNA and cell biology* **19**, 377-382.

Panicker, M. M. & Minkley, E. G., Jr. (1985). DNA transfer occurs during a cell surface contact stage of F sex factor-mediated bacterial conjugation. *Journal of bacteriology* **162**, 584-590.

Panicker, M. M. & Minkley, E. G., Jr. (1992). Purification and properties of the F sex factor TraD protein, an inner membrane conjugal transfer protein. *The Journal of biological chemistry* **267**, 12761-12766.

Pannell, D. & Ellis, J. (2001). Silencing of gene expression: implications for design of retrovirus vectors. *Reviews in medical virology* **11**, 205-217.

Parkhill, J., Dougan, G., James, K. D. & other authors (2001). Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. *Nature* **413**, 848-852.

Paruzynski, A., Arens, A., Gabriel, R. & other authors (2010). Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. *Nature protocols* **5**, 1379-1395.

Paterson, E. S., More, M. I., Pillay, G., Cellini, C., Woodgate, R., Walker, G. C., Iyer, V. N. & Winans, S. C. (1999). Genetic analysis of the mobilization and leading regions of the IncN plasmids pKM101 and pCU1. *Journal of bacteriology* **181**, 2572-2583.

Pattanayak, V., Ramirez, C. L., Joung, J. K. & Liu, D. R. (2011). Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature methods* **8**, 765-770.

Paukner, S., Kudela, P., Kohl, G., Schlapp, T., Friedrichs, S. & Lubitz, W. (2005). DNA-loaded bacterial ghosts efficiently mediate reporter gene transfer and expression in macrophages. *Molecular therapy : the journal of the American Society of Gene Therapy* **11**, 215-223.

Porteus, M. H. & Baltimore, D. (2003). Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763.

Priebe, S. D. & Lacks, S. A. (1989). Region of the streptococcal plasmid pMV158 required for conjugative mobilization. *Journal of bacteriology* **171**, 4778-4784.

Projan, S. J. & Novick, R. (1988). Comparative analysis of five related Staphylococcal plasmids. *Plasmid* **19**, 203-221.

Quenneville, S. P., Chapdelaine, P., Rousseau, J. & Tremblay, J. P. (2007). Dystrophin expression in host muscle following transplantation of muscle precursor cells modified with the phiC31 integrase. *Gene therapy* **14**, 514-522.

Ragonese, H., Haisch, D., Villareal, E., Choi, J. H. & Matson, S. W. (2007). The F plasmid-encoded TraM protein stimulates relaxosome-mediated cleavage at oriT through an interaction with TraI. *Molecular microbiology* **63**, 1173-1184.

Rice, P. A., Yang, S., Mizuuchi, K. & Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87**, 1295-1306.

Rodriguez, H., Snow, E. T., Bhat, U. & Loechler, E. L. (1992). An Escherichia coli plasmid-based, mutational system in which supF mutants are selectable: insertion elements dominate the spontaneous spectra. *Mutation research* **270**, 219-231.

Rolloos, M., Hooykaas, P. J. & van der Zaal, B. J. (2015). Enhanced targeted integration mediated by translocated I-SceI during the Agrobacterium mediated transformation of yeast. *Scientific reports* **5**, 8345.

Rudin, N., Sugarman, E. & Haber, J. E. (1989). Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. *Genetics* **122**, 519-534.

Saadi, S., Maas, W. K., Hill, D. F. & Bergquist, P. L. (1987). Nucleotide sequence analysis of RepFIC, a basic replicon present in IncFI plasmids P307 and F, and its relation to the RepA replicon of IncFII plasmids. *Journal of bacteriology* **169**, 1836-1846.

Sadler, J. R., Tecklenburg, M. & Betz, J. L. (1980). Plasmids containing many tandem copies of a synthetic lactose operator. *Gene* **8**, 279-300.

Saenz, H. L., Engel, P., Stoeckli, M. C., Lanz, C., Raddatz, G., Vayssier-Taussat, M., Birtles, R., Schuster, S. C. & Dehio, C. (2007). Genomic analysis of Bartonella identifies type IV secretion systems as host adaptability factors. *Nat Genet* **39**, 1469-1476.

Sakurai, K., Shimoji, M., Tahimic, C. G., Aiba, K., Kawase, E., Hasegawa, K., Amagai, Y., Suemori, H. & Nakatsuji, N. (2010). Efficient integration of transgenes into a defined locus in human embryonic stem cells. *Nucleic acids research* **38**, e96.

Sambrook, J. & Russell, D. W. (2001). Molecular cloning : a laboratory manual, 3rd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sander, J. D., Dahlborg, E. J., Goodwin, M. J. & other authors (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature methods* **8**, 67-69.

Sander, J. D. & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32**, 347-355.

Saydam, O., Glauser, D. L. & Fraefel, C. (2012). Construction and packaging of herpes simplex virus/adeno-associated virus (HSV/AAV) Hybrid amplicon vectors. *Cold Spring Harbor protocols* **2012**, 352-356.

Schmid, M. C., Schulein, R., Dehio, M., Denecker, G., Carena, I. & Dehio, C. (2004). The VirB type IV secretion system of Bartonella henselae mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells. *Molecular microbiology* **52**, 81-92.

Schmidt, M., Schwarzwaelder, K., Bartholomae, C., Zaoui, K., Ball, C., Pilz, I., Braun, S., Glimm, H. & von Kalle, C. (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nature methods* **4**, 1051-1057.

Schmiederer, M. & Anderson, B. (2000). Cloning, sequencing, and expression of three Bartonella henselae genes homologous to the Agrobacterium tumefaciens VirB region. *DNA and cell biology* **19**, 141-147.

Schmiederer, M., Arcenas, R., Widen, R., Valkov, N. & Anderson, B. (2001). Intracellular induction of the Bartonella henselae virB operon by human endothelial cells. *Infection and immunity* **69**, 6495-6502.

Schroder, G., Krause, S., Zechner, E. L., Traxler, B., Yeo, H. J., Lurz, R., Waksman, G. & Lanka, E. (2002). TraG-like proteins of DNA transfer systems and of the Helicobacter pylori type

IV secretion system: inner membrane gate for exported substrates? *Journal of bacteriology* **184**, 2767-2779.

Schroder, G. & Dehio, C. (2005). Virulence-associated type IV secretion systems of Bartonella. *Trends in microbiology* **13**, 336-342.

Schroder, G. & Lanka, E. (2005). The mating pair formation system of conjugative plasmids-A versatile secretion machinery for transfer of proteins and DNA. *Plasmid* **54**, 1-25.

Schroder, G., Schuelein, R., Quebatte, M. & Dehio, C. (2011). Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen Bartonella henselae. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 14643-14648.

Schulein, R. & Dehio, C. (2002). The VirB/VirD4 type IV secretion system of Bartonella is essential for establishing intraerythrocytic infection. *Molecular microbiology* **46**, 1053-1067.

Schulein, R., Guye, P., Rhomberg, T. A., Schmid, M. C., Schroder, G., Vergunst, A. C., Carena, I. & Dehio, C. (2005). A bipartite signal mediates the transfer of type IV secretion substrates of Bartonella henselae into human cells. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 856-861.

Sclimenti, C. R., Thyagarajan, B. & Calos, M. P. (2001). Directed evolution of a recombinase for improved genomic integration at a native human sequence. *Nucleic acids research* **29**, 5044-5051.

Scott, J. N., Kupinski, A. P. & Boyes, J. (2014). Targeted genome regulation and modification using transcription activator-like effectors. *The FEBS journal* **281**, 4583-4597.

Shen, M. M., Raleigh, E. A. & Kleckner, N. (1987). Physical analysis of Tn10- and IS10-promoted transpositions and rearrangements. *Genetics* **116**, 359-369.

Sherman, J. A. & Matson, S. W. (1994). Escherichia coli DNA helicase I catalyzes a sequence-specific cleavage/ligation reaction at the F plasmid origin of transfer. *The Journal of biological chemistry* **269**, 26220-26226.

Sikora, B., Eoff, R. L., Matson, S. W. & Raney, K. D. (2006). DNA unwinding by Escherichia coli DNA helicase I (TraI) provides evidence for a processive monomeric molecular motor. *The Journal of biological chemistry* **281**, 36110-36116.

Simons, R. W., Hoopes, B. C., McClure, W. R. & Kleckner, N. (1983). Three promoters near the termini of IS10: pIN, pOUT, and pIII. *Cell* **34**, 673-682.

Skaliter, R., Eichenbaum, Z., Shwartz, H., Ascarelli-Goell, R. & Livneh, Z. (1992). Spontaneous transposition in the bacteriophage lambda cro gene residing on a plasmid. *Mutation research* **267**, 139-151.

Smith, R. H. & Kotin, R. M. (2000). An adeno-associated virus (AAV) initiator protein, Rep78, catalyzes the cleavage and ligation of single-stranded AAV ori DNA. *Journal of virology* **74**, 3122-3129.

Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* **317**, 230-234.

Sorrell, D. A. & Kolb, A. F. (2005). Targeted modification of mammalian genomes. *Biotechnology advances* **23**, 431-469.

Souza, D. P., Oka, G. U., Alvarez-Martinez, C. E. & other authors (2015). Bacterial killing via a type IV secretion system. *Nature communications* **6**, 6453.

Street, L. M., Harley, M. J., Stern, J. C., Larkin, C., Williams, S. L., Miller, D. L., Dohm, J. A., Rodgers, M. E. & Schildbach, J. F. (2003). Subdomain organization and catalytic residues of the F factor TraI relaxase domain. *Biochim Biophys Acta* **1646**, 86-99.

Sullivan, T. D., Rooney, P. J. & Klein, B. S. (2002). Agrobacterium tumefaciens integrates transfer DNA into single chromosomal sites of dimorphic fungi and yields homokaryotic progeny from multinucleate yeast. *Eukaryotic cell* **1**, 895-905.

Summers, D., Yaish, S., Archer, J. & Sherratt, D. (1985). Multimer resolution systems of ColE1 and ColK: localisation of the crossover site. *Mol Gen Genet* **201**, 334-338.

Sut, M. V., Mihajlovic, S., Lang, S., Gruber, C. J. & Zechner, E. L. (2009). Protein and DNA effectors control the TraI conjugative helicase of plasmid R1. *Journal of bacteriology* **191**, 6888-6899.

Tato, I., Zunzunegui, S., de la Cruz, F. & Cabezon, E. (2005). TrwB, the coupling protein involved in DNA transport during bacterial conjugation, is a DNA-dependent ATPase. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8156-8161.

Tato, I., Matilla, I., Arechaga, I., Zunzunegui, S., de la Cruz, F. & Cabezon, E. (2007). The ATPase activity of the DNA transporter TrwB is modulated by protein TrwA: implications for a common assembly mechanism of DNA translocating motors. *The Journal of biological chemistry* **282**, 25569-25576.

Tatum, E. L. & Lederberg, J. (1947). Gene Recombination in the Bacterium Escherichia coli. *Journal of bacteriology* **53**, 673-684.

Tebas, P., Stein, D., Tang, W. W. & other authors (2014). Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *The New England journal of medicine* **370**, 901-910.

Thompson, R., Taylor, L., Kelly, K., Everett, R. & Willetts, N. (1984). The F plasmid origin of transfer: DNA sequence of wild-type and mutant origins and location of origin-specific nicks. *The EMBO journal* **3**, 1175-1180.

Thyagarajan, B., Guimaraes, M. J., Groth, A. C. & Calos, M. P. (2000). Mammalian genomes contain active recombinase recognition sites. *Gene* **244**, 47-54.

Thyagarajan, B., Liu, Y., Shin, S. & other authors (2008). Creation of engineered human embryonic stem cell lines using phiC31 integrase. *Stem cells* **26**, 119-126.

Ton-Hoang, B., Siguier, P., Quentin, Y., Onillon, S., Marty, B., Fichant, G. & Chandler, M. (2012). Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences. *Nucleic acids research* **40**, 3596-3609.

Towbin, H., Staehelin, T. & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 4350-4354.

Traxler, B. A. & Minkley, E. G., Jr. (1988). Evidence that DNA helicase I and oriT site-specific nicking are both functions of the F TraI protein. *J Mol Biol* **204**, 205-209.

Truttmann, M. C., Rhomberg, T. A. & Dehio, C. (2011). Combined action of the type IV secretion effector proteins BepC and BepF promotes invasome formation of Bartonella henselae on endothelial and epithelial cells. *Cellular microbiology* **13**, 284-299.

Tsai, M. M., Fu, Y. H. & Deonier, R. C. (1990). Intrinsic bends and integration host factor binding at F plasmid oriT. *Journal of bacteriology* **172**, 4603-4609.

Turan, S. & Bode, J. (2011). Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **25**, 4088-4107.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic acids research* **40**, e115.

Urnov, F. D., Miller, J. C., Lee, Y. L. & other authors (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646-651.

van der Weyden, L., Adams, D. J. & Bradley, A. (2002). Tools for targeted manipulation of the mouse genome. *Physiological genomics* **11**, 133-164.

Villa, L., Garcia-Fernandez, A., Fortini, D. & Carattoli, A. (2010). Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *The Journal of antimicrobial chemotherapy* **65**, 2518-2529.

Wang, P., Zhang, C., Zhu, Y., Deng, Y., Guo, S., Peng, D., Ruan, L. & Sun, M. (2013). The resolution and regeneration of a cointegrate plasmid reveals a model for plasmid evolution mediated by conjugation and oriT site-specific recombination. *Environmental microbiology* **15**, 3305-3318.

Wang, X., Wang, Y., Wu, X. & other authors (2015). Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat Biotechnol* **33**, 175-178.

Ward, P. & Walsh, C. E. (2012). Targeted integration of a rAAV vector into the AAVS1 region. *Virology* **433**, 356-366.

Waters, V. L. & Guiney, D. G. (1993). Processes at the nick region link conjugation, T-DNA transfer and rolling circle replication. *Molecular microbiology* **9**, 1123-1130.

Waters, V. L. (2001). Conjugation between bacterial and mammalian cells. *Nat Genet* **29**, 375-376.

Watson, J. D. & Hayes, W. (1953). Genetic Exchange in Escherichia Coli K(12): Evidence for Three Linkage Groups. *Proceedings of the National Academy of Sciences of the United States of America* **39**, 416-426.

Weinberg, M. S., Samulski, R. J. & McCown, T. J. (2013). Adeno-associated virus (AAV) gene therapy for neurological disease. *Neuropharmacology* **69**, 82-88.

Williams, S. L. & Schildbach, J. F. (2007). TraY and integration host factor oriT binding sites and F conjugal transfer: sequence variations, but not altered spacing, are tolerated. *Journal of bacteriology* **189**, 3813-3823.

Wong, J. J., Lu, J. & Glover, J. N. (2012). Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective. *Molecular microbiology* **85**, 602-617.

Yang, L., Soonpaa, M. H., Adler, E. D. & other authors (2008). Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature* **453**, 524-528.

Zechner, E. L., de la Cruz, F., Eisenbrant, R. & other authors (2000). The horizontal gene pool: Bacterial plasmids and gene spread. London: Harwood Academic Publishers.

Zechner, E. L., Lang, S. & Schildbach, J. F. (2012). Assembly and mechanisms of bacterial type IV secretion machines. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **367**, 1073-1087.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G. M. & Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**, 149-153.

Zupan, J., Ward, D. & Zambryski, P. (2002). Inter-kingdom DNA transfer decoded. *Nat Biotechnol* **20**, 129-131.

Publications

OPEN ACCESS Freely available online

PLoS one

Site-Specific Integration of Foreign DNA into Minimal Bacterial and Human Target Sequences Mediated by a Conjugative Relaxase

Leticia Agúndez[¤a], Coral González-Prieto, Cristina Machón[¤b], Matxalen Llosa*

Departamento de Biología Molecular, Universidad de Cantabria (UC) and Instituto de Biomedicina y Biotecnología de Cantabria, UC-CSIC-SODERCAN, Santander, Spain

Abstract

Background: Bacterial conjugation is a mechanism for horizontal DNA transfer between bacteria which requires cell to cell contact, usually mediated by self-transmissible plasmids. A protein known as relaxase is responsible for the processing of DNA during bacterial conjugation. TrwC, the relaxase of conjugative plasmid R388, is also able to catalyze site-specific integration of the transferred DNA into a copy of its target, the origin of transfer (*oriT*), present in a recipient plasmid. This reaction confers TrwC a high biotechnological potential as a tool for genomic engineering.

Methodology/Principal Findings: We have characterized this reaction by conjugal mobilization of a suicide plasmid to a recipient cell with an *oriT*-containing plasmid, selecting for the cointegrates. Proteins TrwA and IHF enhanced integration frequency. TrwC could also catalyze integration when it is expressed from the recipient cell. Both Y18 and Y26 catalytic tyrosil residues were essential to perform the reaction, while TrwC DNA helicase activity was dispensable. The target DNA could be reduced to 17 bp encompassing TrwC nicking and binding sites. Two human genomic sequences resembling the 17 bp segment were accepted as targets for TrwC-mediated site-specific integration. TrwC could also integrate the incoming DNA molecule into an *oriT* copy present in the recipient chromosome.

Conclusions/Significance: The results support a model for TrwC-mediated site-specific integration. This reaction may allow R388 to integrate into the genome of non-permissive hosts upon conjugative transfer. Also, the ability to act on target sequences present in the human genome underscores the biotechnological potential of conjugative relaxase TrwC as a site-specific integrase for genomic modification of human cells.

Citation: Agúndez L, González-Prieto C, Machón C, Llosa M (2012) Site-Specific Integration of Foreign DNA into Minimal Bacterial and Human Target Sequences Mediated by a Conjugative Relaxase. PLoS ONE 7(1): e31047. doi:10.1371/journal.pone.0031047

Editor: Baochuan Lin, Naval Research Laboratory, United States of America

Received October 25, 2011; **Accepted** December 30, 2011; **Published** January 23, 2012

Funding: This work was supported by grant BIO2008-00133 from the Spanish Ministry of Science and Innovation to ML. CGP was a recipient of a predoctoral fellowship from the University of Cantabria, Spain. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: llosam@unican.es

¤a Current address: King's College London School of Medicine, London, United Kingdom
¤b Current address: Instituto de Investigación Biomédica, Barcelona, Spain

Introduction

Bacterial conjugation is a mechanism for horizontal gene transfer among bacteria. By this process, a DNA molecule of any origin and length can be transferred to a recipient cell, if it contains an origin of transfer (*oriT*); the conjugative machinery can be provided in *trans*. Under laboratory conditions, conjugation to eukaryotic cells has been reported, from bacteria to yeast [1], plants [2] and mammalian cells [3].

Mechanistically, bacterial conjugation can be viewed as a plasmid DNA replication system linked to a secretion channel [4], leading to horizontal rather than vertical transmission of the plasmid. Accordingly, there are three functional modules in the conjugative machinery [5]:

i) the relaxosome, a nucleoprotein complex required for plasmid DNA processing, which is related to rolling-circle replication systems. This complex includes a relaxase protein, often together with accessory proteins which assist its function, and the *oriT*, which is the only DNA sequence required *in cis* for DNA transfer.

ii) the Type IV secretion system (T4SS), a multiprotein complex which forms the transmembranal channel for substrate secretion.

iii) the coupling protein (T4CP), responsible for linking the other two modules by protein-protein interactions.

The relaxase nicks the *oriT* in the DNA strand which is going to be transferred and remains covalently bound to the 5′ end; current models for conjugative DNA transfer propose that this nucleoprotein complex is secreted to the recipient cell through the T4SS, followed by active DNA pumping by the ATPase activity of the T4CP [4]. Once in the recipient cell, the relaxase is functional [6], and presumably recircularizes the transferred DNA.

The relaxosome of the conjugative plasmid R388 comprises the *oriT* (*oriTw*), the relaxase-helicase TrwC, and two helper proteins, R388-encoded TrwA and host-encoded IHF (integration host factor) [7]. The minimal sequence of the *oriTw* essential for *in vivo* mobilization is only 17 bp in length; *in vitro*, it was shown that the sequence 6 nt 5′ and 2 nt 3′ to the *nic* site (6+2) is absolutely required for TrwC binding and for DNA nicking and strand transfer reactions on oligonucleotides [8]. The helper protein TrwA binds to *oriTw*; it acts as a transcriptional repressor of the *trwABC* operon and also increases TrwC nicking activity on *oriTw* [9], [10]. IHF is a host protein which binds specifically to *oriTw*. IHF has been proposed to have a negative effect on TrwC nicking activity, but not TrwC binding [7].

TrwC relaxase activity, responsible for DNA strand cleaving and transfer activities on supercoiled or single-stranded substrates, is contained in its N-terminal 293 residues (N293) [11]. This domain contains catalytic residues Y18 and Y26, which act sequentially [12]: Y18 is the only tyrosine able to act on supercoiled plasmid substrates and so it is responsible for the initial nicking reaction, while Y26 would catalyze the final strand-transfer reaction. The DNA helicase, ATPase and dimerization activities of TrwC are located in the C-terminal residues (C774) [13].

Apart from its role in conjugation, TrwC has been shown to act as a site-specific recombinase between two *oriTw* copies repeated in tandem [14]. This reaction occurs in the absence of conjugation, and thus in the absence of single stranded intermediates. This ability of TrwC is not a general feature of relaxases; TraI of plasmid F, which shares a similar 3D catalytic fold, is not able to catalyze this reaction [15]. TrwC-mediated site-specific recombination is strongly enhanced by TrwA [15]. Host factors which affect DNA topology, such as IHF or transcription through the *oriTw*, affect the reaction [16]. One *oriTw* copy could be narrowed to the core sequence of 17 bp (14 nt 5′+3 nt 3′of *nic*) and still support TrwC-mediated recombination efficiently [15]. A search for possible natural targets sequences for TrwC in the human genome demonstrated that there are at least two possible natural targets on which TrwC can act as a recombinase [17]. The recombinase domain of TrwC locates to the nucleus in human cells, and, by random mutagenesis, a TrwC mutant which enters the nucleus was also obtained [17]. Interestingly, TrwC can catalyze site-specific integration of the incoming DNA into an *oriTw*-containing plasmid in the recipient cell [6]. Overall, these data underscore the biotechnological potential of TrwC as a site-specific integrase which could be introduce into human cells for genomic modification.

In this work we have modified the integration assay in order to address the DNA and protein requirements of the reaction. Our results suggest a model for TrwC-mediated site-specific integration. We have obtained TrwC-mediated site-specific integration on minimal *oriTw* and human target sequences and into the *Escherichia coli* chromosome, underlying the biological significance of this reaction and the potential of TrwC for biotechnological purposes.

Results

Optimization of the integration assay

In earlier work, the ability of TrwC to catalyze site-specific integration was assayed by mobilization of a suicide plasmid to an *oriTw*-containing recipient strain, in a *recA* background [6]. The assay is depicted in **Figure 1A**. The suicide plasmid pR6K::*oriTp oriTw*, containing the *oriTs* of plasmids R388 (IncW) and RP4 (IncP), is mobilized by the conjugative system of plasmid R388 from donor strain CC118 λpir (coding for the Pir protein which allows replication of the R6K replicon) to a recipient strain which contains a plasmid harbouring another *oriTw* copy, where integration takes place. The suicide plasmid cannot replicate in the recipient strain. The integrants are selected in Cm plates, the resistance conferred by the suicide plasmid, and integration subsequently confirmed by PCR amplification of a specific region of the cointegrate molecule with primers P1 and P2 (Fig. 1A). The published integration frequency reached with this assay was 5.4×10^{-6} integrants per donor. It was also reported that the reaction was TrwC-dependent, since no integration events were detected when, in a similar assay, the suicide plasmid was mobilized by the RP4 conjugative relaxase [6].

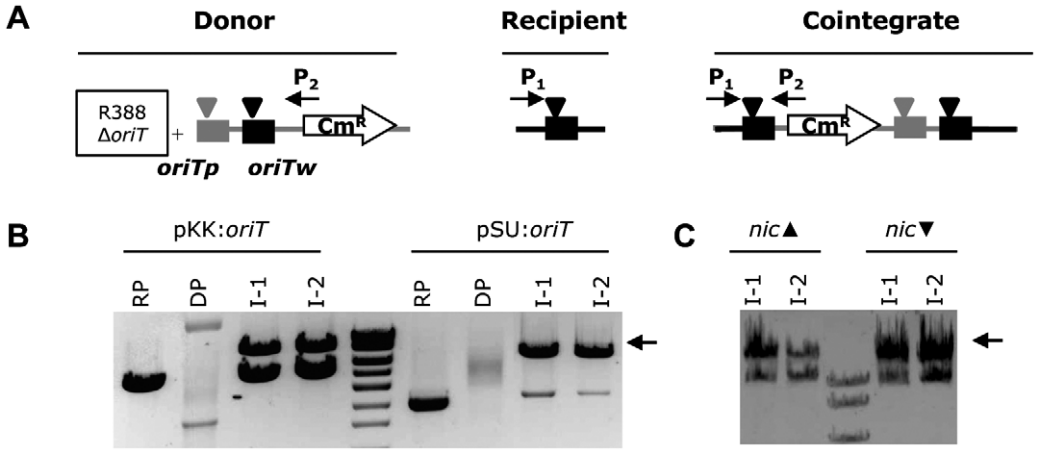

Figure 1. Site-specific integration assays with different recipient plasmids. A. Scheme of the assay and the cointegrate molecule obtained. The suicide plasmid is represented with a grey line and the recipient plasmid with a black line. The *nic* site is indicated by an arrowhead. P1 and P2, oligonucleotides used in the PCR reaction to detect the cointegrates. **B and C**. Restriction analysis with enzymes that cut only once in the recipient plasmid. *NdeI* was used for integrants obtained in pKK:*oriT* (in **B**, to the left of the MW marker), *XcmI* for integrants in pSU:*oriT* (in **B**, to the right of the MW marker), and *BstEII* for integrants in pLA58 and pLA59 (in **C**, *nic*▲ and *nic*▼, respectively). The cointegrate is indicated with an arrow. DP, donor plasmid; RP, recipient plasmid. I1-I2, DNA from two independent integrants obtained on each RP. Sizes in kb of the MW marker from top of the gel: 10-8.0-6.0-5.0-4.0-3.0-2.5.
doi:10.1371/journal.pone.0031047.g001

To improve the integration reaction, we used a different Pir donor strain, II1, which can be handled at 37°C since it is devoid of the λ prophage. As recipient plasmids, we assayed two different replicons: pKK::*oriTw* (pMB8 replicon) and pSU::*oriTw* (p15A replicon). With either recipient plasmid, a significant improvement in the integration frequency was obtained (**Table 1**): ca. 2 logs with pKK::*oriTw* and 1 log with pSU::*oriTw*, compared to earlier work. However, we observed a residual level of integration (2.6×10^{-7} integrants/donor) in the absence of TrwC when using pKK::*oriTw*, for unknown reasons, while no integrants were found in the absence of TrwC when the pSU:*oriTw* recipient plasmid was used (Table 1). Thus, the pSU::*oriTw* recipient plasmid rendered the best integration frequency maintaining TrwC specificity and was used as the recipient plasmid for subsequent integration assays.

We tested the ability of RP4_TraI to promote site-specific integration under similar assay conditions, mobilizing the suicide plasmid by this relaxase to a recipient which contained pSU::*oriTp*. No integrant colonies were found (Table 1, last line, last column), suggesting that RP4_TraI relaxase is not able to catalyze site-specific integration of an incoming DNA. However, transfer of TraI-RP4 to the recipient cell has not been proved experimentally so far. Thus, we cannot discard that the absence of integration could be due to the absence of TraI in the recipient cell.

DNA from the integrants obtained for both systems (pKK and pSU) was analysed by restriction analysis with an enzyme that only cuts the recipient *oriT*-containing plasmid. The results (**Figure 1B**) show that the proportion of cointegrate molecules co-residing with the unaltered recipient plasmid varies significantly depending on the system used. When a pMB1 replicon (pKK system) is used as recipient plasmid, there is roughly 50% of each molecular species, as previously shown [6]. However, when the p15A replicon (pSU system) is used, almost all molecules are cointegrates. This difference could reflect an involvement of the plasmid replication machinery in the reaction. Alternatively, since a similar amount of cointegrate molecules is found in both cases, we reasoned that this could be due to a minimal amount of cointegrate molecules required to resist the selection with Cm25. However, the same integration frequencies and proportion of cointegrate molecules were obtained when applying Cm selection of 10, 25 or 40 μg/ml (data not shown).

It was described that the DNA strand harbouring the *nic* site affected the efficiency of *oriT*-specific recombination mediated by TrwC [15]. When the *nic* site of the *oriT* was in the lagging strand, the percentage of recombined colonies increased 5 times; this was explained by the longer exposure of ssDNA in the lagging strand, which would favour TrwC nicking reaction. In both recipient plasmids used in the integration assays, the *nic* site was located in the lagging strand. We assayed two plasmids with the *nic* site in both orientations with respect to the replication fork, pLA58 (p220::*nic*▲), *nic* site in the lagging strand, and pLA59 (p220.2::*nic*▼), *nic* site in the leading strand. We mobilized the suicide plasmid to a recipient cell containing either pLA58 or pLA59 and no significant differences in the integration frequency obtained were observed (4.7×10^{-6} and 3.2×10^{-6} integrants per donor, respectively). Restriction analysis of the DNA of the integrants showed that the percentage of cointegrate molecules was around 50% when the recipient *nic* site was present in the lagging strand, and around 70–80% when the *nic* site was in the leading strand (**Figure 1C**). A plausible explanation is that TrwC is promoting the resolution of the cointegrate molecules preferentially when both *nic* sites lie on the lagging strand, as described in [15].

Role of relaxosomal proteins TrwA and IHF in integration

The R388 relaxosome is formed by the *oriT* and proteins TrwC, TrwA, and host-encoded IHF. TrwA enhances the nicking activity mediated by TrwC on scDNA, while IHF acts as an inhibitor [7]. With respect to TrwC-mediated *oriT*-specific recombination on scDNA substrates, TrwA increases drastically the efficiency of the reaction in the absence of conjugation, and it also enhances TrwC-mediated site-specific recombination in the recipient cell by almost six times [15]. IHF was shown to inhibit recombination only in the absence of TrwA and on DNA substrates lacking one of the two TrwA and IHF binding sites [16]; in the presence of TrwA, however, the presence or absence of IHF was irrelevant.

To test the effect of TrwA and IHF in site-specific integration, the suicide plasmid (pR6K::*oriTp oriTw*) was mobilized to isogenic IHF+ or IHF− recipient strains containing the recipient plasmid pSU::*oriTw* in the presence or absence of a plasmid which expressed *trwA* (pET3a::*trwA*). The recipient plasmid pSU::*oriTp* was assayed as a negative control in all experimental conditions tested. The integration frequency was compared in wild type or IHF-deficient recipient strains in the presence or absence of TrwA. Results are shown in **Table 2**. Both TrwA and IHF are enhancers of the integration reaction, producing an increase of about 3.5 fold and more than 30 fold, respectively, in the integration frequency. DNA of the integrants was analyzed and the same restriction pattern was obtained from IHF− or IHF+ backgrounds (data not shown). The lack of both proteins in the recipient cell decreases by two logs the frequency of integration mediated by TrwC.

Integration catalyzed by TrwC expressed from recipient cells

The integration assay described (Figure 1A) is based on a conjugation assay, meaning that functional *oriT* and TrwC are required to achieve conjugative transfer of the suicide plasmid, which prevents the study of the DNA and protein requirements of the reaction. We assayed the ability of TrwC to catalyze site-specific integration when expressing *trwC* from the recipient cell as

Table 1. TrwC-mediated site-specific integration into different recipient plasmids.

	Recipient				
Donor	**DH5αλpir[1]**	**pKK**	**pKK::*oriTw***	**pSU::*oriTw***	**pSU::*oriTp***
II1 (pCIG1077+pR6K::*oriTporiTw*)	1.77	$<7\times10^{-6}$	2.4×10^{-4}	5.3×10^{-5}	$<2.8\times10^{-8}$
S17.1 λpir (pR6K::*oriTp oriTw*)	0.32	$<4.9\times10^{-7}$	2.6×10^{-7}	$<2.2\times10^{-8}$	$<2.2\times10^{-8}$

Integration frequencies expressed as integrants/donor. Integrants were confirmed by PCR amplification of a region of the cointegrate, as explained in Experimental Procedures. The recipient strain was DH5α.

[1]DH5α λ*pir* was used as an internal control for conjugation efficiency, expressed as transconjugants/donor.

doi:10.1371/journal.pone.0031047.t001

Table 2. Effect of TrwA and IHF on TrwC-mediated integration.

Recipient plasmid	IHF	TrwA	Integration frequency
pSU::*oriTw*	+	+	2.3×10^{-4}
	+	−	6.5×10^{-5}
	−	+	6.9×10^{-6}
	−	−	1.6×10^{-6}
pSU::*oriTp*	+	+	$<9.4\times10^{-7}$
	+	−	$<9.4\times10^{-7}$
	−	+	$<7.9\times10^{-7}$
	−	−	$<7.9\times10^{-7}$

Donor cells are II1 with pCIG1077 and pR6K::*oriT*$_P$ *oriTw*, and recipient cells are DH5α (IHF+) or CIG1 (isogenic IHF−) containing the indicated recipient plasmid with (+TrwA) or without (−TrwA) plasmid pET3a:*trwA*. Data are the mean of three independent assays. Frequency is given in integrants per donor.
doi:10.1371/journal.pone.0031047.t002

follows. The integration reaction mediated by TrwC also takes place upon the mobilization of the suicide plasmid to a recipient cell which contains a plasmid with an *oriT*, but in this case, the donor plasmid is mobilized by the RP4 conjugative system and *trwC* is being expressed from another plasmid in the recipient cell (**Figure 2**, compare A and B). When plasmid pSU1621 (pET3a:*trwC*) was present in the recipient, we observed an integration frequency of 1.4×10^{-7} (**Table 3**). Thus, *trwC* expressed in the recipient cell is able to locate both *oriT*-containing plasmids and catalyze integration. However, the frequency of this assay is 2–3 logs less than when we mobilized the suicide plasmid attached to TrwC from the donor cell (6.5×10^{-5} integrants per donor; Table 2).

In an attempt to improve this frequency, we also expressed *trwA* in the recipient cell. The integration frequency increased by 30 fold (Table 3). Since the helper plasmid coding for *trwA+trwC* expresses these genes from the putative *trwABC* promoter of R388, the increase could be due to a higher expression level of *trwC* rather than to the presence of *trwA*. To examine this possibility, we expressed *trwC* in the recipient cell under the inducible *tac* promoter. The integration assay was performed on plates supplemented with IPTG to induce the expression of *trwC*. The results demonstrated that an increase in the production of TrwC in the recipient cell is not correlated with higher frequencies; in fact, a slight decrease in the integration frequency was observed (Table 3, last two columns), which could be due to toxicity of *trwC* overexpression. Thus, the improvement in the integration frequency when *trwC+trwA* are expressed in the recipient cell is due to the enhancing role of TrwA in the integration reaction.

DNA of the integrants was analyzed (Fig. 2D, top gel). The proportion of cointegrate molecules was much lower when TrwA is expressed in the recipient. So, TrwA, while enhancing TrwC-mediated integration, also promotes resolution of the cointegrates in the cell. Since *oriT*-specific recombination in the presence of both TrwA and TrwC is more efficient than integration, probably all molecules would be resolved if selection for the cointegrates was not applied.

Since TrwC-mediated site-specific integration can be obtained when TrwC is mobilized from the donor and also when *trwC* is expressed in the recipient, we reasoned that the integration frequency could be improved by joining both TrwC sources, as

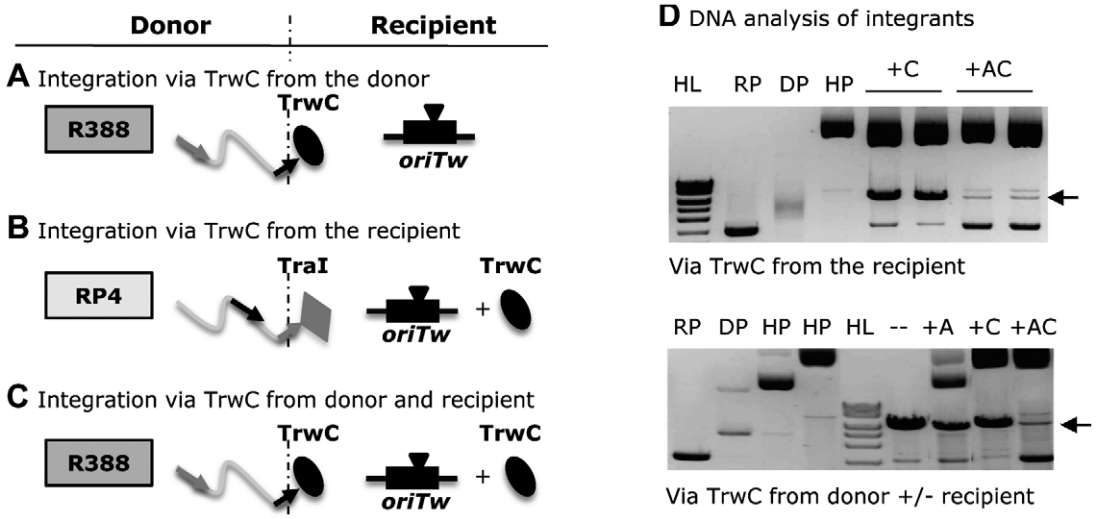

Figure 2. Integration assays expressing *trwC* from the donor and/or the recipient cell. A. Integration assay via TrwC from the donor. The suicide plasmid is mobilized by R388_TrwC relaxase from a II1 donor strain to a DH5α recipient which harbours plasmid pSU::*oriTw*. Integration mediated by TrwC takes place and the cointegrate is formed. **B**. Integration assay via TrwC from the recipient. The suicide plasmid is mobilized by RP4_TraI relaxase from an S17.1 λpir donor strain to a DH5α recipient which harbours pSU::*oriTw* and a plasmid coding for *trwC*. TrwC in the recipient is able to locate both *oriT*-containing plasmids and catalyze integration. **C**. Integration assay via TrwC from the donor and the recipient. The suicide plasmid is mobilized as in A. into a DH5α recipient which harbours plasmid pSU::*oriTw* plus a plasmid coding for *trwC*. R388 and RP4 non-mobilizable conjugative systems are represented with a dark and light grey square, respectively. TrwC is represented by an ellipse and RP4_TraI by a diamond. Arrowheads represent the *nic* site at the recipient *oriTw*. The T-strand of the suicide plasmid is represented by a wavy grey line containing the *oriTw* (black arrow) and the *oriTp* (grey arrow), and the recipient plasmid, with a thick black line. **D**. *XcmI* restriction pattern of integrant DNA. RP, recipient plasmid with R388 *oriT*; DP, suicide donor plasmid; HP, helper plasmid in the recipient. HL: Hyperladder MW marker. −, +A, +C, and +AC refer to the proteins produced by the helper plasmids in the recipient (none, TrwA, TrwC, or both). Top gel: integration assay via TrwC from the recipient (mobilization of donor plasmid with RP4-TraI). Bottom gel: integration assay via TrwC from the donor (mobilization of donor plasmid with R388-TrwC) with different helper plasmids in the recipient cell. The cointegrate molecular species is indicated with a black arrow.
doi:10.1371/journal.pone.0031047.g002

Table 3. Integration assay expressing *trwC*+/−*trwA* in the recipient cell.

		***trwAC* expression in helper plasmid in the recipient**				
Donor relaxase	**Recipient plasmid**	**None**	**P_{T7} *trwC***	**P_{trwA} *trwAC***	**P_{tac}− IPTG *trwC***	**P_{tac}+ IPTG *trwC***
RP4-TraI	pSU::*oriTw*	$<2.2\times10^{-8}$	1.4×10^{-7}	4.2×10^{-6}	6.0×10^{-7}	3.0×10^{-7}
RP4-TraI	pSU::*oriTp*	$<2.2\times10^{-8}$	$<2.2\times10^{-8}$	$<2.2\times10^{-8}$	$<3.2\times10^{-8}$	$<3.2\times10^{-8}$
R388-TrwC	pSU::*oriTw*	5.9×10^{-5}	6.4×10^{-6}	4.4×10^{-6}	NT	NT

Donor cells harbouring the suicide mobilizable plasmid pR6K::*oriT*$_P$ *oriTw* were either S17.1 λpir for TraI-mediated mobilization, or II1 (pCIG1077) for TrwC-mediated mobilization. Recipient strain was DH5α containing the indicated recipient and helper plasmids. The promoter driving the expression of *trwC* or *trwA*+*trwC* is indicated. The inducible Ptac promoter was assayed with (+IPTG) or without (−IPTG) 0,5 mM IPTG in the mating plate. Data are the mean of at least three independent assays. NT, not tested.

doi:10.1371/journal.pone.0031047.t003

shown in Figure 2C. We mobilized the suicide plasmid by TrwC to a recipient harbouring the *oriT*-containing plasmid and a helper plasmid coding for *trwC*. The results, however, showed a decrease in the integration frequency of about 1-log compared to integration rates with no *trwC* expression in the recipient (Table 3, bottom row). When *trwA* was expressed together with *trwC* in the recipient, no recovery of this low frequency was observed (Table 3).

DNA analysis of integrants showed the same pattern observed in previous assays, with most molecules in the form of cointegrates except when TrwA is present in the recipient (Fig. 2D, bottom gel).

TrwC requirements for integrase activity

The integration assay expressing *trwC* in the recipient cell allowed us to test TrwC mutants for their integrase activity. The following experiments were all performed mobilizing the suicide donor plasmid by the RP4 conjugative system.

We tested if the site-specific integration reaction requires the strand transferase activity of TrwC by using a set of point mutants in the catalytic tyrosyl residues: Y18F, Y26F, and the double mutant Y18FY26F. These mutants have been shown to affect conjugation and recombination reactions to different extents: Y18F strongly affects both activities, while Y26F has only a mild effect and the double mutant abolishes both processes [15]. We tested integration catalyzed by these TrwC mutants expressed from the recipient cell and no integrants were found, indicating that the reaction is dependent on both catalytic Tyr residues of the protein (**Table 4**). When TrwA was also produced in the recipient, a few integrants were obtained, allowing analysis of their DNA (**Figure 3A**). We observed that when Y18 was mutated, the resolution reaction was not favoured in spite of the presence of TrwA in the recipient, probably due to the fact that Y18 is the only tyrosine able to act on supercoiled DNA substrates. When only Y26 was mutated, although the integration reaction was reduced to residual levels, resolution of the cointegrates took place.

The C-terminal DNA helicase domain of TrwC is required for conjugative DNA transfer, as it supports unwinding of the DNA to be transferred. Its involvement in the integration reaction was tested using a TrwC point mutant (K502T) affecting motif I of the helicase superfamily I (Walker A box, GxG**K**T); this mutant is deficient in conjugation and proficient in recombination [15]. TrwC(K502T) showed a decrease of less than 3 fold in the integration frequency when compared to TrwC wt (Table 4). A similar slight decrease was observed in the recombination assay, which was attributed to the observation that TrwC(K502T) showed roughly four times less protein product when compared to wild type in a western blot [15]. Thus, the DNA helicase activity is dispensable for TrwC-mediated site-specific integration.

The minimal domain of TrwC able to promote site-specific recombination efficiently was TrwC-N600 [15]. To delimit the minimal TrwC domain sufficient to catalyze integration, we assayed N600 and N450 expressed in the recipient together with TrwA (Table 4). N450 catalyzes integration with about 2-log lower efficiency than wild-type while N600 catalyzes integration about ten times less efficiently than full-length TrwC. This efficiency decreased in about 1-log when TrwA was not expressed in the recipient (Table 4), meaning that TrwA is helping this domain as it helps TrwC-mediated integration. DNA of the integrants was analysed by restriction analysis (**Figure 3B**). In contrast to full-length TrwC, TrwA does not promote the resolution of the cointegrates formed by N600. This result suggests that TrwA can play different roles in integration (enhancing the reaction catalyzed by both TrwC and N600) and in recombination (promoting TrwC-mediated, but not N600-mediated resolution of the cointegrates).

Table 4. TrwC requirements for site-specific integration.

Recipient plasmid	**Helper plasmid in the recipient**		**Integration frequency**
	TrwC	**TrwA**	
pSU::*oriTp*	wt	−	$<6.6\times10^{-9}$
pSU::*oriTw*	wt	−	1.3×10^{-7}
	Y18F	−	$<6.6\times10^{-9}$
	Y26F	−	$<6.6\times10^{-9}$
	Y18F Y26F	−	$<6.6\times10^{-9}$
	N600	−	2.8×10^{-8}
pSU::*oriTp*	wt	+	$<2.9\times10^{-8}$
pSU::*oriTw*	wt	+	4.2×10^{-6}
	Y18F	+	6.6×10^{-8}
	Y26F	+	3.0×10^{-8}
	Y18F Y26F	+	4.0×10^{-8}
	K502T	+	1.5×10^{-6}
	N450	+	3.4×10^{-8}
	N600	+	2.8×10^{-7}

Integration frequencies (integrants per donor) obtained when expressing in the recipient cell the indicated *trwC* derivatives, in the presence or absence of TrwA. Donor cells are S17.1 λpir with pR6K::*oriT*$_P$ *oriTw* and recipient cells DH5α with the indicated recipient and helper plasmids. Data are the mean of at least four independent assays.

doi:10.1371/journal.pone.0031047.t004

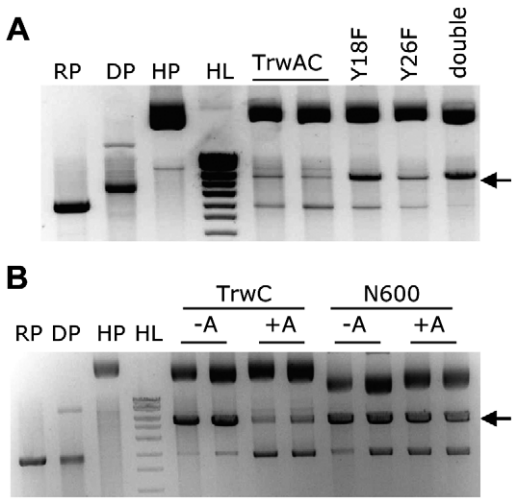

Figure 3. DNA analysis of colonies obtained in the integration assay from the recipient cell expressing different TrwC derivatives. A. TrwC catalytic mutants plus TrwA. **B**. TrwC wt or N600 with or without TrwA. Symbols as in Fig. 2.
doi:10.1371/journal.pone.0031047.g003

oriT requirements for integration

We have analyzed the specificity of TrwC for its integration target, the R388 *oriT* (*oriTw*). The nicking and binding sites of TrwC (**Figure 4A**) are the minimal *oriTw* requirements for different TrwC activities *in vivo* and *in vitro* (see Introduction). To test the requirement of the binding site we have assayed the mutIR *oriT*, in which the sequence of both arms of the inverted repeat IR2 to which TrwC binds was changed by another sequence maintaining the secondary structure (Fig. 4A). This mutation provokes a drastic 5-log reduction of the mobilization frequency [8]. To test the requirement of the nicking site, we used mut23-25 *oriT* (Figure 4A), which has 3 nucleotides changed 5′ to the *nic* site. This mutation almost abolishes mobilization, as this region is critical for the initial nicking reaction [8]. We tested TrwC-mediated integration with recipient plasmids carrying those mutations and expressing *trwC* and *trwA* in the recipient cell, and we did not detect any integration event in any case (integration frequency $<10^{-8}$ integrants per donor), confirming that TrwC also requires its binding and nicking sites to catalyze integration. The same mutations were tested in the integration via TrwC from the donor, i.e. when TrwC enters the recipient cell covalently attached to the transferred DNA strand. The results are shown in **Table 5**. In this case, only a modest decrease (5 to 9-fold) in the integration frequency was observed, indicating different TrwC DNA requirements at the initiation and termination steps of the integration reaction (see Discussion).

In the recombination reaction mediated by TrwC, it was found that one of the *oriT* target copies could be reduced to 17 bp (14+3) containing the *nic* site and the proximal arm of the IR2, maintaining a recombination frequency of 80% [15]. To test if TrwC was also able to promote integration into a recipient plasmid containing this minimal *oriT*, plasmid pSU::*oriT*(14+3) (coordinates 174 to 190 from ref. [18]) was used as recipient plasmid (Fig. 4A). The results (Table 5) show that TrwC is able to promote site-specific integration on this minimal *oriT*; however, the reaction decreases by around 2 logs. When *trwA* was also expressed in the recipient cell, a modest increase in integration frequency was obtained (Table 5). DNA analysis of the integrants showed the same proportion of cointegrates when the recipient plasmid carried full-length or minimal *oriT* (**Figure 4B**).

In a recent report, we demonstrated that TrwC can promote site-specific recombination on substrate plasmids where the *oriT1* copy was replaced by human genomic DNA which contained sequences resembling the essential *oriT* core region (Fig. 4A): two (15+3) sequences, located in chromosomes 5 and X, with no mismatches in the essential core sequence (6+2), worked as targets with the same efficiency as with the canonical 14+3 sequence [17]. We assayed TrwC mediated site-specific integration into recipient plasmids containing these human targets (HuX 15+3(−7) and Hu5 15+3(−10); Fig. 4A) and a helper plasmid which provided TrwA. TrwC catalyzed site-specific integration into both human target sequences; the efficiency of the reaction decreases by only 2–3 times compared to the minimal *oriT*(14+3) (Table 5). When DNA from the integrants was analyzed, we found that 100% of the molecules were cointegrates (Fig. 4B). This suggests that the deviations from the consensus in the human sequences, lying out of the essential nicking region, do not affect significantly termination of the integration reaction, but prevent resolution of the cointegrates formed.

Integration into chromosomal DNA

We tested the ability of TrwC to catalyze site-specific integration into a chromosomal *oriT* copy. In the *E.coli* chromosome we did not find any sequences which resemble the

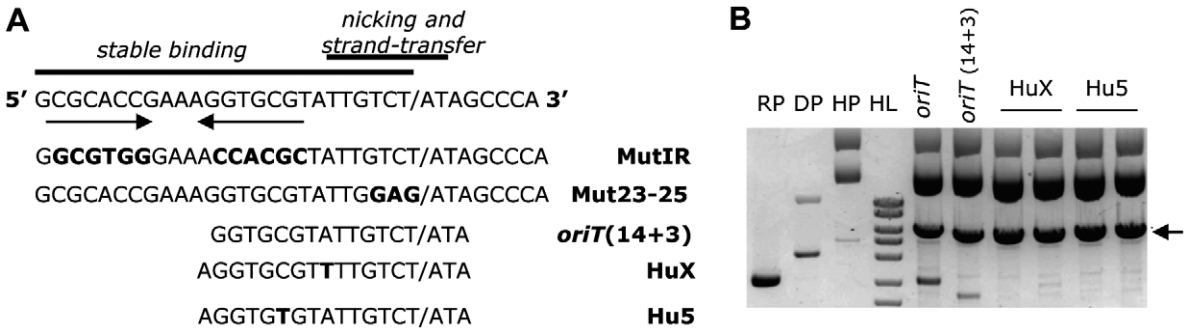

Figure 4. Integration assays on different target sequences. A. DNA sequence of the central R388 *oriT* region, coordinates 201 to 169 from [18]. The arrows show inverted repeat IR2. The *nic* site is indicated by a slash. Horizontal bars indicate minimal sequence requirements for different TrwC activities. Below are shown the DNA sequences of the *oriT* mutations MutIR and Mut23-25, the minimal *oriTw* (*oriT* 14+3), and the human sequences resembling the R388 *nic* site, HuX and Hu5. **B**. DNA analysis of integrants obtained with recipient plasmids containing the indicated target sequences. Symbols as in Fig. 2. RP containing the minimal *oriT* is around 300 bp shorter than RP with full-length *oriT*.
doi:10.1371/journal.pone.0031047.g004

Table 5. DNA targets for TrwC-mediated integration.

Recipient plasmid	TrwA	Integration frequency
pSU::*oriTp*	+	$<3.3\times10^{-8}$
pSU::*oriTw*	+	2.3×10^{-4}
pSU::*oriTw* (mut IR)	−	1.3×10^{-5}
pSU::*oriTw* (mut 23-25)	−	7.3×10^{-6}
pSU::*oriTw* (14+3)	−	7.4×10^{-7}
pSU::*oriT* HuX 15+3*(−7)*	+	5.3×10^{-7}
pSU::*oriT* Hu5 15+3*(−10)*	+	4.6×10^{-7}

Donor cells are II1 with pCIG1077 and pR6K::*oriT*$_P$ *oriTw*, and recipient cells are DH5α containing the indicated recipient plasmid, with (+TrwA) or without (−TrwA) plasmid pET3a::*trwA*. Data are the mean of at least three independent assays. Frequency is given in integrants per donor.
doi:10.1371/journal.pone.0031047.t005

minimal R388 *oriT* sequence required for TrwC binding and nicking. Two recipient strains were constructed containing an *oriTw* copy plus a resistance marker in place of the chromosomal *lacZ* gene (see Experimental Procedures): CMS1 and CMS2, with the *oriTw nic* site lying in the lagging and leading strands with respect to the replication fork, respectively (**Figure 5A**). The integration assay was performed by mobilizing the suicide plasmid (pR6K::*oriTp oriTw*) either with R388-TrwC or with RP4-TraI to recipient strains CMS1 or CMS2, or HMS174 (the isogenic *oriT*- strain) as a negative control. We did not obtain any integrant in any case ($<10^{-8}$ integrants per donor) except when mobilizing the suicide plasmid with TrwC and using CMS1 as recipient. The integration frequency was low (1.1×10^{-7} integrants per donor, mean of five independent assays) but the results were reproducible. DNA analysis of the integrants was done by PCR amplification of the *lacZ* gene (**Figure 5B**). Amplification products of the expected size were obtained in all cases: 3 kb (size of the *lacZ* gene) in HMS174, 1.5 kb in CMS1 and CMS2 (size of the *oriT*-Km cassette in place of the *lacZ* ORF), and around 5 kb for the integrants (which contain the integrated suicide plasmid). It can also be observed that a 1.5 kb product is also visible, indicating that the integration reaction is reversible. In addition, specific amplification of a region of the cointegrate was performed with primers P_A and P_C (**Figure 5C**). This specific 850 bp amplification product was gel-extracted and the DNA sequence was determined, confirming the expected integration structure.

The negative controls (*oriT*- recipient strain, or mobilization with RP4-TraI to any of the strains: $<4.3\times10^{-8}$ integrants per donor in all cases) indicate that the reaction is dependent on TrwC and its target R388 *oriT*. Thus, TrwC is able to catalyze the integration reaction into the chromosome although with 2–3 logs lower efficiency than into an *oriTw*-containing plasmid. Interestingly, no integration events were detected when using CMS2 as a recipient strain. This is likely due to the fact that the *nic* site of the *oriTw* lies in the leading strand of replication, meaning that it could be less accessible to TrwC. So, the exposure of single-stranded DNA, which was shown not to be a requisite in integration into an *oriTw*-containing plasmid, is a relevant factor in order to obtain integration into the host chromosome; in plasmid DNA, this exposure may be easier to obtain by local supercoiling than in the chromosome.

Discussion

The conjugative relaxase TrwC is the protein in charge of piloting the transferred DNA into the recipient cell during bacterial conjugation of plasmid R388. It was reported that this protein acts not only as a conjugative relaxase, but also as a site-specific recombinase and integrase in recipient bacteria [6]. By modifying the donor bacterial strain and the recipient *oriT*-containing plasmid, we have significantly improved the integration frequency reported previously (Table 1). Using different types of integration assays (Fig. 2), we have tried to elucidate the

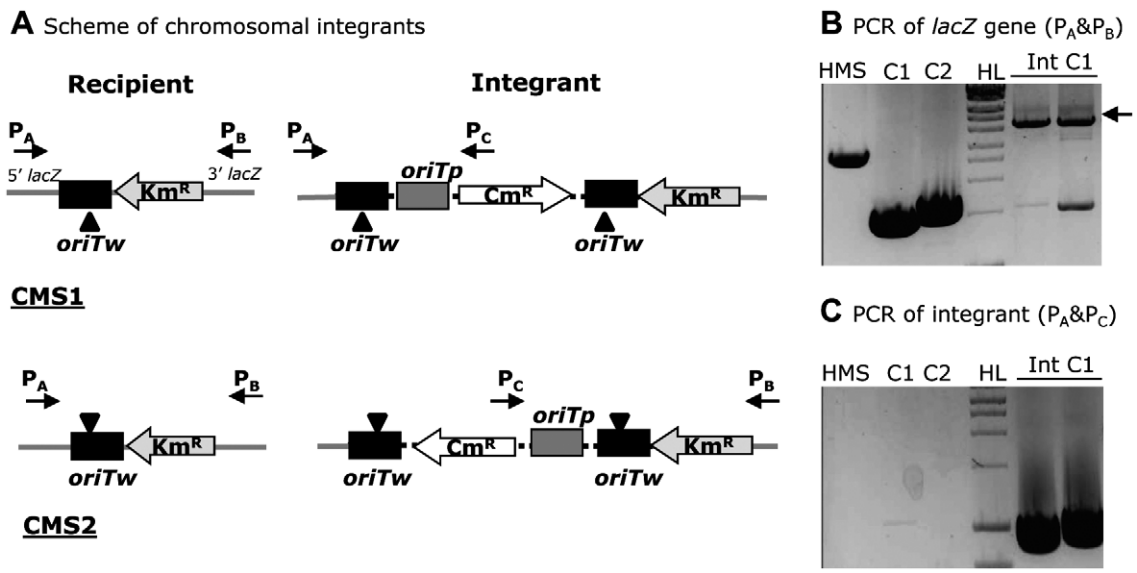

Figure 5. Integration assay in the bacterial chromosome. A. Scheme of the expected integrants in the chromosomal *oriT* copy of recipient strains CMS1 or CMS2. Symbols as in Fig. 1. **B and C**. PCR amplification with oligonucleotides P_A and P_B, flanking the *lacZ* gene (in B), or with P_A and P_C, annealing only to the cointegrate (in C). C1, C2 and HMS, strains CMS1, CMS2, and HMS174, used as a negative control. Int C1, integrants obtained with recipient strain CMS1. The black arrow indicates the amplification product of the suicide plasmid integrated in the chromosomal *oriT* copy. HL: MW marker.
doi:10.1371/journal.pone.0031047.g005

underlying mechanism of the site-specific integration reaction mediated by TrwC in bacteria. Taken together, the results obtained suggest that the reactions mediated by TrwC to accomplish integration mimic those for initiation and termination of conjugation; this assumption is based on the following evidences:

i) TrwC-mediated site-specific integration was favoured by the presence of the R388 relaxosome components IHF and TrwA in the recipient cell (Table 2). This contrasts with TrwC-mediated site-specific recombination tested on plasmids containing two *oriT* copies, where the observed effect of TrwA (enhancer) and IHF (null or repressor under certain conditions) correlated with their described effect on TrwC nicking on supercoiled DNA. This difference probably reflects the fact that recombination is boosted by two initial nicking reactions, while integration in the recipient requires a final strand-transfer reaction. Thus, while a complete relaxosomal complex may be a reluctant substrate for nicking by TrwC, it behaves as an optimal substrate for strand transfer by a TrwC-DNA complex.

ii) We have shown that the integration reaction also took place when *trwC* was expressed only in the recipient, which implies that TrwC is able to locate both targets to perform strand transfer reactions. Curiously, when TrwC-DNA entered from the donor cell, additional expression of *trwC* in the recipient caused a decrease in the integration frequency (Table 3). This could be likely due to the fact that free TrwC molecules are bound to the *oriT* of the recipient plasmid, preventing strand transfer from incoming TrwC-DNA complexes.

iii) A double mutant in both TrwC catalytic residues (Y18FY26F) was inactive, confirming that nicking and strand transfer reactions are required for conjugation, recombination and integration. However, integration was totally dependent on both catalytic residues of TrwC, Y18 and Y26 (Table 4), while the Y26F mutation is well tolerated in recombination [15]. Y18 and Y26 are proposed to act in the initiation and termination of conjugative DNA transfer [12], [19]; it has been proposed that Y26 could have access to the *nic* site on supercoiled DNA once Y18 has formed the covalent complex [19]. Thus, while recombination requires two initiation nicking reactions both catalyzed by Y18, integration requires initiation and termination events.

iv) It has been described that formation of the IR2 hairpin at *oriT* enhances Y18 activity [19]. It can be expected that TrwC requirements for *oriT* recognition are more strict for the initial nicking reaction than for strand-transfer into a second target once the covalent complex is formed with the first target. This model correlates with our observation that changes in the critical area affecting either the *nic* or the binding sites abolished TrwC integration activity when TrwC was only expressed in the recipient cell and thus, it is required to act on the supercoiled mutant *oriT*. In contrast, incoming TrwC-DNA complexes could transfer DNA into acceptor sites with mismatches in the core *oriT* sequence (Table 5).

Figure 6 depicts a model for the integration reaction based on the above mentioned results. A TrwC-DNA complex enters the recipient cell, where a supercoiled *oriT*-containing acceptor plasmid is present. TrwC is attached to the T-strand through Y18, the residue responsible for the initial nicking event on supercoiled DNA (Fig. 6A). The relaxosome formed at the recipient *oriT* with TrwA and IHF would be the preferred

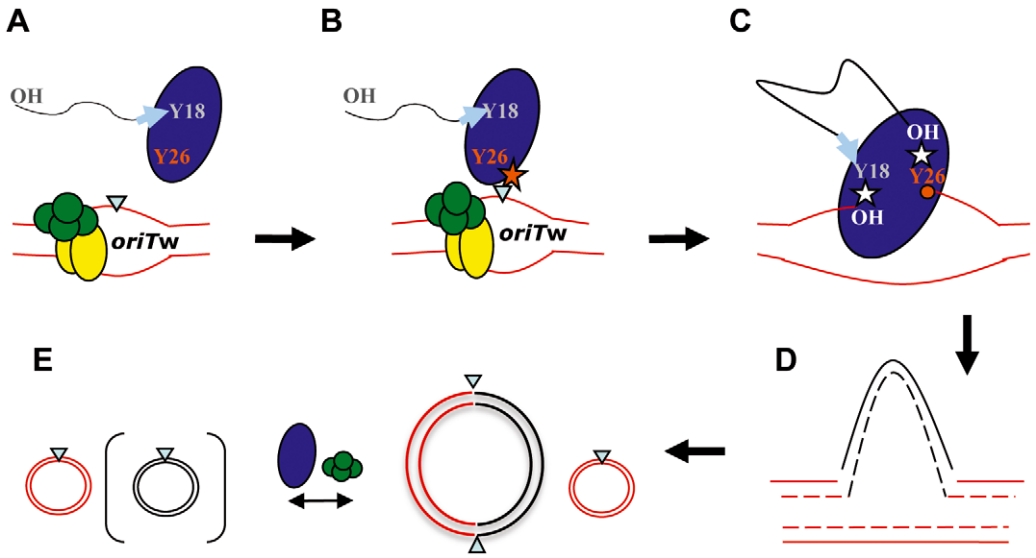

Figure 6. A model for TrwC-mediated site-specific integration. A. TrwC (blue oval) arrives to the recipient cell covalently bound to the suicide plasmid (dashed line) through the Y18 residue. The recipient contains the recipient plasmid (red lines). TrwA (green spheres) and IHF (yellow ovals) sit on the *oriTw* forming a relaxosome conformation which increase the exposure of ssDNA and the *nic* site (blue triangle). **B**. The incoming TrwC-DNA complex has a free Y26 residue. Y26 nicks (orange star) and binds covalently to the recipient *nic* site. **C**. Strand-transfer reactions are produced by the attack of the free –OH groups generated to Y18 covalently bound to the suicide plasmid and to Y26 attached to the recipient plasmid. As a result, the transferred DNA strand is integrated into the recipient plasmid. **D**. The host replication machinery duplicates the integrated DNA. **E**. Two molecules are obtained, the cointegrate and the recipient plasmid (*nic* sites indicated by triangles). The cointegrate can be resolved by TrwC-mediated site-specific recombination (enhanced by TrwA), producing the two initial molecules. The suicide plasmid is lost, since it cannot be replicated in the recipient cell.
doi:10.1371/journal.pone.0031047.g006

substrate for the free Y26 residue in the TrwC-(Y18)-DNA complex, which nicks the recipient *oriT* and forms a covalent intermediate (Fig. 6B); free TrwC molecules, however, would sit on the *oriT* copy and inhibit integration of incoming TrwC-DNA complexes. In this intermediate, there are two free 3′-OH ends which could attack each of the covalent Y-DNA complexes; as a result, the T-DNA strand would be integrated into the recipient *oriT* (Fig. 6C). Integration requires an additional strand-transfer reaction compared to conjugative DNA transfer, which may explain why the Y26F mutation affects integration more than conjugation. The resolution of the covalent intermediate (Fig. 6D) is probably mediated by the host replication machinery, as previously suggested for site-specific recombination [15]. Finally, in the presence of TrwC and TrwA, the cointegrate molecules would be resolved (Fig. 6E), giving rise to a higher proportion of recipient plasmids, as observed (Figs. 2, 3). This model is based on the action of a single TrwC monomer, as previously suggested also for conjugation [19], since the existence of two catalytic Tyr residues allows a monomer to perform both strand-transfer reactions; however, previous evidences of TrwC oligomerization *in vitro* [20] and DNA-independent TrwC transfer into recipients [6] leave open the possibility that TrwC acts as an oligomer both in donor and recipient cells.

Only a few relaxases have been reported to catalyze site-specific recombination, and the only other conjugative relaxase reported to catalyze intermolecular recombination is that of the self-transfer system of the integrative and conjugative element ICE*clc* [21]. This difference among conjugative relaxases has no correlation with taxonomic proximity (according to [22]), or with the number of catalytic Tyr residues. Intriguingly, the relaxase of ICE*clc* can act on two different *oriTs* for ICE transfer with similar efficiencies; however, it is able to catalyze *in vivo* strand-exchange between two plasmids carrying *oriT1* but not when carrying *oriT2* [21]. This result points to differences in the interaction of the relaxase with other relaxosomal proteins or with the *oriT* as factors which could determine the recombinase/integrase activity of relaxases, rather than to intrinsic catalytic differences among these proteins.

No other relaxase has been reported to perform site-specific integration of incoming DNA in recipient bacteria. In addition, we show here that TrwC could integrate the incoming DNA into a chromosomal *oriT* copy (Figure 5). Considering that R388 is a broad host range plasmid, the ability of TrwC to integrate the incoming DNA strand into a recipient chromosomal target may allow integration of R388 into the genome of non-permissive recipient hosts if a suitable target sequence exists in the recipient genome. We have shown that the 17 bp core of the R388 *oriT* was sufficient to act as a target for TrwC-mediated integration, and that DNA requirements are less stringent on the acceptor target site, broadening the possibility of finding natural TrwC targets in any recipient genome. This would confer R388 the possibility to colonize a broader range of microbial hosts in nature.

There have been reports of conjugative DNA transfer into different types of eukaryotic cells, implying that TrwC bound to a DNA molecule of any length could also have access to eukaryotic genomes. In this work, we show that TrwC catalyzed integration into two DNA sequences found in the human genome which resemble the minimal *oriT* (Table 5). Interestingly, integrants formed at these sites did not revert (Fig. 4B), so integration events obtained in this way would be stable, overcoming the limitation of reversible recombinases such as Cre [23]. Our results, together with previous reports on TrwC targeting to the nucleus [17], underscore the potential of TrwC as an integrase for genomic engineering of higher organisms.

Materials and Methods

Bacterial strains and growth conditions

E. coli strains II1 [24] and S17.1 λpir [25] were used as donor cells for the integration assays. As recipients, the following *E. coli* strains were used, as indicated: DH5α λpir [26], as a conjugation control of the suicide plasmid; DH5α [27] and CIG1 [16], as isogenic IHF+ and IHF− strains; and HMS174 [28], CMS1 and CMS2 (this work, see below), as isogenic strains without or with a chromosomal *oriT* copy, in both orientations. Bacteria were grown in Luria-Bertani (LB) broth, supplemented with agar for solid culture. For selection, antibiotics were used at the following concentrations: ampicillin (Ap), 100 μg/ml; chloramphenicol (Cm), 25 μg/ml; erythromycin (Em), 200 μg/ml; kanamycin (Km), 25 μg/ml; nalidixic acid (Nx), 20 μg/ml; rifampin (Rif), 100 μg/ml; streptomycin (Sm), 300 μg/ml. Thymidine (dT) was supplemented when necessary to a final concentration of 0.3 mM. IPTG was added in LB plates to a final concentration of 0.5 mM.

Plasmids and plasmid constructions

Published plasmids used in this work are listed in **Table S1**, and plasmids constructed for this work are detailed in **Table S2**. Plasmids were constructed using standard cloning procedures [29]. Restriction enzymes, Shrimp Alkaline Phosphatase, and T4 DNA ligase were purchased form Fermentas. For PCR amplification, high fidelity Vent DNA polymerase (New England BioLabs) was used. Primers used for PCR were obtained from Sigma-Aldrich. DNA sequences of all cloned PCR segments were determined.

Strain constructions

CMS1 and CMS2 are isogenic strains carrying an R388 *oriT* copy and a kanamycin resistance marker in place of the chromosomal *lacZ* ORF of *E. coli* strain HMS174. These strains were used as recipients to assay integration into the chromosome. The *oriT-km*^{R} cassette was introduced in the chromosome by the method of Dantsenko and Wanner [30] disrupting the *lacZ* gene. Oligonucleotides used for PCR reactions were 5′-ATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACACAGCTATGACCATGATTAC-3′, containing 40 bases of the 5′ region of the *lacZ* gene and annealing towards the *oriT*, and 5′-GACACCAGACCAACTGGTAATGGTAGCGACCGGCGCCAGCTGCTAAAGGAAGCGGAACA-3′, with 40 bases homologous to the 3′ end of the *lacZ* gene and annealing towards the *nptII* gene. Template DNAs were pCMS17 (*oriTw*▲-Km) and pCMS18 (*oriTw*▼-Km) (Table S2). One hundred nanograms of each PCR product were transformed into arabinose-induced HMS174 cells harbouring plasmid pKD20, coding for an L-arabinose-inducible λRed recombinase. Transformants were grown at 30°C and plated on LB+Km. Km resistant colonies were confirmed by their white colour on plates supplemented with X-gal and by PCR analysis using oligonucleotides P_A (5′-ATGACCATGATTACGGATTCA-‘3) and P_B (5′-GACACCAGACCAACTGGT-‘3), annealing to the 5′and 3′ ends of the *lacZ* gene, respectively.

Mating assays

Standard mating assays were performed as described [12], incubating the mating mixture for 4 hours at 37°C. Conjugation frequencies are expressed as number of transconjugants per donor cell.

Integration assays

TrwC-mediated site-specific integration assay described in [6] was optimized as follows. Matings were done using II1 as donor

strain containing a plasmid harbouring an *oriTw*, an *oriTp* and an R6K replicon (only replicates in strains expressing *pir*), used as a suicide plasmid for mobilization into strains lacking the Pir protein. The suicide plasmid was mobilized from II1 by a non mobilizable plasmid (pCIG1077) coding for the transfer region of plasmid R388 except for the *oriT*. DH5α was used as a recipient harbouring a plasmid with R388 *oriT* (*oriTw*) or RP4 *oriT* (*oriTp*), as a negative control. Integrants were selected in Cm plates, as it is the resistance provided by the suicide plasmid upon integration. The frequency is reported as the number of integrants per number of donor cells. Conjugative transfer controls were performed in all experiments using *pir* strains as recipients, and transfer efficiency was always close to 100% transconjugants/donor. As a negative control for integration, the suicide plasmid (containing *oriTp* and *oriTw*) was mobilized from S17.1 λpir by RP4_TraI to the same recipient as in the test. Thus, TrwC is not present in the reaction, but the *oriTw* enters in a single-stranded form, which allow us to rule out other causes of recombination in the recipient cell. In the integration assay via TrwC from the recipient, *trwA* and/or *trwC* were expressed in recipient cells under the control of the T7 promoter, to avoid possible toxic effects of overexpression.

Integration events were also analysed at the molecular level. Plasmid DNA was analysed by restriction analysis with enzymes which cut only once in the recipient plasmid and do not cut the donor plasmid (*BstEII*, *NdeI* or *XcmI*, as appropriate), to estimate the proportion of cointegrates. For PCR analysis, DNA was isolated from the colonies with Instagene (Bio-Rad). Primers P1 (5′-AGCGGATAACAATTTCACACAGGA-3′), annealing to the recipient plasmid, and P2 (5′-GCAGGATCCGC-TAAGCTTTGTCGGTCATTTCGA-3′), annealing to the donor plasmid downstream *oriTw*, were used to obtain an amplicon with a size of 1.2 kb only expected for the cointegrate molecules. As a positive control we used pKK::*oriT-Km* [6], which contains 826 bp of the suicide plasmid, mimicking a possible cointegrate molecule.

The analysis of the integrants obtained by TrwC-mediated site-specific integration into the chromosome were analyzed as follows. When using CMS1 as recipient strain, primers P_A and P_C (5′-GCCTCAAAATGTTCTTTACGA-3′), annealing to 3′ end of the *cat* gene, were used to amplify a specific region (850 bp) of the integrant. To verify integrants obtained when using CMS2 as recipient cell, primers P_C and P_B were used to amplify a band of 1800 bp specific of this type of integrant.

Supporting Information

Table S1 Published plasmids used in this work.
(DOCX)

Table S2 Plasmids constructed for this work.
(DOCX)

Acknowledgments

We are grateful to Maria Lucas c/o Fernando de la Cruz for providing plasmids with *oriT* mutations.

Author Contributions

Conceived and designed the experiments: LA ML. Performed the experiments: LA CGP CM. Analyzed the data: LA ML. Wrote the paper: LA ML.

References

1. Heinemann JA, Sprague GF, Jr. (1989) Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. Nature 340: 205–209.
2. Buchanan-Wollaston V, Passiatore JE, Cannon F (1987) The mob and oriT mobilization functions of a bacterial plasmid promote its transfer to plants. Nature 328: 172–175.
3. Waters VL (2001) Conjugation between bacterial and mammalian cells. Nat Genet 29: 375–376.
4. Llosa M, Gomis-Rüth F-X, Coll M, de la Cruz F (2002) Bacterial conjugation: a two-step mechanism for DNA transport. Mol Microbiol 45: 1–8.
5. Llosa M, de la Cruz F (2005) Bacterial conjugation: a potential tool for genomic engineering. Res Microbiol 156: 1–6.
6. Draper O, Cesar CE, Machon C, de la Cruz F, Llosa M (2005) Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. Proc Natl Acad Sci U S A 102: 16385–16390.
7. Moncalián G, Valle M, Valpuesta JM, de la Cruz F (1999) IHF protein inhibits cleavage but not assembly of plasmid R388 relaxosomes. Mol Microbiol 31: 1643–1652.
8. Lucas M, Gonzalez-Perez B, Cabezas M, Moncalian G, Rivas G, et al. (2010) Relaxase DNA binding and cleavage are two distinguishable steps in conjugative DNA processing that involve different sequence elements of the nic site. J Biol Chem 285: 8918–8926.
9. Moncalián G, Grandoso G, Llosa M, de la Cruz F (1997) oriT-processing and regulatory roles of TrwA protein in plasmid R388 conjugation. J Mol Biol 270: 188–200.
10. Moncalián G, de la Cruz F (2004) DNA binding properties of protein TrwA, a possible structural variant of the Arc repressor superfamily. Biochim Biophys Acta 1701: 15–23.
11. Guasch A, Lucas M, Moncalián G, Cabezas M, Pérez-Luque R, et al. (2003) Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. Nat Struct Biol 10: 1002–1010.
12. Grandoso G, Avila P, Cayón A, Hernando MA, Llosa M, et al. (2000) Two active-site tyrosyl residues of protein TrwC act sequentially at the origin of transfer during plasmid R388 conjugation. J Mol Biol 295: 1163–1172.
13. Llosa M, Grandoso G, Hernando MA, de la Cruz F (1996) Functional domains in protein TrwC of plasmid R388: dissected DNA strand transferase and DNA helicase activities reconstitute protein function. J Mol Biol 264: 56–67.
14. Llosa M, Bolland S, Grandoso G, de la Cruz F (1994) Conjugation-independent, site-specific recombination at the oriT of the IncW plasmid R388 mediated by TrwC [published erratum appears in J Bacteriol 1994 Oct;176(20):6414]. J Bacteriol 176: 3210–3217.
15. César CE, Machón C, de la Cruz F, Llosa M (2006) A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences. Mol Microbiol 62: 984–996.
16. César CE, Llosa M (2007) TrwC-mediated site-specific recombination is controlled by host factors altering local DNA topology. J Bacteriol 189: 9037–9043.
17. Agúndez L, Machón C, César CE, Rosa-Garrido M, Delgado MD, et al. (2011) Nuclear Targeting of a Bacterial Integrase That Mediates Site-Specific Recombination between Bacterial and Human Target Sequences. Appl Environ Microbiol 77: 201–210.
18. Llosa M, Bolland S, de la Cruz F (1991) Structural and functional analysis of the origin of conjugal transfer of the broad-host-range IncW plasmid R388 and comparison with the related IncN plasmid R46. Mol Gen Genet 226: 473–483.
19. Gonzalez-Perez B, Lucas M, Cooke LA, Vyle JS, de la Cruz F, et al. (2007) Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. Embo J 26: 3847–3857.
20. Grandoso G, Llosa M, Zabala JC, de la Cruz F (1994) Purification and biochemical characterization of TrwC, the helicase involved in plasmid R388 conjugal DNA transfer. Eur J Biochem 226: 403–412.
21. Miyazaki R, van der Meer JR (2011) A dual functional origin of transfer in the ICEclc genomic island of Pseudomonas knackmussii B13. Mol Microbiol 79: 743–758.
22. Garcillán-Barcia MP, Francia MV, de la Cruz F (2009) The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev 33: 657–687.
23. Kolb AF (2002) Genome engineering using site-specific recombinases. Cloning Stem Cells 4: 65–80.
24. Demarre G, Guerout AM, Matsumoto-Mashimo C, Rowe-Magnus DA, Marliere P, et al. (2005) A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalpha) conjugative machineries and their cognate Escherichia coli host strains. Res Microbiol 156: 245–255.
25. de Lorenzo V, Timmis KN (1994) Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons. Methods Enzymol 235: 386–405.
26. Platt R, Drescher C, Park SK, Phillips GJ (2000) Genetic system for reversible integration of DNA constructs and lacZ gene fusions into the Escherichia coli chromosome. Plasmid 43: 12–23.
27. Grant SG, Jessee J, Bloom FR, Hanahan D (1990) Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. Proc Natl Acad Sci U S A 87: 4645–4649.

28. Campbell JL, Richardson CC, Studier FW (1978) Genetic recombination and complementation between bacteriophage T7 and cloned fragments of T7 DNA. Proc Natl Acad Sci U S A 75: 2276–2280.
29. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring HarborN.Y.: Cold Spring Harbor Laboratory Press. 3 v p.
30. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97: 6640–6645.

Review

HUH site-specific recombinases for targeted modification of the human genome

Coral González-Prieto[1], Leticia Agúndez[2], Ralph Michael Linden[2], and Matxalen Llosa[1]

[1] Departamento de Biología Molecular (Universidad de Cantabria) and IBBTEC (UC, CSIC, SODERCAN), Santander, Spain
[2] King's College London School of Medicine, and UCL Gene Therapy Consortium, University College London, London, UK

Site-specific recombinases (SSRs) have been crucial in the development of mammalian transgenesis. For gene therapy purposes, this approach remains challenging, because, for example, SSR delivery is largely unresolved and SSR DNA substrates must pre-exist in target cells. In this review, we discuss the potential of His-hydrophobic-His (HUH) recombinases to overcome some of the limitations of conventional SSRs. Members of the HUH protein family cleave single-stranded (ss)DNA, but can mediate site-specific integration with the aid of the host replication machinery. Adeno-associated virus (AAV) Rep remains the only known example to support site-specific integration in human cells, and AAV is an excellent gene delivery vector that can be targeted to specific cells and organelles. Bacterial protein TrwC catalyzes integration into human sequences and can be delivered to human cells covalently linked to DNA, offering attractive new features for targeted genome modification.

Genome modification of human cells through site-specific integration of foreign DNA

Gene therapy aims to treat disease through the alteration of genome content in target tissues and, in most cases, requires stable expression of exogenous DNA. Long-term expression is either achieved through extrachromosomal persistence or by integration of the therapeutic DNA into the human genome; in particular, in proliferating cells. However, random integration in such scenarios has been demonstrated to carry the risk of insertional mutagenesis, potentially leading to tumor growth [1]. An alternative approach, gene targeting via homologous recombination, has recently witnessed promising advances thanks to the design of synthetic nucleases with a high degree of target specificity [2]. However, their design is complex and questions regarding off-target activities have yet to be addressed [3]. Furthermore, methods to predict these events reliably are still missing [4,5].

The use of SSRs (see Glossary), which directly integrate foreign DNA into a specific site in the genome, could help overcome some of these hurdles. There are, however, some potentially limiting prerequisites (unidirectionality of the reaction, existence of a natural target site) as well as additional inherent problems (SSR toxicity, DNA rearrangements). These are discussed here and we attempt to introduce possible solutions, thereby establishing this

Glossary

Adeno-associated virus (AAV): AAV is a small nonpathogenic human parvovirus whose life cycle consists of both a productive replicative phase and latent infection. It is known to need a helper virus for the productive life cycle. AAV Rep proteins, belonging to the HUH family, are essential for initiation of replication of the viral genome and for site-specific integration of the virus into a single target site present in the human genome.
Bacterial conjugation: mechanism of horizontal DNA transfer from a donor to a recipient bacterium. The process involves the generation of ssDNA, which is led into the recipient cell as a nucleoprotein complex by the conjugative relaxase. This HUH protein catalyzes cleavage and strand-transfer reactions at its target site to initiate and end ssDNA transfer. In addition, several relaxases can act as recombinases and integrases on double-stranded (ds)DNA substrates.
HUH protein: defined by the presence of two conserved motifs: a His-hydrophobic-His (HUH) motif required for metal ion binding, and a motif containing one or two catalytic tyrosines for nucleophilic attack of the DNA. HUH family members are strand transferases acting at target sites on a single DNA strand, and are involved preferentially in biological processes involving single-stranded (ss)DNA intermediates, such as rolling-circle replication and transposition, or bacterial conjugation.
Rolling-circle replication (RCR): mechanism used for the replication of some circular molecules, such as plasmids and certain viruses. A Rep protein, belonging to the HUH family, cleaves the target oriV and remains covalently bound to the 5′ end, providing a free 3′-OH end onto which nucleotides are added. This mechanism allows fast production of ss replication products. RCR transposition and processing of DNA during bacterial conjugation are related processes, also based on HUH proteins, which catalyze the initial cleavage and final re-ligation steps.
Site-specific recombinases (SSRs): enzymes that catalyze cut-and-strand transfer reactions on specific DNA sequences, producing rearrangements of DNA segments. A distinctive feature of these recombinases is the formation of a covalent protein–DNA intermediate during the recombination process.
Site-specific recombination (SSrec): also known as conservative site-specific recombination. Recombination process in which DNA exchange takes place between defined DNA sequences possessing only a limited degree of sequence homology, by a mechanism that conserves the phosphodiester bond energy. Depending on the initial arrangement of the two DNA partners, it can result in integration, excision, inversion, resolution, or translocation.
Tyr-SSR and Ser-SSR: two different families of SSRs named after the nucleophilic amino acid residue that they use to attack the DNA, and which becomes covalently linked to it during strand exchange. Although leading to the same practical outcomes, the two families are unrelated to each other, having different protein structures and reaction mechanisms (see Figure 1 in main text).

Corresponding authors: Linden, R.M. (michael.linden@kcl.ac.uk); Llosa, M. (llosam@unican.es).
Keywords: genomic engineering; site-specific recombination; gene therapy; HUH family; AAV Rep; R388 TrwC.

0167-7799/$ – see front matter
 http://dx.doi.org/10.1016/j.tibtech.2013.02.002

Table 1. Comparison of the most relevant SSRs discussed in this paper

SSR	Family[a]	Source	Host	Biological role	Stable integration[b]
Cre	Tyr-SSR	P1 phage	*E. coli*	Resolution of phage genome multimers	No
Int-ΦC31	Ser-SSR	ΦC31 phage	*Streptomyces lividans*	Integration and excision of phage genome	Yes[c]
AAV-Rep	HUH	AAV	*Homo sapiens*	Replication and integration of phage genome	Yes[d]
TrwC	HUH	R388 Plasmid	*E. coli*	Processing and leading DNA during bacterial conjugation	Yes[e]

[a]See text for details and Figure 1 for description of SSrec reaction mechanisms.

[b]Ability of the SSR to catalyze integration of foreign DNA on a target site present in a different genome, without catalyzing its excision subsequently.

[c]Int requires extra factors to catalyze excision.

[d]Rescue of AAV proviruses is thought to be mediated by the initiation of replication of integrated viral genomes. For this the cellular replication machinery and helper virus cofactors are required.

[e]TrwC-mediated reaction is reversible, but integration into the human targets could be irreversible (see text for details).

class of proteins as a viable addition to our tools to modify the human genome.

Site-specific recombination (SSrec): 2+1=2+2

SSRs catalyze the recombination between specific target DNA sequences. SSRs mediate this process in a reaction involving a covalent intermediate with the target DNA. The final product is the recombinant molecule, which, depending on the orientation of the target sequences, can lead to integration, deletion, inversion, resolution, or translocation of the DNA between the crossover sites [6].

In contrast to homologous recombination, SSrec has different biological roles (Table 1). Examples in prokaryotes, which illustrate the potential of SSrec, include the integration–excision cycles of bacteriophages and other mobile genetic elements, the resolution of plasmid multimers, or the control of gene expression through the repositioning of control elements. Among these are the inversion-mediated alternate expression of flagellins, or the assembly of active genes by irreversible deletion of a DNA segment leading to the expression of a reconstructed open reading frame (ORF) in nonvegetative cells, such as cyanobacterial heterocysts or *Bacillus* mother cells during sporulation [6].

SSRs usually refer to the canonical conservative recombinases that act on two target sites and catalyze their recombination by performing four nicks and two strand-exchange reactions. They are divided into two main families, Tyr-SSR (such as Cre recombinase from phage P1) and Ser-SSR (such as Int from the *Streptomyces* phage ΦC31), which differ in their structure, catalytic residues, and mode of action, yet perform essentially identical reactions with the same outcome, that is, recombination (Figure 1). For a detailed description of the recombination mechanisms of Tyr- and Ser- SSR, the reader is referred to a recent review [7].

The integrases responsible for gene cassette insertion/excision in integrons, also termed Int, belong to the family of Tyr-SSR, yet they perform a ss exchange reaction. These SSRs recognize a folded ss intermediate, thus forming an atypical Holliday junction (Figure 1). It has recently been demonstrated that recombination is resolved by the host replication machinery [8]. Thus, a reaction starting with a target dsDNA and a donor ssDNA (2+1) results in a recombination product identical to those of SSRs acting on two dsDNA molecules (2+2). This apparently surprising '2+1=2+2' equation is not so novel: old experiments have shown that Tn7 could switch from conservative to replicative transposition with a single mutation abolishing the second nicking reaction [9]. This observation suggests a remarkable plasticity in the ability to recruit host functions.

Within this scenario, a new unexpected family of SSRs has been found among members of the HUH family of enzymes, defined by the presence of a series of motifs involved in their catalytic activity [10]. These proteins are sequence-specific ss endonucleases that perform strand-transfer reactions on ssDNA, required in processes such as bacterial conjugation, and rolling-circle replication (RCR) or transposition. In addition, some of them have been shown to catalyze site-specific recombination and integration reactions on dsDNA substrates, and thus can also be considered as SSRs. These include conjugative relaxase TrwC of plasmid R388 [11]; rolling-circle replicase Rep from AAV [12]; and transposase TnpA(REP) from *Escherichia coli* K12 [13]. To date, all proposed models involving ss-based SSRs include a replication step, carried out by the host machinery, in order to complete the reaction [11,14,15] (Figure 1).

SSRs are characteristically sequence specific, very efficient, and often initiate recombination without cofactors or host-cell components. These features make them potentially useful tools for genome engineering [16]. Outstanding examples are the Tyr-SSR Cre from phage P1 or the Ser-SSR Int from phage ΦC31. Interestingly, most widely used SSRs come from bacteriophages and have been shown to be functional in the eukaryotic environment. For instance, ΦC31 Int has been shown to catalyze site-specific integration in plant, *Drosophila*, murine, and human cells [17]. This striking promiscuity hints at a considerable evolutionary conservation and possibly an as yet underappreciated function of these mechanisms in higher eukaryotes.

Prerequisites: unidirectionality and target site

SSRs can perform reversible or irreversible reactions. An example of an SSR that catalyzes recombination in both senses with equal efficiency is Cre from phage P1, which has been widely used for the construction of transgenic animals, including the generation of conditional phenotypes [18]. However, the reversibility of the recombination reaction can be a limitation if the goal is to obtain stable

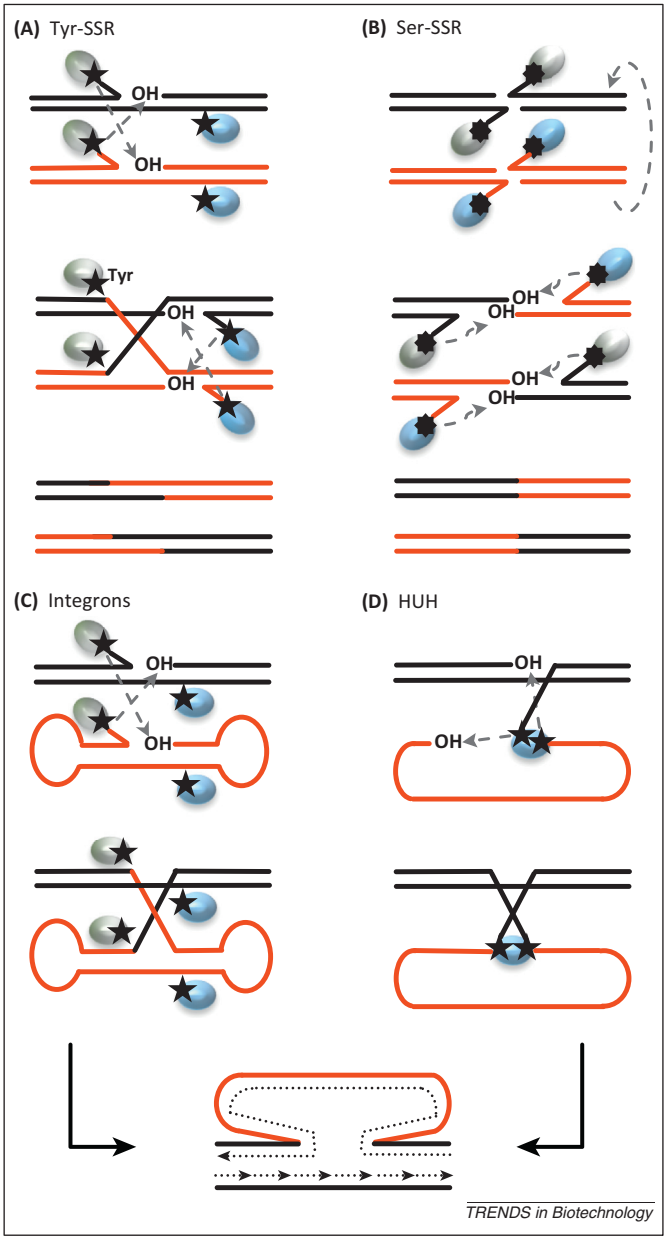

Figure 1. Comparison of recombination reactions catalyzed by site-specific recombinases (SSRs) belonging to different families. **(A)** Tyr-SSRs mediate conservative recombination by sequential action of two pairs of monomers, each catalyzing a strand-transfer reaction. The first crossover leads to the formation of a holiday junction (HJ). **(B)** Ser-SSRs mediate conservative recombination by concerted cleavage of four monomers, followed by a conformational switch of two of them, and both strand exchanges. **(C)** In the case of integrons, Int is an atypical Tyr-SSR which recognizes a folded single stranded (ss) substrate and catalyzes its integration through a single crossover that leads to the formation of an atypical holiday junction, aHJ [8]. This intermediate is resolved by replication. **(D)** Proteins of the His-hydrophobic-His (HUH) family of ss transferases such as R388-TrwC have been shown to catalyze also the integration of ss substrates through two transesterification steps that can be catalyzed by the two catalytic Tyr residues of the protein. The intermediate would be resolved by replication as in (C).

integration of a foreign DNA into the host genome, because the recombinase can catalyze the excision of this DNA at any moment. Some strategies, such as transient expression of the recombinase, can partially overcome this limitation.

By contrast, other SSRs perform a unidirectional reaction. Most SSRs determine the directionality of the reaction by recognizing specific target sites, which, upon recombination, generate new sites that can no longer serve as substrates. This is the case of phage integrases such as Int-ΦC31, which convert the phage attP and host attB sites into two new hybrid sites attL and attR, upon integration. Additional factors are required to catalyze recombination on these newly created sites [19], thus allowing a control of the directionality of the reaction.

One of the main limiting factors for the use of SSR genome modification is the potential absence of a naturally occurring target within the human genome. To date, most proof-of-concept studies use cells that have been engineered to contain respective SSR target sequences; this approach, however, is not applicable for gene therapy purposes.

The integrase of phage ΦC31 catalyzes unidirectional phage integration. This integrase also recognizes target sequences in many eukaryotic genomes, including the human genome [20]. Many pseudo-attP sites have been characterized (Table 2) into which this SSR can integrate any incoming DNA containing an attB motif. However, not all sites are used with the same efficiency: the ψA site in chromosome 8 was reported to be used preferentially [20], and a recent study has found an additional hotspot at 19q13.31 [21].

The use of Int-ΦC31 for integration of exogenous DNA in mammals originated more than a decade ago. The system is very efficient, rendering a high percentage of viable transformed cells. It has been used for genetic correction in mice and also in cultured human cells, including stem cells. Notably, in the past few years this system has allowed phenotypic correction of hemophilia A and B in mice through the expression of human clotting factors [22,23]. It has also been used for targeted integration in human muscle and cardiac progenitor cells [24,25], stem cell lines [26], as well as in approaches to generate mouse iPS cells [27,28]. It was not until recently that the additional protein required for phage excision was determined [19], which will allow further optimization of this tool for mammalian cells [29].

Other Ser-SSR phage integrases have been characterized and used successfully in a mammalian environment. R4 and A118 integrases can integrate an incoming plasmid into endogenous pseudo *att* sites in the human genome, although aberrant chromosomal events have been associated with R4, and for A118, four out of 15 integration events at pseudo *attB* sites showed imperfect junctions [30,31].

AAV is the only known virus capable of targeted integration in human cells. AAV Rep protein, an HUH protein, recognizes a unique target sequence within *AAVS1*, which is located on human chromosome 19 (19q13.3-qter) (Table 2), and several studies have demonstrated that insertion at this site poses no apparent risks [15]. The AAV-Rep-mediated site-specific integration reaction has extensively been studied in tissue culture. The characteristic Tyr of HUH SSR orchestrates the nicking in the target locus, *AAVS1*. Subsequently, a DNA strand exchange between the viral and human chromosomal sequences is proposed to form a covalent junction. This junction formation is then followed by the resolution of the intermediate by the host cell replication machinery [15], as

Table 2. SSR target sites in their host genomes and in the human genome

SSR	Natural target		Targets in human genome		Refs
			bona fide	*pseudo sites*	
Cre	*loxP*	ATAACTTCGTATAGCATACATTATACGAAGTTAT	none	4 ψlox	[64,65]
Int ΦC31	*attP* *attB*	CCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGG GTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCG	none	101 ψattP	[21,66]
AAV Rep	RBE_{itr} RBE_{P5}	GAGCGAGCGAGCGCGC GCCCGAGTGAGCACGC	*AAVS1*	*AAVS2, AAVS3*	[67] [38,39]
TrwC	*nic*	GGTGCGTATTGTCTATA	none	2	[11,59]

outlined in Figure 1D. AAV-Rep-mediated site-specific integration of foreign DNA has been achieved in mouse embryonic stem cells (mESCs) [15] and human embryonic stem cells (hESCs) [32], opening the possibility of using this system in replacement therapies in several human diseases.

A number of conjugative relaxases possess site-specific recombinase activity. Notably, the conjugative relaxase TrwC is also able to integrate the DNA to which it is covalently attached into its target sequence present in the recipient bacterial cell [33]. TrwC has two active Tyr residues that are both involved in the integration reaction [34]. Several sequences resembling its natural target exist in the human genome, and it has been shown that TrwC can catalyze integration into two of these sites [34]. Interestingly, integration is stable, because the target sequences are not substrates for the excision reaction [34]. These features confer potential as a genomic modification tool; however, it remains to be demonstrated whether this activity can be transposed into human cells.

Inherent risks: off-target insertion and genotoxicity

A key issue for the successful use of SSRs in human genomic modification protocols is their sequence specificity. SSRs are generally very specific enzymes, however, they can act on pseudotarget sites with sufficient similarity to their natural targets. This is particularly relevant when protein levels are high, under which conditions these enzymes can introduce nicks and ds breaks, thus possibly initiating unintended recombination events. Therefore, a main concern continues to be the inherent risk of introducing undesired DNA rearrangements during integration. Overall, the genotoxicity of SSRs has not been sufficiently studied to allow for reliable predictions about the safety of their use in humans.

The main drawback of using Int-ΦC31 for gene therapy, for example, is the presence of too many possible insertion sites (pseudo-att sites; Table 2). This problem may be partially overcome by the use of mutant integrases, obtained by directed evolution, that show increased specificity for the main ψA site [35]. Another hurdle is based on the observation that integration may lead to chromosomal rearrangements. This point is open to debate, because there are conflicting reports showing precise integration of Int-mediated transformed cells, whereas others demonstrate the presence of rearrangements upon Int-ΦC31-mediated integration in pseudo-attP sites [36]. It remains to be determined whether these differences are simply due to the cellular levels of the integrase.

AAV-Rep-mediated integration has been reported to be targeted to *AAVS1* in early studies [37,38]. Integration was analyzed in several latently infected human cell lines and 78% of these cells showed integration in *AAVS1*, highlighting the specificity of the system for the human target. Rep-mediated integration into pseudo-AAV sites has been reported [39]. However, because these studies were performed in HeLa cells, it remains to be shown whether these events were due to the well-documented propensity of AAV and AAV vectors to insert into pre-existing ds breaks, rather than due to a Rep-specific off-target event. Other reports have also challenged site-specificity, arguing that Rep-mediated integration was close to random [40]. However, a recent, unbiased analysis of integration sites has provided further evidence for the specificity of Rep-mediated integration [41].

AAVS1 is located within gene *PPP1R12C*, in a very gene-dense region in chromosome 19, making it more likely that possible rearrangements may have deleterious effects for the cell upon integration. An exhaustive study addressing potential adverse effects resulting from *AAVS1* integration has recently been presented. *AAVS1*-targeted mESCs have shown that, despite the resulting rearrangement, the cells maintained multilineage differentiation potential and contributed successfully to mouse development when injected into blastocysts [15]. A potential explanation for the lack of adverse effects in spite of significant associated rearrangements in this stringent model system has been provided by the observation of a duplication resulting from the integration mechanism, thereby preserving functional expression from the disrupted allele. These data underscore the potential of this locus as a suitable safe harbor for therapeutic transgene insertion when Rep is used to mediate integration. Several recent studies have also used *AAVS1* as the target for designed synthetic nucleases [42,43]. These studies further underscore the suitability of *AAVS1* for transgene expression. However, similarly stringent assays have yet to be used to assess whether gene addition to this locus is inert in the absence of Rep-mediated target gene duplication.

As highlighted above, the cellular level of integrase is possibly a key factor in the fidelity of the integration mechanism and thus the associated genotoxicity. Expression of Int-ΦC31 induces a DNA damage response and chromosomal rearrangements in human cells [44], but Int is expressed for only a few hours in both mouse liver and human cultured cells [45], making it unlikely that the integrase would induce such effects. Overexpression of Rep regulates the expression of cellular and viral genes and may induce apoptosis [46], DNA damage and cell cycle arrest [47]. This challenge underlines the necessity to control recombinase expression strictly. In this context, protein delivery by fusing a TAT domain has been reported

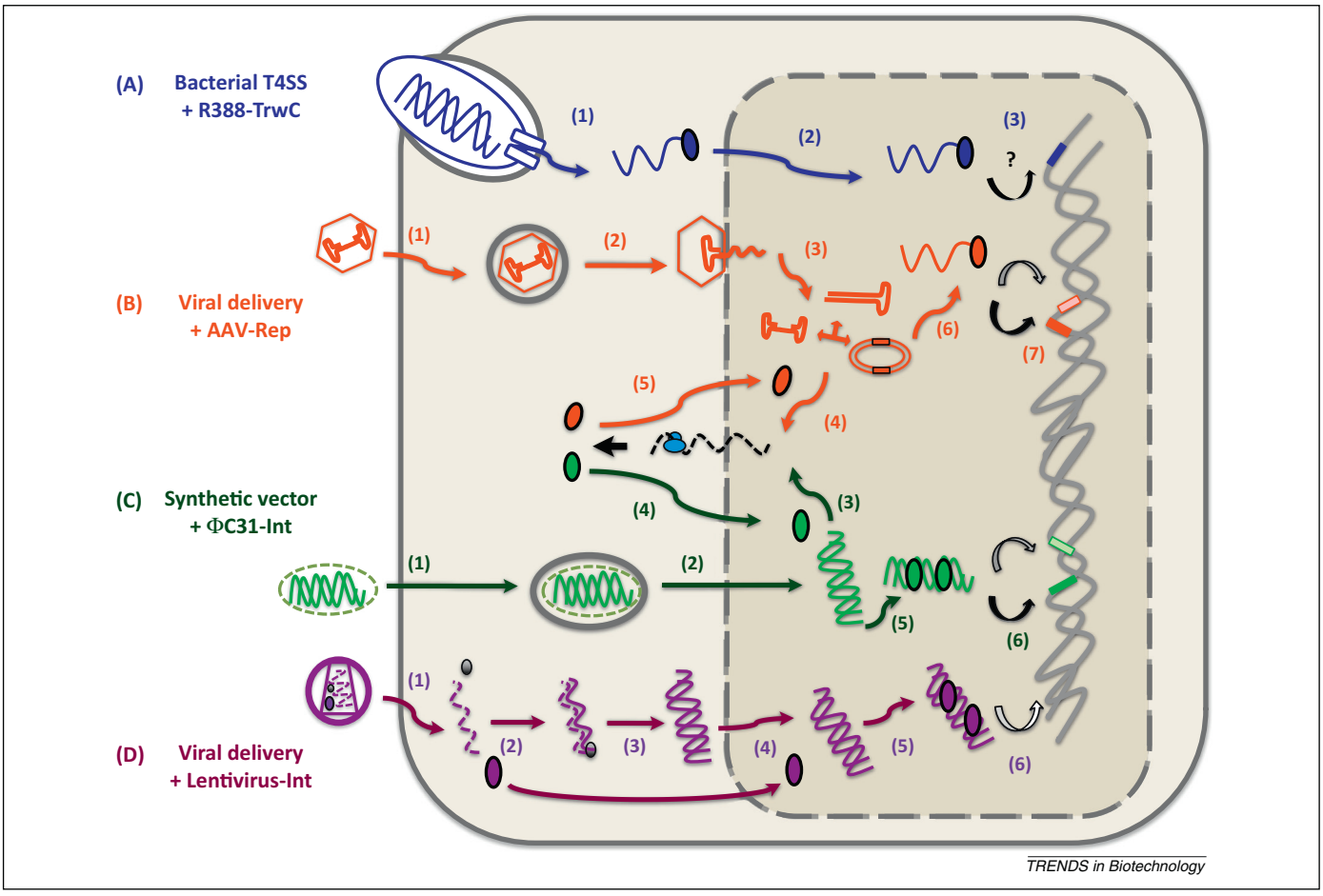

Figure 2. Different pathways for DNA delivery and gene targeting for human genomic modification. Representation of the different tools described in the text, which are under investigation or already being used for gene therapy purposes. **(A)** Bacterial T4SS + R388-TrwC: (1) The human pathogen *Bartonella henselae* transfers a plasmid that codes for the gene of interest together with the R388 conjugative relaxase TrwC. TrwC (blue sphere) covalently attached to the single-stranded (ss)DNA is secreted via the bacterial Type IV Secretion System (T4SS) [56]. (2) Once in the cytoplasm, the TrwC–ssDNA complex has to reach the nucleus; TrwC can be targeted to the nucleus [59]. (3) It is expected that the integrase catalyses a site-specific integration reaction (curved arrow in black) of the attached DNA into either Hu5 or HuX sites (represented by a blue square); both previously shown to be targets for TrwC integrase in *in vitro* assays [34], but this activity has yet to be proven *in vivo*. **(B)** Viral delivery + AAV Rep: (1) AAV binds to the cell using different surface glycans as receptors and specific co-receptors for efficient infection [68]. (2) AAV is internalized by endocytosis via clathrin-coated vesicles, followed by escape from the vesicles [69]. It is likely that AAV injects its genome into the nucleus. (3) The ssDNA genome is replicated to double-stranded (ds)DNA, required for gene expression. During replication, intermediates such as circular ds molecules are presumably assembled, allowing episomal persistence to the viral genome [70]. (4) RNA synthesis (broken black line) and subsequent translation (ribosome represented in light blue) is necessary to provide Rep integrase (red sphere). (5) A nuclear localization signal (NLS) targets Rep to the nucleus [58]. (6) The integrase targets its viral origin (presumably ssDNA [71]) in order to form protein–DNA complexes. (7) Site-specific integration (curved arrow in black) mediated by Rep occurs into *AAVS1* (red square, [15]). Other pseudosites (light red square and curved arrow in gray) have also been reported to act as targets with lower efficiency [39]. **(C)** Synthetic vector + φC31 Int: (1) DNA coding for the transgene and φC31 Int can be transfected directly to the target cell or introduced with synthetic vectors such as polymers or liposomes that interact with cellular receptors to achieve internalization through endocytosis. (2) After escaping from the vesicles, the vector is disassembled and the dsDNA (double line in green) reaches the nucleus. (3) This is followed by transcription and translation of the DNA. (4) φC31 Int (green sphere) is expressed in the cytoplasm and the protein has to find its way towards the nucleus. (5) Once inside, the integrase binds its target dsDNA and (6) catalyzes its site-specific integration [72] (curved arrow in black) into specific hotspots (green square), as well as into many pseudosites (curved arrow in gray and light green square) in the human genome [21]. **(D)** Viral delivery + Lentivirus Int: (1) The lentivirus RNA genome (broken purple line) is contained in the capsid (purple trapezoid) together with the reverse transcriptase (black spheres) and the viral integrase (purple sphere). Upon cell entry through binding to receptors and co-receptors, capsid proteins are uncoated, resulting in the release of the RNA genome together with the viral proteins into the cytoplasm [73]. (2) Reverse transcription takes place, giving rise to an RNA–DNA hybrid structure, which is subsequently converted to dsDNA (3) (double purple line). The dsDNA enters the nucleus, and the integrase leaded by an NLS [74] is imported to the nucleus (4). (5) The integrase binds the viral origins within the dsDNA and (6) catalyzes random integration (curved arrow in white) in the human genome [75].

for Int-ΦC31 [48], but this approach requires protein purification, and renders lower recombination efficiency.

Delivery of SSRs and DNA to the human cell

Any strategy for genome modification must include a way to deliver foreign DNA and the integration system to the target cells. DNA can be introduced into human cells by a variety of methods, including naked DNA, synthetic vectors, viral vectors, or bactofection [49]. SSRs from any source can be introduced by any of these methods. Figure 2 compares the entry pathways for various integrases.

Int-ΦC31 is routinely introduced by plasmid transfection of cultured cells, or by hydrodynamic tail-vein injection in mice. Introduction with adenovirus vectors allows for site-specific integration of large DNA fragments with low genotoxicity [50].

A key advantage of the AAV-Rep system is that the virus naturally infects human cells and the site-specific integration potential has been retained throughout evolution.

In addition, AAV poses no known safety issues. The main hurdle of AAV as a targeting vector is the strict space limitation imposed by its capsid. In order to overcome this problem, attempts have been made to use dual infections [51] or to incorporate Rep into other viruses, such as adenovirus or herpesvirus [52,53]. An additional benefit of AAV is that recombinant vectors have been widely used in preclinical as well as some clinical trials [54]. In addition, the availability of capsids from different AAV serotypes allows for targeting specific tissues, such as skeletal muscle, liver, central nervous system, retina, and heart [54]. Furthermore, the addition of a mitochondrial targeting sequence to AAV capsids redirects the virus inside the organelle, achieving correction of Leber's hereditary optic neuropathy in a mouse model [55].

The conjugative relaxase TrwC also shows a potential for *in vivo* delivery into human cells. A recent report has shown that this SSR can be delivered to specific human cells using the Type IV Secretion Systems (T4SS) of pathogenic bacteria [56]. T4SSs are encoded by many human pathogens, each targeting specific cell types, thereby introducing the potential for some tissue specificity for *in vivo* gene therapy [57]. A major advantage, however, might be that the SSR enters the cell in a covalent complex with the transgene and thus overcomes the need for recombinase expression in the cell and favors irreversible integration of the incoming DNA.

Many artificial delivery systems can reach the cytoplasm, yet integration takes place within the nucleus of the cell. Accordingly, AAV Rep has a nuclear localization signal (NLS) for nuclear targeting [58]. SSRs from bacteria or phages are not expected to target the nucleus, but it appears feasible to engineer an approach that includes nuclear import. For example, although TrwC localizes to the cytoplasm, a mutant shows nuclear localization [59]. Similarly, the addition of an NLS to Int-ΦC31 increases integration [60]. In addition, a TAT–Int–NLS has been shown to recombine more efficiently than TAT–Int in mammalian cells [48]. By contrast, nuclear localization provides little or no benefit to φC31 integrase for liver-directed gene therapy, even in the absence of cell division [61], suggesting that Int nuclear entry is not the limiting factor for integration. A small fraction of the integrase can enter the nucleus bound to the DNA, as occurs in the case of retroviral integrases [62].

Future directions

The main conclusions and outstanding questions on the use of SSRs for human genomic modification are outlined in Box 1. Taken together, no simple solution has been put forward to address the challenge that any system ideally would: (i) efficiently deliver DNA to the target tissue *in vivo*; (ii) allow there for efficient integration; (iii) express only transiently the required exogenous recombinase; and (iv) evade significant immune detection, and thus ensure the survival of the modified target cell. SSRs are a potential tool for *in vivo* and *ex vivo* genome modification. Although tailor-made nucleases show great potential, we propose that the inherent and unique characteristics of SSRs might provide distinct benefits that warrant further investigation.

Box 1. Conclusions and outstanding questions

Conclusions

- SSRs are valuable tools for human genomic modification.
- SSRs belonging to the HUH protein family catalyze integration of ssDNA with the aid of the host replication machinery.
- AAV Rep catalyzes integration into *AAVS1* with no known additional effect on the recipient genome, partly due to reconstruction of the target gene by partial duplication upon integration.
- Several conjugative relaxases have been shown to act as SSRs, providing new sources of potential integration sites. TrwC from plasmid R388 can integrate DNA into two sequences from the human genome, which resemble its natural target.
- TrwC can be delivered as a protein–DNA complex into specific human cell types through bacterial T4SSs. These machines are present in bacteria targeting different tissues.
- AAV is an excellent vector for delivery of the transgene and Rep. It can be targeted to different cellular types and even to mitochondria.
- The introduction of the SSR protein in place of the gene may be the best way to avoid genotoxicity.

Outstanding questions

- Will it be possible to target different human cellular types through the T4SSs of different human pathogens?
- Can TrwC integrate foreign DNA into its specific targets in the human genome? If so, what will be the effect of integration into these sites?
- Is it possible to obtain Rep-mediated modification of human iPS cells, allowing correction of diseases such as X-linked severe combined immunodeficiency?
- What will be a suitable system to transiently deliver AAV Rep to target cells in order to mediate site-specific integration?

Single-strand-dependent SSRs from the HUH family, such as AAV Rep, have been shown to be as efficient in integration as conservative SSRs, and proof-of-principle together with evidence for safety and utility of this approach have been provided in mESCs and hESCs [15,32]. Rep-mediated modification of human iPS cells is ongoing in our laboratory. The underlying strategy is that iPS cells can be obtained from patients and subsequently the genetic defect can conceivably be corrected by AAV-mediated site-specific integration of the unmutated gene where appropriate, resulting in a suitable cell population for differentiation and subsequent transplantation.

Bacterial conjugative relaxases may represent promising new tools due to their site-specific integrase activity and the presence of potential target sites within the human genome. In addition, new substrate specificities can be engineered [63], thus broadening the possibility to find the adequate insertion site. *In vivo* delivery through bacterial T4SSs as a covalent protein–DNA complex constitutes a unique feature conferring added value. However, to date, proof-of-concept for bacterial SSR-mediated site-specific integration into the human genome has yet to be provided.

HUH recombinases may represent a family of moonlighting proteins evolutionarily selected to perform site-specific integration, in addition to their role in viral replication, bacterial conjugation, or transposition; this molecular strategy has been preserved from bacteria, to plants, to mammalian viruses. The intricacies of this approach include efficient cooperation with host enzymes (thus only one exogenous protein is required) and, in the

case of AAV Rep, a mechanism that includes partial gene duplication through which functional expression from both target alleles is retained. Among this new family of recombinases is the possibility to overcome such problems as target infidelity and thereby genotoxicity. The main limiting factor for the use of SSRs is the presence of a target sequence in the human genome. However, these enzymes are highly prevalent in bacteria and viruses and, as TrwC demonstrates, it is likely that suitable candidates with human target sequences can be identified.

Acknowledgments

Work in our laboratories is supported by grant BIO2010-11623E from the Spanish Ministry of Science and Innovation to ML and UK Medical Research Council grant 1001764 to RML. CGP was a recipient of a predoctoral fellowship from the University of Cantabria (Spain).

References

1 McCormack, M.P. and Rabbitts, T.H. (2004) Activation of the T-cell oncogene LMO2 after gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* 350, 913–922
2 Mussolino, C. and Cathomen, T. (2012) TALE nucleases: tailored genome engineering made easy. *Curr. Opin. Biotechnol.* 23, 644–650
3 Handel, E.M. and Cathomen, T. (2011) Zinc-finger nuclease based genome surgery: it's all about specificity. *Curr. Gene Ther.* 11, 28–37
4 Gabriel, R. *et al.* (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* 29, 816–823
5 Pattanayak, V. *et al.* (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods* 8, 765–770
6 Craig, N.L. *et al.* (2002) *Mobile DNA II*, ASM Press
7 Hirano, N. *et al.* (2011) Site-specific recombinases as tools for heterologous gene integration. *Appl. Microbiol. Biotechnol.* 92, 227–239
8 Loot, C. *et al.* (2012) Replicative resolution of integron cassette insertion. *Nucleic Acids Res.* 40, 8361–8370
9 May, E.W. and Craig, N.L. (1996) Switching from cut-and-paste to replicative Tn7 transposition. *Science* 272, 401–404
10 Ilyina, T.V. and Koonin, E.V. (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res.* 20, 3279–3285
11 Cesar, C.E. *et al.* (2006) A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences. *Mol. Microbiol.* 62, 984–996
12 Smith, R.H. and Kotin, R.M. (2000) An adeno-associated virus (AAV) initiator protein, Rep78, catalyzes the cleavage and ligation of single-stranded AAV ori DNA. *J. Virol.* 74, 3122–3129
13 Ton-Hoang, B. *et al.* (2012) Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences. *Nucleic Acids Res.* 40, 3596–3609
14 Mendiola, M.V. *et al.* (1994) Differential roles of the transposon termini in IS91 transposition. *Proc. Natl. Acad. Sci. U.S.A.* 91, 1922–1926
15 Henckaerts, E. *et al.* (2009) Site-specific integration of adeno-associated virus involves partial duplication of the target locus. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7571–7576
16 Turan, S. and Bode, J. (2011) Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications. *FASEB J.* 25, 4088–4107
17 Geisinger, J.M. and Calos, M.P. (2013) Site-specific recombination using PhiC31 integrase. In *Site-directed Insertion of Transgenes* (Renault, S. and Duchateau, P., eds), pp. 211–239, Springer
18 Sorrell, D.A. and Kolb, A.F. (2005) Targeted modification of mammalian genomes. *Biotechnol. Adv.* 23, 431–469
19 Khaleel, T. *et al.* (2011) A phage protein that binds phiC31 integrase to switch its directionality. *Mol. Microbiol.* 80, 1450–1463
20 Thyagarajan, B. *et al.* (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol. Cell. Biol.* 21, 3926–3934
21 Chalberg, T.W. *et al.* (2006) Integration specificity of phage phiC31 integrase in the human genome. *J. Mol. Biol.* 357, 28–48
22 Keravala, A. *et al.* (2011) Long-term phenotypic correction in factor IX knockout mice by using PhiC31 integrase-mediated gene therapy. *Gene Ther.* 18, 842–848
23 Chavez, C.L. *et al.* (2012) Long-term expression of human coagulation factor VIII in a tolerant mouse model using the phiC31 integrase system. *Hum. Gene Ther.* 23, 390–398
24 Quenneville, S.P. *et al.* (2007) Dystrophin expression in host muscle following transplantation of muscle precursor cells modified with the phiC31 integrase. *Gene Ther.* 14, 514–522
25 Lan, F. *et al.* (2012) Safe genetic modification of cardiac stem cells using a site-specific integration technique. *Circulation* 126, S20–S28
26 Thyagarajan, B. *et al.* (2008) Creation of engineered human embryonic stem cell lines using phiC31 integrase. *Stem Cells* 26, 119–126
27 Ye, L. *et al.* (2010) Generation of induced pluripotent stem cells using site-specific integration with phage integrase. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19467–19472
28 Karow, M. *et al.* (2011) Site-specific recombinase strategy to create induced pluripotent stem cells efficiently with plasmid DNA. *Stem Cells* 29, 1696–1704
29 Farruggio, A.P. *et al.* (2012) Efficient reversal of phiC31 integrase recombination in mammalian cells. *Biotechnol. J.* 7, 1332–1336
30 Keravala, A. *et al.* (2006) A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. *Mol. Genet. Genomics* 276, 135–146
31 Olivares, E.C. *et al.* (2001) Phage R4 integrase mediates site-specific integration in human cells. *Gene* 278, 167–176
32 Yang, L. *et al.* (2008) Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature* 453, 524–528
33 Draper, O. *et al.* (2005) Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16385–16390
34 Agundez, L. *et al.* (2012) Site-specific integration of foreign DNA into minimal bacterial and human target sequences mediated by a conjugative relaxase. *PLoS ONE* 7, e31047
35 Sclimenti, C.R. *et al.* (2001) Directed evolution of a recombinase for improved genomic integration at a native human sequence. *Nucleic Acids Res.* 29, 5044–5051
36 Ehrhardt, A. *et al.* (2006) Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. *Hum. Gene Ther.* 17, 1077–1094
37 Kotin, R.M. *et al.* (1990) Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. U.S.A.* 87, 2211–2215
38 Kotin, R.M. *et al.* (1992) Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J.* 11, 5071–5078
39 Huser, D. *et al.* (2010) Integration preferences of wildtype AAV-2 for consensus rep-binding sites at numerous loci in the human genome. *PLoS Pathog.* 6, e1000985
40 Drew, H.R. *et al.* (2007) Increased complexity of wild-type adeno-associated virus-chromosomal junctions as determined by analysis of unselected cellular genomes. *J. Gen. Virol.* 88, 1722–1732
41 Ward, P. and Walsh, C.E. (2012) Targeted integration of a rAAV vector into the AAVS1 region. *Virology* 433, 356–366
42 Hockemeyer, D. *et al.* (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* 27, 851–857
43 Lombardo, A. *et al.* (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat. Methods* 8, 861–869
44 Liu, J. *et al.* (2009) PhiC31 integrase induces a DNA damage response and chromosomal rearrangements in human adult fibroblasts. *BMC Biotechnol.* 9, 31
45 Chavez, C.L. *et al.* (2010) Kinetics and longevity of PhiC31 integrase in mouse liver and cultured cells. *Hum. Gene Ther.* 21, 1287–1297
46 Schmidt, M. *et al.* (2000) Adeno-associated virus type 2 Rep78 induces apoptosis through caspase activation independently of p53. *J. Virol.* 74, 9441–9450
47 Berthet, C. *et al.* (2005) How adeno-associated virus Rep78 protein arrests cells completely in S phase. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13634–13639
48 Zhang, M.X. *et al.* (2009) TAT-phiC31 integrase mediates DNA recombination in mammalian cells. *J. Biotechnol.* 142, 107–113
49 Seow, Y. and Wood, M.J. (2009) Biological gene delivery vehicles: beyond viral vectors. *Mol. Ther.* 17, 767–777

50 Robert, M.A. *et al.* (2012) Efficacy and site-specificity of adenoviral vector integration mediated by the phage PhiC31 integrase. *Hum. Gene Ther. Methods* 23, 393–407
51 Ghosh, A. *et al.* (2008) A hybrid vector system expands adeno-associated viral vector packaging capacity in a transgene-independent manner. *Mol. Ther.* 16, 124–130
52 Saydam, O. *et al.* (2012) Construction and packaging of herpes simplex virus/adeno-associated virus (HSV/AAV) Hybrid amplicon vectors. *Cold Spring Harb. Protoc.* 2012, 352–356
53 Sitaraman, V. *et al.* (2011) Computationally designed adeno-associated virus (AAV) Rep 78 is efficiently maintained within an adenovirus vector. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14294–14299
54 Asokan, A. *et al.* (2012) The AAV vector toolkit: poised at the clinical crossroads. *Mol. Ther.* 20, 699–708
55 Yu, H. *et al.* (2012) Gene delivery to mitochondria by targeting modified adenoassociated virus suppresses Leber's hereditary optic neuropathy in a mouse model. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1238–E1247
56 Fernandez-Gonzalez, E. *et al.* (2011) Transfer of R388 derivatives by a pathogenesis-associated type IV secretion system into both bacteria and human cells. *J. Bacteriol.* 193, 6257–6265
57 Llosa, M. *et al.* (2012) New perspectives into bacterial DNA transfer to human cells. *Trends Microbiol.* 20, 355–359
58 Cassell, G.D. and Weitzman, M.D. (2004) Characterization of a nuclear localization signal in the C-terminus of the adeno-associated virus Rep68/78 proteins. *Virology* 327, 206–214
59 Agundez, L. *et al.* (2011) Nuclear targeting of a bacterial integrase that mediates site-specific recombination between bacterial and human target sequences. *Appl. Environ. Microbiol.* 77, 201–210
60 Andreas, S. *et al.* (2002) Enhanced efficiency through nuclear localization signal fusion on phage PhiC31-integrase: activity comparison with Cre and FLPe recombinase in mammalian cells. *Nucleic Acids Res.* 30, 2299–2306
61 Woodard, L.E. *et al.* (2010) Effect of nuclear localization and hydrodynamic delivery-induced cell division on phiC31 integrase activity. *Gene Ther.* 17, 217–226
62 Whittaker, G.R. and Helenius, A. (1998) Nuclear import and export of viruses and virus genomes. *Virology* 246, 1–23
63 Gonzalez-Perez, B. *et al.* (2009) Changing the recognition site of a conjugative relaxase by rational design. *Biotechnol. J.* 4, 554–557
64 Abremski, K. and Hoess, R. (1984) Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J. Biol. Chem.* 259, 1509–1514
65 Thyagarajan, B. *et al.* (2000) Mammalian genomes contain active recombinase recognition sites. *Gene* 244, 47–54
66 Groth, A.C. *et al.* (2000) A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5995–6000
67 Feng, D. *et al.* (2006) A 16bp Rep binding element is sufficient for mediating Rep-dependent integration into AAVS1. *J. Mol. Biol.* 358, 38–45
68 Henckaerts, E. and Linden, R.M. (2010) Adeno-associated virus: a key to the human genome? *Future Virol.* 5, 555–574
69 Ding, W. *et al.* (2005) Intracellular trafficking of adeno-associated viral vectors. *Gene Ther.* 12, 873–880
70 Duan, D. *et al.* (1998) Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. *J. Virol.* 72, 8568–8577
71 Daya, S. *et al.* (2009) Adeno-associated virus site-specific integration is mediated by proteins of the nonhomologous end-joining pathway. *J. Virol.* 83, 11655–11664
72 Thorpe, H.M. and Smith, M.C. (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5505–5510
73 Sakuma, T. *et al.* (2012) Lentiviral vectors: basic to translational. *Biochem. J.* 443, 603–618
74 Bouyac-Bertoia, M. *et al.* (2001) HIV-1 infection requires a functional integrase NLS. *Mol. Cell* 7, 1025–1035
75 Engelman, A. and Cherepanov, P. (2012) The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat. Rev. Microbiol.* 10, 279–290