UNIVERSIDAD DE CANTABRIA

PROGRAMA DE DOCTORADO EN BIOLOGÍA MOLECULAR Y BIOMEDICINA

Tesis doctoral:

Relaxasas conjugativas como vehículos de la translocación de proteínas y ADN a través de Sistemas de Secreción Tipo IV: implicaciones biológicas y biotecnológicas

PhD thesis

Conjugative relaxases as drivers of protein and DNA translocation through Type IV Secretion Systems: biological and biotechnological implications

Presentada por: **Dolores Lucía Guzmán Herrador**

Dirigida por: **Matxalen Llosa Blas**

Escuela de Doctorado de la Universidad de Cantabria

Santander, 2021

Dña. **Matxalen Llosa Blas**, Catedrática de Genética, perteneciente al Departamento de Biología Molecular de la Universidad de Cantabria,

CERTIFICA: Que Dña. **Dolores Lucía Guzmán Herrador** ha realizado bajo mi dirección el presente trabajo de Tesis Doctoral titulado: **Relaxasas conjugativas como vehículos de la translocación de proteínas y ADN a través de sistemas de secreción tipo IV: implicaciones biológicas y biotecnológicas.**

Considero que el trabajo de Dolores se encuentra terminado y reúne los requisitos necesarios para su presentación como Memoria de Doctorado al objeto de poder optar al grado de Doctor con opción a mención Internacional por la Universidad de Cantabria.

Y para que conste y surta los efectos oportunos, expido el presente certificado en Santander, a 26 de febrero de 2021.

Fdo. Matxalen Llosa Blas

El presente trabajo ha sido realizado en el Departamento de Biología Molecular de la Universidad de Cantabria, bajo la dirección de la profesora doctora Matxalen Llosa Blas.

Durante el periodo de la tesis doctoral Dolores Lucía Guzmán Herrador realizó dos estancias de investigación. En el laboratorio del Dr. Miguel A. Álvarez en el Instituto de Productos Lácteos de Asturias (Asturias, España), durante un mes. En el laboratorio del doctor David Bikard en el Instituto Pasteur (París, Francia), durante tres meses.

This work was conducted in the Department of Molecular Biology, University of Cantabria, under the supervision of professor Dr. Matxalen Llosa Blas.

Part of the experiments presented in this thesis were performed by Dolores Lucía Guzmán Herrador during two short stays. In the laboratory of Dr. Miguel A. Álvarez, at the Diary Lactic Products of Asturias, (Asturias, Spain) for one month. In the laboratory of Dr. David Bikard at the Institute Pasteur (Paris, France) for three months.

Esta Investigación ha sido financiada por una ayuda para contratos predoctorales en el área de la Biomedicina, Biotecnología y Ciencias de la Salud de la Universidad de Cantabria: 7665391046 Y0SC001170.

El trabajo en el laboratorio de Matxalen Llosa Blas ha sido financiado por el Ministerio de Economía, Industria y Competitividad de España: BIO2013-46414-P y BIO2017-87190-R.

Las estancias de investigación en el Instituto de Productos Lácteos de Asturias (IPLA) y en el Instituto Pasteur se realizaron gracias a sendas ayudas de la Universidad de Cantabria.

This research was supported by a predoctoral appointment from the University of Cantabria: 7665391046 Y0SC001170.

The work in Matxalen Llosa Blas's laboratory was financed by the Spanish Ministry of Economy, Industry and Competitiveness: BIO2013-46414-P and BIO2017-87190-R.

The research stays at the Diary Products Institute of Asturias (IPLA) and Institut Pasteur were supported by fellowships from the University of Cantabria.

Agradecimientos

En primer lugar, quiero agradecer a mi directora de tesis, Matxalen, por haber confiado en mí, y abrirme las puertas del laboratorio dándome la oportunidad de realizar esta tesis doctoral. Gracias por tu dirección y supervisión, durante estos años he aprendido muchísimo.

En segundo lugar, agradecer a todas las personas que a lo largo de estos años han formado parte del laboratorio. Gracias, por vuestros consejos y por el apoyo, por las risas y horas compartidas dentro y fuera del labo, todos habéis ayudado en el desarrollo de este trabajo, desde Coral quien me inició en el mundo del IBBTEC hasta las que somos ahora. Especialmente a ti Sara, gracias por haberme ayudado, escuchado y apoyado, sobre todo este último tramo.

To David Bikard, thank you for the great opportunity to join your lab in Paris. It was an incredible experience and I learnt a lot! También a todas esas personas que conocí en París, tres meses dan para mucho! Gracias a Miguel A. por acogerme en el IPLA e iniciarme en el mundo de los "lactos".

Gracias a los del laboratorio de al lado (¡aunque ya tenéis colonizado medio Instituto!). A las que ya estaban cuando llegué, a Sheila, Ana y Rachel, por todas esas dudas resueltas. A Mariadel, Alfonso e Irene, por preguntar siempre, por fiaros de mi criterio científico y por todas las risas ¡mucho ánimo con lo que queda! Gracias a Mapi, por todas las visitas a la sala de becarios estos meses de escritura, por tus charlas, y en general por toda tu ayuda. A los miembros de los Micromeetings por todos vuestros consejos, en especial a David, por ser mi compi de laboratorio durante una temporada, y seguir viniendo a visitarnos desde la 3ª planta. También agradecer a todo el servicio de técnicos del IBBTEC, especialmente a Mati, por haberme ayudado tanto estos años, sobre todo con la purificación de proteínas y los clones imposibles. En general, gracias a todo el mundo del Instituto que ha estado ahí durante estos 5 años, que se ha preocupado, me ha aconsejado y me ha ayudado a crecer en esta etapa predoctoral.

A Jorge y a Carol. Porque fuisteis los primeros en acogerme cuando llegué y porque me habéis acompañado durante todo el camino. Gracias por todas las horas y

aventuras compartidas. He aprendido muchísimo de los dos y me siento muy afortunada de haberos encontrado. GRACIAS.

Gracias a los que están ahí desde hace mucho y a todas esas personitas que me ha regalado Cantabria durante estos años. Gracias por todos los momentos, me habéis ayudado a desconectar y a llegar cada día con fuerza al laboratorio.

Para terminar, pero no menos importante, gracias a mi familia, a toda ella. Por haberme enseñado tanto, por haber estado siempre ahí, aunque me encontrase en la otra punta del país y por preguntar qué tal se portaba la ciencia y el norte conmigo. Carlos, gracias también por la ayuda de última hora, definitivamente el arte no es lo mío. Carmen, gracias por llamarme siempre que lo he necesitado. Gracias en especial a mis padres y mis hermanos. Por apoyarme en esta aventura norteña, por guiarme y por creer siempre en mí.

Gracias a Laro por estar siempre ahí, por apoyarme, animarme y escucharme cuando lo he necesitado. Gracias por hacer que todo sea más llevadero y por tener paciencia por los dos.

GRACIAS

Lolu

Contents

Abbreviations

aa	Amino acid
AAV	Adeno-associated virus
Ap	Ampicillin
aTc	Anhydrotetracycline
Bep	*Bartonella* effector protein
BFS	Bovim fetal serum
BID	Bep intracelular delivery
Bp	Base pair
Cas	CRISPR associated protein
Cfu	Colony-forming unit
Cm	Chloramphenicol
CP	Coupling protein
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR-RNA
DAP	Diaminopimelic acid
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleotide acid
dNTP	Deoxyribonucleotide triphosphate
DSB	Double-strand break
dsDNA	Double stranded DNA
Dtr	DNA transfer and replication
eGFP	Enhanced green fluorescent protein
Em	Erythromycin
FACS	Fluorescence activated cell sorting
gDNA	Genomic DNA
Gm	Gentamicin sulphate
gRNA	Guide-RNA
HGT	Horizontal gene transfer
His	Histidine

HR	Homologous recombination
HUH	Histidine-hydrophobic-histidine
Hyg	Hygromycin B
ICEs	Integrated conjugative elements
IHF	Integration host factor
IMC	Inner membrane complex
Inc	Incompatibility group
IPTG	Isopropyl-β-D-thiogalactopyranoside
Is	Integration site
Kb	Kilobase
kDa	Kilodalton
Km	Kanamycin monosulphate
LAB	Lactic acid bacteria
LB	Luria-Bertani broth
Log	Variation in 10 times magnitude
MOB	Mobility
MOI	Multiplicity of infection
Mpf	Mating pair formation
Mps	Mating pair stabilization
NHEJ	Non-homologous end joining
nic	Nicking site in the *oriT*
NLS	Nuclear localization sequence
nt	Nucleotide
Nx	Nalidixic acid
OD	Optical density
OMC	Outer membrane complex
o/n	Overnight
ORF	Open reading frame
oriT	Origin of transfer
$\boldsymbol{oriT_P}$	RP4 *oriT*
$\boldsymbol{oriT_w}$	R388 *oriT*

oriV	Origin of replication
PAM	Protospacer adjacent motif
PBS	Phosphate-buffered saline
PCR	Polimerase chain reaction
RCR	Rolling-circle replication
RNA	Ribonucleic acid
rpm	Revolutions per minute
scDNA	Supercoiled DNA
Sm	Streptomycin
ssDNA	Single-stranded DNA
SSI	Site-specific integrase
SSR	Site-specific recombinase
T4CP	Type IV coupling protein
T4S	Type IV secretion
T4SS	Type IV secretion system
TALEN	Transcription activator-like effectors nuclease
Thy	Thymidine
TRA	Conjugative transfer región
tracrRNA	Trans-activating crispr RNA
TS	Translocational signal
Tyr	Tyrosine
U	Enzyme unit
Wt	Wild-type
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

1. Introduction

1. Introduction

1.1. Bacterial conjugation

Bacterial conjugation is one of the main mechanisms of Horizontal Gene Transfer (HGT) among prokaryotes (de la Cruz and Davies, 2000). During conjugation, DNA is transferred from a donor to a recipient bacterium in physical contact through a conjugative apparatus. Conjugation is a very promiscuous process. It has been described in natural sources and under laboratory conditions between different bacterial species, between Gram-positive and Gram-negative bacteria (Trieu-Cuot *et al*., 1987; Aviv *et al*., 2016), and even between prokaryotic and eukaryotic cells (Lacroix and Citovsky, 2018). Bacterial conjugation has biological relevance, as it generates genetic variability (de la Cruz and Davies, 2000); from the clinical point of view, it contributes to the spread of virulence factors (Christie and Vogel, 2000) and antibiotic resistance (Mazel and Davies, 1999). Conjugative elements could be also used with biotechnological purposes as genetic modification tools (Gonzalez-Prieto *et al*., 2013).

The machinery utilized during conjugation to transfer DNA is usually encoded by conjugative plasmids or other mobile genetic elements such as integrated conjugative elements (ICEs) (Guglielmini *et al*., 2011). Conjugative systems carry two set of genes: mobility genes (MOB), involved in conjugative DNA processing, and mating pair formation genes (MPF), which provide the conjugative channel or type IV secretion system (T4SS) between donor and recipient cells. MOB genes codify a short DNA sequence required for plasmid mobility, the origin of transfer (*oriT*); a relaxase, which is the protein that catalyzes the first and last steps of conjugation; accessory proteins which contribute to the relaxase action; and a Type IV coupling protein (T4CP) to interconnect DNA processing with DNA transport.

According to their mobilization ability, plasmids can be classified as conjugative, mobilizable and non-mobilizable. Conjugative plasmids contain the two sets of genes necessary for their own transfer. Within mobilizable plasmids, there are different groups: there are mobilizable plasmids, such as RSF1010, without MPF genes which can also lack the T4CP, and they use the T4SS of a co-resident self-transmissible element.

Also, there are mobilizable plasmids which only harbor an *oriT* sequence and they need to hijack the MPF and MOB genes of other systems. Plasmids unable to conjugate or to be mobilized are called non mobilizable (Smillie *et al*., 2010; Ramsay and Firth, 2017).

Plasmids have been defined by their transfer range as broad or narrow host range plasmids depending on their ability to conjugate to a wide or to a narrow number of different recipient cells. This range definition also includes the replication range, as the measurement is performed by transconjugants detection, which requires the replication of the plasmid in the recipient cell (Suzuki *et al*., 2010). Therefore, the conjugation host range is probably underestimated and in fact, it is known that the ability to transfer is usually broader than the ability to replicate of a plasmid (Kishida *et al*., 2017; Samperio *et al*., 2021). Furthermore, under laboratory conditions, the generation of shuttle vectors has confirmed this broader conjugation range by mobilizing DNA even to yeast (Moriguchi *et al*., 2013).

1.1.1. Model for bacterial conjugation

Bacterial conjugation systems could be considered as the result of the merging of two ancient bacterial processes: rolling-circle replication (RCR) and a T4SS (Llosa *et al*., 2002) . This assumption is based on the high sequence similarities between relaxases and their target sequences (*oriT*) with RCR Rep proteins and their targets (*oriV*) (Waters and Guiney, 1993), on one side, and between the MPF conjugative genes and the family of protein transporters known as T4SS (Christie, 2001). These two ancient processes might became connected by the T4CP, which couples the plasmid replication machinery to the secretion system in the membrane (Llosa *et al*., 2002; Llosa and Alkorta, 2017).

The conjugative machinery is composed by three different modules (Llosa and de la Cruz, 2005; Cabezon *et al*., 2015):

1. The T4SS is a multiprotein complex that expands from the inner to the outer membrane of bacteria cells, forming a transmembrane conduit.
2. The relaxosome is a nucleoprotein complex formed by a relaxase, an *oriT* sequence and one or more accessory nicking proteins. It is responsible of DNA processing.

3. The coupling protein (T4CP) is an ATPase which drives the relaxosome to the T4SS. It connects both parts of the conjugative machinery.

For bacterial conjugation to occur, the donor cell has to be in contact with a recipient cell through an appendix known as the conjugative pilus, which requires the activation of the *mpf* genes. On the donor cell, the relaxase localizes the *oriT* (target) in the DNA to be transferred, it performs a specific single-stranded DNA (ssDNA) cleavage at the nicking site in the *oriT* (*nic* site), and it makes a covalent bound within its 5' end (**Figure 1**, (1)). This nucleoprotein complex is recruited by the T4CP of the T4SS. Meanwhile, replication starts from the 3' end of the cleaved strand using as template the uncleaved strand (2). The unwinding of the DNA is produced, originating ssDNA to be transferred. The nucleoprotein complex is transferred through the T4SS channel, helped by the T4CP pumping activity (3).

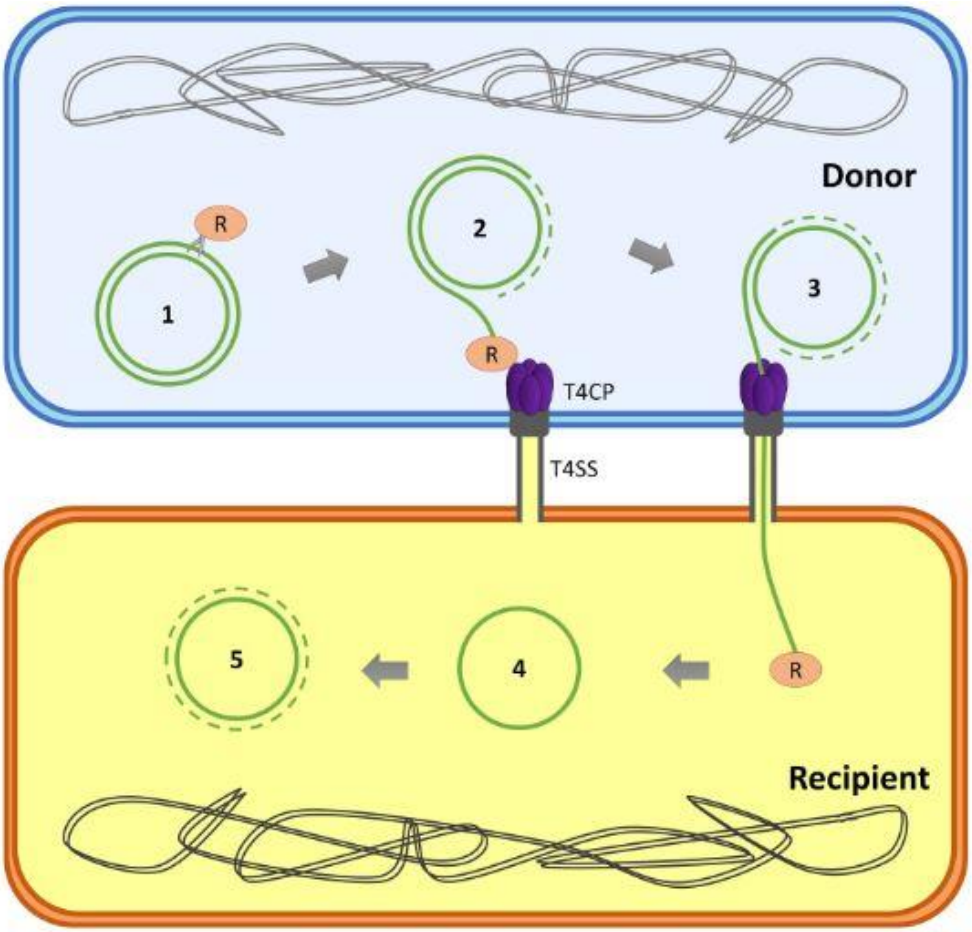


Figure 1. General scheme of bacterial conjugation R, relaxase. (See text for more details). Taken from (Getino and de la Cruz, 2019)

Once in the recipient cell, the relaxase is active and catalyzes the recircularization of the transferred DNA strand (4). The transferred ssDNA is converted into a double-stranded DNA (dsDNA) plasmid (5). Now, the recipient cell can act as a donor for a new cycle of conjugation (Getino and de la Cruz, 2019).

Bacterial conjugation implies the translocation of DNA through a T4SS, however, it is unclear how the DNA is transported from the donor to the recipient cell. The shoot and pump model proposed a two-step mechanism, using the conjugative plasmid R388 as paradigm (Llosa *et al.*, 2002). During the first step, the relaxase will be the active substrate of the T4SS and will act as a pilot protein, and the DNA, which is covalently bound to the relaxase, will be passively transported into the channel. On the second step, when the relaxase is shooted from the channel, the DNA would be actively pumped out through the T4SS, presumably by the T4CP (**Figure 2**). While the transport of the relaxase has been thoroughly documented in different conjugative systems (Luo and Isberg, 2004; Draper *et al.*, 2005; Dostal *et al.*, 2011), until now there is no evidence of the involvement of the T4CP ATPase activity or any other T4SS ATPase in the DNA transfer process.

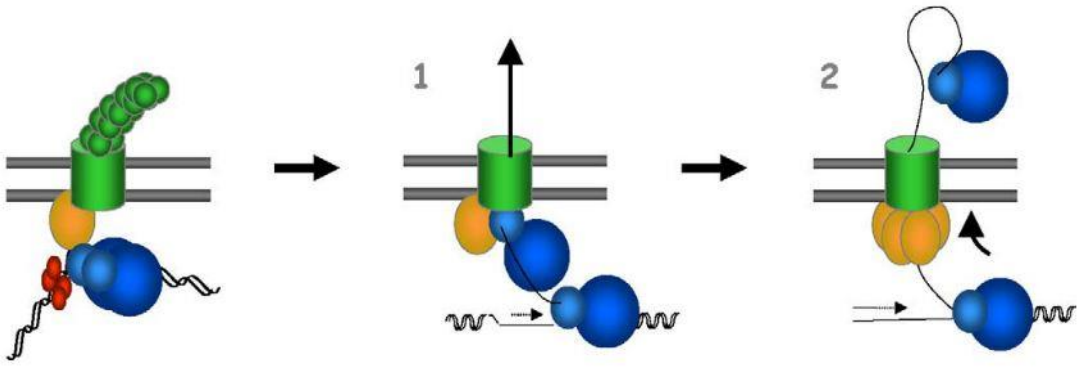


Figure 2. Shoot and pump model for conjugal DNA transfer. The hypothesis divided the DNA transfer process in 2 steps. (1) Shooting step: the relaxase is actively transported through the channel, while the ssDNA covalently bound to it is passively transported. (2) Pumping step: the remaining DNA is actively pumped across the conduit by the T4CP. Taken from (Llosa and de la Cruz, 2005).

1.1.2. Type IV Secretion systems

Bacterial T4SS are a highly diverse superfamily of macromolecule transporter systems. They are multiprotein nanomachines described in Gram-positive and Gram-negative bacteria. Their high plasticity is demonstrated at a functional level by all the different activities that they are involved in, the different substrates they can translocate, and the various possible destinations of the cargo (Grohmann *et al*., 2018). T4SS are divided in three different subfamilies depending on the function that they are involved in (**Figure 3**) (Li *et al*., 2019).

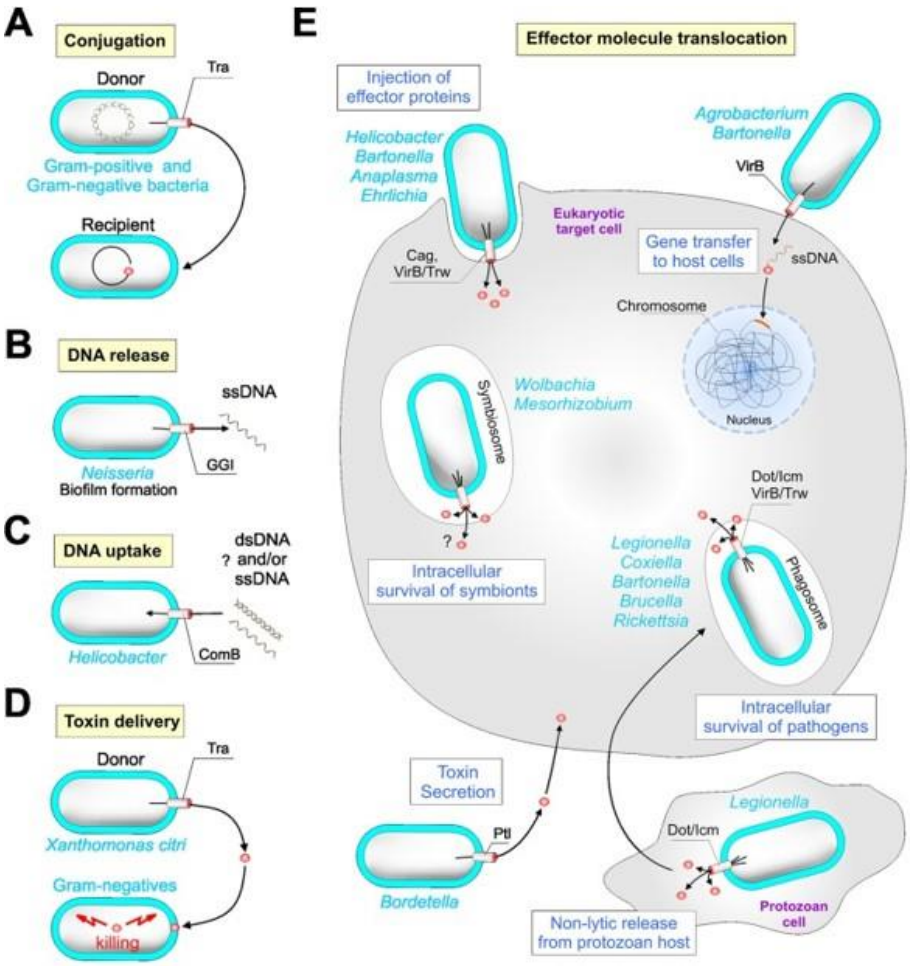


Figure 3. T4SS functional diversity in T4SS of Gram-negative bacteria. T4SS can perform different biological roles which involve the uptake or translocation of different substrates into the media or into eukaryotic or prokaryotic recipient cells. **a**) Conjugative T4SS translocate DNA from a donor bacterium into various recipient. **b and c)** DNA release and uptake systems facilitate an exchange of DNA with the extracellular space**. d)** T4SS can deliver protein toxin to kill neighboring bacterial competitors. e**)** T4SS from pathogenic bacteria can deliver effector proteins or DNA–protein complexes into their host. Taken from (Grohmann *et al*., 2018).

-T4SS involved in conjugation transfer a nucleoprotein complex from a donor to a recipient cell during conjugation. They have an important clinical impact because they contribute to the antibiotic resistance spread (Davies and Davies, 2010). This family includes R388 or RP4 conjugative systems.

-T4SS involved in protein translocation to recipient cells are used by bacterial pathogens to translocate effector proteins to the host cells during bacterial infection. This family includes the T4SS of relevant human pathogens such as *Bartonella* spp or *Legionella* spp. This subfamily also includes T4SS involved in bacterial killing. Although this activity was thought to be exclusive of Type VI Secretion Systems (T6SS), it has been shown that, for example, *Xanthomonas* spp. use T4SS to translocate toxin components of toxin-antitoxin systems to kill neighboring bacteria (Souza *et al*., 2015).

-T4SS involved in DNA export and import mediate DNA transfer between the bacteria and the external medium. They are involved in different activities such as DNA exchange with the media (contributing to the genetic exchange between bacteria) or biofilm formation (Grohmann *et al*., 2018). This subfamily includes the ComB T4SS of *Helycobacter pylori* involved in DNA uptake, and the Tra system of *Neisseria gonorrhoeae* involved in DNA export. These systems will not be further described here.

Independently of their biological function, T4SS in Gram-negative bacteria are classified into two different subfamilies according to their sequence similarity: Type IVA and Type IVB (**Figure 4**) (Christie *et al*., 2017). *Agrobacterium tumefaciens* VirB/D4 T4SS is the paradigm of the first group and provides a unifying genetic nomenclature for all T4ASS. The VirB/D4 has been characterized extensively (Cabezón *et al*., 2015). It is composed of 12 subunits: VirB proteins from 1 to 11 and VirD4 protein (the T4CP). These subunits form the different parts of the T4SS: the pilus, the core channel complex, the inner membrane platform and the three hexameric ATPases at the base of the channel, which are in charge of supplying the energy necessary for the pilus biogenesis and the substrate transport. Members of this family include the conjugative T4SS of plasmids R388 or RP4, or the T4SS VirB/D4 involved in effector translocation of the human pathogen *Bartonella henselae*. Type IVB are composed by more than 25 proteins. Most of these proteins (over 20) are specific of T4BSS, and not related to VirB/D4 units. The paradigmatic T4BSS is the Dot/Icm T4SS of the human pathogen *Legionella*

pneumophila. Other members of this family are the conjugative T4SS of F plasmid or the Dot/Icm T4SS of the human pathogen *Coxiella burnetii* (Voth *et al*., 2012).

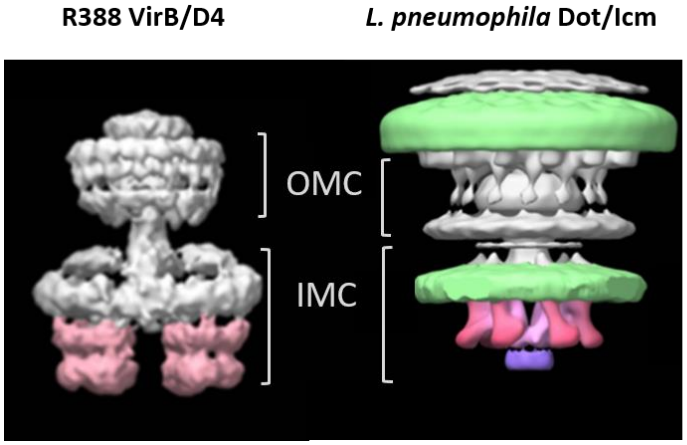


Figure 4. Three-dimensional (3D) structures of the VirB/VirD4 from plasmid R388 and Dot/Icm from *L. pneumophila* T4SS. Hexameric barrels of the VirB4 ATPase are colored in pink. The bacterial membranes are in green and the DotO and DotB hexameric ATPases are in shades of pink and purple, respectively. OMC: outer membrane complex/core. IMC: inner membrane complex. Adapted from (Li *et al*., 2019).

1.1.2.1. T4SS involved in effector translocation

Many bacterial pathogens use T4SS to deliver effector proteins to the host cell, contributing to the virulence of pathogens such as *Legionella pneumophila*, *Brucella melitensis, Helicobacter pylori,* or *Bartonella henselae* (Grohmann *et al*., 2018). Some T4SS translocate only one or few proteins, such as *H. pylori Cag* T4SS, which translocates only one substrate, and others translocate hundreds of different effectors, such as *L. pneumophila* Dot/Icm. T4SS involved in protein translocation recognize their substrates through different signals located on the C-terminal part of the protein (Grohmann *et al*., 2018).

The VirB/D4 T4SS of B. henselae

Bartonella is a genus of α-proteobacteria, which includes different facultative intracellular pathogens producing hemotrophic infection in different mammalian species, including humans (Wagner and Dehio, 2019). *B. henselae* is a worldwide zoonotic pathogen. Cats are its natural host, where the pathogen causes sub-clinical

intra-erythrocytic bacteremia and from which the bacteria can infect humans. In humans, *B. henselae* promotes different symptoms, depending on the immune status of the human host. There are a range of clinical manifestations, including cat-scratch disease in immunocompetent individuals, or bacillary angiomatosis and peliosis in immunocompromised patients (Dehio, 2005).

Although there are three different T4SS described in the *Bartonella* genus (Trw, Vbh/TraG, and VirB/VirD4), in this work we are only going to refer to the VirB/D4 T4SS (Saenz *et al*., 2007). The T4SS VirB/D4 of *B. henselae* is one of the principal virulence factors of this pathogen (Padmalayam *et al*., 2000; Schmiederer and Anderson, 2000). *B. henselae* VirB/D4 is composed by an operon of 10 genes with strictly conserved gene order (*virB2-B11*) plus another operon with the T4CP virD4 (**Figure 5**) (Schulein and Dehio, 2002). There are different studies which established the role of VirB/D4 and the effector translocated in the regulation of cellular functions in nucleated mammalian cells, such as human endothelial cells *in vitro,* and their importance for reaching and colonialization of the blood-seeding niche *in vivo* (Harms and Dehio, 2012).

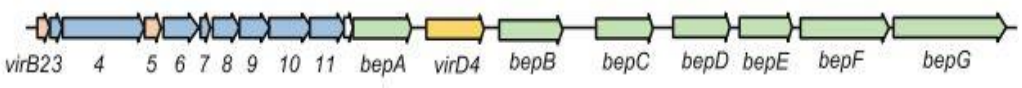


Figure 5. Genetic organization of the *virB/virD4/bep* locus in *B. henselae.* The *virB* components (*virB2-virB11*; in light pink: pilus associated components, in blue: rest of the component of the T4SS), the coupling protein (*virD4*, in yellow), Bep proteins (*bepA-bepG,* in green) Taken from (Dehio and Tsolis, 2017).

VirB/D4 T4SS is an important virulence factor because it translocates *Bartonella* effector proteins (Beps) into host cells (Schulein *et al*., 2005). Beps are multidomain proteins composed of an N-terminal effector domain and a C-terminal bipartite translocation signal. This signal is composed by a Bep Intracellular Delivery (BID) domain, essential for the intracellular delivery of the Bep proteins, and an unconserved positively charged tail sequence which mediates the T4SS recognition by the T4CP. Apart from its role in substrate recognition, BID domains may also have evolved to play a role related to effector function within host cells (Wagner and Dehio, 2019).

It has been reported than VirB/D4 from *B. henselae* could also translocate conjugative relaxases, covalently bound to DNA, into human cells (Schulein *et al*., 2005; Fernández-González *et al*., 2011). This point is detailed in Section 1.2.3.3.

1.1.2.2. T4SS involved in conjugation

T4SS associated to conjugative systems are the most widely distributed subfamily of T4SS. They can be found in Gram-positive and Gram-negative bacteria, and even in archaea (Alvarez-Martinez and Christie, 2009). They are essential in bacterial conjugation, previously described in Section 1.1. (Grohmann *et al*., 2018). One of the best known conjugative T4SS is the one of the conjugative plasmid R388, which is described below.

The conjugative T4SS of plasmid R388

Plasmid R388 was first isolated from *E. coli* in 1972 (Datta and Hedges, 1972). R388 is a broad host range plasmid; it is transferred to most proteobacterial species (Fernández-López *et al*., 2006) and even to cyanobacteria (Encinas *et al*., 2014). Its pilus is rigid, which entails better transfer frequencies in solid surfaces (Bradley *et al*., 1980). R388 has a size of 33,926 bp and its sequence has been divided in functional modules, corresponding to all basic functions implicated in survival and spreading of the plasmid (**Figure 6**) (Fernández-López *et al*., 2006).

The TRAw region contains the genes responsible for the synthesis and the assembly of the T4SS and the genes responsible for DNA processing and mobilization (Fernández-López *et al*., 2006). This region is subdivided in two functional units (MPF and MOB region). The MFP region contains the genes *trwD* to *trwK,* which encode the proteins necessary for the formation of the T4SS and the conjugative pilus. The MOB region (also known as Dtr region) contains the operon *trwABC* and the *oriT* sequence (Llosa *et al*., 1994a).

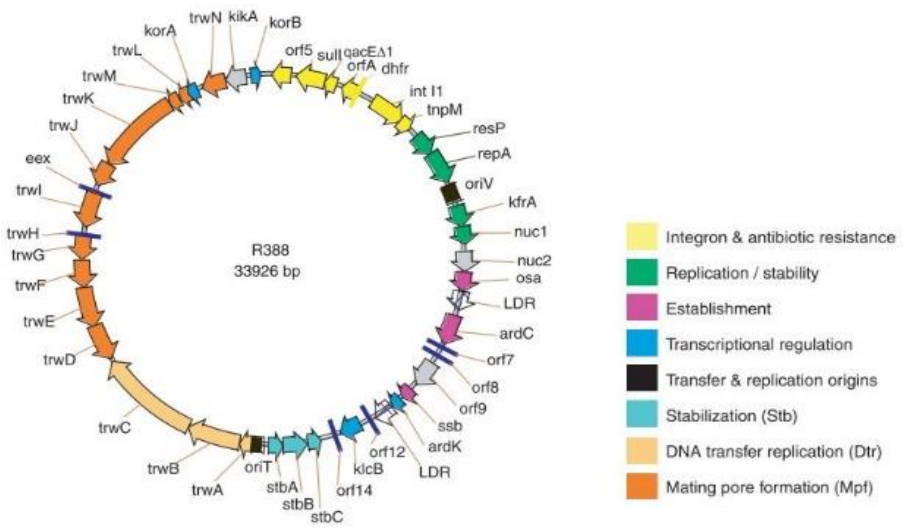


Figure 6. Genetic map of plasmid R388. The figure shows the organization in modules of the plasmid, shown by the color-code. Genes involved in conjugation (T4SS formation and DNA processing and mobilization) are shown in orange. Taken from (Fernández-López *et al*., 2006).

oriT is the minimal sequence required in a DNA molecule to be transferred efficiently by conjugation. The R388 *oriT (oriT$_w$)* was first described as a sequence of 402 bp (Llosa *et al*., 1991) and then delimited to 63-330 bp (César *et al*., 2006). It contains the TrwA and TrwC binding sites. The *nic* is where TrwC introduces a nick at nucleotide (nt) 176 (bottom strand) (Llosa *et al*., 1995). As the transferred strand is the nicked one, and the 5' end is the first transferred, *trw* genes enter the recipient in the last place.

TrwA is the relaxosome accessory protein. It is dispensable for conjugation, although its absence decreases drastically the conjugation frequency (Llosa *et al*., 1994a). TrwA binds specifically to two regions within the *oriT*. It also enhances TrwC relaxation activity *in vitro* (Moncalián *et al*., 1997), TrwC-mediated site-specific recombination (César *et al*., 2006), and integration (Agúndez *et al*., 2012), and the ATPase activity of the T4CP TrwB (Tato *et al*., 2007).

TrwB is the coupling protein of R388. It connects the relaxosome and the T4SS during the conjugative transfer of R388 (Llosa *et al*., 2003). TrwB is involved in different functions: it has a presumed role in DNA transport, it interacts with TrwA (Llosa *et al*., 2003) and it is also necessary for TrwC translocation even in the absence of DNA transfer (Draper *et al*., 2005), suggesting that TrwB is needed for protein substrate recruitment.

TrwC is the relaxase-helicase of R388. It is described in detail in Section 1.2.3.

1.2. Conjugative relaxases

Conjugative relaxases are well-characterized proteins responsible for initiating and terminating DNA processing during the conjugative DNA transfer process (Guzmán-Herrador and Llosa, 2019). They are endonucleases with site- and strand- specific activity, which catalyze a transesterification reaction, acting specifically at the *nic* site, in the *oriT* sequence (Zechner *et al*., 2017).

Current relaxases classification divided them into 9 MOB families: $MOB_{F,}$ MOB_P, MOB_Q, MOB_V, MOB_C, MOB_H, MOB_T, MOB_M and MOB_B (Garcillán-Barcia *et al*., 2020). Previously, it was thought that all relaxases belonged to the histidine-hydrophobic-histidine (HUH) superfamily and that all shared similar structural homology. In fact, the resolution of the 3D structure of different relaxases revealed the conservation of the HUH fold (Chandler *et al*., 2013). However, an increasing number of relaxase families with non-canonical HUH motifs, have revealed a diversity of proteins motifs and catalytic alternatives (Guzmán-Herrador and Llosa, 2019). Actually, from the 9 MOB families described, 6 of them contains an HUH fold: MOB_F, MOB_Q, MOB_P, MOB_V, MOB_H and MOB_B (**Figure 7**).

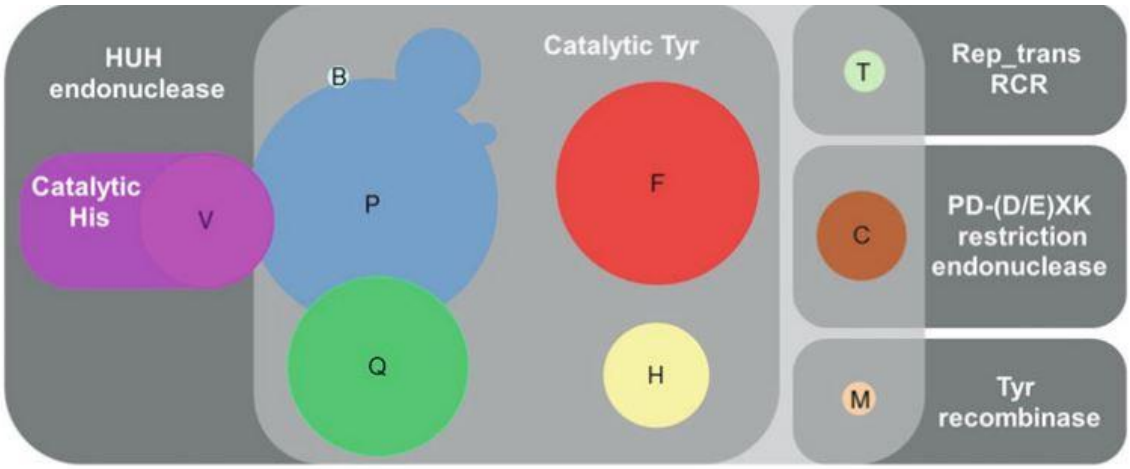


Figure 7. Scheme of the relationships between the relaxase protein families. Each superfamily to which relaxase MOB belongs are clustered in dark gray boxes (HUH endonucleases, tyrosine (Tyr) recombinases, Rep_trans RCR initiators, and PD-(D/E)XK restriction endonucleases). MOB families with 1 or 2 active catalytic Tyr> are shadowed by a light gray box. Light magenta box contains the MOB family which uses a catalitic active His residue, instead a Tyr residue. Each MOB family is represented by colored circle. The size of the circle is proportional to the number of known relaxases it includes. Overlapping circles indicate homology between profiles. Taken from (Garcillán-Barcia *et al*., 2020).

There are three amino acid sequence motifs typically present in the HUH superfamily (Pansegrau *et al*., 1993a):

- Motif I: it contains the catalytic tyrosine(s) residue(s), which form the covalent complex with the nicked DNA.
- Motif II: it contains a conserved serine, and it is involved in maintaining DNA-protein contact.
- Motif III: it is characterized by a set of 3 histidine residues, the HUH signature. It is important for coordinating the metal cation required for the nucleophilic attack by the active Tyr residue.

Canonical HUH relaxases perform a reaction which produces a covalent intermediate (**Figure 8**). A catalytic Tyr residue performs the nucleophilic attack on the target DNA sequence, leading to the covalent binding of the Tyr residue to the 5' end of the cleaved DNA strand. After the transesterification reaction, the relaxase and the nicked dsDNA with a free 3'-OH group are covalently bound through the 5' end via a phosphotyrosil linkage. This reaction is reversible by a second transesterification reaction. The nicked DNA will initiate the conjugative DNA transfer. DNA processing during conjugation terminates with the recircularization of the transferred DNA by a second transesterification reaction (Byrd and Matson, 1997).

As previously mentioned, some variants in the HUH motifs of the relaxases have been found. Similarities and differences between conjugative relaxases have been recently reviewed by (Guzmán-Herrador and Llosa, 2019). For example, the MOB_F relaxases carry two Tyr residues instead of one (as TrwC, the relaxase of R388) (Grandoso *et al*., 2000). In MOB_P members (ColE1_ MbeA), a HEN motif replaces the conserved 3-His (Varsaki *et al*., 2003). Also, the MOB_V relaxase MobM (from pMV158) uses a catalytic His residue in place of the Tyr (Fernandez-Lopez *et al*., 2013). Other relaxases do not have HUH motifs, such as the relaxase TcpM from pCW3 plasmid (MOB_M) (Wisniewski *et al*., 2016), or Orf20 from Tn916 (MOB_T) (Rocco and Churchward, 2006), which carries a motif related to Tyr-Recombinases or Rep-trans proteins . In these last two atypical families of relaxases, no covalent complex between DNA and relaxase has been detected, with the substantial change of paradigm that this could imply. So far,

no detailed molecular mechanism has been put forward to explain the DNA transfer process in the absence of a relaxase-DNA covalent intermediate.

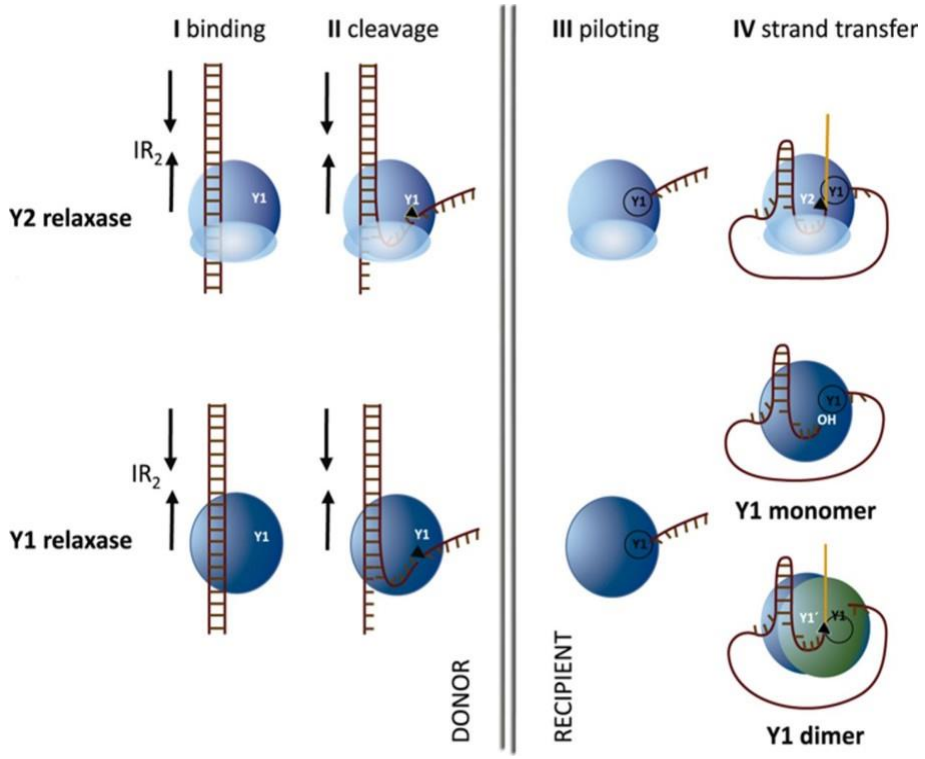


Figure 8. Models of conjugative DNA processing by relaxases with one catalytic tyrosine (Y1) or two (Y2). I) Conjugation started when the relaxase recognized the proximal arm IR_2 adjacent to the *nic* site in the *oriT.* II) Relaxase binding allows the formation of ssDNA U-turn. The *nic* site is positioned at the relaxase active site and *nic*-cleavage by the relaxase takes place. Then, a covalent phosphotyrosine bond between the cleaved DNA strand and the relaxase is formed. III) Subsequent DNA strand displacement produces the DNA single strand that is translocated into the recipient cell covalently bound to the relaxase. IV) In Y2 relaxases, a second tyrosine attacks the newly formed *nic* site to generate a free 3'OH end able to recircularize the transferred plasmid DNA. In Y1 relaxases, two different situations could happen, or the free 3'OH is released in the donor cell (monomeric Y1 model) or a second free tyrosine is provided by a second relaxase molecule (dimeric Y1 model). Taken from (Carballeira *et al.*, 2014).

1.2.1. Target specificity

Specificity of conjugative relaxases in binding to their substrate (*oriT*) were biochemically characterized for different relaxases such as RP4_TraI (MOB_P) (Pansegrau *et al.*, 1993b), F_TraI (MOB_F) and R388_TrwC (MOB_F) (Zechner *et al.*, 2017). Despite its specificity for the targets, some relaxases recognize (with lower efficiency) heterologous sequences. This recognition promiscuity varies from one relaxase to other. For example, the relaxase from pSC101, which recognizes an *oriT* with high homologous sequence

with the one from the plasmid R1162, cannot cleave the R1162 *oriT*, while R1162_MobA can act on both *oriTs* (Becker and Meyer, 2003). Also, relaxases R388_TrwC and Ptw_PtwC are closely related, although their *oriT* region has no significant homology. In spite of this, TrwC it is able to mobilize a plasmid containing Ptw *oriT* (Fernandez-Gonzalez *et al*., 2016).

This trans-mobilization phenomenon is more frequent than previously thought. It has been found that numerous plasmids initially considered as non-mobilizable, are mobilized by a relaxase present in a co-resident plasmid. They are called "orphan" plasmids, and they contain short *oriT* sequences which resembles the target sequences of other relaxases. These plasmids have been found in many Gram-positive bacteria such as *Staphylococcus aureus, Leuconostoc* or *Lactobacillus* species. Plasmids from these bacteria, such as pA1, pCI1411 or pT48, have been found to contain sequences which resemble the RS_A *oriT* of pMV158 (Ramsay and Firth, 2017). This promiscuity of some relaxases mobilizing heterologous *oriT* shows the biological relevance of this activity in horizontal genetic transfer.

1.2.2. Moonlighting relaxases

Conjugative relaxases have been phylogenetically classified according to their catalytical domains, present in the N-terminal 300 residues. However, relaxases are often multidomain enzymes. These additional functional domains could play a role in the conjugative transfer process (such as oligomerization, DNA binding, or the DNA the helicase domain of MOB_F relaxases (Zechner *et al*., 2017)), or they could add additional functions independent of the conjugation process. Some relaxases can perform site-specific recombination reactions (Llosa *et al*., 1994b), or site-specific integration (SSI) reactions, where DNA could be integrated into an *oriT* copy (in a plasmid or in the chromosome) (Draper *et al*., 2005; Agúndez *et al*., 2012). Other activities are involved in plasmid replication (such as the primase domain of MobA from RSF1010 (Henderson and Meyer, 1996, 1999)) or even in the regulation of the plasmid copy number, as in pMV158_MobM (Lorenzo-Diaz *et al*., 2017).

1.2.2.1. Relaxases with recombinase and integrase activity

Some conjugative relaxases have *oriT*-specific recombinase and integrase activity (reviewed by (Wawrzyniak *et al*., 2017)). This activity is independent of the conjugative process. R388_TrwC is able to perform this reaction (Llosa *et al*., 1994b). Other unrelated relaxases such as R64_NikB (MOB_P), pAD1_TraX, and the relaxases of pAMα1 (MOB_C) and ICE*clc* element (MOB_H), have also been reported to have recombinase activity (Francia and Clewell, 2002b, 2002a; Furuya and Komano, 2003; Miyazaki and van der Meer, 2011). RSF1010_MobA is able to catalyze the recombinase reaction on ssDNA but not on supercoiled DNA (scDNA) (Meyer, 1989). In contrast, relaxase TraI from RP4 plasmid cannot perform this reaction (Agúndez *et al*., 2012).

Nowadays, it is unknown which factors determine the recombinogenic properties of a relaxase. There are different elements which can influence this reaction. For example, the presence of accessory proteins affects this activity. In the case of R64_NikB, when the reaction is on dsDNA, it is needed the presence of the accessory protein NikA, whereas when the substrate is ssDNA, only NikB is required (Furuya and Komano, 2003). In the case of R388_TrwC, it has been shown that the presence of TrwA increases the efficiency of the reaction (César *et al*., 2006). Also, the host-encoded replication/machinery could be involved (César *et al*., 2006; César and Llosa, 2007). The *oriT* sequence itself also seems to play an important role, since the MOB_H relaxase of ICE*clc* catalyzes recombination only on one of the two *oriT* copies present in this ICE, while it can act on both *oriT* for conjugal DNA transfer (Miyazaki and van der Meer, 2011).

Another related reaction catalyzed by some relaxases is *oriT*-specific integration. This activity has been reported for some conjugative relaxases such as TrwC (Draper *et al*., 2005) and the relaxase Mob02281 from the *Bacillus thuringiensis* plasmid pBMB0228 (Wang *et al*., 2013). Other relaxases related to TrwC were tested, such as F_TraI or TraI_RP4, but no integrants were obtained (Agúndez *et al*., 2012). The SSI activity is less described between relaxases than the recombinase activity, and it is unclear why some relaxases can perform this activity. It is also unknown which makes a relaxase able to perform the integration reaction.

The ability of relaxases to act not only on its target for DNA mobilization, but also for other activities and on other targets, could have important biological implications**.** These activities will contribute to increasing the genomic plasticity, as depicted in **Figure 9.**

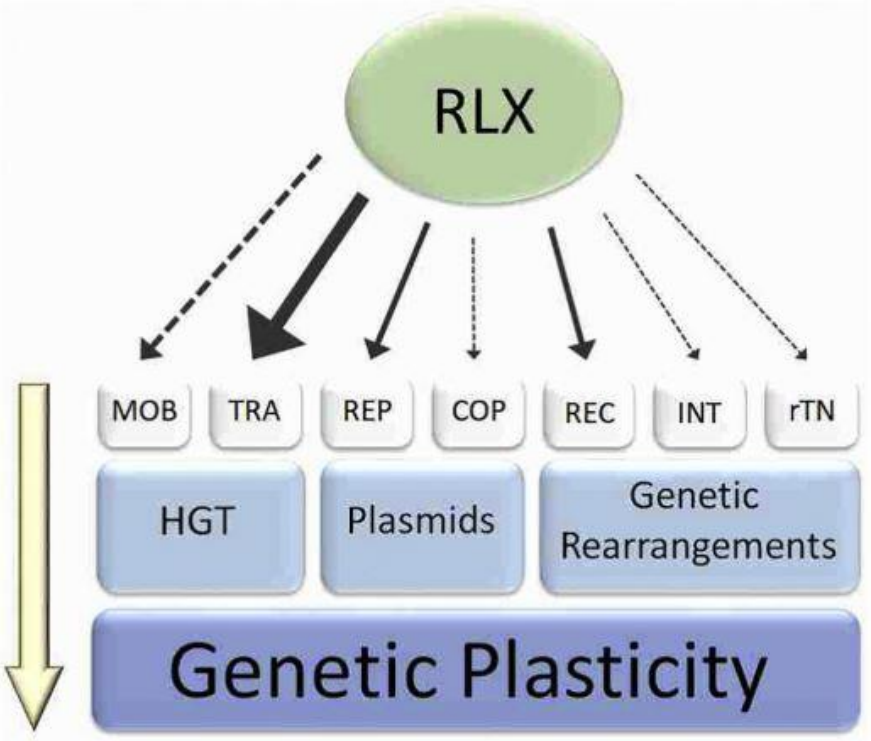


Figure 9. Schematic of biological functions of relaxases and their contribution to genetic plasticity. Black arrows point to the biological functions reported for conjugative relaxases. The thickness of the arrow depends on the dedication of relaxases to the function. Functions based on specificity of relaxases to their target, and functions derived from their activities on non-cognate targets are represented with solid or dotted arrows respectively. RLX, Relaxases; MOB, Mobilization; TRA, self-transfer; REP, Replication; COP, Copy number; REC, site-specific recombination; INT, SSI; rTN, Retrotransposition; HGT, Horizontal gene transfer. Taken from (Guzmán-Herrador and Llosa, 2019).

1.2.3. <u>The relaxase TrwC</u>

TrwC is the relaxase of the conjugative plasmid R388. It is an HUH conjugative relaxase, and it belongs to the MOB_F family (Garcillán-Barcia *et al*., 2009). TrwC has 966 aa and a molecular weight of 108 kDa. It recognizes and cleaves its target (the *oriT* sequence at the *nic* site) in the DNA strand to be transferred, making a covalent bond with its 5'end. As previously mentioned, TrwC has two active catalytic Tyr residues, Y18 and Y26. Y18 is the only tyrosine able to act on super-coiled plasmid (Grandoso *et al*.,

2000), so it is responsible for the initial cleavage reaction, leading to the covalent complex with the DNA. This nucleoprotein complex is recruited by the T4SS and translocated to the recipient cell, where the relaxase is active(Draper *et al*., 2005). This is significant, since during its translocation trough the T4SS, TrwC has to be unfolded and refolded (Trokter and Waksman, 2018). In the recipient, the second catalytic Tyr residue (Y26) catalyzes the recircularization of the transferred DNA strand (Gonzalez-Perez *et al*., 2007). Apart from DNA mobilization, TrwC also catalyzes *oriT*-specific recombination and integration reactions and promotes non-specific DNA integration in human cells.

1.2.3.1. TrwC catalytic activities and functional domains

The relaxase TrwC can perform different activities. Functional mapping of TrwC has assigned the different activities to particular domains of the protein. Catalytic activities and functional domains are detailed below.

Relaxase activity

The relaxase domain is located in the N-terminal 293 aa of the protein (Guasch *et al*., 2003). Although N293 fragment is required for scDNA nicking, a fragment containing the N-terminal 275 aa is enough to perform cleavage and strand transfer reactions on ssDNA (Llosa *et al*., 1996; Guasch *et al*., 2003). The three motifs described for the HUH superfamily are located in this domain (Ilyina and Koonin, 1992).

-Motif I: it comprises the active tyrosil residues involved in the nucleophilic attack to the *nic* site (Grandoso *et al.*, 2000). There are four conserved tyrosines, Y18, Y19, Y26 and Y27. By mutagenesis, it was determined that Y18 and Y26 are essential for DNA processing during conjugation. They are directly involved in the cleavage and strand transfer reactions. Both tyrosines promote cleavage and strand transfer reactions of ssDNA containing a *nic* site. However, only Y18 can cleavage supercoiled *oriT* containing DNA (Grandoso *et al*., 2000). Mutant in Y18 decreases conjugation frequency in 500-fold and Y26 mutant decreases it in 10 times. The double mutant, Y18+Y26, completely abolished conjugation (Grandoso *et al*., 2000).

-Motif II: it contains an aspartic residue, D85, which is believed to activate the tyrosine hydroxyl group by proton abstraction (Boer *et al*., 2006).

-Motif III: it harbors the histidine triad (H150, H161 and H163) which coordinates the metal ion for the nicking reaction (Guasch *et al*., 2003).

The two catalytic tyrosines act sequentially (Grandoso *et al*., 2000) and they are responsible for the sequence-specific cut and strand-transfer reactions. These reactions occur in *oriT*-containing scDNA *in vivo and* can be also observed on *oriT*-containing ssDNA *in vitro*.

scDNA nicking activity. TrwC presents sequence-specific endonuclease activity and it can 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 minimal core sequence of the *oriT* for TrwC relaxase activity is 17 bp (which comprises *nic* site and IR_2, the TrwC binding site) (Guasch *et al*., 2003). Once the DNA is nicked, the relaxase remains bound to the 5' (covalently) and to the 3' end (not covalently). The reaction reaches the equilibrium in 5 minutes. The reaction activity is increased by TrwA or TrwB addition, and decreased with Integration host factor(IHF) addition (Moncalián, 2000).

ssDNA nicking *in vitro.* TrwC is able of nicking oligonucleotides containing the *nic* site and performing strand transfer to a second oligonucleotide also containing the *nic* site (Llosa *et al*., 1996). The *oriT* sequence (6+2) (6 nt 5' to the *nic* site and 2 nt 3' to the *nic* site) is absolutely required for single-strand nicking and strand transfer reactions (Lucas *et al*., 2010).

DNA helicase activity

TrwC shows DNA helicase activity. The helicase domain is located in the C-terminal region of TrwC (192-966 aa) and it is known as C774 domain. It has the seven motifs characteristic of the family of DNA helicases (Matson *et al*., 2001), including Walker A and Walker B nucleotide binding motifs. It contains the putative dimerization domain of TrwC (Llosa *et al*., 1996). DNA helicase activity is dependent on ATP, Mg^{2+} and ssDNA. The unwinding direction is 5' to 3'. Helicase activity efficiency in fragments bigger

than 100 bp *in vitro* decreases, which means that purified TrwC is not a processive helicase (Grandoso *et al*., 1994).

Site-specific recombinase activity

TrwC has site-specific recombinase activity. The recombinase domain is known as N600 domain. It includes the N-terminal 600 residues of the protein and is the smallest polypeptide able to perform recombination reaction efficiently (César *et al*., 2006). In 1994, Llosa and collaborators showed that TrwC could catalyze a site-specific recombinase reaction *in vivo* between two *oriT* copies repeated in tandem in supercoiled plasmid DNA, in the absence of conjugation (Llosa *et al*., 1994b). Later, César and collaborators demonstrated TrwC recombinase activity in the donor and after translocation into the recipient cell. They constructed a recombination cassette with a *ntpII* and *laqI*^{q} flanked by two R388 *oriT* copies and recombination was measured based on *lacZ* gene expression (blue colonies) (**Figure 10**)**.** This new system allowed a thorough study of the reaction (César *et al*., 2006; César and Llosa, 2007).

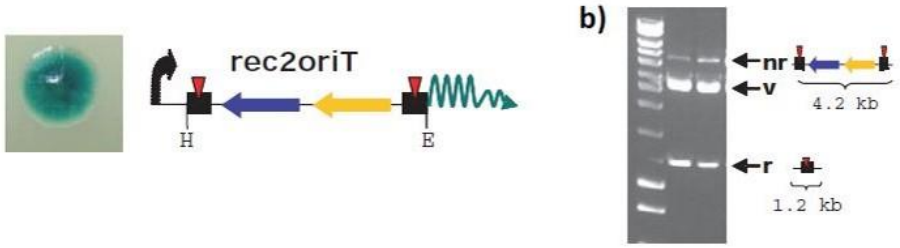


Figure 10. TrwC-mediated *oriT* recombination activity assay. At the left, the structure of the recombination cassette re2*oriT*. A recombinant colony is shown after co-transformed R388 with the recombination cassette into DH5α and plated on X-Gal containing medium. Black boxes represent *oriT* and the red triangle the *nic* site. *lacI*^{q}, purple; *ntpII* yellow. Blac arrow, lactose promoter. Arrow points the direction of the transcription. H and E, *HindIII* and *EcoRI* sites respectively. At the right, a restriction digestion using *EcoRI+HindIII* of recombinant plasmids is shown. nr, not recombinant; v, vector; r, recombinant. Taken from (César *et al*., 2006).

TrwC relaxase is essential for the site-specific recombination reaction, as it cannot be performed by a TrwC mutant (Llosa *et al*., 1994b). It is also known that other proteins affect the reaction. TrwA is needed for efficient site-specific recombination while IHF proteins decreases the recombination efficiency (César *et al*., 2006; César and Llosa, 2007).

DNA sequence requirements for *oriT*-specific recombination are different from the ones for conjugation and *oriT1* and *oriT2* have different tolerance for changes in their sequence (César *et al*., 2006). *oriT1* sequence can be reduced to 33 bp, comprising the *nic* site and TrwC binding site, while *oriT2* needs the complete region 3' to the *nic* site for efficient recombination frequencies.

Site-specific integrase activity

TrwC has site-specific integrase activity in prokaryotes. It can integrate the transferred DNA strand into an *oriT* containing plasmid or into a chromosomal *oriT* copy. The recombinase domain N600 is the minimal domain able to catalyze the reaction efficiently (Agúndez *et al*., 2012). Draper and collaborators demonstrated the SSI activity by mobilizing an *oriT*-containing suicide plasmid into recipient cells which contained a plasmid carrying an *oriT* copy (Draper *et al*., 2005). The mobilizable plasmid was Pir-dependent for replication, so it could not be maintained in the recipient strain (without *pir*) unless it is integrated. The suicide plasmid used also contains an RP4 *oriT*, used as negative control. Integrants were detected by selecting the mobilizable plasmid in the recipient cell. They performed the experiments in parallel using TrwC or RP4_TraI to mobilize the plasmid. Only when TrwC was present, integrants were detected. No integrants were detected using TraI, demonstrating that the reaction was totally dependent on TrwC presence (Draper *et al*., 2005).

Using an optimized integration assay, Agúndez and collaborators detected TrwC integration activity also when *trwC* was expressed in the recipient cell and the suicide plasmid was mobilizable with RP4_TraI (**Figure 11**) (Agúndez *et al*., 2012). Finally, in order to test if TrwC was able to perform the integration reaction into a chromosomal *oriT* copy, they constructed two recipient strains with the *oriT* integrated in both orientations in the chromosome. TrwC was able to catalyze the integration reaction, but only in the strain with the *nic* site in the lagging strand.

The integration reaction depends on both tyrosine residues and the helicase domain is dispensable for TrwC integration activity. As happened with the

recombination activity, TrwA and IHF proteins affected the reaction, although in this case, both enhanced it.

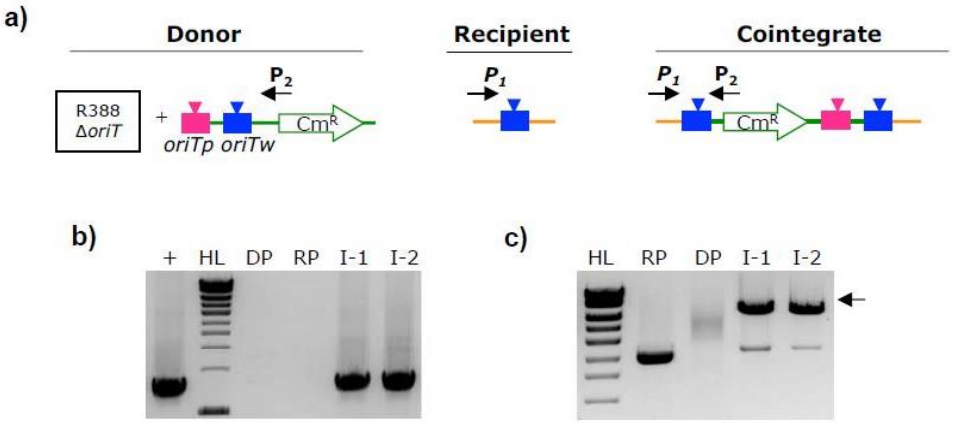


Figure 11. SSI assay. a) Scheme of the donor and recipient plasmids and the cointegrate molecule obtained. The green line represents the suicide plasmid while the orange line represents the recipient plasmid. $oriT_P$ (from RP4) and $oriT_W$ (R388) are boxed in pink and blue respectively. *nic* sites are indicated by arrowheads. P1 and P2 are the oligonucleotides used to detect cointegrants by PCR. **b)** PCR amplifications of cointegrants. **c)** Restriction digestion using *XcmI*, which only cut once in the recipient *oriT*-containing plasmid. The cointegrates are indicated with an arrow. HL, Hyperladder, DP, donor plasmid; RP recipient plasmid; I1-I2, DNA from two integrants obtained. Modified from (Agúndez *et al*., 2012).

TrwC was able to perform the integration reaction on the minimal *oriT* sequence (17 bp), although the efficiency decreases 2 logs. By using different *oriT* mutants, the authors analyzed the DNA requirements for the reaction. Interestingly, changes in the *oriT* affecting either the *nic* site or the binding site abolished TrwC integration activity when TrwC was only expressed in the recipient cell and it was required to act on scDNA containing the mutant *oriT*. However, incoming TrwC-DNA complexes could integrate DNA into acceptor sites with mismatches in the core *oriT* sequence (17 bp). They also showed that TrwC could catalyze the integration reaction into two DNA sequences of human origin with a single mismatch from the minimal *oriT* sequence with a decreased efficiency of 2-3 fold compared to the 17 bp original sequence (Agúndez *et al*., 2012). This data showed that DNA requirements for TrwC are more flexible in the acceptor target site.

Oligomerization

The oligomerization ability of TrwC and its domains has been studied using different techniques such as gel filtration chromatography and analytical centrifugation. TrwC has been shown to form dimers in the presence of 550 mM NaCl by gel filtration chromatography analysis, however it is possible that this dimer assembles into more complex oligomeric forms under lower salt conditions (Grandoso *et al*., 1994).

The relaxase domain N293 has been thoroughly study by analytical centrifugation and its sedimentation coefficient reveals that it formed monomers in the absence of its ssDNA *oriT* target. When N293 was studied in the presence of its target (a ssDNA containing an *oriT* sequence), the complexes obtained were formed by one molecule of ssDNA and one molecule of protein, which means that it also behaved as a monomer (Lucas *et al*., 2010).

The oligomerization profile of the recombinase domain N600 has been studied by gel filtration chromatography and it has been showed to form monomers under the same conditions where TrwC run as a dimer (César *et al*., 2006).

The helicase domain C774 oligomerization ability was also studied by gel filtration chromatography assay. It has showed to elute as a dimer in the presence of 550 mM Nacl so it contains the putative dimerization domain of TrwC, in the 495 C-terminal residues (Llosa *et al*., 1996). **Figure 12** summarized the characteristics, activities and oligomerization abilities of the TrwC domains (the N293 relaxase domain, the N600 recombinase domain, and the C774 DNA helicase domain).

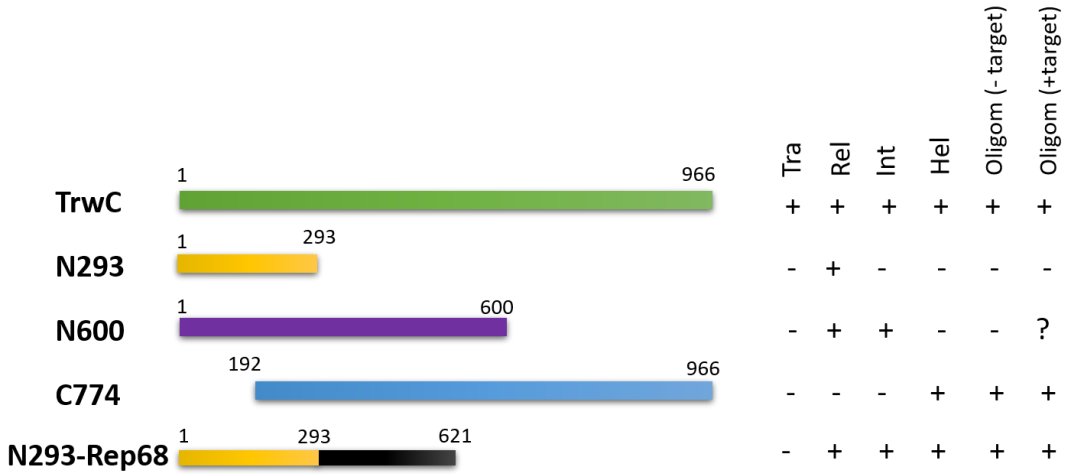


Figure 12. Functional domains of TrwC and the chimera protein N293-Rep68. N293, relaxase domain. N600, recombinase domain. C774, helicase domain, N293-Rep68 chimera protein. Tra, conjugative DNA transfer. Rel, *in vitro* relaxase activity. Int, *in vivo*, site-specific integrase activity. Hel, DNA helicase activity. Oligom (-target and + target), oligomerization ability without and with the DNA target. –, + and ? means without, with or unknow activity/ability respectively (Llosa *et al*., 1996; César *et al*., 2006; Lucas *et al*., 2010) (Agúndez *et al*., 2018). Adapted from (Agúndez *et al*., 2011).

1.2.3.2. TrwC-Rep68 chimera

Rep 68 is a multi-domain protein which belongs to the HUH superfamily. It is a rolling-circle replicase (Rep) and it is involved in both replication and SSI of the human adeno-associated virus (AAV) in the human genome. The N-terminal domain (origin binding domain, OBD) contains the HUH motifs. It is the domain responsible for DNA binding and it recognizes the Rep binding sites (RBS) and the nicking site, also known as terminal resolution site (*trs*), present in the viral inverted terminal repeats (ITR) and in the chromosomal integration sites (Linden *et al*., 1996). This domain is also enough to catalyzes site-specific endonuclease activity (Yoon *et al*., 2001). The central domain of the protein possesses the ATPase and 3'-to-5' DNA helicase activities (James *et al*., 2003). The linker connecting the OBD and the helicase domain (from the 215 to 224 amino acid) has been shown to be required for functional oligomerization of the protein (Zarate-Perez *et al*., 2012).

Rep68 and TrwC are two distantly related HUH superfamily members. They both perform the nicking reaction required to mediate SSI by similar mechanism (Guasch *et al*., 2003; Hickman *et al*., 2004). Agúndez and collaborators engineered a chimeric

protein consisting of the N-terminal TrwC relaxase domain and the C-terminal AAV Rep68 helicase domain (including the amino acids required for the oligomerization of Rep68 protein) (**Figure 13**) (Agúndez *et al*., 2018).

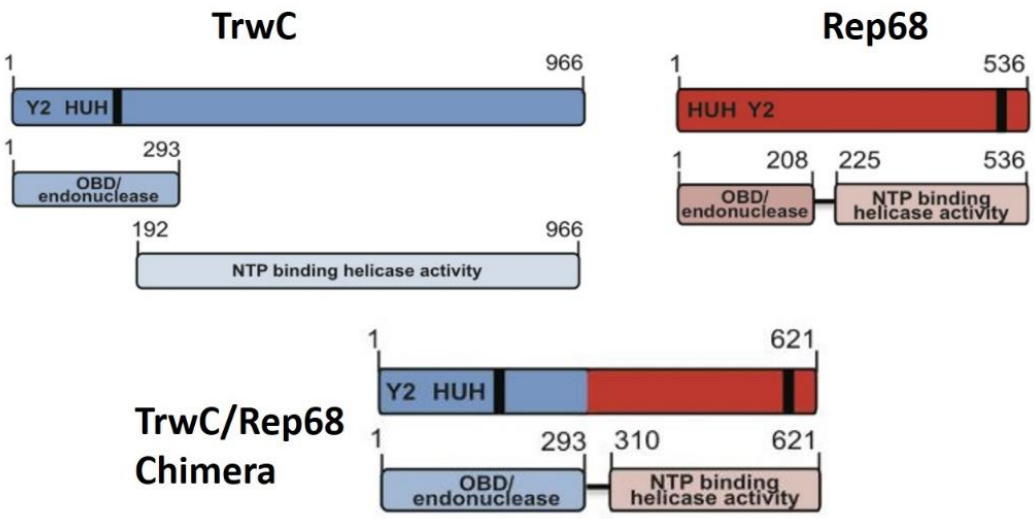


Figure 13. Schematic representation of TrwC, Rep 68 and TrwC/Rep68 chimera. The black lines indicate the NLS. Amino acid positions are indicated above each protein. TrwC and Rep68 domains are represented in blue and red, respectively. Taken from (Agúndez *et al*., 2018).

The chimeric protein was tested to determine its biochemical characteristics. It maintained the helicase activity and fully supports dsDNA unwinding. TrwC/Rep68 presented a high affinity for its substrate: it recognized and bound to the *oriT* substrate specifically and it did not bind to a mutated binding site or to the Rep68 substrate (AAV *ori*).

They also tested the oligomerization ability of TrwC/Rep68. N293 relaxase domain of TrwC behaves as a monomer, irrespectively of the presence of its target DNA (Lucas *et al*., 2010). Rep68 oligomerization profile is complex, it is composed by two populations, one with monomers and dimers in slow equilibrium and a second one consisting of a mixture of multiple-ring structures of seven and eight members (Zarate-Perez *et al*., 2012) and it assembles into a ring-shaped double octamer in the presence of ssDNA (Mansilla-Soto *et al*., 2009). When TrwC/Rep68 oligomerization was tested in the absence of its *oriT* substrate, it behaves as a monomer, with a sedimentation coefficient value ranged from≈ 3.8S to 4S (**Figure 14**).

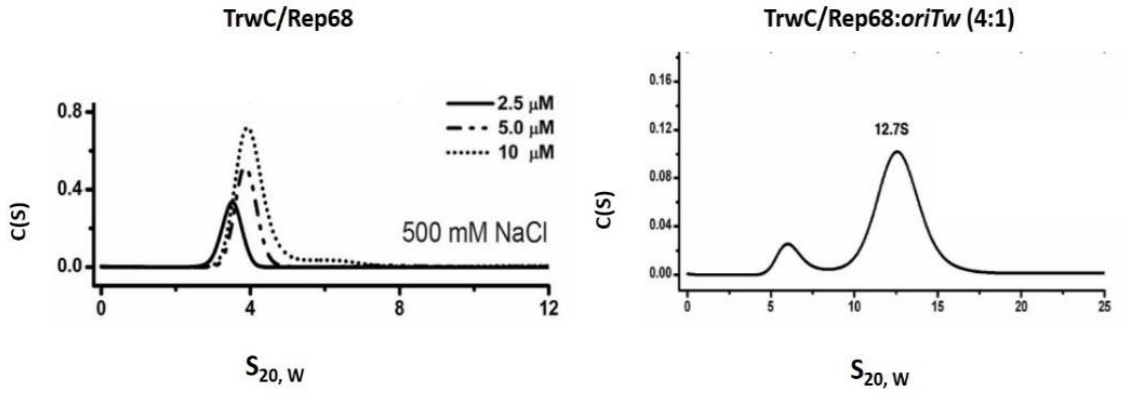


Figure 14. TrwC/Rep68 chimera oligomerization ability. At the left, the sedimentation velocity analysis of TrwC/Rep68 in 500 mM NaCl at 3 protein concentrations (2.5, 5 and 10 μM). At the right, the sedimentation profile of TrwC/Rep68 incubated with an *oriT* (25+8) oligonucleotide. The ratio used was 4:1 (protein:oligonucleotide). C(S); sedimentation coefficient distributions. $S_{20}W$, sedimentation coefficient. Modified from (Agúndez *et al*., 2018).

Interestingly, when the chimeric protein was incubated with its substrate (a fluorescently labelled *oriT* (34 mer) oligonucleotide), the sedimentation profile shows a major species sedimenting at 12.7 S (Figure 14). This result suggests that in the presence of DNA, the chimera forms a different oligomeric complex than its parental proteins (Figure 12).

TrwC/Rep68 catalyzes SSI in bacteria. Agúndez and collaborators performed an integration assay as it was previously described for TrwC (Agúndez *et al*., 2012). A suicide plasmid containing the *oriT* of R388 and RP4 was mobilized by RP4_TraI from the donor cell to a recipient cell harboring a plasmid with the *oriT* of R388 and expressing N293-TrwC, TrwC or TrwC/Rep68. Unexpectedly, the chimeric protein was able to catalyze the integration reaction with a frequency very similar to TrwC (**Figure 15**). This result was surprising, as TrwC N293 is not enough for performing efficient integration reaction. TrwC/Rep68 was also assayed to determine integration activity in human cells using the same requirements than the ones for Rep68. However, no integration activity was detected (Agúndez *et al*., 2018).

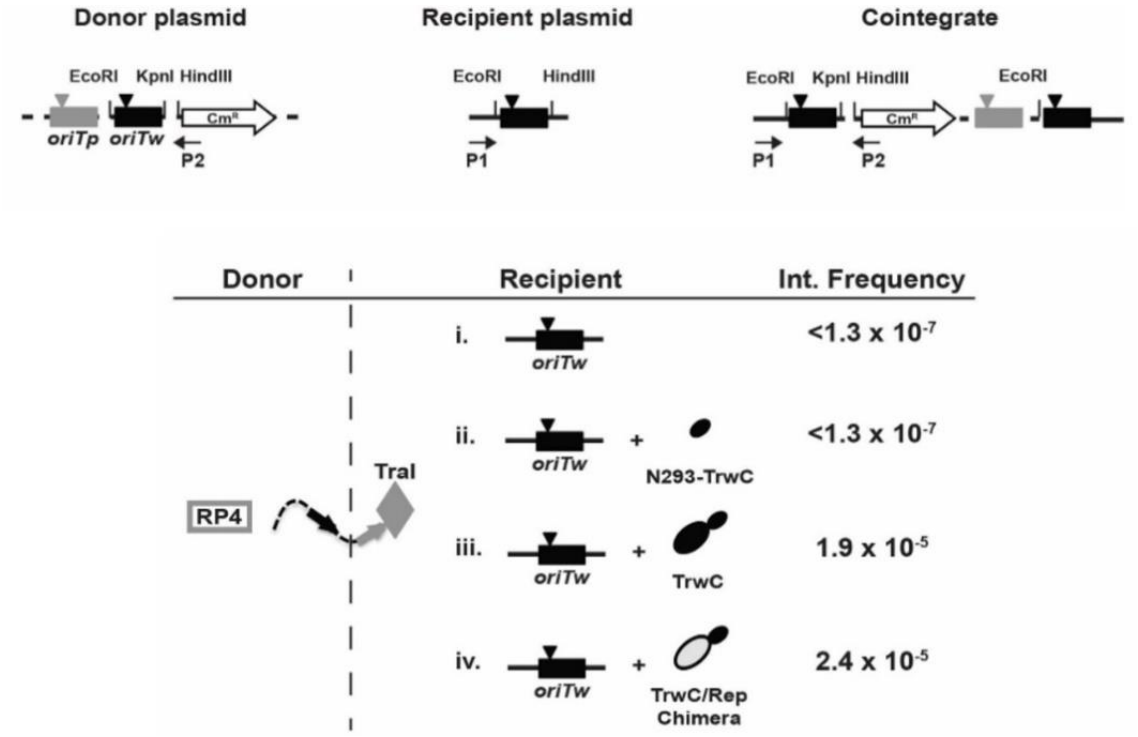


Figure 15. SSI assay. At the top, scheme of the plasmids used in the assay. The suicide plasmid is represented with a dotted line. *oriT*$_P$ and *oriT*$_W$, origins of transfer of RP4 and R388 system respectively, are represented as grey and black boxes. The *nic* sites are indicated as triangles. Cm^R correspond to the chloramphenicol resistance cassette. At the bottom, the representation of the integration assay. The suicide plasmid is mobilized by RP4_TraI (grey diamond) into a recipient cell. Taken from (Agúndez *et al*., 2018).

1.2.3.3. TrwC recruitment by T4SS

During conjugation, TrwC must be recruited and transferred by the R388 T4SS. It is also known than TrwC can be translocated by the T4SS of the human pathogen *B. henselae* VirB/D4.

TrwC translocation by the T4SS of B. henselae

The T4SS of the human pathogen *B. henselae* has been used to transfer DNA from bacteria to human cells (Fernández-González *et al*., 2011; Schröder *et al*., 2011). These reports proved that a T4SS involved in pathogenesis can also translocate a nucleoprotein complex (DNA-relaxase) via a process resembling bacterial conjugation, underscoring the versatility of T4SS for macromolecular substrate transfer. Two different systems

were assayed: the cryptic plasmid pBGR1 of *B. henselae* with its relaxase Mob fused to the BID signal, and the conjugative system R388 with the relaxase TrwC. In both cases, the DNA transfer was dependent on the presence of the conjugative relaxase.

The DNA transfer was detected by the expression of an enhanced green fluorescent protein (eGFP) cassette encoded on a bacterial mobilizable plasmid containing the *oriT* and the conjugative genes from each system. When pBGR1 was used with wild type bacteria, 0.02% of eGFP was detected (**Figure 16a**) (Schröder *et al*., 2011). For R388 system, 1-2% of GFP positive cells were detected (Figure 16b) (Fernández-González *et al*., 2011).

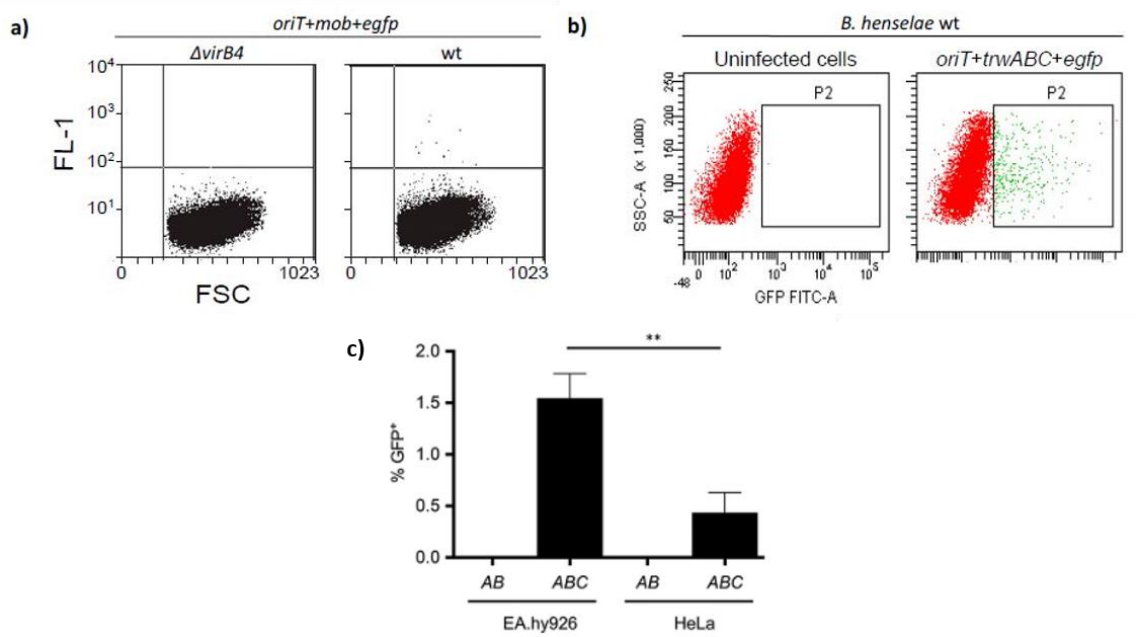


Figure 16. DNA transfer through the *B. henselae* VirB/D4 T4SS. Fluorescence-activated cell sorting (FACS) to detect GFP positive cells. a) DNA transfer of pBGR1 derivatives with the relaxase Mob. 0.02% of GFP positive cells were detected when infections where performs using a wild type *B. henselae.* No GFP positive cells were detected when infections were perform using a T4SS-deficient *Bartonella* strain. b) DNA transfer of R388 derivatives with the relaxase TrwC. 1-2% of GFP positive cells were detected. Uninfected cells were used to determine GFP background. c) DNA transfer of R388 derivatives using different cell lines. The graph shows the percentage of GFP positive cells. *AB* and *ABC* indicated the presence in the transferred plasmid of *oriT trwAB* or *oriT trwABC* respectively. Error bars indicate standard deviations. **, $P < 0.01$. SSC-A, side scatter. FITC-A, GFP fluorescence intensity. Modified from (Fernández-González *et al*., 2011; Schröder *et al*., 2011; Gonzalez-Prieto *et al*., 2017).

It is important to underline that when infections where performed using EA.hy926 cells, derived from fusion of A549 lung carcinoma cells with human vascular

endothelial cells, which are the natural target for *Bartonella*, instead of HeLa cells, the DNA transfer efficiency was higher (Figure 16c) (Gonzalez-Prieto *et al*., 2017).

The addition of the BID signal recognized by the T4SS of *B. henselae* could increases the relaxase transfer. While adding this signal to TrwC the DNA transfer is slightly increased, adding this signal to Mob produced an increase of 100-fold (Schröder *et al*., 2011). This data confirmed that TrwC is a better substrate for VirB/D4 T4SS than Mob.

TrwC recruitment signals

In order to be recruited by the different T4SSs, TrwC must possess a translocation signal. Relaxases translocational signal (TS) in T4SS have been mapped to various internal positions (Parker and Meyer, 2007; Lang *et al*., 2010). In contrast, protein effectors translocated though T4SS are mainly recruited through their C-termini, however, sometimes other elements are necessary for effectors recruitment, as for example the BID domain in *B. henselae* (Schulein *et al*., 2005).

Alperi and collaborators found two putative translocation sequence motifs in TrwC, TS1 (GDTIRIT at positions 796 to 802) and TS2 (GDRMKVV at positions 813 to 819) (Alperi *et al*., 2013). Mutations in these sequence motifs were tested in bacterial conjugation (where TrwC is recruited by its own T4SS) and in DNA transfer to human cells (where it is recruited by the VirB/D4 of *B. henselae*). Different results were obtained for each assay, indicating that TrwC could be recruited by two different T4SS through different signals (**Figure 17**). Mutations in TS1 drastically affected conjugation frequencies, while DNA transfer to human cells was less affected, showing a 50% reduction. TS2 mutations only showed a slight decrease.

On the other hand, they evaluated the effects of different C-termini in TrwC translocation trough both T4SS (Figure 17). Mutations in the C-terminal residues from TrwC and the addition of different C-terminal fusion peptides showed different effects in TrwC recruitment. While there was no substantial effect in conjugative DNA transfer through R388 T4SS, the TrwC variants showed significant effects in DNA transfer through VirB/D4 T4SS to human cells (Alperi *et al*., 2013).

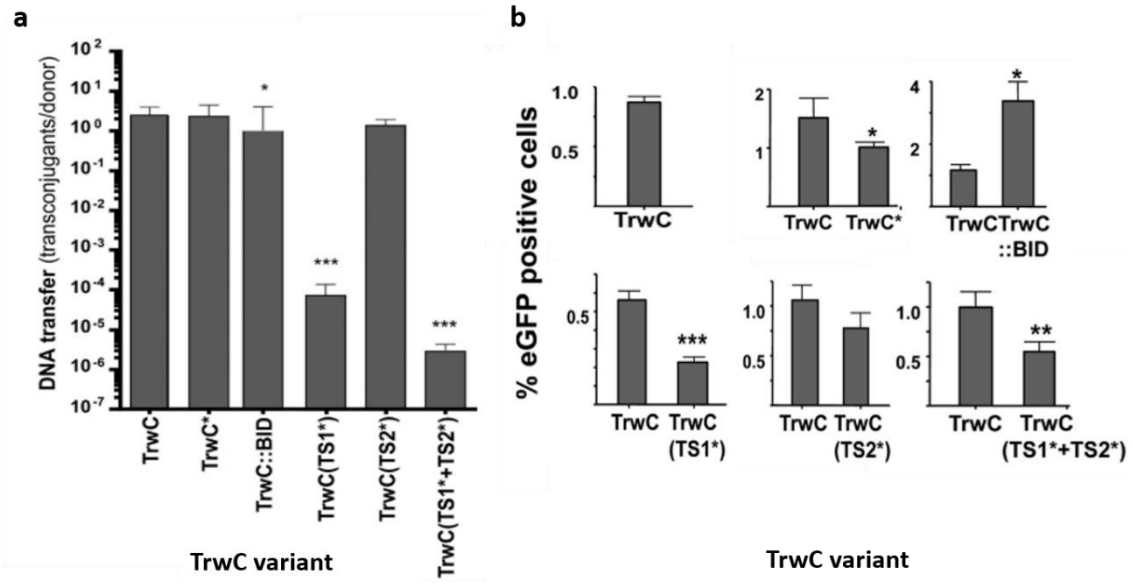


Figure 17. TrwC variants recruitment by the R388 T4SS (a) and by VirB/D4 T4SS (b). a) Conjugative DNA transfer mediated by TrwC derivatives. Donors contained the plasmid coding for the MOB region of R388 (*oriT*:*trwABC*) and a helper plasmid (pSU4058) to provide R388 T4SS. The TrwC variant in each construct is indicated at the bottom of the graphic: TrwC*, a variant with different C-terminal 32 residues; variants in TS motifs are indicated with asterisks. DNA transfer is expressed as the number of transconjugants per donor. The bars represent means of 5 independent experiments. The error bars indicate standard errors of the mean. Student's t test was used to analyze the data referred to the positive control, TrwC. *, P< 0.01, ***, P< 0.0001. **b)** Percentages of eGFP-positive EA.hy926 cells infected by *B. henselae* carrying the indicated TrwC variant compared with its own positive control. TrwC variants corresponded to the ones explained for Figure 17a. The bars represent means from at least 3 independent experiments done in triplicate. The error bars indicate standard errors of the mean. Student's t test was used to analyze the data referred to the positive control, TrwC. *, P< 0.01; **, P< 0.001; ***, P< 0.0001. Modified from (Alperi *et al*., 2013).

In summary, it seems that the relaxase TrwC recruiting motifs needed for R388 recruitment during conjugation are different from the ones required by the VirB/D4 T4SS from *B. henselae*

1.2.3.4. TrwC activity in human cells

As explained in Section 1.2.3.3., TrwC-DNA complexes can be introduced in human cell through the T4SS of the human pathogen *B. henselae*.

Considering the SSI activity of TrwC in recipient cells, the possibility that TrwC promoted integration of the covalently attached DNA strand in the human genome was

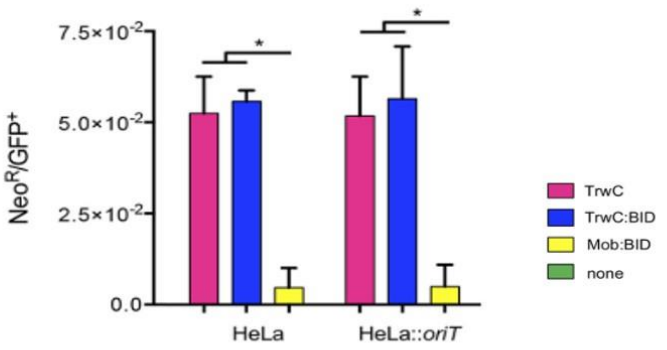
explored. González-Prieto and collaborators performed an assay were they measured the integration ratio in human cells of the DNA transferred by TrwC (Gonzalez-Prieto *et al*., 2017). They constructed a plasmid which contains R388 MOB region (*oriT trwABC*), a eukaryotic *gfp* expression cassette, and a neomycin phosphotransferase eukaryotic expression cassette. As negative controls, plasmids without relaxases were used**.** They performed a DNA transfer assay from *B. henselae* into different human cell lines and they selected integration events of the transferred plasmids by antibiotic treatment with G418 during several weeks. They performed the experiments in parallel with the Mob relaxase from a cryptic plasmid of *B. henselae*, fused to a BID signal, which is recognize by *B. henselae* T4SS (Schulein *et al*., 2005). The integration rate was calculated as Neo^r colonies/GFP^+ colonies. The integration rate of TrwC was 1:20 while Mob:BID was 1:250, very similar to the ones obtain for random integration by cell transfection (1:300). Also, there were no differences in the integration rates between HeLa and HeLa::*oriT.* These data revealed that TrwC enhanced integration of the transferred DNA in the human cell chromosome (**Figure 18a**).

In order to characterize the integration pattern, they performed a LAM-PCR. Only one of the thousands of integration events analyzed was integrated at the *nic* site, meaning that it was the product of a SSI reaction. This integration event occurred in a region of the genome showing 8 bp identity with the *oriT* at the 5' end of the *nic* site (Figure 18b). None of the rest of the integration junctions analyzed were at this point, which means that they were random integration events (Figure 18c). Finding one SSI event could mean that TrwC has SSI activity in human cells, however, host-mediated random integration is as least 3 logs more efficient. Complete *oriT* sequences were found during the sequence analysis, which indicates that TrwC is active in the recipient cell and it is recircularizing the plasmid in the human cell.

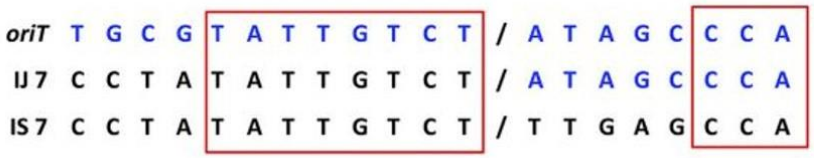
These results revealed that although TrwC catalyzes *oriT*-SSI in bacteria, in human cells TrwC promoted the unspecific integration of the transferred DNA. Also, it is important to highlight that the unique SSI event in human cells was found in a region showing only 8 bp identity with the minimal sequence of the *oriT*, supporting the idea of less stringent requisites in the recipient cell are needed for DNA integration. The reason why TrwC could be acting as an enhancer of the integration of the transferred

DNA is not known, but one possibility is that its covalent binding to the DNA ends could be protecting the DNA from degradation, leading to long-term presence in the nucleus which could favor its subsequent random integration by the host machinery (Gonzalez-Prieto *et al*., 2017).

a)



b)



c)

		Integration junction		
IS	Sequence (5′ to 3′)	Chr	Genomic locus	Plasmid
2	AAAATGAGGACAGTT/ATATTTTTTAAATGT	11	35225119	*nic*+1
3	CCAGATCGTGCCACT/GCATTCCAGCCTGGC	2	37383046	*nic*−15
4	TGGGAAACAAATGAA/GAAACAACCCTGCTG	2	111118923	*nic*−16
5	GTTTCCATGGACATT/TGCCACCCCGGCTTC	6	9173423	*nic*−18
6	CGGGTTAGAAACCAA/GCACCCAAGCCGGCG	12	28128063	*nic*−16
8	CACTTGCTGGGCTCA/GAGACAACCCAGCCC	16	68832248	*nic*−16
9	GTTGTAACTGCCTAA/GATTGACCAACCCTA	19	18303607	*nic*−14
10	GTCACATGATAAAAA/GATTATTTCATTTTG	15	60623276	*gfp*_73
11	ATTTAATCCAAATAG/AAATAAGTTTCAGAT	15	60724330	*trwA*_353
oriT	AGGTGCGTATTGTCT/ATAGCCCAGATTTAA			*nic*

Figure 18. TrwC promotes non-specific integration of the transferred DNA into the human cell. a) The graphic represents the integration ratio of the assayed relaxases. The integration ratio is calculated dividing the neomycin resistant cell by the GFP positive cell after an integration assay. HeLa and HeLa::*oriT* cells were assayed as recipient cell. The different bar represents the different relaxases under study. Error bars indicated standard deviations. *, $P< 0.05$. **b)** Characterization of the integration event where the integration junction occurs at the *nic* site. IJ, integration junction, IS, integration site. The *nic* site and the insertion sites are marked as /. Homology regions between plasmid and genomic sequences are boxed in red. **c)** Analysis of the integration events and integration junctions. All of them are random integration events. The integration sites (or the *nic* site in the *oriT* sequence) are represented as a /. IS, integration site. Modified from (Gonzalez-Prieto *et al*., 2017)

1.3. Genomic engineering

Genomic engineering techniques are intended to perform a permanent genetic modification of a cell. This methodology is equally useful for research purposes or for direct applications. When applied to humans, this technology is used with therapeutic purposes (gene therapy), allowing to counteract genetic defects, or the modification of the cell to combat disease symptoms (Li *et al*., 2020).

The genetic modification of the target cell requires the stable expression of exogenous DNA. Nowadays there are two main challenges to get successful genomic modification: an efficient *in vivo* DNA delivery method, and the site specificity of the modifications (gene targeting).

1.3.1. Foreign DNA delivery

Any strategy for genome modification should include an effective way to deliver the foreign DNA into the desired cell. The strategy is very different depending on the target cell. The introduction of DNA in a bacterial cell is simpler (yet still challenging) than targeting cells within a multicellular eukaryote. We will briefly enumerate the many different methodologies used.

1.3.1.1. Introduction of foreign DNA in bacteria

There are different mechanisms to introduced DNA into bacteria cell. Generally they can be divided in mechanisms which transiently disrupt the bacteria cell wall, such as electroporation and chemical transformation, mechanisms which inject the DNA across the bacteria wall (conjugation or transduction) and mechanism which induces the uptake machinery already present in the cell (natural competence) (Waller *et al*., 2017).

Electroporation has been one of the most widely used methods, due to its simplicity, efficiency and wide applicability, especially in laboratory strains. However, efficiencies vary strongly among species and protocols need to be optimized for each of them. The use of wild-type strains is increasing due to the needed for novel compounds to use as antimicrobials, probiotics or food additives, as for example happens with many lactobacilli species. The transformation of these strains is usually challenging and even

impossible (Zeaiter *et al*., 2018; Börner *et al*., 2019a). Bacterial conjugation (see Section 1.1.) could be an alternative to the electroporation. It is a naturally efficient and promiscuous mechanism of horizontal gene transfer, which operates among all main bacterial types. This mechanism allows the introduction of the foreign DNA into recipient bacteria which can be difficult or impossible to transform (Samperio *et al*., 2021).

1.3.1.2. Introduction of foreign DNA *in vivo* in human cells

There are different techniques to introduce DNA into human cell, which are divided in physical, chemical, or biological methods.

Physical delivery techniques, such as electroporation or microinjection, have a better control of the dose, and they avoid the size limitation. However, their use as *in vivo* vehicles is quite limited since they do not scale up well. Chemical methods are based on the modification of the deliverable itself by lipid encapsulation, use of Cell-Penetrating-Peptides (CPPs) or use of gold nanoparticles, among others. Their use is very promising for *in vivo* applications; however, some limitations are an effective encapsulation or the stability of the complex under physiological conditions (Li *et al*., 2018a; Chen *et al*., 2020). Biological delivery includes viral infections and bactofection delivery. Bactofection is based on the engulfment of bacteria carrying the genetic material of interest by a eukaryotic cell, which causes bacterial lysis and DNA release. This technique allows the delivery of intact bacterial artificial chromosomes (BAC) containing therapeutic genes into human cells (Pálffy *et al*., 2006). Viral vectors, specifically adeno-associated virus (AAV), are one of the most extended methods for *in vitro* and *in vivo* delivery because of their efficiency. However, there is an important limitation with the packaging capacity (Xu *et al*., 2019).

As explained in Section 1.2.3.3., DNA can also be delivered into human cells using the T4SS of specific pathogens. This alternative have been proposed as a new tool for targeted DNA delivery (Llosa *et al*., 2012b). The observed increase in DNA transfer by adding the BID signal to the relaxase opens up the possibility of translocating other proteins by the addition of BID signal. Moreover, the addition of different signals recognized by other T4SS could be used to extend the use to T4SS of pathogens

specifically targeting different cell lines *in vivo*. For example, the addition of the signal RalF to TrwC allowed its translocation covalently bound to DNA through the T4SS of the pathogen *L. pneumophila* (Guzmán-Herrador *et al*., 2017).

In addition, conjugative-like DNA transfer does not have a theoretical size limit; in fact, DNA of different lengths have been delivered from *B. henselae* (Fernández-González *et al*., 2011); and DNA could be translocated together with other proteins of interest into the recipient cell.

1.3.2. Targeted integration of exogenous DNA

The development of SSI tools is essential for targeted genomic modification. There are different tools which allow the specific integration of exogenous DNA into a genome, such as meganucleases, Zinc Finger nucleases (ZFN) or transcription activator-like effector nucleases (TALEN) (Gaj *et al*., 2013; Zhang *et al*., 2019). However, in the last decade a new technique is prevailing because of its simplicity and feasibility: the CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated proteins) systems.

All the techniques are based on the same principle: the use of a site-specific endonuclease which recognizes a specific sequence in the DNA and targets it, generating double-strand breaks (DSBs). In mammalian cells, the presence of the DSB attracts the host DNA repair systems, leading to either mutations (deletions or insertions) by non-homologous end joining (NHEJ) pathways when non template DNA is available, or insertion of an homologous DNA template by homologous recombination (HR) (**Figure 19**) (Adli, 2018). The main advantage of using CRISPR-Cas systems is that the Cas proteins can be easily directed with an RNA molecule to act on any desired site in the genome, while the other site-specific endonucleases have to be designed by protein engineering to recognize each target sequence (Doudna, 2020).

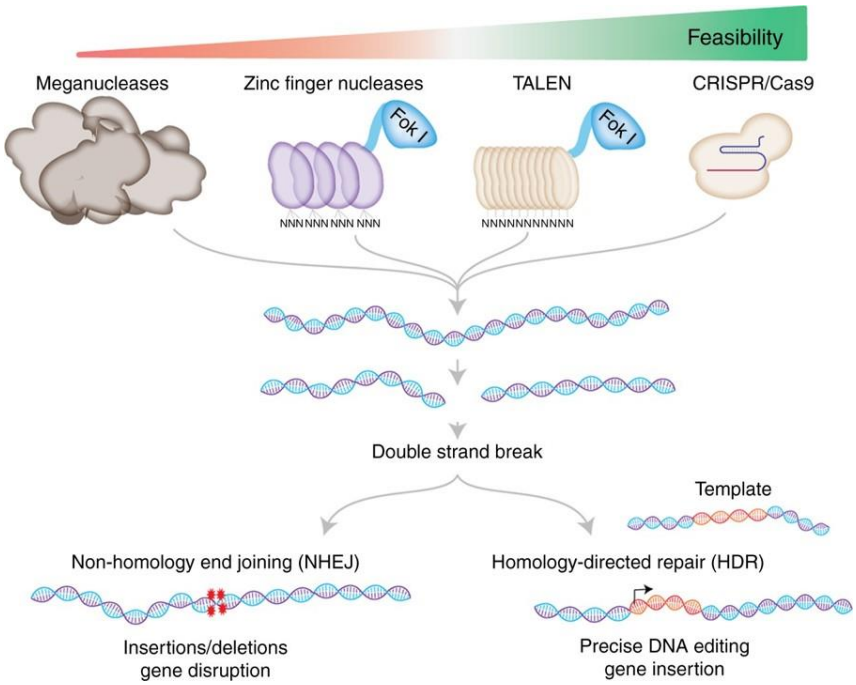


Figure 19. Genetic modification scheme. Different genome editing tools (with different feasibilities) will target specific DNA sequence in the genome, generating DSBs. They will be repaired by the cell DNA repair system by NHEJ, when non homologous template is available, or by homologous recombination, when a homologous template is present. Taken from (Adli, 2018).

In bacteria cell, the generation of DSBs in the chromosome could also be repaired by NHEJ and HR pathways (Shuman and Glickman, 2007). Due to most bacteria lack an NHEJ system, DSBs in bacterial chromosome usually produce the kill of the cell when a homologous template is not present. However, it has been reported that cell death is not the only possible outcome when no homologous template is provided. It has been reported that homologous recombination through distal homologous sequences or between micro-homologies in a RecA dependent manner, could repaired these DSBs allowing bacteria to survive (Cui and Bikard, 2016) .

1.3.2.1. CRISPR-Cas systems

CRISPR-Cas systems provide prokaryotes an adaptive immunity which confers them resistance to invading nucleic acids (Mojica *et al.*, 2005; Barrangou *et al.*, 2007; Marraffini, 2015). Although the first report of a CRISPR-array was in 1987 (Ishino *et al.*, 1987), it was not until 2000 when Mojica and collaborators realized that these CRISPR-arrays were present in many bacteria and archaea, and contained sequences from

invading mobile genetic elements, suggesting a defensive function (Mojica *et al*., 2000). Later on, the deciphering of the RNA-guided site specificity of the Cas proteins led to their use as easily customized endonucleases for genomic editing, and nowadays CRISPR-Cas systems have become an essential tool for the targeted genome modification in many organisms.

A CRISPR locus is formed by a cluster of short spacer sequences separated by a repeated DNA sequence, and the accompanying Cas proteins, coded in an operon (**Figure 20a**) (Marraffini, 2015). Each spacer sequence encodes a guide RNA (gRNA) which targets an exogenous DNA sequence (often of viral or plasmid origin) previously incorporated into the CRISPR array, during the immunization stage. During an immunity response, when the exogenous DNA enters the cell, the gRNA targeting the invading nucleic acids is bound by the effector Cas protein, which is guided by the gRNA to the target sequence. There, the Cas endonuclease will recognize the protospacer-adjacent motif (PAM) and will cleavage the DNA sequence generating a double strand break (DSB) (Figure 20b) (Makarova *et al*., 2020).

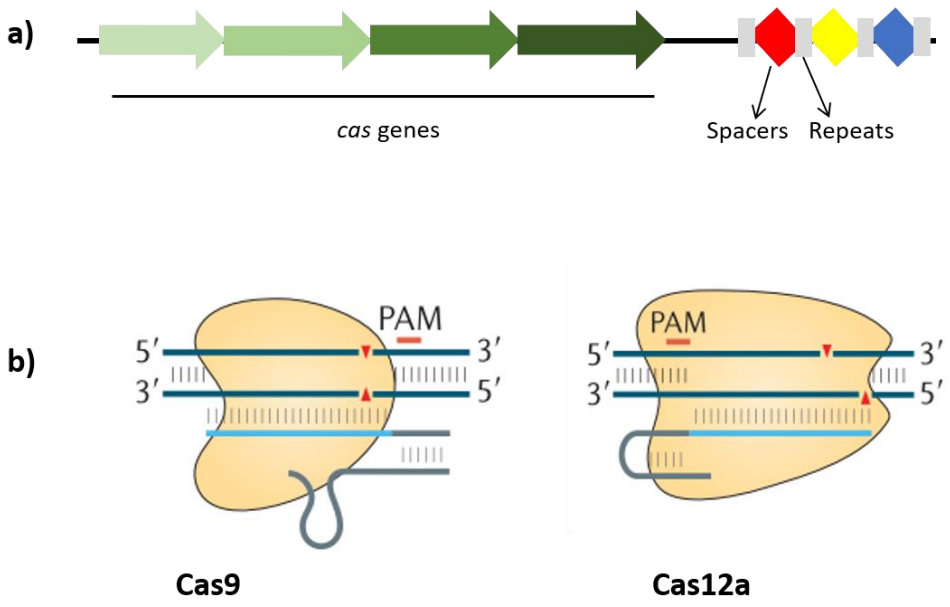


Figure 20. General description of a CRISPR-Cas system. **a)** Structure of a CRISPR locus. In blue, *cas* genes, coding for Cas proteins. The repeat and the spacer sequences are represented in grey and colored boxes respectively**. b)** CRISPR-Cas9-sgRNA-DNA complex (left) and CRISPR-Cas12a-gRNA-DNA complex. See text for more details. The PAM sequence is represented in red. The small red triangles show the position of the cuts. Adapted from (Jiang and Doudna, 2017) (Shmakov *et al*., 2017).

The last classification divides the CRISPR-Cas systems into 2 classes (class 1 and class 2) (Makarova *et al*., 2020). Class 2 systems encode a single large Cas protein that binds to the gRNA and to the target sequence and performs the specific cleavage (Makarova *et al*., 2020), making these proteins the ideal candidates for genomic engineering purposes. While the CRISPR-Cas9 system was the first to be characterized and used (Jinek *et al*., 2012), other systems such as Cas12a (Zetsche *et al*., 2015) (also known as Cpf1) are emerging and increasing their importance.

Cas12a

Cas12a, also known as Cpf1, was discovered in 2015 (Zetsche *et al*., 2015). It belongs to the V-A subtype within Class 2 CRISPR-Cas systems. The structure of the protein is known; it adopts a bilobed structure and it contains a characteristic Ruv-C like endonuclease domain (Zetsche *et al*., 2015).

Even though Cas9 and Cas12a belong to the same Class 2 CRISPR-Cas family, there are significant differences between them (Figure 20b). The majority of Cas9 variants are bigger than Cas12a analogs. Also, Cas9 requires two RNA molecules: trans-activating crispr RNA (tracrRNA) and a crispr RNA (crRNA), although for biotechnological purposes both RNAs have been fused into one single guide RNA (sgRNA), whereas Cas12a requires only a small single RNA molecule, the crRNA. Cas9 cannot process the gRNAs on its own and it needs the formation of a different complex and the assistance of an RNase III. In contrast, Cas12a processes its own crRNA in its ribonuclease catalytic site on its own. This makes Cas12a an attractive candidate for multiplex gene regulation. Both enzymes have similar types of specificities and tolerate similar mismatches *in vitro*, although Cas12a has been proved to have less off-target activity in *vivo* (Swarts and Jinek, 2018; Paul and Montoya, 2020).

Other differences between both proteins are regarding the target site. For Cas9 targeted DNA sequences, the PAM is situated downstream of the spacer sequence on the non-template strand, and it is typically 5’-NGG-3’. In contrast, Cas12a recognizes a PAM, typically 5’- TTTV-3’, located upstream of the spacer (Swarts and Jinek, 2018). Cas9 cuts of the target sequence produce a blunt DSB, while Cas12a produces a staggered DSB. Cas9 cleavage occur three base pairs upstream from the PAM, therefore a normal

outcome is the inhibition of Cas9 activity by the insertion or deletion of single nucleotides produced by NHEJ-mediated which prevent further cleavage of the genomic target site. In contrast, Cas12a cleavage occurs after the 18th nt on the non-targeted strand and after the 23rd base on the targeted strand (Zetsche *et al*., 2015) (these cleavage sites could varies various nucleotides depending on the size the DNA substrate containing the target site (Stella *et al*., 2017)). Therefore, several rounds of Cas12a cleaving and NHEJ could happen without disrupting the PAM which could increase the likelihood of recovering homologous recombinants and decrease the likelihood of recovering NHEJ-mediated mutants (Swarts and Jinek, 2018).

Finally, both Cas proteins have been used for bacteria editing (Vento *et al*., 2019). However, Cas9 have been described to showed toxicity in some bacteria, while Cas12a has less toxic effects and can efficiently edit genomes of these bacteria (Ungerer and Pakrasi, 2016; Jiang *et al*., 2017).

Biotechnological applications

CRISPR-Cas systems can be used in both eukaryotic and prokaryotic cells. Generally, their use as genomic editing tools is based on the co-expression in the cell to be modified of a gRNA targeting the desired genomic site and the Cas protein. The endonuclease will produce a DSB in the target sequence. Then, the cell will try to repair these DSBs through different repairing pathways by the generation of mutations (deletions or insertions) or by using an homologous DNA template (Cong *et al*., 2013). New modifications of the system are constantly appearing, which means that specific genome modification tools are continually arising. The technique can be used for genome editing, base editing, transcriptional control and even epigenetic modifications in eukaryotic cells (Doudna, 2020).

Although CRISPR applications for the genomic modification of eukaryotes have been prevalent, applications in prokaryotic cells are on the rise (Yao *et al*., 2018). This technology has allowed metabolic engineering of different bacteria such as *Escherichia coli, Clostridium spp.* or *Cyanobacteria spp.* (Jakočiunas *et al*., 2016; Ungerer and Pakrasi, 2016; McAllister and Sorg, 2019), improving their use as cell factories. They have also been used for biomedical research purposes to study different pathogens such as

Mycobacterium tuberculosis, Yersinia pestis o Klebsiella pneumoniae (Yao *et al*., 2018). CRISPR-Cas can also target specific bacterial or plasmid populations, allowing their use as antimicrobials or as antibiotic resistance tools (Marraffini and Sontheimer, 2008; Bikard *et al*., 2014; Fagen *et al*., 2017).

Limitations and challenges

Despite its enormous impact in the biotechnology field, CRISPR-Cas technology has limitations that must be overcome for its correct use. There are two principal aspects that must be enhanced: to decrease the off-target activity of Cas proteins, and to improve the delivery system *in vivo* into the recipient cell. To decrease the off-target effect, different algorithms have been designed that predict the potential off target sites and also the system has been modified to improve its specificity (Bae *et al*., 2014; Stemmer *et al*., 2015; Kleinstiver *et al*., 2016, 2019; Slaymaker *et al*., 2016).

The second aspect is to improve the delivery system into the cell (**Figure 21**). There are different reviews on the CRISPR-Cas delivery systems (Fagen *et al*., 2017; Rui *et al*., 2019; Wilbie *et al*., 2019; Xu *et al*., 2019; Chen *et al*., 2020). CRISPR-Cas system can be delivered using different vehicles with different cargos Briefly, Cas could be delivered as DNA, mRNA or protein. Nucleic acid delivery has been addressed in Section 1.3.1). The use of proteins as delivery cargoes has some disadvantages such as the high cost of the purification process, or bacterial endotoxin contamination. However, it is the most straightforward approach, because it avoids the need of transcription and translation of the *cas* gene. Moreover, it decreases the off-target effect because of its transient expression, and eliminates the problem of DNA integration, which are relevant disadvantages of using DNA as the delivered cargo (Glass *et al*., 2018).

.

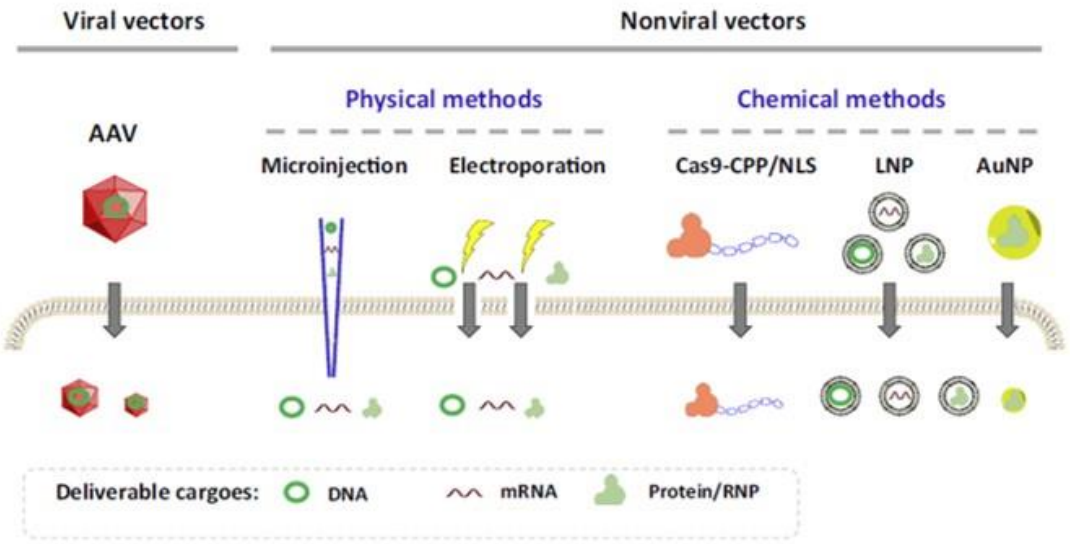


Figure 21. CRISPR-Cas system delivery methods. The system could be delivery as DNA, mRNA or protein/Ribonucleoprotein complex (RNP). There delivery strategy could be using viral vectors or by physical or chemical methods. Taken from (Glass *et al*., 2018).

2. Aims and scope

2. Aims and scope

Conjugative relaxases are indispensable for initiation and termination of DNA processing during bacterial conjugation. Often, they are multidomain proteins which can catalyze other reactions not necessarily associated to conjugative DNA transfer, such as DNA integration. Nowadays, it is unclear what makes a relaxase able to catalyze or promote an integration reaction; the existing data made us think that the ability to oligomerize on the DNA substrate could be the key. It is also not known which biological role this reaction could play; we hypothesized that it could contribute to the colonization of non-permissive hosts, which would have profound biological implications.

The relaxase TrwC is able to catalyze integration of the transferred DNA into a second *oriT* copy present in the recipient cell (in a plasmid or in the chromosome). Moreover, TrwC can pilot DNA and promote its integration into human cells; this integration is not site-specific, yet it has an interesting biotechnological potential. In summary, TrwC can be translocated into a recipient cell (eukaryotic or prokaryotic), alone or covalently linked to an *oriT*-containing DNA, where it can promote the integration of the transferred DNA into the host genome. Therefore, TrwC can be used as a vehicle for specific delivery of DNA or other proteins into a recipient cell. We reasoned that the ability of TrwC to deliver DNA and promote its integration into the human genome, could be combined with other systems which are able to perform site-specific genomic edition, in order to obtain a tool for *in vivo* delivery and targeted genomic edition.

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

1. To study and compare the integrase activity of conjugative relaxases.
    a. To study the relationship between SSI activity and oligomerization ability in the presence of target ssDNA, using the relaxase TrwC as a model.
    b. To study the possible biological role of relaxase-driven integration on colonization of non-permissive hosts after conjugative DNA transfer.
    c. To analyze and compare different relaxases in their ability to promote DNA integration in human cells.

2. To use relaxases as protein/DNA delivery systems for biotechnological purposes.
    a. To use a TrwC-Cas12a fusion protein as a Cas12a delivery method in prokaryotic cells
    b. To prove TrwC-Cas12a activity once it is translocated through the T4SS into a prokaryotic recipient cell
    c. To construct Cas12a fusions and validate their activity in human cells

With these approaches, we aim to gain knowledge on the integrase ability of conjugative relaxases and its biological implication. We also aim to explore their possible application as *in vivo* delivery vehicles for DNA/proteins into recipient cells which are difficult to target by other means, and in particular, their contribution to genomic editing in combination with the CRISPR-Cas systems.

3. Experimental procedures

3. Experimental procedures

3.1. Bacterial strains

Bacterial strains used in this work and their relevant genotype are listed in **Table 1.**

Table 1. Bacterial strains used in this work

Strain	Genotype	Reference
Bartonella henselae		
RSE247	Sm^R spontaneous mutant of ATCC 49882	(Schmid *et al*., 2004)
Escherichia coli		
C41	*F⁻ dcm ompT hsdS (r_B^- m_B^-) gal λ (DE3)*	(Miroux and Walker, 1996)
D1210	Sm^R*; recA hspR hsdM rpsl lacIq*	(Sadler *et al*., 1980)
DH5α *pir*	*endA1 hsdR17 glnV44 thi-1 recA1 gyrA96 relA1 φ80dlacΔ(lacZ)M15 Δ(lacZYA-argF)U169 zdg-232::Tn10 uidA::pir+*	(Platt *et al*., 2000)
DH5α T1 phage resistant	Nx^R; *F- ϕ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(rk−, mk+) phoA supE44 λ-thi-1 gyrA96 relA1 tonA*	(Killmann *et al*., 1996)
FD3	MG1655*::sacB*	Depardieu and Bikard, unpublished
MG1655	*F- lambda-ilvG-rfb-50rph-1*	(Blattner *et al*., 1997)
S17.1	Sm^R; *F⁻* RP4-2-Tc::Mu *aph*::Tn*7* recA	(Simon *et al*., 1983)
β2150	*ΔdapA::(erm-pir) thrB1004, pro, thi, strA, hsdS, lacZ ΔM15,(FΔ lacZ ΔM15 lacIq, traD36, proA+,proB+)*	(Demarre *et al*., 2005a)
Lacticaseibacillus casei		
Lacticaseibacillus casei 393	Laboratory strain	(Hansen and Lessel, 1971)

3.2. Plasmids

Table 2 summarizes the plasmids used in this work that have been previously constructed. **Table 3** summarizes the plasmid constructed for this work.

Table 2. Previously constructed plasmids used in this work

Plasmid	Description	Reference
pAA12	pHP159::*trwC-ralF TS*	(Alperi *et al*., 2013)
pAA58	RSFK::*egfp*	(Guzmán-Herrador *et al*., 2017)
pBBR6	Cloning vector derived from pBBR1-MCS	(Vergunst *et al*., 2000)
pCIG1099	pET3a::*trwC* (N600)	(César *et al*., 2006)
pCMS11	pSW23 ($oriV_{R6K}$)::$oriT_w$+$oriT_p$	(Draper *et al*., 2005)
pCOR31	pHP159 Neo^R	(Gonzalez-Prieto *et al*., 2017)
pCOR35	pHP181 Neo^R	(Gonzalez-Prieto *et al*., 2017)
pCOR48	Shuttle vector *E. coli* and *Lactobacillus oriV*. Ap^R, Em^R, $oriT_w$	(Samperio *et al*., 2021)
pCOR49	Shuttle vector *E. coli* and *Lactobacillus oriV*. Ap^R, Em^R, $oriT_p$	(Samperio *et al*., 2021)
pCOR50	Suicide vector *E. coli oriV*. Ap^R, Em^R, $oriT_w$	Coral González-Prieto
pCOR51	Suicide vector *E. coli oriV*. Ap^R, Em^R, $oriT_p$	Coral González-Prieto
pHP159	pBBR6::*oriT trwABC+egfp*	(Fernández-González *et al*., 2011)
pHP161	pBBR6::*oriT trwABC+egfp*	(Fernández-González *et al*., 2011)
pHP181	pBBR6::*oriT trwAB+egfp*	(Fernández-González *et al*., 2011)
pLA24	pHP159*::BID*	(Fernández-González *et al*., 2011)
pMTX708	Contains Hyg^R cassette	(Gonzalez-Prieto *et al*., 2017)
pMTX808	pAA58::*ap (mobA-)*	Matxalen Llosa

pMTX821	pHP159 Km^R Gm^S	Matxalen Llosa
pMTX822	pHP181 Km^R Gm^S	Matxalen Llosa
pOSIP-CO-RBS-library-dCas9	Contains *Ptet*	(Cui *et al*., 2018)
pRS130	pBGR::*mob:BID+gfp* Neo^R	(Schröder *et al*., 2011)
pSU711	R388*::ΔoriT*	(Demarre *et al*., 2005b)
pSU1445	R388::*tn5tac1* in *trwC*	(Llosa *et al*., 1994b)
pSU1588	pET3a:: *trwC* (N293)	(Boer *et al*., 2006)
pSW27	pSW23*::oriT*$_{R388}$*+oriV*R6K	(Demarre *et al*., 2005a)
pUC8	Cloning vector	(Vieira and Messing, 1982)
pUC18	Cloning vector	(Norrander *et al*., 1983)
pY010	pcDNA3.1:: *hAscas12a*	(Zetsche *et al*., 2015)
pZA31-sulA-GFP	pZA31::*pSOS gfp*	(Cui and Bikard, 2016)
RSFK	RSF1010 Km^R Sm^S	(Lessl *et al*., 1993)

Table 3. Plasmids constructed in this work.

			Construction			
Plasmid	**Description**	**Phenotype**	**Method[1]**	**Vector**	**Insert**	**Oligonucleotides sequence (5' to 3')/ Restriction sites [2]**
pLG01	pBBR6::*oriT trwABC-cas12a+egfp*	Gm^R	IA	pAA12	pY010	**Insert$_1$**: AATTAACCCTCACTAAAGGGAACAAAAGCTGGGTACCGGGCCCCCCC**TTAGGCATAGTCGGGGAC** **Insert$_2$:**GAAGGCCGCCAGAGAAGCCGAGCGCGGCATGGAGGCCGGAAGGTCA**ATGACACAGTTCGAGGGC** **Vector:** XhoI
pLG02	pY010::*cas12a-BID*	Ap^R	IA	pY010	pLA24	**Insert$_1$**: TTATGCATACCCATATGATGTCCCCGACTATGCC**CCCCTCTACGAAGGAG** **Insert$_2$:** CGAGCGGCCGCCACTGTGCTGGATATCTGCAGAATTC**TTACATACCAAAGGCCA** **Vector$_1$: GGCATAGTCGGGGACAT** **Vector$_2$: GAATTCTGCAGATATCCA**
pLG03	RSF1010K::*mobA-::ApR::HygR::egfp*	Ap^R, Km^R, Hyg^R	IA	pMTX808	pMTX708	**Insert$_1$**: TCCAGATGTATGCTCTTCTGCTCGGCGCGCC**TTTCGTCTCGAGGCAGTG** **Insert$_2$:**TGCGATGATAAGCTGTCAAACAGGCGCGCC**GTCAGTTAGGGTGTGGAAAG** **Vector:** SgsI
pLG04	RSFK1010::*egfp::HygR*	Km^R, Hyg^R	IA	pAA58	pMTX708	**Insert$_1$**: TCCAGATGTATGCTCTTCTGCTCGGCGCGCC**TTTCGTCTCGAGGCAGTG** **Insert$_2$:**TGCGATGATAAGCTGTCAAACAGGCGCGCC**GTCAGTTAGGGTGTGGAAAG** **Vector:** SgsI

[1]IA, isothermal assembly; RC, restriction cloning. [2]For plasmids constructed by isothermal assembly, nucleotides annealing to the template during PCR amplification are shown **in bold**. When the vector was linearized by restriction digestion, the enzymes used are shown. For plasmids constructed by restriction cloning (RC), restriction sites used for cloning are underlined. Oligonucleotides or restriction enzymes used for the insert or vector obtention are described as "Insert" or "Vector" for each construction.

Plasmid	Description	Phenotype	Construction			
			Method[1]	**Vector**	**Insert**	**Oligonucleotides sequence (5' to 3')/ Restriction sites [2]**
pLG05	pBBR6::*oriT trwABC::HygR::KmR*	Km^R, Hyg^R	RC	pMTX821	pMTX708	**Insert$_1$:** CCAAACATCGAT**GTCAGTTAGGGTGTGGAAAG** **Insert$_2$:**CCAAACATCGAT**CTTTCGTCTCGAGGCAGTG** **Vector:** ClaI
pLG06	pBBR6::*oriT trwAB::HygR::KmR*	Km^R, Hyg^R	RC	pMTX822	pMTX708	**Insert$_1$:** CCAAACATCGAT**GTCAGTTAGGGTGTGGAAAG** **Insert$_2$:** CCAAACATCGAT**CTTTCGTCTCGAGGCAGTG** **Vector:**ClaI
pLG07	pBBR6::*cas12a-BID*	Gm^R	IA	pBBR6	pLG02	**Insert$_1$:** TTTAACGCGAATTTTAACAAAATATTAACGC**TTACATACCAAAGGCCA** **Insert$_2$:** TGAGCGGATAACAATTTCACACAGGAAACAGCT**ATGACACAGTTCGA GGG** **Vector$_1$: GCGTTAATATTTTGTTAAAATTCG** **Vector$_2$: AGCTGTTTCCTGTGTGAAA**
pLG08	pY010::*trwC-cas12a*	Ap^R	IA	pY010	pAA12	**Insert$_1$:** TTAAGCTTGGTACCGCCACC**ATGCTCAGTCACATGGTATTG** **Insert$_2$:** AAGCCCTCGAACTGTGTCAT**CCTTCCGGCCTCCAT** **Vector$_1$:ATGACACAGTTCGAGGG** **Vector$_2$: GGTGGCGGTACCAAG**
pLG11	pY010::*mobA-cas12a*	Ap^R	IA	pY010	pAA58	**Insert$_1$:** TTAAGCTTGGTACCGCCACC**ATGGCGATTTATCACCTT** **Insert$_2$:**AAGCCCTCGAACTGTGTCAT**CATGCTGAAATCTGGCC** **Vector$_1$: ATGACACAGTTCGAGGG** **Vector$_2$:GGTGGCGGTACCAAG**

Table 3. Plasmids constructed for this work (continued)

			Construction			
Plasmid	**Description**	**Phenotype**	**Method[1]**	**Vector**	**Insert**	**Oligonucleotides sequence (5' to 3')/ Restriction sites [2]**
pLG14	pBBR6::*cas12a*	Gm^R	AI	pBBR6	pY010	**Insert$_1$:** TGAGCGGATAACAATTTCACACAGGAAACAGCT**ATGACACAGTTCG AGGG** **Insert$_2$:** TTTAACGCGAATTTTAACAAAATATTAACGC**GGCATAGTCGGGGACA** **Vector$_1$: GCGTTAATATTTTGTTAAAATTCG** **Vector$_2$: AGCTGTTTCCTGTGTGAAA**
pLG15	pUC8::*lacZ*	Ap^R	RC	pUC8	*lacZ$_{gRNA}$*[3]	**Insert:** GAATTCGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGATCCGACCGCA CGCCGCATCCAGCGCTGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGATAAGCTT **Vector**: EcoRI-HindIII
pLG19	pUC18:: *sacB$_{gRNA}$*	Ap^R	RC	pUC18	*sacB$_{gRNA}$*[3]	**Insert:**GAATTCGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGATGGACAGCTGGC CATTACAAAACGGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGATAAGCTT **Vector**: EcoRI-HindIII
pLG22	pBBR6::*trwC*	Gm^R	IA	pBBR6	pHP159	**Insert$_1$:**GGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCT**ATGCTCAGTC ACATGGTATTGA** **Insert$_2$:**TAACAAAAATTTAACGCGAATTTTAACAAAATA**TTAACGCTTACCTTCCGG CCTCCAT** **Vector$_1$:GCGTTAATATTTTGTTAAAATTCGC** **Vector$_2$: AGCTGTTTCCTGTGTGAAATTGT**

Table 3. Plasmids constructed for this work (continued)

Plasmid	Description	Phenotype	Construction			
			Method[1]	Vector	Insert	Oligonucleotides sequence (5' to 3')/Restriction sites [2]
pLG24	pBBR6::*trwC-cas12a*	Gm^R	IA	pLG14	pY010	**Insert$_1$:**TATCCGGAGGCATATCAAATGACCTAGTTAGGAGGCAAAA**ATGCTCAGTCACATGGTATTGA** **Insert$_2$:**TCACCTGATACAGGTTGGTAAAGCCCTCGAACTGTGTCATTGA**CCTTCCGGCCTCCAT** **Insert$_3$: TTAAGACCCACTTTCACATTTAAG** **Insert$_4$: TTTTGCCTCCTAACTAGGTCAT** **Insert$_5$: ATGACACAGTTCGAGGGCT** **Insert$_6$: GCGTAGCACCAGGCGT** **Vector$_1$: ACCAATAGGCCGACTGCGAT** **Vector$_2$:** GGATTAGAAAAACAACTTAAATGTGAAAGTGGGTCTTAA**TTAGGTGGCGGTACTTGGGTCG**
pLG27	pSW27::P*tac::sacB** homologous recombination cassette	Cm^R	IA	pSW27	FD3 strain gDNA	**Insert$_1$:**CATTTTCGCCAAAAGTTGGCCCAGGGCTTTTGACAATTAATCATCGGCTCGTATAATG**TGCGTACATAAAAAAGGAGACAT** **Insert$_2$: TTGTAATGGCCAGCTGTCCCATTAGTCCAGGCCTTTTGCA** **Insert$_3$: TGGGACAGCTGGCCATTACAA** **Insert$_4$: CTGTTGATACCGGGGTCAATA**GAAGTTTCGCCGACTTTTTGA **Vector$_1$: AAGCCCTGGGCCAACTTTTG** **Vector$_2$: TATTGACCCCGGTATCAACAGGG**

Table 3. Plasmids constructed for this work (continued)

3.3. Molecular biology techniques

3.3.1. DNA extraction and purification

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

-For plasmid DNA extraction, GenElute Plasmid miniprep kit (Sigma Aldrich) was used. GeneJet Gel extraction kit (Thermo Scientific) was used for gel extraction and DNA purification. GeneJet PCR purification kit (Thermo Scientific) was used for PCR product purifications.

-For total DNA extraction of lactobacilli, a colony was resuspended in 50 µl of TE buffer (10 mM Tris·HCl; pH 8.0, 1 mM EDTA). Then, 50 µl of chloroform were added and mixed thoroughly. The mixture was centrifuged 10 minutes at 4°C. The top phase containing the genomic DNA was collected carefully and used directly for PCR analysis (Samperio *et al*., 2021).

-GeneElute Mammalian Genomic DNA Miniprep Kit (Merck) was used for purification of genomic DNA from human cells.

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

3.3.2. DNA electrophoresis

DNA was analyzed by agarose gel electrophoresis. 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. Agarose gels were stained with SYBR safe (Invitrogen). DNA samples were diluted in 6x Loading buffer (bromophenol blue 0.25 % (w/v), sucrose 40 % (w/v) in TBE). GeneRuler 1kb DNA ladder (Thermo Scientific) was used as a molecular weight marker. Electrophoresis were performed using a horizontal BioRad electrophoretic system (with constant voltage between 80-120 V). Gel Doc2000 UV system was used for agarose gel visualization and images were analyzed with Quantity One software (BioRad).

3.3.3. Cloning procedures

The constructions listed in Table 3 were constructed by standard restriction cloning procedure (RC) or by Isothermal assembly (IA) (also called Gibson assembly).

3.3.3.1. Standard restriction cloning procedure

Standard restriction cloning procedures were performed as described in (Sambrook and Russell, 2001). Details of restriction sites used for each plasmid constructed are detailed in Table 3.

PCR amplifications for cloning procedures were performed with Phusion High-Fidelity DNA Polymerase (Thermo Scientific™) or with PCRBIO HiFi (PCR Biosystems), following manufacturer's recommendations. Oligonucleotides were designed with tails containing the recognition site for the desired restriction enzyme and 4-5 extra base pairs, needed for an efficient cleavage. Restriction enzymes were used for DNA restriction to obtain sticky ends in both, plasmid and insert. Digestions were performed with Thermo Scientific restriction enzymes, following manufacturer's recommendations. Digestions usually were performed in 20 µl total mix volume at 37°C for 1h.

Dephosphorylation of the vector was performed to avoid vector religation. After restriction digestion, FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific) was added. Samples were incubated for 10 min at 37°C. The enzyme was inactivated at 75°C for 5 min.

Ligation reactions were performed using the T4 DNA ligase (Thermo Scientific). Different molar ratios (insert/vector) were used. Samples were incubated overnight at 22°C in a cold room (at 4°C). For each ligation, the same reaction without insert DNA was used as negative control. The ligase was inactivated at 70°C for 5 min. Samples were dialyzed for 30 min using 0.05 µM filters pore size nitrocellulose (Millipore GS). The samples were then electroporated into electrocompetent cells, as is described in Section 3.4.2.

Colony PCR to screen bacterial colonies with the desired plasmid product were performed with Kapa Taq Polymerase (Kapa Biosystems), following manufacturer's recommendations.

3.3.3.2. Isothermal assembly

Isothermal assembly method (Gibson *et al*., 2009) is a one-reaction cloning method which is based on the homology between the ends of the fragments to be assembled. Insert fragments were obtained by PCR amplification using Phusion High-Fidelity DNA Polymerase (Thermo Scientific™) or PCRBIO HiFi (PCR Biosystems). The oligonucleotides used for amplification contained a 20-40 bases homology sequence as tail. Vector linearization was obtained by PCR amplification or by restriction digestion. Vector DNA fragments were digested with FastDpnI (Thermo Scientific) restriction enzyme at 37°C for 15 min to eliminate possible template background. For the isothermal assembly, different molar ratios (inserts/vector) were used. A volume of 5 µl of insert-vector mixture was mixed with 15 µl of the Gibson buffer (1M Tris-HCl pH 7.5, MgCl2 2M, dNTPs 100mM, DTT 1M, 1,5 g de PEG 8000, NAD 100mM). This buffer also contains T5 exonuclease (Epicentre), Phusion polymerase (Thermo Scientific) and Taq ligase (New England BioLabs). The reaction was incubated 1 hour at 50°C. Samples were microdialyzed for 30 min using a 0.05 µM wide pore nitrocellulose filter (Millipore GS) and electroporated in electrocompetent cells (Section 3.4.2.).

Colony PCR to screen bacterial colonies with the desired plasmid product were performed with Kapa Taq Polymerase (Kapa Biosystems), following manufacturer's recommendations.

3.3.4. DNA sequencing

DNA sequences were determined by Sanger DNA sequencing (STAB VIDA (Caparica, Portugal)).

3.3.5. Protein purification

3.3.5.1. TrwC N600 and N293 purification

For purification of TrwC fragments N293 and N600, we followed the procedure previously described in (Boer *et al*., 2006; César *et al*., 2006). C41 cells carrying plasmids pCIG1099 or pSU1588 were grown in 1L of LB supplemented with antibiotics at 37°C to an OD_{600}= 0.5 was reached. Then, IPTG was added to 500 µM final and the cultures were induced for 4 h at 37°C. Cells were centrifuged 10 min at 4000 rpm at 4°C and the pellets stored at -80°C at least overnight. Cells were resuspended in buffer Tris 100mM, NaCl 400 mM, EDTA 1mM, PMSF 0.001%, sonicated and ultracentrifuged (4°C, 20 min at 40,000 rpm). Afterwards, samples were diluted with dilution buffer (Tris-HCl 100mM pH7.5, EDTA 1mM, PMSF 0.001%). A first ion exchange chromatography was performed on a phosphocellulose-P11 column (Whatman). The resin was activated according to the manufacturer's instructions and 15 ml were packed with a 2.5 cm internal diameter adapter (BioRad). Lysates were loaded into the column previously equilibrated with buffer A (Tris-HCl 100 mM, NaCl 150 mM, EDTA 1mM, PMSF0.001%) and washed with 5 column volumes of buffer A. Proteins were eluted with buffer B (100 mM Tris-HCl pH7,5, NaCl 700 mM, EDTA 1mM, PMSF 0.001%). Fractions containing the desired TrwC fragments (N293 or N600) were collected, concentrated, and applied to a Superdex 75HR column (GE Healthcare) for a second chromatographic step. 3-5 volumes of equilibration buffer (25 mM Tris-HCl, NaCl 200mM, EDTA 0.1 mM) were applied to the column; proteins were eluted, and the different fractions containing the proteins were collected and concentrated. Finally, glycerol was added to the samples to 5% final concentration for cryoconservation, and they were stored at -80ºC.

During the purification process, samples were collected at different points and loaded on SDS-PAGE gels (Section 3.3.6) to monitor the process. Different polyacrylamide gel concentrations were used for N293 and N600 electrophoresis (12% and 10% respectively).

3.3.6. Protein electrophoresis

Protein samples were analyzed by SDS-polyacrylamide electrophoresis (Sambrook and Russell, 2001). SDS-polyacrylamide gels at 9-12% (acrylamide:bysacrylamide 29:1) were used for TrwC detection. Electrophoresis was carried out using a Mini-PROTEAN II system (BioRad) in 6.1 cm x 1 mm gels. As molecular weight marker, Protein Dual Color Standards (BioRad) or NZYColour Protein Marker II (NZYtech) were used. Electrophoretic run was performed at 200 V for 2 h in TGS buffer (Tris 25 mM, glycine 250 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.

3.3.7. Western Blot analysis

3.3.7.1. Preparation of samples

Total protein extracts were obtained as described in (Towbin *et al*., 1979). For TrwC detection in prokaryotes, *E. coli* D1210 cells containing the indicated plasmids were grown overnight. The cultures were diluted 1:20 and induced with IPTG 500 µM or aTc 100 ng/ml for 3 hours. 1 ml of each culture was collected, centrifuged, and resuspended in 1/10 volume of 2xSDS-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 -20°C for at least overnight. For TrwC detection in eukaryotic cells, HEK293T cells were seeded into a 6-well plate at a density of 6.25 x 10^5 cells per well. 24 h later, 2.5µg of pY010 or pLG8 were transfected into HEK293T using Lipofectamine 3000 reagent (Invitrogen), following the manufacturer's protocol, and incubated for 3 days prior to analysis. After 3 days of transfection, cells were collected, centrifuged, and resuspended in 1/10 volume of 2xSDS-gel loading buffer. Samples were stored at -20°C for at least one night.

3.3.7.2. Western Blot

Samples were boiled for 5 min and loaded on 9% acrylamide SDS-PAGE gels (Section 3.3.6.). NZYColour Protein Marker II (NZYtech) and Precision Plus Protein Standards Dual Color (BioRad) were used as molecular weight markers. After the run, proteins were transferred to nitrocellulose membranes in TGM buffer (Tris 25 mM, glycine 192 mM, methanol 20 % (v/v), pH 8.3) during 2 h at 180 mA and 4°C. After the transfer, the membranes were washed with TBST buffer (Tris HCl 1 M, NaCl 5 M, 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 10 %(w/v) in TBS (Tris HCl 1 M, NaCl 5 M)) overnight at 4°C. Then, they were washed 3 times in TBST during 10 min at room temperature. Incubation with primary antibody was carried out during 1 h at room temperature. Anti-TrwC (Grandoso *et al*., 1994) was used as primary antibody and it was diluted 1:10000 in (BSA 2% (w/v) in TBS). After the incubation, filters were washed 3 times with TBST during 10 min at room temperature. Secondary antibody (IRDye 800CW anti-rabbit IgG, Li-Cor) was diluted as anti-TrwC antibody. 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).

3.3.8. Analytical ultracentrifugation

Sedimental velocity experiments were carried out by analytical ultracentrifugation assays. These experiments were performed by the Molecular Interaction Facility (Centro de Investigaciones Biológicas, Madrid). A Beckam XL-I analytical ultracentrifuge was used and a An-50Ti rotor. Samples were loaded into double sector 12mm Epon charcoal-filled cell. Then, they were centrifuged at 48000 rpm at 20°C. A range of protein concentrations from 1 to 10 µM was used to evaluate the potential oligomerization of the TrwC domains. The DNA used was a HPLC purified oligonucleotide containing the sequence of the *oriTw* (25+8): 5'-GCGCACCGAAAGGTGCGTATTGTCTATAGCCCA-3' (Merck). The oligonucleotide concentration used was 1 µM, while the protein concentration depended on the ratio tested (1 µM, 4 µM, 5 µM or 10 µM). For the formation of the complex, the oligonucleotide was incubated with the protein 20 min at 20°C in buffer 25 mM Tris-HCl,

NaCl 200mM, EDTA 0.1 mM, glycerol 0.5M. Sedimentation profiles for the oligonucleotide and proteins were recorded using UV absorption (260 nm and 280 nm respectively) scanning optics. The SEDFIT 16.1c program (Schuck, 2000) was used to analyze the results. The S_{20W} coefficient was calculated to correct the values into standard conditions (water, 20°C).

3.4. Microbiological techniques

3.4.1. Growth conditions and selection media

Escherichia coli. Strains were grown at 37°C in Luria-Bertani broth (LB: 10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl; Pronadisa), supplemented with agar 1.5 % for solid culture. Selective media included the following antibiotics (Apollo Scientific or Sigma Aldrich) at the indicated concentrations: chloramphenicol (Cm) 20 µg/ml or 25 µg/ml ; ampicillin (Ap) 100 µg/ml; kanamycin monosulfate (Km) 20 µg/ml; streptomycin (Sm) 300 µg/ml; gentamicin sulfate (Gm) 10 µg/ml; and nalidixic acid (Nx) 20µg/ml. aTc (anhydrotetracycline hydrochloride) 100ng/ml, IPTG (Isopropyl β- d-1-thiogalactopyranoside) 500 µM, X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) 40 µg/ml and 1% sucrose were added when needed.

To store *E. coli* strains, a stationary phase culture was centrifuged and resuspended in peptone-glycerol (peptone 0.75 % (w/v), glycerol 50% (v/v)). Strains were kept at -20 and -80°C.

Bartonella henselae was grown in Columbia blood agar (CBA) (Pronadisa) Petri plates with 5% sheep blood (Oxoid) for 3-4 days at 37ºC in a 5 % CO2 atmosphere. Selective media included Gm, 10 µg/ml; Sm, 100 µg/ml; Km, 50 µg/ml.

To maintain *B. henselae* strains, bacteria from a grown CBA plate were collected with a cotton swab and resuspended in LB-glycerol (LB 50 % (v/v), glycerol 50 % (v/v)). Strains were kept at -20 and -80°C.

Lacticaseibacillus casei 393 was grown in Man, Rogosa and Sharpe (MRS, Oxoid), supplemented with agar 2 % for solid culture. Bacteria were grown at 37°C without aeration for 24 h. Selective media included erythromycin (Em) 5 µg/ml.

To preserve *L. casei*, 10 ml of culture were centrifuged and resuspended in 500 µl of MRS and glycerol was added to a final concentration of 23%. Strains were stored ad -80°C.

3.4.2. Electroporation

Plasmids were introduced in bacteria by electroporation.

Electroporation of *E. coli*

Electrocompetent cells were prepared as follows: overnight cultures were diluted (1/20), grown to OD_{600}=0.5-0.7 and pelleted by centrifugation at 3,500 rpm for 20 min at 4°C. Cells were washed 4 times with 1 volume ice-cold milliQ water (Millipore Corporation), centrifuged and the supernatant removed. A final wash was made with 1/50 volume ice-cold glycerol 10%. Finally, cells were resuspended in 1/400 volume ice-cold glycerol 10% and aliquoted in 50 µl samples. Aliquots were kept at -80°C until usage. Aliquots were mixed with 1-100 ng of DNA in a tube and the mixture was transferred into a 0.2 cm Gene Pulser cuvette (BioRad) and subjected to an electric pulse (2.5 kV/cm, capacitance 25µF and 200 Ω) in a MicroPulser TM (BioRad). 1 ml LB was added to the electroporated cells, which were incubated at 37°C to allow antibiotic-resistance gene expression for 1 hour. After incubation, cells were plated on antibiotic containing media.

Electroporation of *B. henselae*

B. henselae strains were transformed by electroporation using a protocol based in the one described in (Grasseschi and Minnick, 1994). For preparing electrocompetent cells, the content of a 3-day-old CBA plate was collected with a cotton swab into 950 µl of ice-cold PBS. Cells were centrifuged at 4,000 rpm for 5 min at 4°C and the pellet was resuspended in 950 µl of ice-cold glycerol 10%. Cells were centrifuged again in the same conditions. Cells were washed 3 times with 950 µl of ice-cold glycerol 10%, centrifuged and the supernatant removed. Then, the pellet was resuspended in 100 µl of ice-cold glycerol 10%. A volume of 40 µl of competent cells were transferred to a new precooled tube and mixed with 300 ng/µl of DNA. Mixture was incubated on ice for 15 min, placed into a cooled BioRad 0.2 cm Gene Pulser cuvette and it was subjected to an electric pulse (2.5 kV/cm, capacitance 25 µF and 200 Ω) in a MicroPulser TM. Then, 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)) at room temperature was added. The mixture was incubated for 3.5 h at 37°C under 5 % CO_2. Then, cells were centrifuged at 4,000 rpm for 4 min and the pellet was resuspended in 40 µl SB broth and plated on a CBA plate with the appropriate antibiotics.

3.4.3. Bacterial conjugation

All matings in this work were performed on solid media. In brief, donors and recipients were mixed, washed, centrifuged, and transferred to a conjugation filter (0.2 µm cellulose acetate filter, Sartorius) on an agar plate. After incubation, filters were introduced into 2 ml of liquid media and appropriate dilutions were plated on selective media for donors, transconjugants and recipients (as indicated). Transconjugant and donor cells were counted, and the frequency of conjugation was expressed as the number of transconjugants per donor cell.

The growth conditions of bacterial cultures, the mating incubation time, the washing and conjugation media used and specific induction conditions, depend on the strains used and on the conjugation experiment, and they are detailed below.

- **Conjugation from E. *coli* to *E. coli***

Standard mating assays were performed as described in (Grandoso *et al*., 2000). DH5αT1^R or D1210 were usually used as donor and recipient cells. Both were grown until stationary phase. 100 µl of each overnight culture were used. They were centrifuged and washed with LB. Then, donors and recipients were mixed, centrifuged, and resuspended in 20 µl of LB. The mixture was transferred into a cellulose acetate filter, on a LB agar plate. The mating plate was incubated 1 h at 37°C. Then, the filter was introduced into 2 ml of LB and vortexed. Different dilutions were plated in LB agar supplemented with corresponding antibiotics.

In matings to test for TrwC-Cas12a function in conjugation, D1210 and DH5α were used as donor and recipient bacteria respectively. Induction conditions were as follows: overnight cultures of donor strains were diluted 1/20 and grown for 3 h in the presence

of aTc 100 ng/ml or IPTG 500 µM, depending on the construction used. A volume of 100µl were mixed with 100 µl of overnight recipient culture and placed in a cellulose acetate filter on a LB agar plate supplemented with aTc 100 ng/ml for 3 hours.

For mating assays to test for Cas12a activity in recipient cells, D1210 was used as donor and MG1655 or MG1655::*sacB* as recipient strains. Overnight cultures of donor and recipient strains were diluted 1/20 and grown for 3 h in the presence of aTc 100 ng/ml (donor) and IPTG 500 µM (recipient). The mixture was placed in a cellulose acetate filter on a LB agar plate supplemented with aTc 100 ng/ml and IPTG 500 µM for 3 hours. Transconjugants were also selected on plates supplemented with antibiotics and with 1% sucrose, in order to detect sucrose-resistant mutants.

- **Conjugation from *E. coli* to *L. casei***

Donor *E. coli* strains were grown on liquid LB supplemented with antibiotics overnight. Recipient *L. casei 393* were grown on liquid MRS without antibiotics. 100 µl of donor and recipient cells were used. They were centrifuged and washed with BHI media (Oxoid). Both strains were mixed, centrifuged, and resuspended in 20 µl of BHI. The mixture was transferred into a cellulose acetate filter, on a BHI 2% agar plate. The mating plate was incubated at 37°C for 24 hours. Then, the filter was resuspended on 2 ml of BHI and appropriate dilutions were made and plated on selective media for donors (LB agar with antibiotic), recipients (MRS agar) and transconjugants (MRS agar with Em 5 µg/ml).

3.4.4. SOS detection assay

To detect induction of the SOS response upon translocation of TrwC-Cas12a, the plasmid pZA31-sulA-GFP (Table 2) was introduced by electroporation into the recipient strains MG1655 or FD3. Matings were performed (as it is detailed in Section 3.4.3.) and after 3h incubation, the mating plates were introduced directly into an Azure Biosystems c400 Imaging System (Azure Biosystems), in order to directly detect Green Fluorescent Protein (GFP) levels on the conjugation plates.

Next, conjugation was stopped by introducing the filter in 2 ml LB broth. 100 µl of each sample were added on a 96 well black flat microtiter plate. GFP signal (excitation

filter: 475 nm and emission filter: 515 nm) and bacterial cell density (OD 600nm) were measured with a TECAN infinite M200 Pro plate reader. GFP signal/OD600 ratios were calculated.

3.5. Cellular biology techniques

3.5.1. Cell culture

Human cell lines **(Table 4)** were routinely grown in Dulbecco's modified Eagle medium (DMEM; Lonza or Gibco), supplemented with 10% heat inactivated fetal bovine serum (FBS; Lonza). Cells were incubated at 37°C under 5% CO_2. When needed, hygromycin B (Invitrogen), 300 µg/ml or G418 disulfate salt (Sigma Aldrich), 500 µg/ml were used. 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 rpm Supernatants were discarded, and cells were resuspended in 2 ml of freezing medium (DMEM 71.5 %, FBS 21.5 %, DMSO 7 %). Aliquots were stored at -80°C for a week and then transferred to -140°C.

Table 4. Human cell lines used in this work.

Cell line	ATCC number	Description
EA.hy926	CRL-2922	Fusion of HUVEC cells and adenocarcinomic human alveolar basal epithelial A549 cells
HEK293T	CRL-3216	Derivative of human embryonic kidney 293 cells, and contains the SV40 T-antigen
HeLa	CCL-2	Epithelial human cells of cervix adenocarcinoma

3.5.2. Transfection

For routine transfection of HeLa or HEK293T cells, JetPei transfection reagent (Polyplus Transfection) was used. Transfection conditions were adjusted depending on the cell culture format used, following the manufacturer's instructions.

For Cas12a and Cas12a fusions activity detection assay in eukaryotes, transfections were performed with Lipofectamine 3000 (Thermo scientific), according to the manufacturer's instructions.

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

EA.hy926 and HeLa cells were infected with *B. henselae*. *B. henselae* strains containing the appropriate plasmids were grown on CBA plates for 3 to 4 days (first passage). Infections were performed using passages 3 or 4. Human cells were seeded 1 day before infection. For routine infections, cells were seeded in 6-well plates (80,000 cells per well) in 3 ml of 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 100-mm tissue culture dishes seeded with 450,000 cells in 12 ml of media. The day of infection, DMEM was replaced by M199 medium (Gibco) supplemented with 10% FBS and appropriate antibiotics to select for the plasmids carried by the *B. henselae* strains to be added. 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} of 1 corresponds to 10^9 bacteria/ml (Kirby and Nekorchuk, 2002). The multiplicity of infection (MOI) used was 400 bacteria per host cell. The dishes or plates were incubated for 72 h at 37°C under 5% CO_2.

3.5.4. Fluorescence microscopy

Infections, cell morphology and eGFP expression of transfected and infected cells were observed using fluorescence microscopy. A Nikon Eclipse Ti microscope was used for these purposes. The following filters were used for GFP (excitation and emission spectra): 450-490 nm and 520 nm.

3.5.5. Flow cytometry

eGFP levels expressed in human cells were quantified and detected by flow cytometry. At the indicated hours post infection or post transfection, the medium was removed, and cells were washed with PBS and trypsinized. Cells were centrifuged 10

min at 1,000 rpm and the pellet was resuspended in 400 µl PBS in special flow cytometry tubes. A Cytomics FC500 flow cytometer (Beckman Coulter) was used to analyze the samples. For each sample, 20.000 events were analyzed. An untransfected/uninfected control was used to adjust the GFP background level, delimiting the start point of the population of GPF positive cells.

3.5.6. Detection of stable integrants

At 72 hours post infection, 300 µg/ml of hygromycin B or 500 µg/ml of G418 were added to HeLa infected cells in order to select stable integrants. The selection was maintained for 4 to 5 weeks. Antibiotic resistant cells were counted using a Nikon Eclipse Ti microscopy.

Integration experiments were performed in parallel with infections to measure GFP positive cells by flow cytometry. The percentage of GFP positive cells was extrapolated to the number of cells in the 100-mm plate used in the integration assays. To calculate the integration rate, the number of resistant colonies was divided by the inferred number of GFP positive cells.

3.5.7. Detection of Cas12a-mediated cleavage in human cells (SURVEYOR assay)

An U6::gRNA$_{dnmt1}$ eukaryotic expression cassette was generated by PCR amplification of the U6 promoter from HEK293T gDNA, using oligonucleotides U6_crRNA_F and DNMT1_crRNA3_R (**Table 5**). These oligonucleotides also carried a tail sequence with the gRNA targeting *dnmt1*, therefore the resultant amplicon harbours this sequence downstream the promoter. Lipofectamine p3000 (ThermoFisher) was used to cotransfect HEK293T cells with 750 ng of the gRNA cassette and 250 ng of plasmid DNA encoding the nuclease. After 3 days, genomic DNA was extracted using the GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma Aldrich).

Cas12a gRNA-directed cleavage was detected using the Surveyor Mutation Detection Kit (IDT) (Qiu *et al*., 2004). This kit allows the detection of the small insertion-deletions (indels) formed by the non-homologous end joining (NHEJ) pathways of the eukaryotic cell trying to repair the double strand breaks (DSB) produced by the Cas12a

cleavage when a homologous template is not present. The assay consists in four steps: PCR amplification of the targeted region, denaturalization of the strands and rehybridization to allow for the mutant and wild-type strands to anneal, treatment with Surveyor nuclease, which acts on the heteroduplexes, and analysis of DNA by electrophoresis. The gDNA of the edited sample would contain a mixture of wild type sequences and edited sequences (with indels). The targeted region is amplified by PCR and the amplicons (which will be a mixture of wild type and edited sequences) are hybridized to form heteroduplex complex (formed by the wild-type sequence and the indel-containing sequence). The Surveyor nuclease detects this heteroduplex and cleaves the DNA. The resulting fragments can be visualized on a DNA electrophoresis.

The process is as follows. First, the genomic region flanking the expected cleavage site was amplified using Phusion High-Fidelity DNA Polymerase (Thermo Scientific™) with the primers (*dnmt1_F* and *dnmt1_R*) (Table 5), and the amplicons were gel-purified. Secondly, 20 µl of 700 ng of each PCR purification were run in a iCycler (BioRad) thermocycler, in order to obtain the heteroduplex formation, with the following program: 95°C 10 min, 95°C to 85°C (-2.0°C/sec), 85°C 1 min, 85°C to 75°C (-0.3°C/sec), 75°C 1 min, 75°C to 65°C (-0.3°C/sec), 65°C 1 min, 65°C to 55°C (-0.3°C/sec), 55°C 1 min, 55°C to 45°C (-0.3°C/sec), 45°C 1 min, 45°C to 35°C (-0.3°C/sec), 35°C 1 min, 35°C to 25°C (-0.3°C/sec), 25°C 1 min, 4°C. Then, samples were treated with 2 µl of Surveyor Nuclease S and 1 µl of Surveyor Enhancer S, vortexed and incubated 1 hour at 42°C. To each sample, 2,3 µl of STOP solution were added and they were run on a 2% agarose gel.

Table 5. Oligonucleotides used for Cas12a activity detection in eukaryotes

Oligonucleotide	Sequence 5' to 3'[1]
U6_crRNA_F	**GAGGGCCTATTTCCCATGATTCCT**
DNMT1_crRNA3_R	GAGTAACAGACATGGACCATCAGATCTACAAGAGTAGAAATTA**CGGTGTT TCGTCCTTTCCACAAG**
dnmt1_F	**CTGGGACTCAGGCGGGTCAC**
dnmt1_R	**CCTCACACAACAGCTTCATGTCAGC**

[1]Nucleotides annealing to the template during PCR amplification are shown **in bold**

3.6. Computer analysis

3.6.1. Statistical analysis.

For representing the data and performing the statistical comparisons, the software GraphPad Prism 8.00 (San Diego, CA) was used. Student's t-test were used for data comparison. In the graphs through the text, the significance is indicated by one or more asterisks and the corresponding p-value is indicated.

3.6.2. Software

Vector NTI Advance and SnapGene. They are a sequence analysis and design tools that can be used to view, analyze, transform, create, annotate, and share nucleotide and protein sequences. There is no free version available, but a trial version can be downloaded from their websites.

BLAST. The Basic Local Alignment Tool finds regions of local similarity between sequences. It compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST also can be helpful to the identification of members of gene families https://blast.ncbi.nlm.nih.gov/Blast.cgi

Chromas Lite. It is a DNA sequence viewer, allowing the visualization of sequencing chromatogram files. It is a free software.

GraphPad Prism. It is a 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. https://www.graphpad.com/demos/

4. Results

4. Results

4.1. Study of the integrase activity of conjugative relaxases

Some relaxases promote the integration of their attached DNA strand in the recipient cell. However, it is still unclear which factors allow a relaxase to perform the reaction, if this activity is extended among relaxases, and which biological role could this activity play. In this chapter, we attempt to answer some of these questions.

4.1.1. Exploring the relationship between site-specific integrase activity and DNA-dependent oligomerization

The relaxase TrwC was the first described to have site-specific integrase activity (Draper *et al*., 2005). It is able to integrate a transferred DNA strand into an *oriT* containing plasmid (Draper *et al*., 2005) or into a chromosomal *oriT* copy (Agúndez *et al*., 2012). The recombinase domain N600 is the minimal domain able to perform this reaction with high efficiency. The N293 relaxase domain, which contains the catalytic Tyr residues responsible for the cut-and-strand transfer reactions, is not able to perform this activity (Agúndez *et al*., 2012). Thus, the region of TrwC between aa 293 and 600 is providing some function which is essential for the protein to act as an integrase.

Agúndez and collaborators generated a TrwC-Rep68 chimera by fusing the N293 domain of TrwC, with the C-terminal domain of the replicase Rep68 (see Section 1.2.3.2) (Agúndez *et al*., 2018). Surprisingly, the chimera was able to catalyze SSI reaction in bacteria with an integration frequency similar to TrwC. The chimera protein had a different oligomerization behavior than the parental proteins: TrwC-Rep68 did not form oligomers, but after its incubation with the $oriT_w$ (25+8), the sedimentation profile showed oligomer formation corresponding to probably hexamers. These data suggested that the ability to oligomerize in the presence of DNA could be contributing to the integration ability of the chimera.

These results prompted us to determine if the integration activity of N600 could be due by its ability to oligomerize on the target DNA. Both the relaxase and recombinase domains (N293 and N600) behave as monomers in gel-filtration

chromatography (César *et al*, 2006). N293 also behaves as a monomer in analytical centrifugation experiments, both in the absence and presence of DNA (Lucas *et al*, 2010). But the oligomerization properties of N600 in the presence of its target DNA had not been analyzed.

We compared the oligomerization profile of the N600 and N293 domains in the absence/presence of their DNA target. If there was a relation between both characteristics, we would be able to detect oligomerization changes in N600 in the presence of its DNA target. As DNA target, we used the *oriT* sequence 25+8, which comprises the sequence: 5'-GCGCACCGAAAGGTGCGTATTGTCTATAGCCCA-3'. This is the oligonucleotide used by Agúndez *et al* (2018) to show DNA-dependent oligomerization of the N293-Rep68 chimera. The N293 and N600 domains were purified, and the sedimentation velocity was determined by analytical centrifugation.

4.1.1.1. Protein purification

We purified N293 and N600 domains from pSU1588 and pCIG1099 plasmids respectively. N293 has a molecular weight of 33kDa and N600 of 60 kDa. During the purification process, we loaded samples on a polyacrylamide gel to make sure that the purification was successful and to verify which fractions of each protein should be conserved. As explained in Experimental procedures (Section 3.3.5), a first ion exchange chromatography was performed on a phosphocellulose-P11 column. We loaded samples in a gel and we collected and concentrated fractions 53-65 from the N293 domain and 49-55 from the N600 domain **(Figure 22a)**. During each step of the purification, we also collected control samples after and before induction, or after lysis.

The second chromatography step was performed on a Superdex75HR column (Figure 22b). We loaded samples in a gel, and we collected, concentrated, and stored fractions 14-22 from N293 and 9-15 from N600. In this chromatography, both N293 and N600 eluted with the expected size as monomers, as previously reported (César *et al*, 2006).

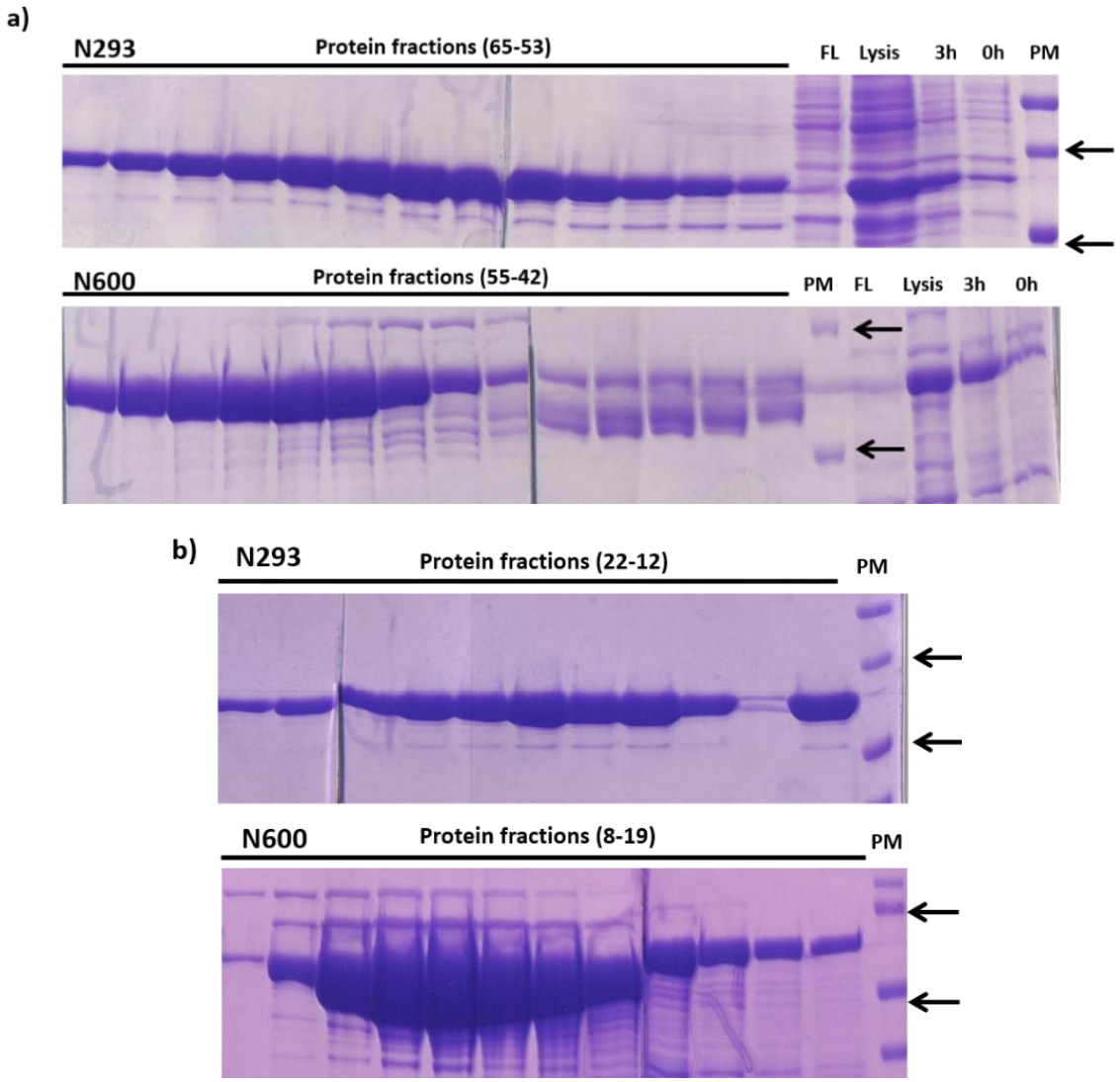


Figure 22. **Purification of N293 and N600. a)** After the purification steps through a P11 column, samples were loaded to determine the fractions which contain the best concentration of proteins. On the top and on the bottom, the fractions corresponding to N293 and N600 respectively. Arrows indicate the molecular weight of 37 and 25 kDa (N293) and 75 and 50 kDa (N600). PM, protein marker; FL, Flow through; Lysis; 3h: 3 hours after induction; 0h, 0 hours after induction. **b)** After the second purification through a Superdex75HR column, fractions containing the proteins were loaded in a gel. On the top and on the bottom, the fractions corresponding to N293 and N600 respectively. Arrows indicates the molecular weight of 37 and 25 kDa (N293) and 75 and 50 kDa (N600). PM, protein marker.

4.1.1.2. Sedimentation velocity

We performed sedimentation velocity experiments to determine the oligomerization profile of both domains in the absence and presence of the *oriT* (25+8) target. These experiments were performed in collaboration with Dr Germán Rivas, by the Molecular Interaction Facility (Centro de Investigaciones Biológicas, Madrid).

Results are summarized in **Table 6** and **Figure 23.** As previously described, both proteins without the oligonucleotide showed a sedimentation coefficient value corresponding to a monomer (2.1S and 2.8S values were obtained for N293 and N600 respectively) (Figure 23a).

Table 6. Sedimentation coefficients obtained from the sedimentation velocity experiments.

Sample	Protein (µM)	ssDNA (µM)	Sedimentation profile coefficient (S)	Normalized sedimentation profile coefficient (S_{20w})
***oriT* (25+8)**	-	1 µM	**1.6S**, 3.4S	**1.9S**, 3.9S
N293	10 µM	-	**2.1S,** 5.7S	**2.6S**, 6.9S
	1 µM	1 µM	1.6S, **3.1S**, 5.0S	1.9S, **3.7S**, 6.0S
	4 µM	1 µM	1.6S, **3.0S**, 4.1S	1.9S, **3.7S**, 4.9S
	5 µM	1 µM	**3.1S**, 5.6S	**3.7S,** 6.7S
	10 µM	1 µM	2.2S, **3.1S,** 5.5S	2.7S, **3.7S**, 6.6S
N600	10 µM	-	**2.8S,** 6.5S	**3.4S,** 7.8S
	1 µM	1 µM	1.5S, 2.3S, **3.7S**, 6.0S	1.9S, 2.8S, **4.5S**, 7.2S
	4 µM	1 µM	1.5S, 2.5S, **3.7S**, 5.7S, 6.9S	1.8S, 3.06S, **4.5S**, 6.8S, 8.3S
	5 µM	1 µM	2.9S, **3.8S**, 5.8S	3.5S, **4.6S**, 7.0S
	10 µM	1 µM	2.9S, **3.9S**, 6.2S	3.4S, **4.7S**, 7.4S

S, experimental sedimentation coefficients. S_{20w}, Normalized sedimentation coefficients (normalized in water and at 20ºC). The coefficient marked in **bold** corresponds to the value of the oligonucleotide, N293, N600, or the complex formed by a molecule of protein and a molecule of DNA.

The oligonucleotide had a sedimentation coefficient of 1.6S. To determine if there were differences in the oligomerization profile in the presence of the target ssDNA, we incubated the oligonucleotide and the proteins using different ratios of ssDNA:protein (1:1, 1:4, 1:5 and 1:10). Figure 23b shows the sedimentation coefficients obtained under the different conditions. For the N293 domain, we obtained values ranged from 3.0-3.1S, while for N600 we obtained values between 3.7-3.9S. Smaller species could be observed, which corresponded with the oligonucleotide or with the protein alone. The traces of bigger species observed are not compatible with higher complex formed by the protein and the oligonucleotide. Thus, the sedimentation profiles corresponded to one molecule of DNA with one molecule of protein in all cases.

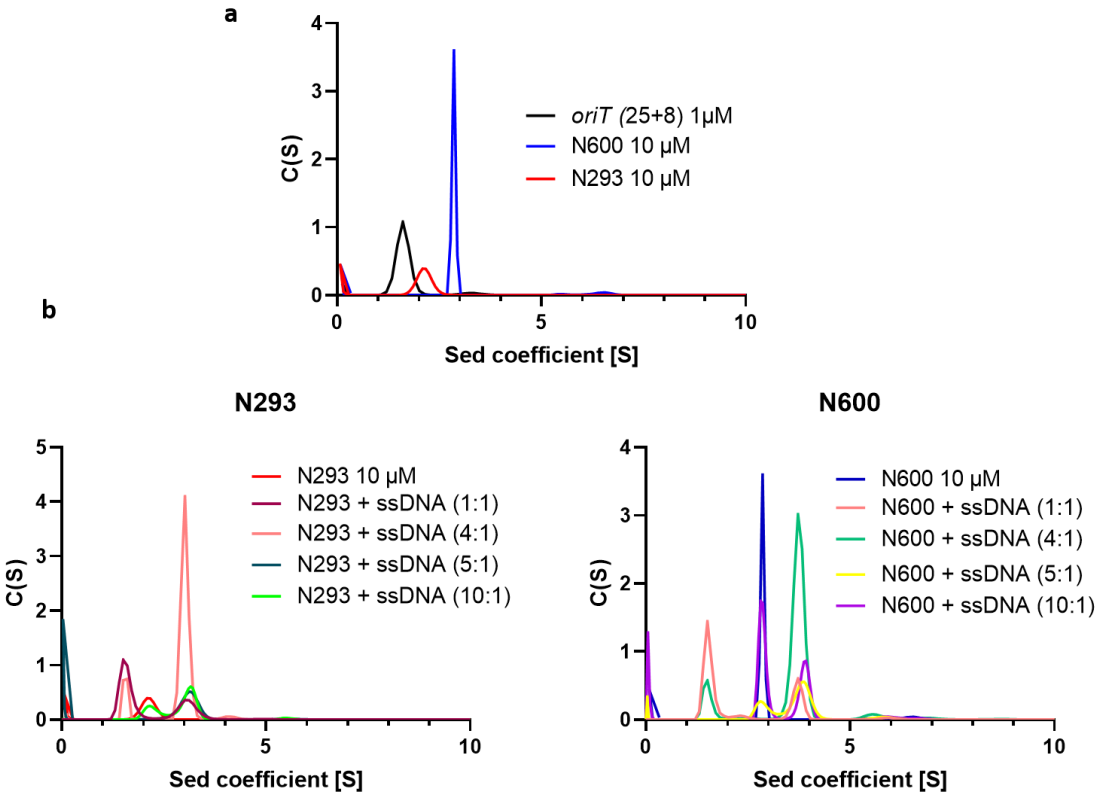


Figure 23. Sedimentation velocity analysis of N293, N600 and the *oriT* 25+8 target. a) Sedimentation profiles of the oligonucleotide *oriT* 25+8 (1µM), N293 (10µM) and N600 (10µM). **b)** Sedimentation profiles of N293 (left) and N600 (right) incubated with its specific substrate *oriTw* (25+8) oligonucleotide. Ratios 1:1, 4:1, 5:1 and 10:1 (protein:oligonucleotide) were tested. C(S); sedimentation coefficient distributions, (S); coefficient of sedimentation. Data were obtained using the SEDFIT program.

In conclusion, there were no oligomerization differences between N293 and N600 in the absence and presence of their *oriT* substrate. This means that, at least under the conditions tested, N600 did not show any DNA-dependent oligomerization ability.

4.1.2. Possible biological role of the integrase activity of conjugative relaxases

The ability of some relaxases to catalyze integration of the transferred DNA could have an important biological role. The transfer range of plasmids is usually broader than the replication range (Kishida *et al.*, 2017). Therefore, integration activity would facilitate the colonization of non-permissive hosts by allowing the integration of the mobilizable elements in the chromosome.

In order to test this hypothesis, we aimed to compare integration of the conjugatively transferred DNA into a non-permissive host, such as Gram-positive

bacteria, using in parallel relaxases known to promote such integration (R388-TrwC) and not reported to do so (RP4-TraI). For this purpose, we had to establish a conjugation protocol in the first place, since there have been no reports of R388 conjugation into Gram-positive bacteria.

The selected recipient was *Lacticaseibacillus casei* 393 strain (previously known as *Lactobacillus casei* (Zheng *et al*., 2020)). These bacteria are widely used in the food industry (Börner *et al*., 2019b) and they have important human biomedical applications (Wang *et al*., 2016). However, the genetic modifications tools available for them are very limited (Bosma *et al*., 2017).

4.1.2.1. DNA mobilization from *E. coli* to *L. casei*

Conjugative transfer using RP4 system from *E. coli* to several Gram-positive bacteria was described long ago (Trieu-Cuot *et al*., 1987). However, no reports of DNA mobilization from *E. coli* into lactobacilli strains had been reported until now.

In order to determine if R388_TrwC and TraI_RP4 were able to mobilize a plasmid from *E. coli* to *L. casei*, we used two shuttle plasmids (pCOR48 and pCOR49) **(Figure 24**). Both plasmids carried an ampicillin resistance cassette for selection in *E. coli*, and an erythromycin resistance cassette for selection in *L. casei*. They also carried two different origins of replication: pBBR322 *oriV* (*oriV1*) and P8014-2 *oriV (oriV2)*, for *E. coli* and *L. casei* respectively, allowing its replication in both bacteria. Finally, pCOR48 carried and *oriTw* (R388) and pCOR49 and *oriTp* (RP4). The *oriT* is the only element required in *cis* for mobilization. The rest of the conjugative machinery was provided in the donors in *trans*. For TraI mobilization, S17.1 strain was used as donor, as it contains the RP4 conjugative system integrated in its chromosome. For TrwC mobilization, the helper plasmid pSU711 (carrying the R388 conjugative system without an *oriT*) was used in D1210 donor cells. As negative controls, we tested in parallel the donor cells without providing the conjugative system in *trans*.

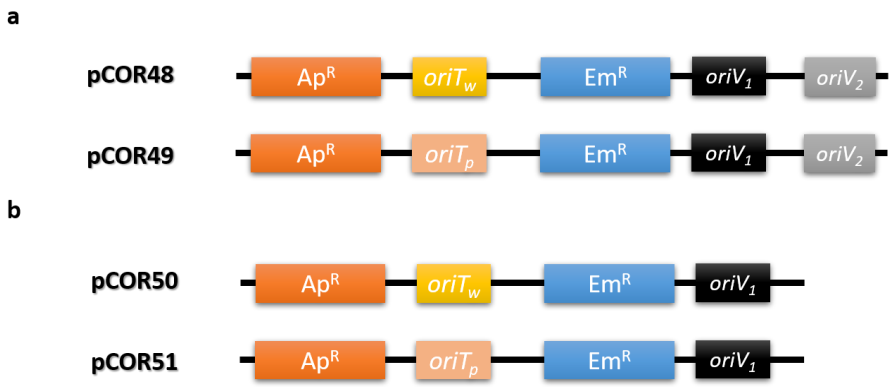


Figure 24. Scheme of plasmids used for DNA mobilization and DNA integration assays. a) pCOR48 and pCOR49 shuttle plasmids used for DNA mobilization**.** Both plasmids carry two *oriV* to replicate in *E. coli* (pBBR322 *oriV*, *oriV1*) and *L. casei* (P8014-2 *oriV, oriV2*), an ampicillin resistance cassette, an erythromycin resistance cassette. pCOR48 carries the origin of transfer *oriTw*(R388) and pCOR49 of *oriTp* (RP4). **b)** pCOR50 and pCOR51 suicide plasmids used for DNA integration. pCOR50 carries the *oriTw* of R388 and pCOR51 of RP4. Both plasmids carry *oriV,* to replicate in *E. coli* (pBBR322 *oriV*, *oriV,* an ampicillin resistance cassette and an erythromycin resistance cassette. Black and grey boxes: *E. coli* (pBBR322 *oriV*) and *L. casei* (P8014-2 *oriV*). Orange boxes: ampicillin resistance cassette. Blue boxes: erythromycin resistance cassette. Yellow and light orange boxes: *oriTw* and $oriT_p$.

First, we needed to set up a conjugation protocol from *E. coli* to *L. casei* using both conjugative systems. For this, we adapted the protocol routinely used for conjugative DNA transfer among Gram-negative bacteria on solid media (Grandoso *et al*., 2000) and tested both systems. Conjugation to *L. casei* was assayed under different conditions, in collaboration with the laboratory of M.A. Alvarez (IPLA, Asturias). Several factors were analyzed in order to obtain an optimized conjugation protocol, such as mating times (3h, 6h and 24h), ratio of recipient per donor (1:1 and 5:1), growth phases of donor and recipient (stationary and exponential phase) and growth media (LB, MRS and BHI). No significant differences were found between the different factors, with the exception of the growth media used and the mating times. Growth media in this protocol played an important role, as it was used to select specifically donors (*E. coli* grows in LB) or recipients (*L. casei* grows in MRS). Also, BHI medium was used for the washing and mating plates, as both bacteria could grow on it. For the final protocol:

- Donor and recipient bacteria were grown overnight using their optimal growth conditions.

- 100 µl of each culture were washed with BHI for several times in order to remove antibiotics and growth media, and then mixed together.

- The mixture was resuspended in 20 µl of BHI and placed in a conjugation filter on the conjugation plate (BHI agar). The mating was incubated for 24 hours at 37°C.

- Bacteria were selected using different media and antibiotics. Donor cells were grown in LB agar plates supplemented with the appropriate antibiotics. Recipients and transconjugants were grown in MRS agar (supplemented with erythromycin for transconjugants selection).

Donor bacteria grew after 24 hours, while recipients and transconjugants grew after 72 hours.

In order to determine if shuttle plasmids were functional, firstly we performed DNA mobilization using the *E. coli* strain DH5α as recipient. Then, mobilization assays using *L. casei* as recipient were performed. Both systems were able to mobilize an *oriT* containing plasmid from *E. coli* to *L. casei*, although RP4 could do it with a higher frequency ($2.76x10^{-5}$, compared to $1.17x10^{-6}$ obtained for R388). With these data we demonstrated that TrwC and TraI could mobilize plasmids containing *oriT* from *E. coli* to *L. casei*. It is noteful that this is the first report of conjugative DNA transfer from *E. coli* to Lactobacilli (Samperio *et al*, 2021). **Table 7** and **Figure 25** summarize mobilization frequencies obtained for each system.

Table 7. Conjugation from *E. coli* to *L. casei* using R388 and RP4 conjugative systems

Recipient	Donor (*E. coli*)	Conj. System [1]	Shuttle plasmid (*oriT)*	Conj. frequency
E. coli [2]	D1210 (pSU711)	R388	pCOR48 ($oriT_w$)	$3.4\ x10^{-3}$ ($\pm1.63x10^{-3}$)
	D1210	none	pCOR48 ($oriT_w$)	$<2.74x10^{-7}$ ($\pm1.02x10^{-7}$)
	S17.1	RP4	pCOR49 ($oriT_p$)	$4.6\ x10^{-2}$ ($\pm2.52x10^{-2}$)
	D1210	none	pCOR49 ($oriT_p$)	$<6.51x10^{-7}$ ($\pm4.32x10^{-7}$)
L. casei	D1210 (pSU711)	R388	pCOR48 ($oriT_w$)	$1.17x10^{-6}$ ($\pm1.63x10^{-6}$)
	D1210	none	pCOR48 ($oriT_w$)	$<3.04x10^{-7}$ ($\pm6.02x10^{-7}$)
	S17.1	RP4	pCOR49 ($oriT_p$)	$2.76x10^{-5}$ ($\pm4.30x10^{-5}$)
	D1210	none	pCOR49 ($oriT_p$)	$<2.47x10^{-7}$ ($\pm5.99x10^{-7}$)

[1]The conjugation system was provided by the helper plasmid pSU711 for R388 or by the S17.1 chromosome for RP4. [2]Conjugation into the *E. coli* recipient strain DH5α. Data represent the mean of at least 3 independent experiments.

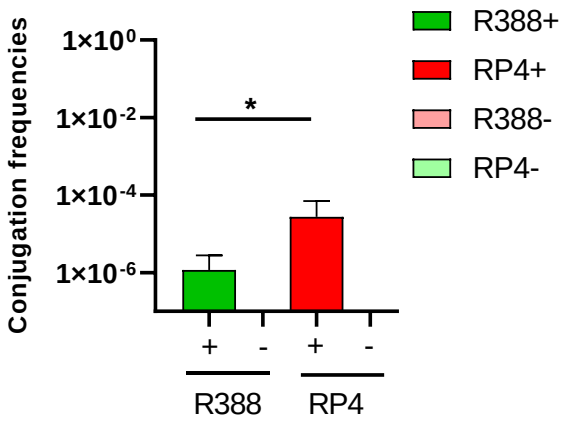


Figure 25. Conjugation frequencies from *E. coli* to *L. casei* 393. Frequencies are shown as transconjugants per donor. Data represent the mean of at least 3 independent experiments. *, p<0.05.

Transconjugants obtained were analyzed to confirm their identity. First, the *oriT* sequence was amplified by PCR from total genomic DNA of transconjugants **(Figure 26)**. The amplifications from all transconjugants analyzed reveals that they all carried the expected *oriT* sequence (*oriT*$_w$, 300 bp; *oriT*$_p$, 280 bp).

Also, 16S rRNA gene was amplified from each transconjugant and sequenced, confirming that they were *L. casei*. Taking these data together, we confirmed that the transconjugants obtained were *L. casei* and that they carried the mobilizable plasmids pCOR48 or pCOR49.

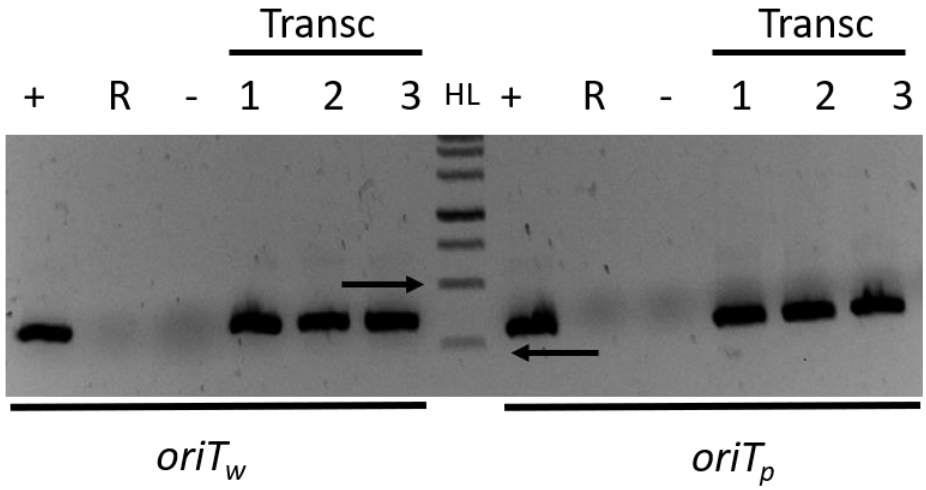


Figure 26. *oriT* amplification from gDNA of the *L. casei* 393 transconjugants obtained. On the left, PCR amplifications of transconjugants obtained using R388 system. On the right, PCR amplifications of transconjugants obtained using RP4 system. +, control PCRs from pCOR48 (left) or pCOR49 (right) plasmid DNA. R, control PCR from recipient *L. casei* gDNA. -, PCR negative control (no DNA template). 1, 2 and 3, PCR amplifications of three different transconjugants obtained in each mating. HL, hyperladder. Top arrow, 500 pb, bottom arrow, 250 pb.

4.1.2.2. Chromosomal integration of DNA transferred from *E. coli* to *L. casei*

After confirming DNA mobilization from *E. coli* to *L. casei* using both conjugative systems, the next step was to test the ability of relaxases to promote the integration of the transferred DNA. To this end, we used the mobilizable suicide plasmids pCOR50 and pCOR51, which shared the same characteristics as pCOR48 and pCOR49, but were incapable of replicating in *L. casei* as they only harbored the origin of replication for *E. coli* (pBBR322 *oriV, oriV*$_1$) (Figure 24b). Therefore, erythromycin-resistant *L. casei* colonies obtained from the mating assays would be the result of integration events.

Similarly as for DNA mobilization assayed, we first mobilized suicided plasmid into the *E. coli* strain DH5α, to determine the mobilization frequency. Then, we performed integration assays using the same conjugation protocol as described before and using *L. casei* as recipient. The integration rate was calculated as the integration frequency divided by the conjugation frequency. We assayed TrwC and TraI integration reactions in parallel. Since RP4_TraI does not catalyze SSI reaction, we expected to determine if this activity was involved in the generation of integrants. The results of the integration assays are summarized in **Table 8** and **Figure 27**.

Integrants were obtained at low frequency using both conjugative systems. Although more integrants were obtained when using RP4, since conjugation frequencies are also higher, no significant differences were found in the integration rates obtained by both relaxases. Analysis of integrants by PCR amplification of the corresponding *oriT* revealed complete *oriT* sequences in all the cases. These results suggested that the integrants obtained were the result of random integration events; otherwise the *oriT* copies would be truncated at the *nic* site.

In conclusion, TrwC did not promote integration of the transferred DNA in prokaryotes when the target *oriT* was not present in the recipient cell.

Table 8. Integration assays from *E. coli* to *L. casei*.

Recipient	Conj. system[1]	Suicide plasmid	Integration		
			Conjugation frequency [3]	Integ. Frequency [4]	Integration rate[5]
E. coli [2]	R388	pCOR50 (*oriTw*)	$1.51x10^{-3} \pm 1.27x10^{-3}$	NA[6]	NA[6]
	none	pCOR50 (*oriTw*)	$<7.81x10^{-7} \pm 4.21x10^{-7}$	NA[6]	NA[6]
	RP4	pCOR51 (*oriTp*)	$7.1x10^{-3} \pm 4.11x10^{-3}$	NA[6]	NA[6]
	none	pCOR51 (*oriTp*)	$<1.44x10^{-7} \pm 1.01x10^{-7}$	NA[6]	NA[6]
L. casei	R388	pCOR50 (*oriTw*)	$1.39x10^{-6} \pm 2.24x10^{-6}$	$1.24x10^{-8} \pm 2x10^{-8}$	$3.58x10^{-2} \pm 4.3x10^{-2}$
	none	pCOR50 (*oriTw*)	$<8.52x10^{-8} \pm 5.1x10^{-8}$	$<5.25x10^{-8} \pm 1x10^{-8}$	NA[6]
	RP4	pCOR51 (*oriTp*)	$3.22x10^{-5} \pm 6.05x10^{-5}$	$2.73x10^{-8} \pm 5.45x10^{-8}$	$1.14x10^{-2} \pm 2.79x10^{-2}$
	none	pCOR51 (*oriTp*)	$<7.14x10^{-7} \pm 3.12x10^{-7}$	$<6.21x10^{-8} \pm 2x10^{-8}$	NA[6]

[1]The conjugation system was provided by the helper plasmid pSU711 for R388 or by the S17.1 strain for RP4. [2]Conjugation into the *E. coli* recipient strain DH5α. [3]Frequencies represent transconjugants per donor. [4]Integration frequencies are shown as integrants per donor. [5]Integration rate was calculated as integration frequency per conjugation frequency. Data represent the mean of at least 3 independent experiments. [6]NA, not applicable

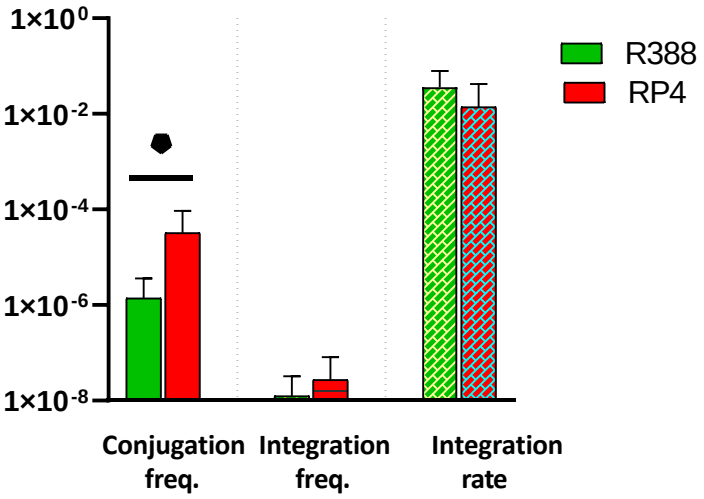


Figure 27. Integration rates from *E. coli* to *L. casei* using R388 and RP4 conjugative systems. The graphic shows the conjugation and integration frequencies calculated as transconjugants per donor and integrants per donor respectively. Integration rate was calculated as integration frequency per conjugation frequency. Data represent the mean of at least 3 independent experiments. *, p<0.05.

4.1.3. Analysis of the ability of different relaxases to promote DNA integration in human cells

Finally, we wanted to test if the ability of promoting the integration of the transferred DNA into a recipient human cell is extended between relaxases. González-Prieto and collaborators showed that TrwC promoted the integration of the transferred DNA into the human cells after been translocated through the T4SS of *B. henselae* (Gonzalez-Prieto *et al*., 2017). They mobilized a plasmid carrying a eukaryotic GFP cassette expression (to detect DNA transient expression) and a neomycin phosphotransferase expression cassette (to detect permanent expression) (**Figure 28**).

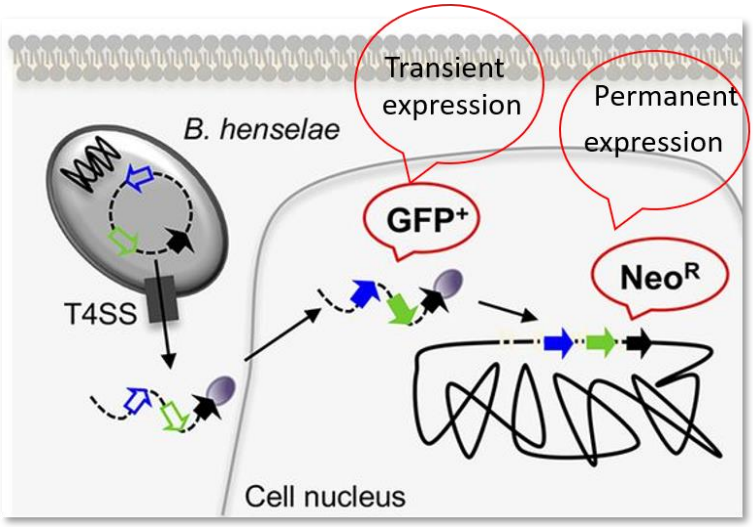


Figure 28. Scheme of mobilization and integration assay to detect transient expression or stable integration of the DNA transferred by *B. henselae* T4SS. After infection of human cell lines with *B. henselae*, relaxase-DNA complex will be translocated through the T4SS and will get to the eukaryote nucleus where genes will be expressed. At 3 days post infection, transient expression of *gfp* can be measured by flow cytometry. Neomycin treatment was applied for long-term selection of neomycin-resistant colonies, to detect stable integration events. Adapted from (Gonzalez-Prieto *et al*., 2017).

In this work we wanted to evaluate if the T4SS of *B. henselae* could recognize and translocate other relaxases and if once in the recipient cell, they could promote DNA integration. In order to test this, we assayed in parallel TrwC and Mob:BID, previously tested (Gonzalez-Prieto *et al*., 2017), with the promiscuous relaxase of the plasmid RSF1010, MobA.

RSF1010_MobA belongs to a family of plasmids which hijacks the T4SS of co-residing conjugative plasmids, so it can be translocated through various T4SS. In

addition, it is known that it can be translocated through the T4SS of *A. tumefaciens* (Vergunst *et al*., 2005). This relaxase also has a primase domain, which is required for plasmid replication (Henderson and Meyer, 1996). MobA is also able to catalyze *oriT*–*oriT* recombination on single-stranded substrates but not on supercoiled plasmid substrates (Meyer, 1989).

4.1.3.1. Plasmids construction

The mobilizable plasmids used by (Gonzalez-Prieto *et al*., 2017) contained the R388 MOB region, an eukaryotic *gfp* cassette, and a neomycin phosphotransferase expression cassette in order to be able to select for stable chromosomal integration events. We constructed a derivative of plasmid RSF1010K, which encodes a kanamycin resistance gene, carrying and a eukaryotic *gfp* cassette (plasmid pLG04). For the negative control, we generated the same plasmids without the MobA relaxase (pLG03). In order to avoid recombination problems between the neomycin cassette and the kanamycin cassette, a hygromycin cassette was inserted instead of the neomycin cassette. R388 derivates carrying the hygromycin cassette were also constructed (pLG05 and pLG06) (Table 3, Section 3.2.). The plasmid pRS130 carrying the *gfp* cassette, the neomycin resistant cassette and the Mob:BID relaxase of the cryptic plasmid of *B. henselae* pBGR1 (Schröder *et al*., 2011) was used as a control in the integration experiments.

4.1.3.2. DNA mobilization through the T4SS of *B. henselae*

The different plasmids were electroporated into *B. henselae* wild type bacteria. Infections were carried out as described in Section 3.5.3.. Firstly, we wanted to determine if the relaxase MobA could be recognized by the T4SS of *B. henselae*. We performed DNA-transfer assays with TrwC and MobA relaxases in parallel using as recipient cell the human cell lines EAhy.926 and HeLa cells, previously demonstrated to be infected by *B. henselae* (Gonzalez-Prieto *et al*., 2017). After three days post infection, we measured DNA-transfer by flow cytometry to detect GFP levels. **Figure 29** shows representative plots of the assays.

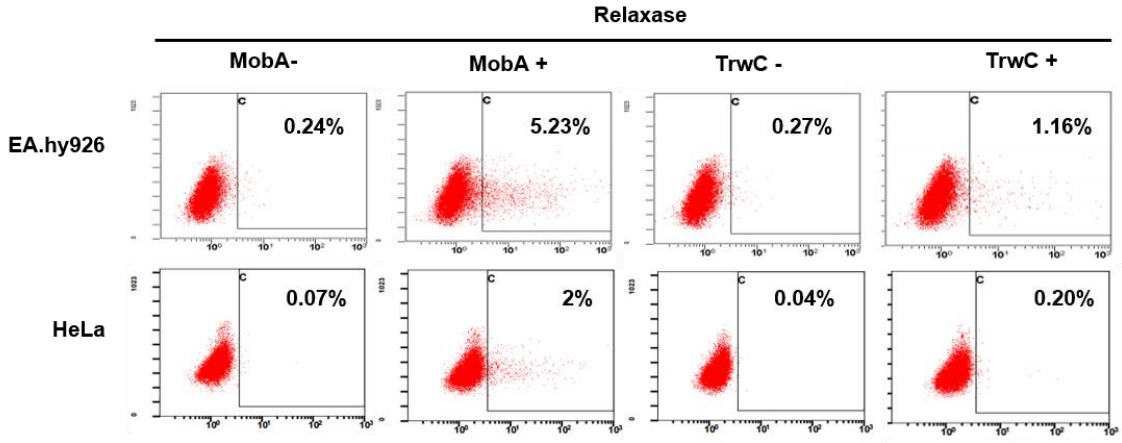


Figure 29. **Expression of the DNA transferred by MobA or TrwC relaxases**. Representative plots (cell granularity versus GFP intensity). The square marks the population considered as positive. Percentage of GFP positive cells is indicated inside each square. The relaxase present/absent in each experiment is indicated on top of the panel.

Table 9 and Figure 30 show the compilation of the results of DNA transfer experiments. As expected, TrwC could mobilize DNA from *B. henselae* to EA.hy and HeLa cells. This transfer happened at a higher frequency in EA.hy cells. When MobA was used to mobilize DNA, GFP positive cells were also detected, meaning that MobA could be also recognized by *B. henselae* T4SS and translocated to the recipient cell. Just as happens with TrwC, DNA transfer frequencies with MobA were higher when EA.hy cells were used. Therefore, we detected GFP positive cells with both relaxases. It is noteful that DNA transfer rates were notably higher when we used MobA compared to TrwC.

Table 9. Rates of DNA transfer to human cells through the VirB/D4 f *B. henselae*

Transfer system	Relaxase	Infected cells	(GFP^+)%
RSF1010	MobA	EA.hy926	5,72 ± 1,37
RSF1010	----	EA.hy926	0,29 ± 0,07
R388	TrwC	EA.hy926	1,00 ± 0,09
R388	----	EA.hy926	0,14 ± 0,19
RSF1010	MobA	HeLa	2,00 ± 1,48
RSF1010	----	HeLa	0,07 ± 0,05
R388	TrwC	HeLa	0,20 ± 0,03
R388	----	HeLa	0,04 ± 0,06

Data from flow cytometry (right column) show the percentage of GFP positive cells (mean ± SD from two and four independent experiments).---- means that no relaxase was present in the experiment.

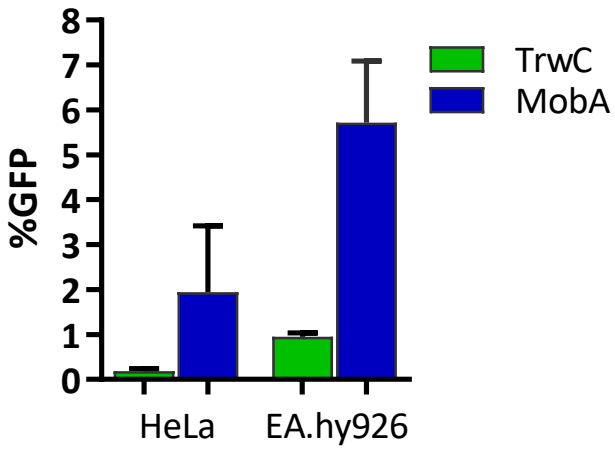


Figure 30. Efficiency of DNA transfer to human cells. DNA transfer was measured as GFP positive cells by flow cytometry. Data of both relaxases (TrwC in green and MobA in blue) mobilization are shown in the graphic. The recipient human cell line is indicated at the bottom. Means from two to four independent experiments are shown.

4.1.3.3. DNA integration activity of translocated relaxases

After demonstrating that both relaxases were recognized and translocated through VirB/D4 of *B. henselae*, we wanted to determine the integration ability of MobA. Previously, Gonzalez-Prieto and collaborators showed that TrwC promoted the integration of the transferred DNA into eukaryotic cells. They also demonstrated that Mob:BID integration rate was similar to the integration rate obtained after transfecting cells with DNA, which means that Mob:BID did not promote integration (Gonzalez-Prieto *et al*., 2017).

In order to determine the integration ability of MobA relaxase, the assays were performed using HeLa cells, as EA.hy cells were previously shown to have a low viability in the integration assay (Gonzalez-Prieto *et al*., 2017). HeLa cells were infected with *B. henselae* carrying the plasmids pLG04 (*mobA*), pLG05 (*trwC*) or pRS130 (*mob:BID*) and the negative control pLG03 (without *mobA*) and pLG06 (without *trwC*). We performed infections as previously explained and after 3 days the GFP levels were determined by flow cytometry. Only when a relaxase was present, GFP expression was detected. After confirming that DNA transfer and expression of the mobilizable plasmid had taken place, antibiotic treatment with hygromycin o neomycin (only in the case of Mob:BID) was carried out for 4-5 weeks to select for stable integration events of the transferred

plasmids into the genome of HeLa cells. After selection, we calculated the number of hygromycin or neomycin-resistant cell colonies by counting using an optical microscope. **Table 10** summarizes the data obtained for the integration assays.

Table 10. Transient and permanent expression of DNA transferred to HeLa cells.

	Relaxase in mobilizable plasmid			
	None	**TrwC (pLG04)**	**MobA (pLG05)**	**Mob:BID (pRS130)**
# Cells	8.80×10^6	8.80×10^6	8.80×10^6	8.80×10^6
GFP$^+$	0	$2.62 \times 10^4 \pm 1.47 \times 10^4$	$4.97 \times 10^4 \pm 3{,}11 \times 10^3$	$9.68 \times 10^3 \pm 8.80 \times 10^2$
%GFP$^+$	0	0.30±0.17	0.57± 0.04	0.11±0.01
[1]Ab^R	0	$2.52 \times 10^2 \pm 1.32 \times 10^1$	$1.17 \times 10^2 \pm 2.19 \times 10^0$	$2.32 \times 10^0 \pm 4.02 \times 10^0$
Ab^R/cells	$<5 \times 10^{-8}$	$2.86 \times 10^{-5} \pm 1.50 \times 10^{-6}$	$1.33 \times 10^{-5} \pm 2.49 \times 10^{-7}$	$2.64 \times 10^{-7} \pm 4.57 \times 10^{-7}$
Ab^R/ GFP$^+$	-	$1.23 \times 10^{-2} \pm 7.12 \times 10^{-3}$	$2.36 \times 10^{-3} \pm 1.03 \times 10^{-4}$	$2.20 \times 10^{-4} \pm 3.81 \times 10^{-4}$

[1]Ab^R: Antibiotic (Hygromycin for TrwC and MobA assays or neomycin for Mob:BID assays).
Data are the mean of 2 independent experiments.

The graphic in **Figure 31** represents the integration ratio calculated by dividing the hygromycin or neomycin resistant cells by the GFP positive cells obtained for each relaxase. The results indicated that the integration rate when MobA mobilized the plasmid was approximately one log higher than in the case of Mob:BID. This data indicated that MobA promoted the integration in the recipient cell. The data also confirmed that, although MobA had a higher efficiency transferring the DNA, TrwC showed a stronger effect in promoting the integration of the DNA (five-fold higher approximately).

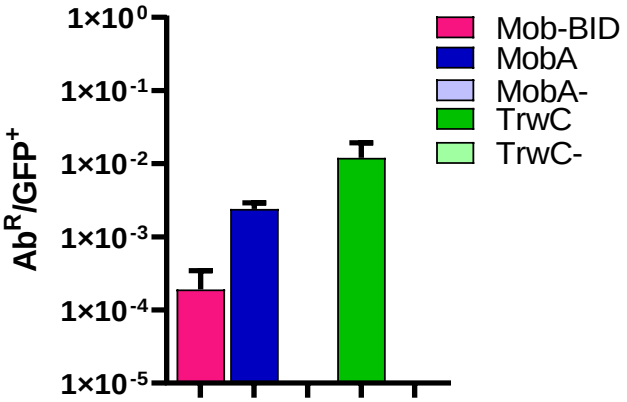


Figure 31. DNA integration rate for different relaxases. The graph shows the ratio between the number of HeLa cells resistant to hygromycin or neomycin and the GFP positive HeLa cells. Each relaxase is shown with a different color. The data are the result of two independent experiments. Ab^R: hygromycin when TrwC and MobA plasmids were assayed. Neomycin when Mob:BID plasmid was assayed.

4.2. Relaxases as protein/DNA delivery systems for biotechnological purposes

The use of TrwC with a biotechnological purpose has been previously proposed (Agúndez *et al*., 2012; Llosa *et al*., 2012b; Gonzalez-Prieto *et al*., 2013, 2017). **Figure 32** summarizes the principal characteristics of the relaxase that make it a good alternative as a genetic modification tool.

TrwC can be translocated into a recipient bacterial or human cell (Fernández-González *et al*., 2011; Llosa *et al*., 2012b). The relaxase can be translocated covalently bound to a ssDNA molecule recognized by the *oriT* sequence (1). Although TrwC has shown to localize in the cytoplasm of human cells, TrwC can also enter the nucleus (Agúndez *et al*., 2011) (2). It was shown that human cells harbor in their genomes natural targets that could be recognized by TrwC (Agúndez *et al*., 2012), however, TrwC does not promote SSI in human cells (3). Despite this, it has been proved that TrwC promotes non-specific integration in human cells (Gonzalez-Prieto *et al*., 2017). Thus, TrwC can deliver *in vivo* to human cells any DNA molecule and promote its integration but lacks the site-specificity required for gene targeting.

In this work, we proposed the use of TrwC as a vehicle to deliver site-specific nucleases which would promote the SSI reaction in the genome. The use of CRISPR-Cas systems as genomic modification tools has been expanded in the last years. However, some limitations must be solved in order to improve their use (see Section 1.3.2.1.). Here we studied the possibility of combining them with relaxases as an *in vivo* delivery method, bypassing the need to express the Cas nuclease from the target cell and thus minimizing off target or toxicity problems of CRISPR-Cas systems.

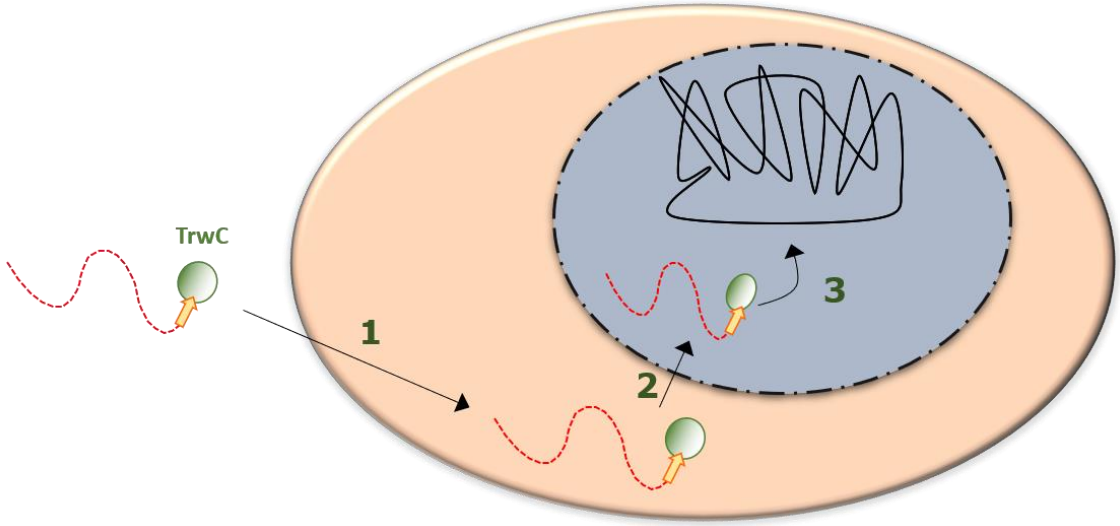


Figure 32. Schematic representation of TrwC characteristics as a genomic modification tool. See text for details.

4.2.1. Generation and validation of a TrwC-Cas12a construction in prokaryotic cells

For the generation of a Relaxase-Cas fusion, we have chosen the CRISPR-Cas endonuclease Cas12a (see Section 1.3.2.1.). Specifically, we selected the variant AsCas12a, whose activity has been described in eukaryotes and prokaryotes (Zetsche *et al.*, 2015). We fused AsCas12a to the C-terminus of TrwC, generating a fusion protein TrwC-Cas12a. We used a humanized *cas12a* sequence, obtained from plasmid pY010 (Addgene), encoding a C-terminal NLS (Nuclear Localization Signal) and a 3xHA tag.

We designed the construction in a prokaryotic expression vector. It is important to mention that this construction did not harbour an *oriT* sequence, meaning that it was a non-mobilizable plasmid. Therefore, our aim was to translocate during conjugation the fusion protein, and not the plasmid containing the sequence of the fusion protein, avoiding its expression in the recipient cell. The election of the promoter had to be fine-

tuned. First, we constructed the fusion under the control of the native P*trwA* promoter from the plasmid R388. To test Cas12a activity, we measured its lethality in bacteria when cotransformed with a gRNA targeting a chromosomal gene, since the introduced DSB are lethal. Bacteria do not usually harbour NHEJ repair pathways, which means that effective DSB cuts in their chromosome without an homologous template lead to bacterial death (Cui and Bikard, 2016). Thus, we could measure Cas12a activity by the decrease in the number of viable cells in the target bacterial population. The *E. coli* strain D1210 was co-electroporated with a plasmid codifying a gRNA targeting the *lacZ* gene (pLG15), and the plasmid encoding the nuclease: pBBR6 vector as a negative control, pAA12 (*ptrwA::trwABC*), pLG14 (P*lac::cas12a*), or pLG01(*ptrwA::trwABC-cas12a)* as the test. After co-electroporations, we grew transformations on LB plates supplemented with IPTG, as the pLG15 carried the $lacZ_{gRNA}$ under a P*lac* promoter and with appropriated antibiotics (**Figure 33**). When the plasmid codifying for Cas12a was coelectroporated with pLG15, we observed a decreased in the number of transformants. However, no decrease was observed with the plasmid encoding TrwC-Cas12a. Therefore, we were not able to detect Cas12a activity within the fusion using the pLG01 plasmid. This could be due to the lack of functionality of Cas12a in the fusion protein, or to a low level of expression from the *ptrwA* promoter.

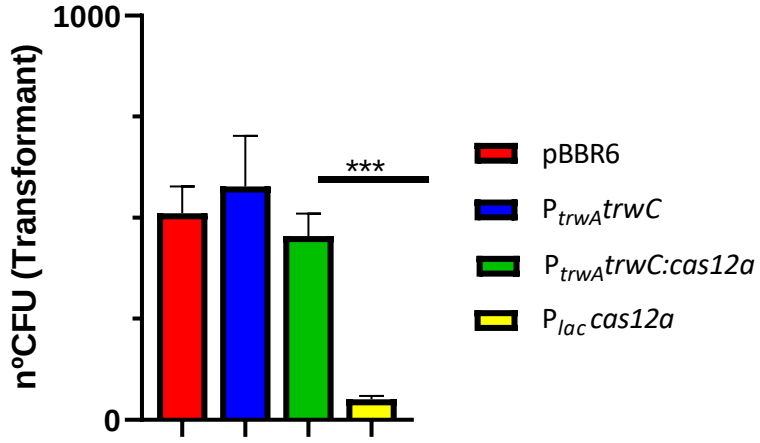


Figure 33. TrwC-Cas12a nuclease activity in prokaryotic cells. The graphic represents the numbers of CFU transformants. D1210 cells were co-electroporated with plasmids codifying different proteins (color-coded) and with the pLG15 plasmid carrying the $lacZ_{gRNA}$. Bacteria were plated under induction condition (IPTG) as the $lacZ_{gRNA}$ is under the control of a lactose promoter. The legend shows the different promoter for each construction: pBBR6 (empty plasmid), $p_{trwA}trwC$(pAA12), $p_{trwA}trwC$-*cas12a* (pLG01), *Plac::cas12a* (pLG14) Data of three independent experiments are represented. P< 0.05.

Then, we tried to generate a construct under the lactose promoter; however, the construction was not obtained, suggesting it could be lethal for the bacteria. This could be due to toxicity of the fusion protein produced from background expression of the lactose promoter. Finally, we chose the P*tet* promoter. This promoter allowed the overexpression of the protein with a tight control of the expression in induction conditions with aTc. The stability of the fusion protein was tested by a western blot using an anti-TrwC antibody.

The western blot results are shown in the **Figure 34**. The fusion protein had an expected molecular weight of 263 kDa. The blot showed a band with this size. However, the western blot also revealed degradation products of the fusion protein, indicating that it was not very stable. The main degradation product corresponded with the size of TrwC. In all, this result showed that despite the big size of the construction and its partial degradation, we were able to obtain the full-size fusion protein (dotted arrow in Figure 34).

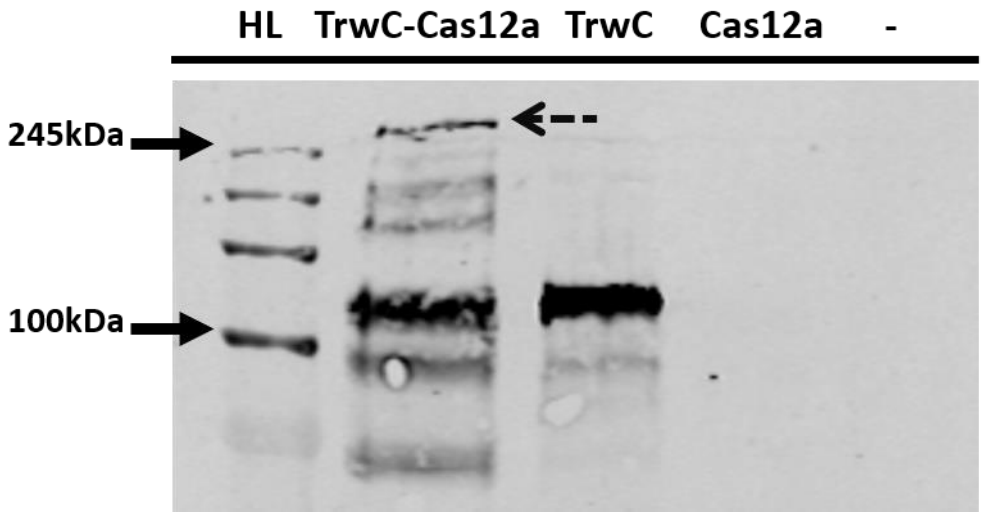


Figure 34. Stability of TrwC-Cas12a protein. A Western Blot was performed using an anti-TrwC antibody. D1210 cell lysate was used as negative control (-). TrwC, Cas12a and TrwC-Cas12a correspond with D1210 lysates carrying pLG22, pLG14 and pLG24 respectively, after 3 hours of induction with IPTG (for TrwC and Cas12a) or aTc (for TrwC-Cas12a). The expected sizes for the proteins are: TrwC 108kDa, Cas12a 187 kDa, TrwC-Cas12a 263 kDa. Left lane, molecular weight marker. Black arrows indicate 245 kDa (top) and 100 kDa (bottom) bands. The dotted arrow indicates full-size TrwC-Cas12a protein.

4.2.1.1. Site-specific endonuclease activity of TrwC-Cas12a

Once we obtained the fusion protein, we evaluated its activity. Each element of the fusion protein was tested to determine if they maintained their activity within the protein. Cas12a activity was first tested.

As we mentioned before Cas12a activity could be measured by the decreased in the number of viable cells in the target bacterial population. To test the lethality of TrwC-Cas12a produced under the control of P*tet* promoter, we coelectroporated pLG24 (P*tet*::*trwC-cas12a*) with plasmids expressing gRNA under the P*lac* promoter (inducible by IPTG), targeting two different genes: *lacZ (*pLG15) and *sacB* (pLG19). As host cell we used *E. coli* D1210 strain, which contains in its chromosome the gene *lacZ* but not *sacB*. We selected the transformants on plates containing IPTG (for gRNA expression) and with or without aTc (induction and non-induction conditions respectively for TrwC-Cas12a). The results are shown in the **Figure 35**. In all the cases we obtained transformants with good efficiency, except when pLG15 was present and the transformation was plated under induction conditions (aTc), i.e. when $lacZ_{gRNA}$ was co-expressed with *trwC-cas12a*. In this case, no transformants were obtained (Figure 35a). Figure 35b summarizes the number of transformants obtained in all the assays.

These data confirmed that TrwC-Cas12a could target specific genomic sequences when an appropriate guide RNA was present in the cell. Also, the data demonstrated that the expression of the protein under the control of P*tet* promoter could be well regulated by the aTc inductor.

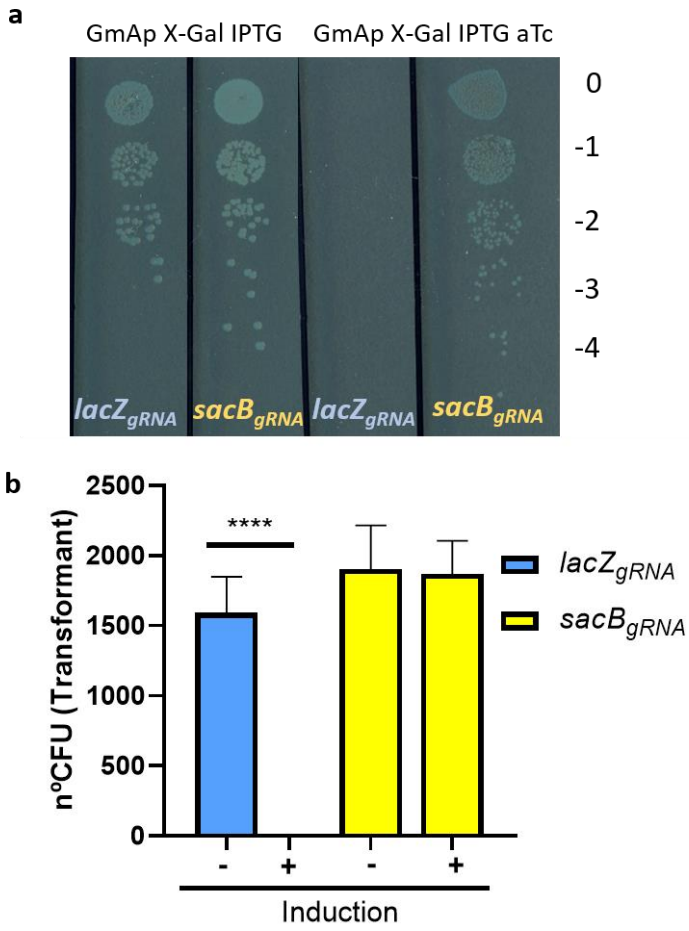


Figure 35. TrwC-Cas12a nuclease activity in bacterial cells. a) D1210 coelectroporated with gRNAs targeting either *lacZ* or *sacB* were plated under non-induction or induction (aTc) conditions for the expression of *trwC-cas12a*. 10 µl of each dilution (0 to -4) were plated. **b)** Number of transformants after coelectroporation of E coli D1210 with pLG24 (P*tet*::*trwC-cas12a*) and the indicated gRNAs, under conditions of induction (+) or non-induction (-) of *trwC-cas12a* expression with aTc. Data correspond with the mean of at least 3 independent assays. ****, P<0.0001.

4.2.1.2. Translocation of TrwC-Cas12a by the T4SS of the conjugative plasmid R388

The next step was to determine if TrwC was active as a relaxase in the fusion protein, and if the fusion was recognized and translocated through the R388 T4SS. It is known that the relaxase activity is needed to finish the conjugation process in the recipient cell (Garcillán-Barcia *et al*., 2007). Therefore, the detection of transconjugant cells after performing a mating, implied that the TrwC moiety of TrwC-Cas12a was active in the recipient cell.

We performed mating assays using as donor cell D1210 carrying pSU1445 (a R388 derivate with no *trwC*). We complemented this plasmid using pLG22 (P*lac*::*trwC)*

or pLG24 (*Ptet::trwC-cas12a*). As negative control, we used pSU1445 without plasmid complementation. Conjugation frequencies obtained were very similar in both cases and no significant differences were detected (**Table 11**).

Table 11. Complementation assays with TrwC or TrwC-Cas12a

Mobilizable plasmid	Complementing plasmid	Relaxase	Conjugation frequency[1]
pSU1445 (R388 *trwC*-)	pLG22	TrwC	$5.48x10^{-1}$ ±$2.7x10^{-1}$
	pLG24	TrwC-Cas12a	$4.1\ x10^{-1}$ ±$2.6\ x10^{-1}$
	none	none	$<1.8x10^{-8}$

Donor and recipient *E. coli* strains were D1210 and DH5α T1-resistant, respectively. Data represent the mean +/- SD of 4 independent assays. [1] Transconjugants per donor.

These data showed that TrwC was active within the fusion protein and that it could be translocated through the T4SS. However, due to the degradation of the fusion protein observed in the western blot (Figure 34), this result was not conclusive of translocation of TrwC-Cas12a, since the relaxase activity in the recipient could derive from translocation of partially degraded fusion proteins. Thus, Cas12a activity in the recipient cell must be detected in order to confirm that the fusion protein was fully-translocated to the recipient cell.

4.2.2. Validation of Cas12 activity after translocation through the T4SS

Detecting Cas12a activity in the recipient cell was the next step to confirm that the system could be used *in vivo* for endonuclease delivery. Detecting the activity of the CRISPR endonuclease in the recipient would imply the translocation of the whole fusion protein through the T4SS. Also, it would mean that Cas12a recovered its activity after being transported during the conjugation process. It is known that TrwC needs to be unfolded for its transport through the T4SS (Trokter and Waksman, 2018). It was unclear what could happen with Cas12a and if it would recover its activity after the transport.

We decided to detect TrwC-Cas12a activity in the recipient cell using two different strategies: detection of an increase of the SOS signal as a consequence of the generation of DSBs (indirect detection of Cas12a activity), and detection of mutations in

the gene targeted by the gRNA (direct detection of Cas12a activity). Both strategies are represented in **Figure 36**.

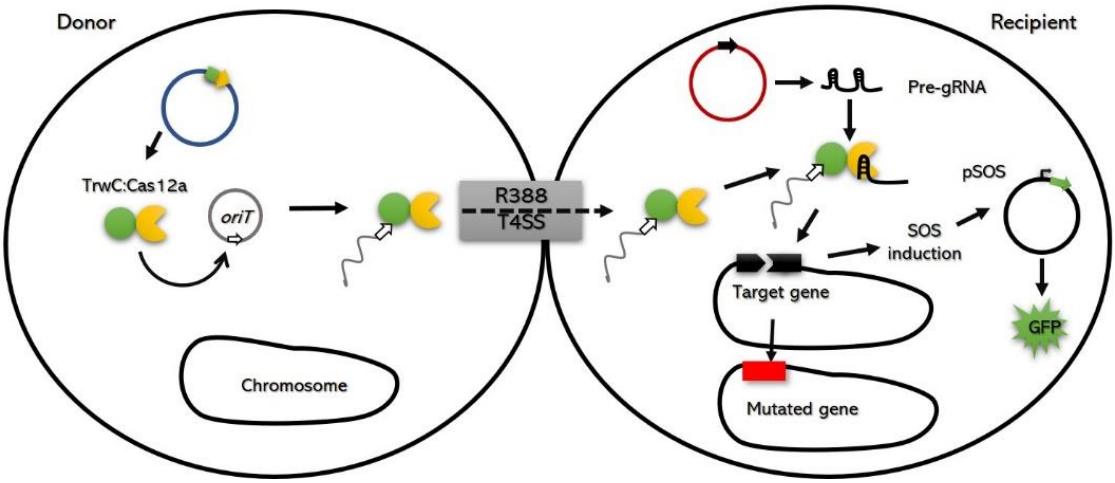


Figure 36. Schematic representation of Cas12a activity assays in the recipient cell. In the donor cell, pLG24 (in blue) will express TrwC-Cas12a. Thanks to its relaxase activity, the fusion protein will cleave and bind covalently to the *oriT* (white arrow), and the complex will be recruited and translocated through the T4SS into the recipient cell. In the recipient, pLG15 or pLG19 (plasmid in red) will produce a gRNA targeting a gene on the chromosome. Due to its site-specific endonuclease activity, the incoming TrwC-Cas12a will process the gRNA generating a complex, which will be guided to the target gene, where it will produce a DSB. This cleavage will activate the SOS signal, which will induce the SOS promoter on the pZA31-sulA-GFP plasmid (in green), thus producing GFP because of the TrwC-Cas12a cleavage activity. The DSB will be repaired by bacterial host pathways, producing mutations of the target gene (in red).

4.2.2.1. SOS detection assay

SOS response is the result of DNA damage and replication in some bacteria such as *E. coli*. It is a bacterial stress response which is regulated by a complex pathway involving different proteins such as LexA or RecA. Under normal conditions, LexA is repressing the genes involved in the SOS response. However, when an abnormal rate of ssDNA is present in the cell, the SOS response is induced. This ssDNA is the substrate of RecA, needed for its polymerization. RecA forms a complex with the ssDNA and these complexes induces the self-cleavage of LexA, leading to the expression of SOS response genes (**Figure 37**) (Walker, 1996; Simmons *et al*., 2008).

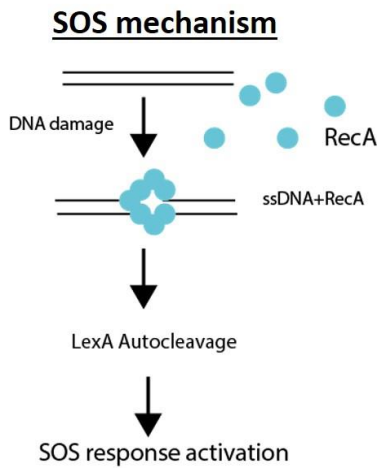


Figure 37. SOS response mechanism. DNA damage induces the formation of RecA+ssDNA complex which produce the autocleavage of the SOS response repressor LexA, inducing the SOS response activation.

There are different mechanisms that induce this response. Cui and Bikard showed that DSBs produced by Cas9 in the bacterial chromosome induce this signal (Cui and Bikard, 2016). They generated the reporter plasmid pZA31-sulA-GFP. This plasmid encodes the green fluorescent protein gene (*gfp*) under the control of a SOS-inducible promoter (*PsulA*). Using this reporter plasmid, they showed an increase in the GFP levels when Cas9 produced DSBs in bacteria chromosome. Although there are no reports on the induction of a SOS response by Cas12a, this endonuclease, as Cas9, also produces DSB on the target DNA.

We have adapted this assay to detect Cas12a activity in the recipient cell after being translocated during conjugation. This was done in collaboration with Dr David Bikard (Institute Pasteur). D1210 cells harbouring the *trwC* deficient R388 derivative pSU1445 complemented with plasmid pLG24, which produces TrwC-Cas12a, were used as donor cells. As recipient, we used the RecA proficient strain MG1655, harbouring the reporter plasmid pZA31-*sulA:gfp* (where the SOS-inducible *sulA* promoter drives the expression of *gfp*) and a second plasmid responsible for the IPTG-inducible expression of the gRNA. If the translocated TrwC-Cas12a shows Cas12a activity once in the recipient cell, it would activate the SOS response and thus produce an increase of GFP levels. We used two recipients in parallel, one with a gRNA targeting a chromosomal gene (pLG15; $lacZ_{gRNA}$) or other with a gRNA targeting a gene not present in the chromosome (pLG19;

sacB$_{gRNA}$). After the matings, the level of GFP was visualized directly on the mating plate (**Figure 38a**).

Despite a GFP background was detected in all the cases (probably due to the conjugation induction of the SOS response; (Baharoglu *et al*., 2010)), we have detected a significant increase in fluorescence when TrwC-Cas12a was translocated into the recipient expressing the gRNA against the *lacZ* gene. The GFP levels were later measured and expressed relative to the bacterial cell density (calculated with the OD_{600}) (Figure 38b). The graphic confirms a significant increase in GFP levels when the recipient cell harbored a gRNA targeting a chromosomal gene (*lacZ*$_{gRNA}$). With these data we showed that TrwC-Cas12a was being translocated through the T4SS into the recipient cell, and that the Cas12a moiety was active after being transported, being able to introduce DSB when a gRNA guides it to a target DNA. In addition, we showed that DSBs produced by Cas12a cleavage induce the SOS signal.

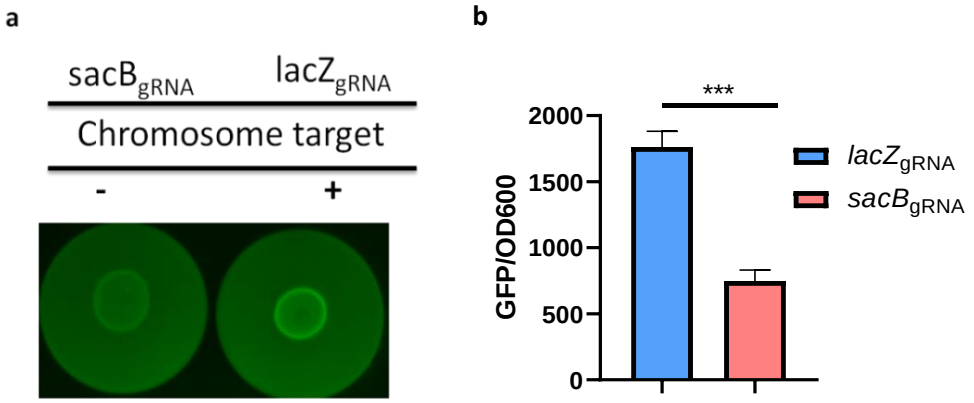


Figure 38. Detection of Cas12a cleavage activity in the recipient cell by SOS response induction. a) GFP detection in conjugation filters after 3h of mating under induction conditions. Filter on the right (gRNA with a target on the chromosome) shows increased fluorescence level in comparison with the filter on the left (gRNA without target on the chromosome). **b)** The graphic shows the GFP fluorescence relative to OD_{600} levels of the conjugation mix after 3 hours of mating, measured with a TECAN Infinite M200 Pro. Data correspond to the mean of 3 independent assays (***, P<0.0005).

4.2.2.2. Detection of mutations in the region targeted by the gRNA

Next, we wanted to detect mutations caused by the cleavage of Cas12a in the target sequence. Usually, bacteria lack the presence of NHEJ pathways. However, it has been reported that cell death is not the only possible outcome of efficient CRISPR-Cas

chromosome cleavage when no homologous template is provided. Cui and Bikard reported that recombination through distal homologous sequences or between micro-homologies could occur after Cas cleavage in a RecA dependent manner, allowing bacteria to survive (Cui and Bikard, 2016). In order to test this, we decided to detect Cas12a activity guided to the *sacB* gene in a *sacB*-containing strain, selecting transconjugants resistant to sucrose. The expression of this gene in the presence of sucrose is lethal in bacteria (Reyrat *et al*., 1998), so transconjugants containing *sacB* and plated in sucrose would only survive if there was a mutation inactivating the gene. Therefore, Cas12a activity targeting this gene would increase the *sacB* mutant rate.

Matings were performed using donor D1210 cells harbouring the *trwC* deficient R388 derivative pSU1445, complemented with either plasmid pLG24, which produces TrwC-Cas12a, or pLG22, which produces TrwC. MG1655 and FD3 strains (kindly provided by David Bikard (Institut Pasteur, France)) were used as recipient cells (Table 1, Section 3.1). FD3 is the MG1655 strain with a *sacB* gene copy integrated in the chromosome. This strain was constructed using an integrative vector based on pOSIP-KL (St-Pierre *et al*., 2013) which carries the *yhhX* target sequence on the non-template strand between a constitutive promoter and the *mCherry* reporter gene followed by a *sacB* counter-selection marker. The plasmid was integrated at the lambda *attB* site in the chromosome of *E. coli* MG1655 and the backbone was flipped out using the pE-FLP plasmid (St-Pierre *et al*., 2013). Both recipients carried the pLG19 plasmid, which codified $sacB_{gRNA.}$ We performed matings in induction conditions (Section 3.4.3.). Transconjugants were plated under standard selection conditions to obtain conjugation frequencies, and also supplemented with 1% sucrose to detect sucrose mutants. Our hypothesis was that if Cas12a cleaved the *sacB* gene, we would be able to detect sucrose mutants in the presence of the fusion protein, while less or non-sucrose mutants should be detected in its absence. Mating results are shown in **Table 12**. Conjugation frequencies were similar between TrwC and TrwC-Cas12a. As expected, transconjugants from the matings using as recipient the strain MG1655 (without *sacB*) grew in selective media supplemented with 1% sucrose (Table 12, $Sucrose^R$ transcs. column). Also, the results obtained showed the appearance of several sucrose-resistant MG1655::*sacB* transconjugants when TrwC-

Cas12a was translocated into the recipient cell, while a parallel mating using TrwC as the relaxase rendered no surviving cells (Sucrose^R transcs. column).

Table 12. Detection of *sacB* mutants in recipient cells.

Relaxase (donor)[1]	Recipient[2]	Conjugation frequency[3]	Sucrose^R transcs. frequency[4]	Sucrose^R transcs.[5]
TrwC	MG1655	$7x10^{-1}$	Nd[6]	lawn
	MG1655::*sacB*	$1.23x10^{-3}$	$<8.92x10^{-7}$	0
TrwC-Cas12a	MG1655	$1.3x10^{-3}$	Nd	lawn
	MG1655::*sacB*	$1.27x10^{-1}$	$5.82x10^{-6}$	11

[1]Donor cells were D1210 harboring pSU1445 (R388 without *trwC*) complemented with pLG22 (P*lac*::*trwC*) or pLG24 (*Ptet*::*trwC-cas12a*)
[2] Both strains carried the plasmids pZA31-*sulA::gfp* and pLG19 (P*lac*:*sacB*$_{gRNA}$)
[3] Expressed as transconjugants per donor
[4] Expressed as sucrose-resistant transconjugants per donor
[5] Number of sucrose-resistant transconjugants. The mating mix was directly plated (dilution 0) on LB agar with antibiotics to select for transconjugants, supplemented with 1 % sucrose.
[6] Not determined. The data represent the results of one experiment.

The 11 sucrose resistant transconjugants obtained were analyzed by PCR. The oligonucleotides sacB_F (5'-CTACCGCACTGCTGGCAG-3') and sacB_R (5'-GATGCTGTCTTTGACAACAG-3') were used to amplify the 5' region of *sacB* which contained the gRNA target sequence. The amplicon contained 1343 bp of *sacB*, starting 204 bp upstream the PAM. 5 of the 11 mutants did not amplify any sequence (**Figure 39**). This could be due to big deletions involving the amplified region. In the remaining 6 colonies, all the amplicons had the expected size (Figure 39).

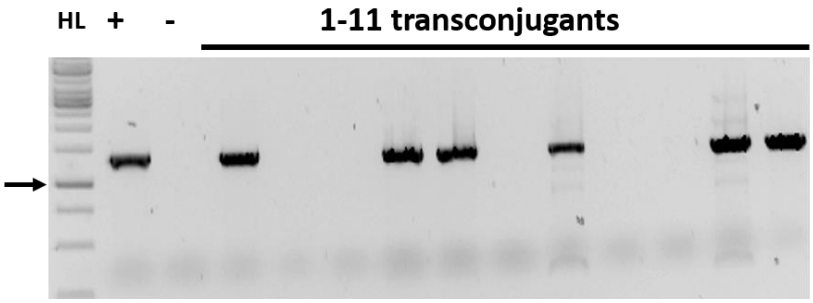


Figure 39. **PCR amplification of the *sacB* targeted region of the 11 sucrose resistant transconjugants.** HL: Hyperladder +: PCR amplifications of the MG155*::sacB* recipient strain. -: PCR negative control (no template DNA). 1-11: transconjugants resistant to sucrose. The arrow indicates 1 kb fragment.

We determined the sequence of the amplicons to detect any mutations produced by Cas12a cleavage (**Figure 40a**). One of these transconjugants (TC4) showed no mutation in the target sequence, so it could probably be a spontaneous *sacB* mutant unrelated to Cas12a cleavage. The remaining 5 transconjugants showed mutations very close to the Cas12a cleavage site. These mutations showed deletions of 1 to 4 nt in the gRNA target DNA sequence.

Figure 40. Analysis of the *sacB* mutations in sucrose-resistant colonies. a. Alignment of the *sacB* region close to the PAM and target sites for Cas12a-gRNA. The *sacB* region was PCR-amplified from sucrose-resistant transconjugants (TC1 – TC6) and from a sucrose-resistant transformant (TF1). The *sacB* sequence in strain MG1655::*sacB* was also determined and is shown at the top for comparison (WT). The PAM sequence and the spacer sequence are shown at the top. The red triangle marks Cas12a cleavage site in the shown DNA strand. Nucleotides in red mark the site where the DNA sequence splits into two (see text). **b**) Amino acid sequences of SacB variants resulting from the different mutations. The deletions are indicated at the left, in blue. The amino acids shown in green and underlined are encoded by the PAM sequence. Stop codons are shown as *.

In order to confirm that these deletions were produced solely by the Cas12a cleavage activity of the fusion protein, and that TrwC activity was not interfering in the cleavage pattern, we also sequenced the *sacB* amplicon of a sucrose resistant mutant obtained by coeletroporating pLG14 (P*lac:cas12a*) and pLG19 ($sacB_{gRNA}$) into MG1655::*sacB* strain (TF1 line in Figure 40). The sequence obtained was similar to the ones obtained for the transconjugants resistant to sucrose (3 nt deletion).

The amino acid sequence of the mutants was also checked to confirm that these small mutations in the Cas12a target sequence produced a null SacB variant. All the deletions produced an early stop codon on the *sacB* ORF (Figure 40b). These results provided direct proof of Cas12a RNA-guided cleavage activity in the recipient cell after translocation of TrwC-Cas12a through the T4SS.

Intriguingly, when we looked at the chromatograms, in 3 of the 6 transconjugants analysed, a mixture of sequences appeared after the Cas12a cleavage site (**Figure 41**), which corresponded to a mixture of two different sequences carrying two different deletions on the *sacB* sequence (showed in red in the Figure 40a).

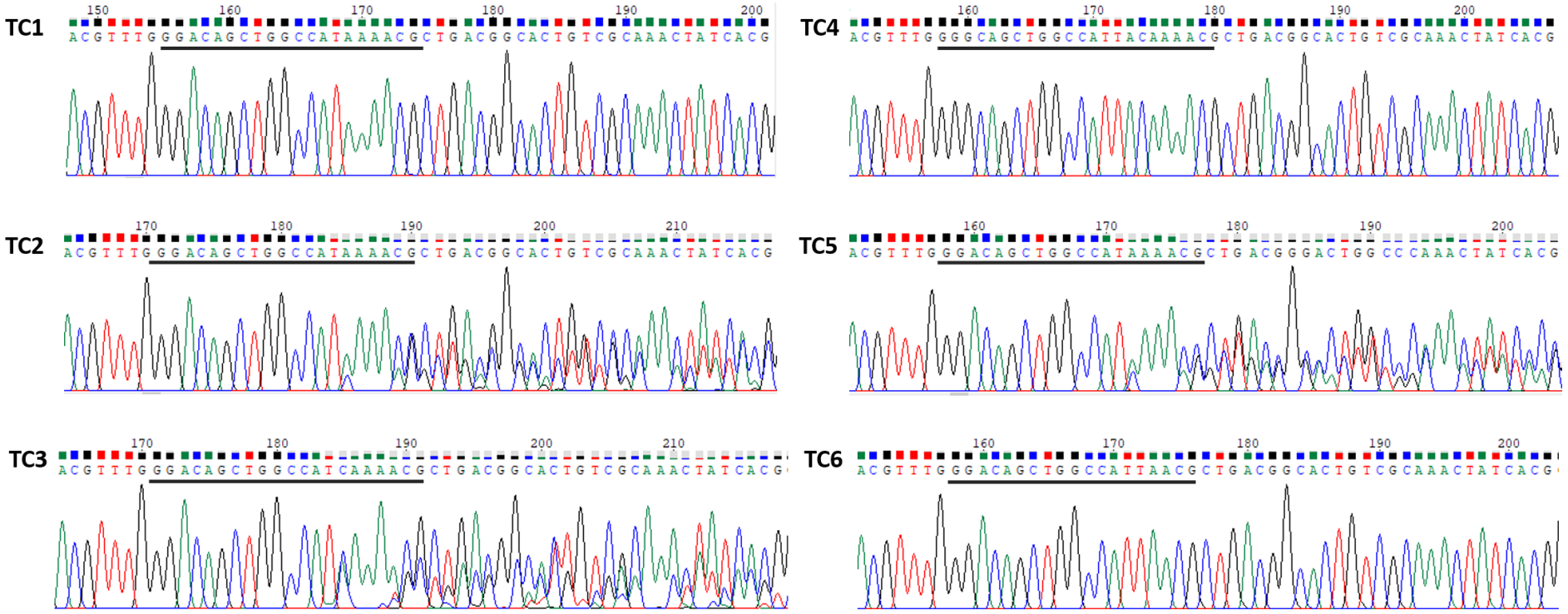


Figure 41. Chromatograms of the sacB amplified region from the transconjugants resistant to sucrose. The sequence targeted by the gRNA is underline in black. A mixture of sequences was detected in the same region for three *sacB* sequences (TC2, TC3 and TC5). (TC1-TC6; transconjugants 1-6)

4.2.3. Incorporation of site-specific mutations using a homologous recombination cassette

One of the advantages of CRISPR-Cas editing system is the possibility of introducing a homologous template carrying a mutation of the targe gene to specifically edit a cell without leaving a scar. The use of relaxases as a vehicle to deliver CRISPR-Cas system into the recipient cell, theoretically allows the co-delivery of the editing template DNA. Therefore, we wanted to determine if TrwC-Cas12a could be used to edit the recipient cell by translocating a homologous recombination cassette covalently bound to the fusion protein.

For this purpose, we constructed plasmid pLG27. This plasmid is based on the mobilizable suicide plasmid pSW27 (Demarre et al., 2005a) which contains an $oriT_{w,}$ a chloramphenicol resistance cassette and a R6K *oriV* (which needs the presence of the Pir protein for its replication). We have inserted a *sacB* homologous recombination cassette under the control of a Tac promoter (**Figure 42**).

This homologous cassette contained a 430 bp sequence of the 5'region of the *sacB* gene from the MG1655::*sacB* strain. This region includes the sequence targeted by the gRNA coded by the pLG19 plasmid. The PAM sequence in the homologous recombination cassette was mutated to generate an early STOP codon in the gene. Consequently, incorporation of the mutation also abolished the PAM sequence, preventing other Cas12a-mediated cleavage events. The site of the mutation was surrounded by a left homology arm of 277 bp and a right homology arm of 150 bp.

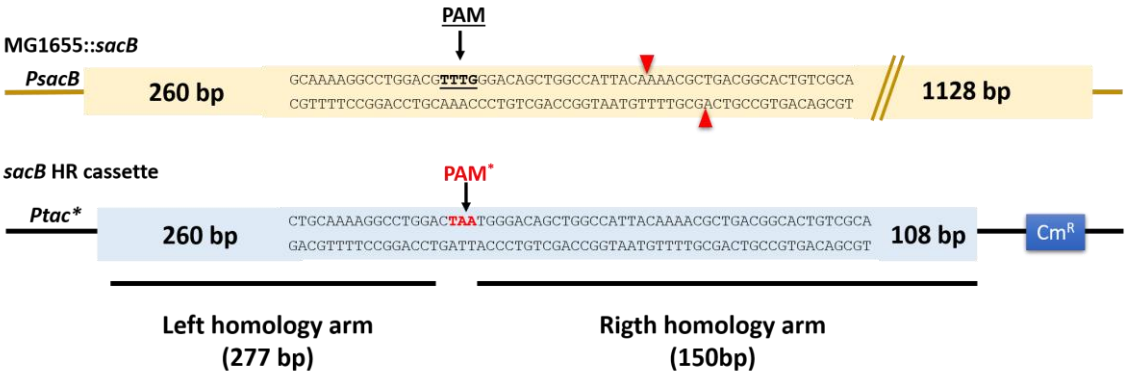


Figure 42. Design of the *sacB* homologous recombination cassette. On the top, the genome sequence of the MG1655*::sacB* strain used for the construction of the HR cassette. In bold and underlined, the PAM sequence. The red triangles point to Cas12a cleavage sites. The brown line represents the bacteria genome. The two slashes symbolize the discontinuity of the full *sacB* sequence. On the bottom, the design of the *sacB* recombination cassette inserted in the suicide mobilizable plasmid pSW27. The *Ptac* promoter was placed upstream the *sacB* HR cassete. The asterisk represents the unexpected mutation found in the promoter (see text for details). The HR cassette sequence contains 430 bp of the 5' region of the *sacB* gene. This region includes the *sacB* region targeted by the gRNA encoded in pLG19 plasmid. The PAM sequence was mutated to generate an early STOP codon (PAM*, in red). The black line represents the plasmid pLG27 sequence. Cm^R in the blue box represents the chloramphenicol resistance cassette.

Once the plasmid was constructed, we tested if it was efficiently mobilized to a recipient cell by TrwC-Cas12a. We performed mating assays using as donor cell DH5αpir carrying the plasmids pLG27 and pSU1445 (R388 derivate with no *trwC*). We complemented pSU1445 plasmid using pLG22 (P*lac::trwC)* or pLG24 (*Ptet::trwC-cas12a*). As negative control, we used pSU1445 without plasmid complementation. As recipient cell we used β2150, an *E. coli* strain *pir*$^+$*,* which allowed the replication of pLG27. Conjugation frequencies obtained were very similar in both cases and no significant differences were detected: $6.5x10^{-1}$ $\pm 1.2x10^{-1}$ transconjugants/donor for TrwC, and 5.3 $x10^{-1}$ ± 1.0 $x10^{-1}$ for TrwC-Cas12a (mean of 3 independent assays), with no transconjugants obtained in the negative control. Therefore, pLG27 plasmid could be mobilized by TrwC-Cas12a to a recipient cell.

The next step was to determine if the editing efficiency increased in comparison with the recombination background when TrwC-Cas12a translocated the HR cassette containing plasmid into a recipient cell expressing the $sacB_{gRNA}$. Therefore, we mobilized pLG27 by TrwC-Cas12a into the recipient cell MG1655::*sacB* (where pLG27 could not replicate) containing $sacB_{gRNA}$. After translocating TrwC-Cas12a covalently bound to the

sacB HR cassette into the recipient, TrwC-Cas12a would cleavage chromosomal *sacB* gene, generating DSBs. This DSB would be repaired by homologous recombination pathways using the *sacB* homologous template, leading to the integration of the pLG27 plasmid at the Cas12a cleavage site (**Figure 43a**). Resulting integrants would be chloramphenicol resistant (as pLG27 carries a Cm^R cassette), and sucrose resistant, since both *sacB* copies produce truncated products. A subsequent recombination event between the homologous sequences would generate Cm^S recombinants carrying the desired mutation or the *sacB* wild type copy, depending on the crossover site (Figure 43a, bottom).

If the initial recombination between the incoming plasmid and the genomic *sacB* sequence was not mediated by Cas12a cleavage, the outcome would depend on the crossover site. Since the left HR arm is larger, it was more probable than the recombination occurred at this point (Figure 43b). In that case, the resulting integrant would carry a truncated copy of the edited *sacB**, and a wild type *sacB* under the P*tac*. Therefore, the integrant would be sucrose sensitive. However, although this was our initial design, in order to reduce the background recombination, after the construction of pLG27, a deletion of 3 nt was detected (5'-TTG-3') which affected the -35 region of the promoter. So, we do not know if those integrants generated by HR are sucrose sensitive or resistant.

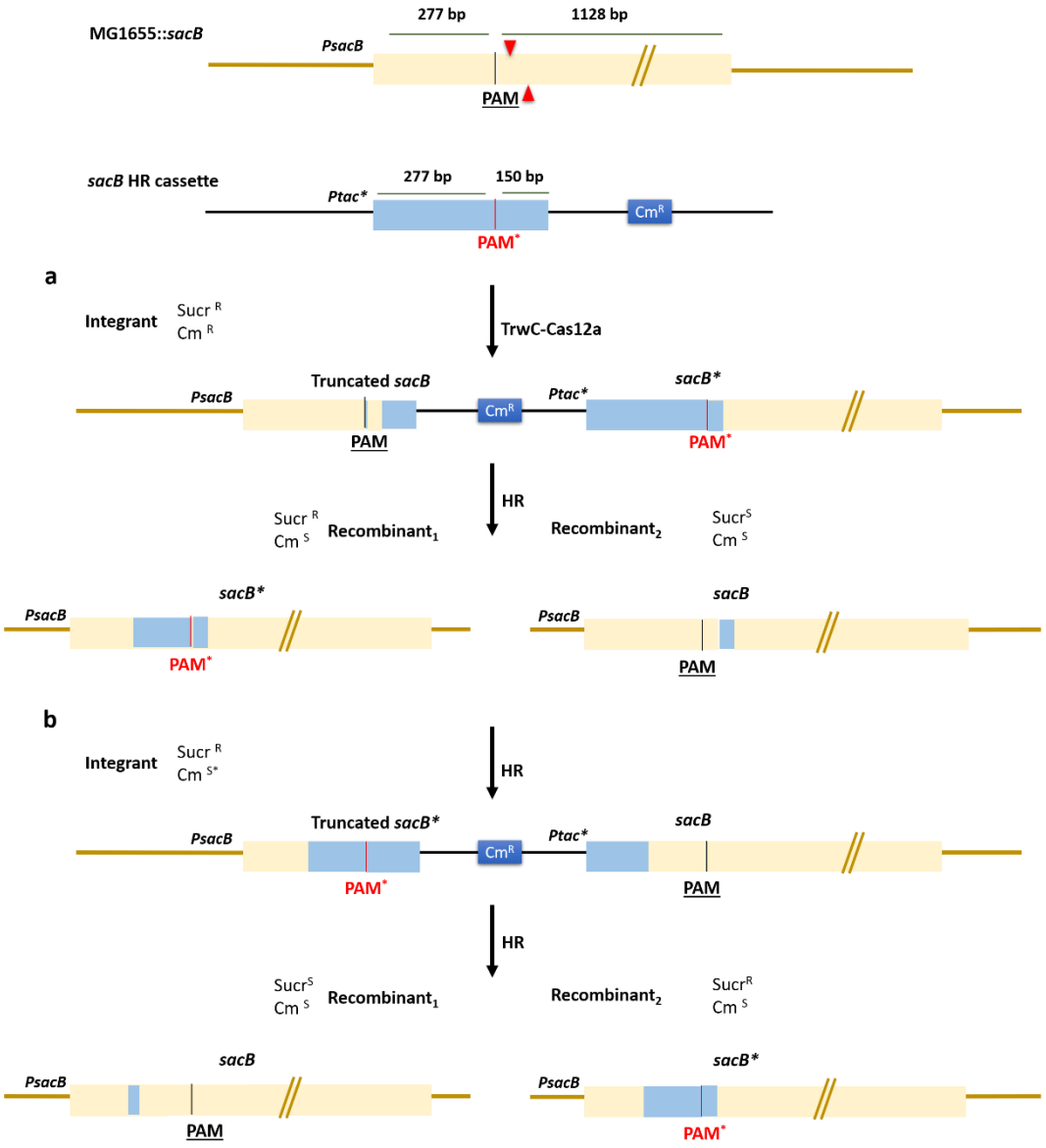


Figure 43. Introduction of seamless mutations using a HR cassette. a) TrwC-Cas12a mediated edition. Cas12a cleavage in the chromosomal *sacB* copy would generate a DSB which will be repaired by the HR pathways of the cell. The resulting integrants would carry a *sacB* copy with the wild type PAM and the right arm of the HR cassette (thus encoding a truncated *sacB*), and a second *sacB* copy with the mutated PAM (red vertical bar,) and the full *sacB* sequence. The second recombination could occur 5' from the PAM sequence, generating a recombinant (Recombinant$_1$) with a *sacB* copy containing the mutated PAM (left), or it could occur 3' from the PAM sequence, generating the recombinant$_2$, with a non-edited *sacB* copy. **B)** Edition non-mediated by TrwC-Cas12a. Since the left HR arm is larger, the recombination would occur with more probability at this point. The resulting integrant would carry a truncated copy of the edited *sacB**, and a wild type *sacB* under the P*tac* (the integrant would be Sucr^S), however this P*tac* may not be functional (P*tac**). The second recombination could occur 5' from the PAM sequence, generating a recombinant (Recombinant$_1$) with a non-edited *sacB* copy (left), or it could occur 3' from the PAM sequence, generating the recombinant$_2$, with a *sacB** edited copy. Red triangles: Cas12a cleavage sites. Light orange boxes: chromosomal sequences; light blue boxes: HR cassette sequences. Black vertical bar, wild type PAM. Red vertical bar, mutated PAM. *sacB**: *sacB* containing the desired mutation. Cm^R/Cm^S: chloramphenicol resistant/sensitive. Sucr^R/Sucr^S: sucrose resistant/sensitive. The two brown slashes represent the rest of the full *sacB* sequence (not shown to scale).

We performed matings using as donor cell DH5αpir and mobilizing pLG27 (encoding the *sacB* HR cassette) with TrwC (pLG22) or TrwC-Cas12a (pLG24). As recipient cells we used MG1655*::sacB* harboring pLG19 ($sacB_{gRNA}$) or the empty vector (pUC8). Integrants were selected by supplementing media with Cm and Ap. In parallel, recipients were selected with Ap (pLG19 and pUC8 carry an Ap^R cassette), with or without sucrose, to calculate the rate of mutation (recipients resistant to sucrose/total recipients). The results are summarized in **Table 13**.

Table 13. Homologous recombination assays

Relaxase (donor)[1]	Recipient[2]	Integration frequency[3]	$Sucrose^R$ frequency[4]
TrwC	$sacB_{gRNA}$	$5.6x10^{-4}±5.6x10^{-4}$	$3.8x10^{-5}±3.5x10^{-5}$
	no gRNA	$1.3x10^{-3}±1.0\ x10^{-3}$	$9.4x10^{-6}±5.2x10^{-7}$
TrwC-Cas12a	$sacB_{gRNA}$	$2.8x10^{-6}±3.6x10^{-6}$	$1.7x10^{-5}±1.9x10^{-5}$
	no gRNA	$1.3x10^{-5}±1.9x10^{-5}$	$9.7x10^{-6}±4.2x10^{-6}$

The data represent the results of two to six experiments. [1]Donor cells were DH5αpir harboring pLG27 (with *sacB* homologous recombination cassette), pSU1445 (R388 without *trwC*) complemented with pLG22 (P*lac*::*trwC*) or pLG24 (*Ptet*::*trwC-cas12a*). [2] MG1655*::sacB* was used as recipient cell, harbouring plasmid pLG19 ($sacB_{gRNA}$) or pUC8 (no gRNA). [3] Expressed as Cm^R Ap^R integrants per donor. [4] Expressed as sucrose-resistant recipients per recipient.

We observed a decrease in the number of integrants obtained when TrwC-Cas12a was present in the assays (Table 13, Integration frequency column). This decrease was independent of the presence of the $sacB_{gRNA}$ in the recipient cell, so we don´t have an explanation for this result. When we analysed the number of colonies resistant to sucrose, we did not observe an increase in the frequency when TrwC-Cas12a was present in the cell (Table 13, $Sucrose^R$ frequency column).

A total of 90 $Sucr^R$ colonies were analysed: 30 from TrwC + gRNA, 30 from TrwC-Cas12a -gRNA, 30 from TrwC-Cas12a +gRNA conditions. First, we replicated them in Cm plates in order to discern if they were integrants or double recombinants. Then, to determine if the Cm^S colonies had incorporated the PAM mutation from the recombination cassette, we amplified a fragment of 357 bp containing the PAM mutated region using the oligonucleotides (5'- CTACCGCACTGCTGGCAG-3') and (5'-AGCTCCACCGCGGTGGCGGCCGCTCTAGAACTGTCAATAGAAGTTTCGCCGA-3'). All the

amplifications had the expected size, and their sequence were determined. **Table 14** summarizes the results of the colony analysis, and **Figure 44** shows the editing ratio, calculated as the number of chloramphenicol sensitive colonies which had incorporated the desired mutation, divided by the total number of $Sucr^R$ colonies analysed for that condition.

Table 14. Colony analysis after integration assays

Relaxase	*sacB*$_{gRNA}$[1]	$Sucr^R$ colonies	Cm^S edited colony	Editing ratio[2]	n
TrwC-Cas12a	-	16	1	0.06	1
		14	1	0.07	2
TrwC	+	16	1	0.06	1
		14	1	0.07	2
TrwC-Cas12a	+	16	4	0.25	1
		7	5	0.71	2
		5	1	0.20	3
		2	2	1.00	4

[1]The recipient strain MG1655*::sacB* carried either pUC8 (*sacB*$_{gRNA}$ (-)) or pLG19 (*sacB*$_{gRNA}$ (+)). [2]The editing ratio was calculated by the number of Cm^S edited colonies per Suc^R colony analyzed. The data represent the results of two and four independent experiments.

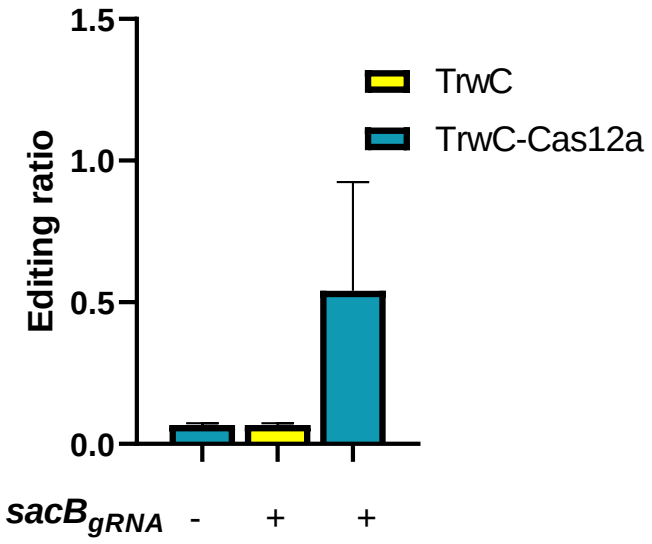


Figure 44. Editing rate of TrwC and TrwC-Cas12a. The editing rate is calculated as the number of chloramphenicol sensitive colonies which had incorporated the *sacB* mutations from the homologous recombination cassette, per the total number of $Sucr^R$ analyzed colonies. *sacB*$_{gRNA}$ presence in the recipient cell is indicated with a + or a -. The data represent the result of two to four independent experiments.

We observed that, when TrwC-Cas12a and the gRNA targeting *sacB* gene were present, the percentage of edited cells was more than 8- fold higher than when TrwC or TrwC-Cas12a without gRNA were assayed.

4.2.4. Generation and validation of Cas12a fusions in human cells

As it has been mentioned, TrwC can be recognized by the T4SS of *B. henselae* and it can be translocated into human cells, together with the transferred DNA (Llosa *et al*., 2012b). We have shown that TrwC-Cas12a could be translocated by the T4SS of plasmid R388, and the protein was fully functional in the recipient cell. In the future, it could be tested if we could use T4SS as an *in vivo* delivery tool to translocate CRISPR-Cas proteins to human cells. In order to open this research line, we have constructed and validated several Cas12a fusion proteins which could be used to test their activity in human cells upon translocation through the T4SS of *B. henselae.*

4.2.4.1. Generation of fusion protein constructions

We have constructed the fusion protein TrwC-Cas12a (as it was described in the Section 4.2.1), but in this case under the control of the constitutive promoter CMV, to allow its expression in human cells. The construction was named pLG08. As it was done for prokaryotic cells, we checked TrwC-Cas12 stability by western blot. It was performed from HEK 293T lysates after transfecting the plasmids encoding Cas12a (pY010) or TrwC-Cas12a (pLG08) into HEK293T cells. To detect the proteins, we used an anti-TrwC antibody (**Figure 45**).

The results showed a main band with the expected size for the fusion protein, without major degradation products. This result demonstrated that the fusion protein TrwC-Cas12a was stable in human cells.

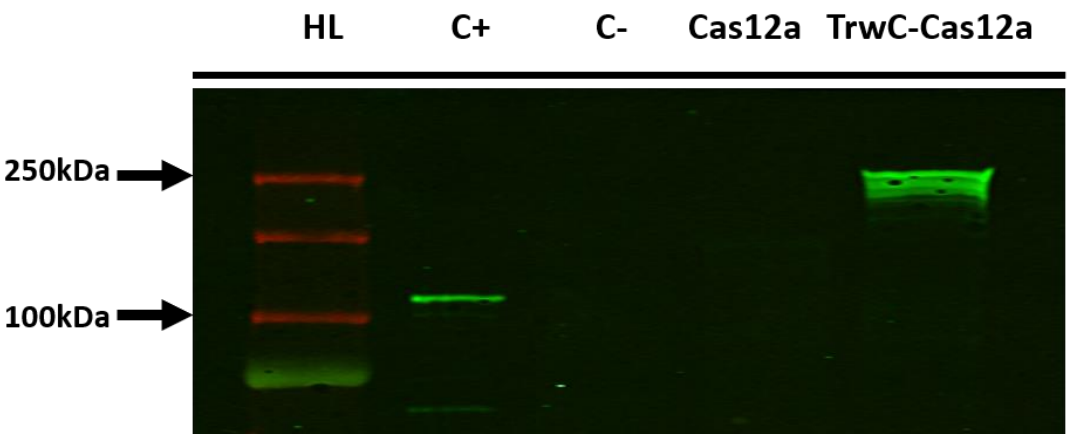


Figure 45. TrwC-Cas12a stability in HEK293 cells. A Western Blot was performed using an anti-TrwC antibody. (C+): purified TrwC protein was used as positive control. (C-): HEK293T lysate was used as negative control, Cas12a and TrwC-Cas12a correspond with HEK293T lysates after being transfected with pY010 and pLG08 respectively. HL: Protein weight marker. The expected sizes for the proteins were: TrwC (108kDa), Cas12a (187 kDa), TrwC-Cas12a (263 kDa).

In parallel, we also constructed a MobA-Cas12a fusion. As we showed in Section 4.1.3, MobA could also be recognized and translocated through the *B. henselae* VirB/D4 T4SS. Furthermore, we showed that MobA could be better recruited than TrwC by this T4SS. These results encouraged us to construct a MobA-Cas12a fusion. MobA-Cas12a was constructed similarly to TrwC-Cas12a fusion. The *cas12a* sequence contained a humanized Cas12a (hAsCas12a) with a terminal NLS (Nuclear Localization Signal) and a 3xHAtag, and we fused it to the C-terminus of *mobA*. The construction expression was controlled by a pCMV promoter. The construction was named as pLG11.

In addition to the relaxase-Cas12a fusions, we also constructed a Cas12a-BID fusion. The main pending question for the possible utility of the previous fusions to target human cells is their recruitment by the *B. henselae* T4SS, since the C-terminus of the substrate affects recognition. BID is the sequence necessary for effector translocation by the T4SS of *B. henselae* (see Section 1.1.2.1.). This means that this fusion would be recognized by the T4SS of *B. henselae* and translocated to human cells as an effector protein. The *cas12a* sequence used was the same as for the other fusions. The BID signal was added to the C-terminus of *cas12a*. The fusion was under control of a pCMV promoter. The construction was named as pLG02.

4.2.4.2. Validation of Cas12a activity in human cells

We evaluated Cas12a activity of the different fusions in human cells. In contrast to prokaryotic cells, eukaryotic cells have NHEJ pathways which repair DSBs in the absence of a homologous DNA template. These repairs produce small indels (insertions-deletions). Therefore, appearance of indels would be the consequence of Cas12a activity. We used the Surveyor Mutation Detection kit, which provides an endonuclease that recognizes and cleaves heteroduplexes formed by reannealing of DNA which contains some molecules with indels (**Figure 46**) (Ran *et al*., 2013).

This assay is widely used for detection of CRISPR-Cas activity in eukaryotic cells. Briefly, we extracted the gDNA of the cells previously transfected with DNA encoding the nuclease and the gRNA. If the nuclease cleaves the target sequence, this would lead to indels. However, Cas12a cleavage efficiency is not 100%. Therefore, a mixture of edited and non-edited cells carrying the indels sequence or original sequence would be present in the sample. Then, the gDNA is amplified with primers flanking the target sequence. This PCR amplification would generate a mixture of DNA sequences containing the wild type sequence or the mutated sequence. In the next step, samples are mixed and hybridized to form heteroduplex complexes. Finally, the samples are digested with the Surveyor nuclease which recognizes and cleaves heteroduplexes. Samples are analysed by agarose gel electrophoresis.

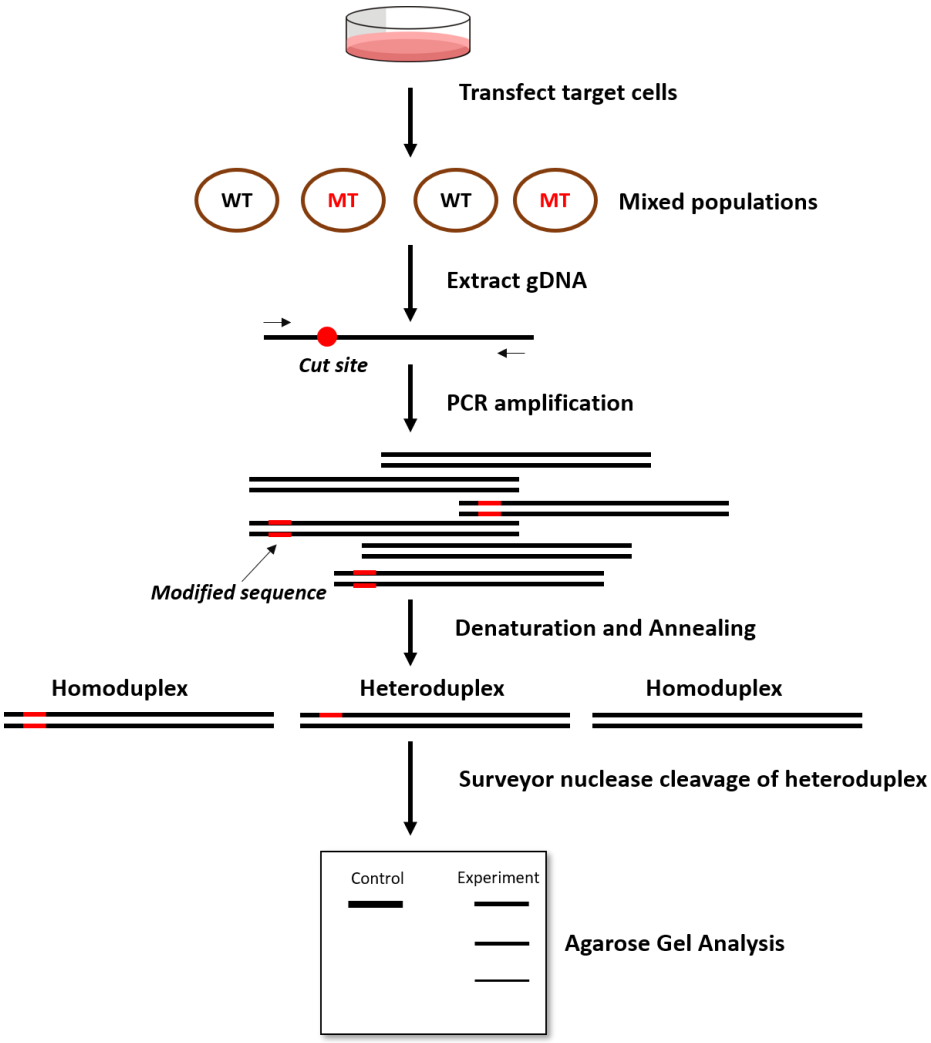


Figure 46. Surveyor nuclease assay scheme. The Surveyor nuclease assay was used to detect Cas12a activity in human cells. See text for more details.

For our purpose, we co-transfected human HEK293T cells with plasmids expressing either Cas12a, TrwC-Cas12a, MobA-Cas12a or Cas12a-BID under the control of a pCMV promoter, and with a PCR-amplified expression cassette to generate a gRNA targeting the *dnmt1* gene from a U6 promoter (Section 3.5.7.). This gene was present in HEK293T cells. After 3 days, the genomic DNA was extracted, the *dnmt1* region was amplified and the PCR products were hybridised and treated with the Surveyor Mutation Detection kit. **Figure 47** shows that the surveyor endonuclease was able to detect indels and digest the *dnmt1* amplifications only in the cases where the gRNA targeting the gene and the plasmid encoding Cas12a protein (wild type or fusion proteins) were co-transfected together. With this result, we proved that the Cas12a component of the

different fusion proteins maintained its activity, being able to efficiently recognize its PAM, bind to the gRNA and cleave the target DNA in human cells. These constructions will be useful for future experiments to check their T4SS-mediated translocation and subsequent activity in human cells.

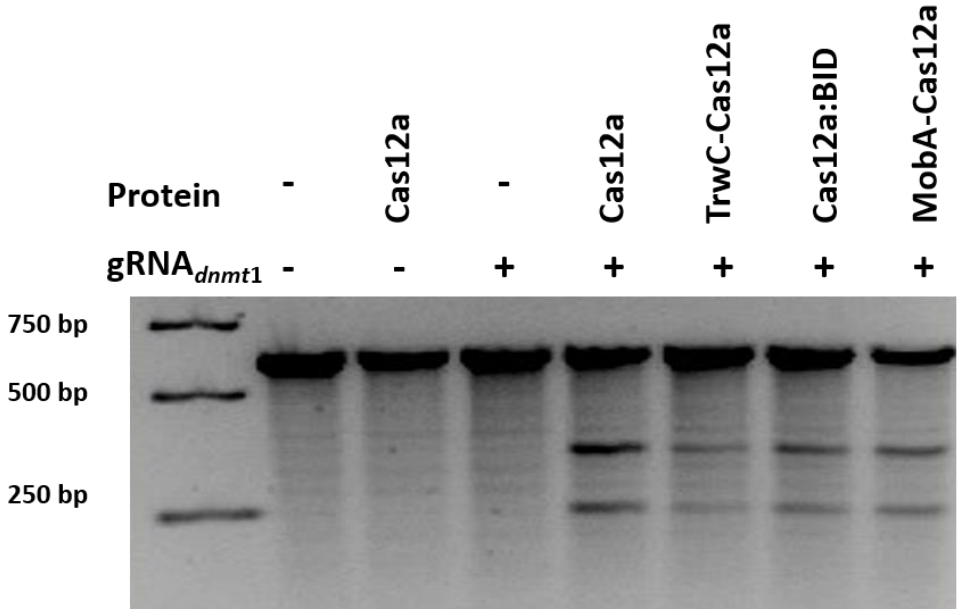


Figure 47. Cas12a activity in human cells. 2% agarose gel showing the results of the Surveyor assay digestion on *dnmt1* PCR products. HEK293T were transfected with DNA encoding Cas12a (pY010), TrwC-Cas12a (pLG08), Cas12a-BID (pLG02) or MobA-Cas12a (pLG11) and cotransfected with the *dnmt1*$_{gRNA}$ cassette expression as indicated (gRNA + or -).

5. Discussion

5. Discussion

Conjugative relaxases are well characterized enzymes. They perform the first and last reaction of the conjugative DNA transfer process by site- and strand-specific nucleophilic cleavage and strand transfer reactions. Although most of them have been classified within the superfamily of HUH endonucleases, many exceptions are arising. Many relaxases are multifunctional proteins which perform other roles not necessarily involved in conjugation, such as SSI. These other functions have been validated in many different systems, suggesting that they could play an important biological role. The paradigm is changing, and relaxases could not only be key enzymes for horizontal gene transfer, but also be implicated in other functions which contribute to prokaryotic genetic plasticity.

TrwC was the first relaxase described to perform SSI reactions (Draper *et al*., 2005). It can mediate integration of the conjugatively transferred DNA into a resident *oriT* copy once it is transferred to the recipient bacteria. TrwC can also be translocated through the T4SS of a bacterial human pathogen into a human cell, where it promotes random integration of the incoming DNA into the human genome. These characteristics confer TrwC a high biotechnological potential.

The moonlighting nature of many conjugative relaxases continues to be an intriguing fact. In this work, we aimed to dig into the molecular basis, biological role, and biotechnological potential of the integrase activity of conjugative relaxases, using as a model the relaxase TrwC.

5.1. Study of the integrase activity of conjugative relaxases

The ability to catalyze *oriT*-specific integration reactions has been previously reported for some relaxases. The additional ability shown by TrwC to promote unspecific integration of the covalently attached DNA in the human genome was, so far, unique. Many relevant questions do not have an answer yet, such as: what makes a relaxase able to catalyze SSI? Why some relaxases can catalyze the reaction while others closely related cannot? And what could be the biological role of promoting integration (site-

specific or random) of the transferred DNA, which is apparently not involved in conjugative DNA transfer itself?

Our first goal was to address the biochemical basis conferring SSI activity to a relaxase. TrwC has been analyzed to determine which was the minimal domain able to catalyze the SSI reaction; the relaxase domain N293, which contains the catalytic Tyr residues mediating DNA cleavage and strand-transfer, was not enough by its own to catalyze the reaction, and the minimal domain found to perform efficiently the reaction was N600 (Agúndez *et al*., 2012). Since the 3D structure of this region of the protein (N293-N600) is not solved, and there are no apparent conserved domains, there was little clue as to what was this extra domain providing to the relaxase domain in order to be able to accomplish integration.

Agúndez and collaborators generated a chimeric protein formed by the N293 relaxase domain of TrwC and the C-terminal region of the AAV Rep 68 protein, which contains the DNA helicase domain and the interdomain linker necessary for Rep68 oligomerization (Agúndez *et al*., 2018). Both TrwC and Rep 68 are members of the HUH protein family and both show SSI activity, however they are distantly related. Surprisingly, the chimera protein N293-Rep68 was able to perform the SSI reaction in bacteria, using an *oriT*$_{w}$ as target. This result was unexpected as N293 was not able to catalyze the reaction on its own. Therefore, it seemed that the C-terminal domain of Rep68 was contributing to the SSI activity of the chimera. Additionally, the oligomerization profile of the chimera when it was incubated with its *oriT* target (25+8) was different when compared to the profile of both parental (N293 and Rep68) separately, as it formed a high order oligomeric complex of at least six subunits. The C-terminal region of the chimera contained the helicase domain of Rep68, however it was unlikely that it was affecting the reaction as the helicase domain of TrwC is dispensable for SSI (Agúndez *et al*., 2012). On the other hand, this C terminus contained a region necessary for Rep68 oligomerization, which seems to play an important role in the SSI of the virus (Bardelli *et al*., 2016) as it may be involved in host protein interactions. In summary, the data made us suspect that maybe the ability to oligomerize on the target DNA was the necessary requisite for a relaxase to act as an integrase.

In order to elucidate if DNA-dependent oligomerization could affect the SSI activity of a relaxase, we expected that if N600 was the minimal domain able to perform the reaction efficiently, it should form an oligomeric complex in the presence of its *oriT* target. N600 has been previously described as a monomer by gel filtration chromatography, however, no sedimentation velocity experiments have been performed (César *et al*., 2006).

To test our hypothesis, we decided to compare the oligomerization profile of N293 and N600 incubated in the presence of its target (the *oriT* sequence) by analytical centrifugation techniques. First, we purified N293 and N600 domains in two steps and SDS-PAGE gel showed that both domains were purified correctly (Figure 22). We performed sedimentation velocity experiments to determine the sedimentation coefficients, related with the size and form of the complex and molecules.

The sedimentation coefficient for N293 and N600 alone corresponded to the size of one molecule, confirming the previous observations that these proteins behaved as monomers in solution (César *et al*, 2006). To determine the type of complex formed for each domain in the presence of the ssDNA substrate, the proteins were incubated with the oligonucleotide *oriT* (25+8) which carries the IR_2 (TrwC binding site) and the *nic* site. Although TrwC can catalyze the integration reaction on a smaller *oriT* sequence (17 bp), we decided to choose this oligonucleotide as it was the one used for the chimera oligomerization assays (Agúndez *et al*., 2018). We tested different ratios of protein:target, including a 4:1 protein:DNA ratio used by Agúndez and collaborators to observe the DNA-dependent oligomerization of the N293-Rep chimera. In all the conditions tested, for the N293 domain, the sedimentation profiles obtained corresponded to one molecule of DNA with one molecule of protein (Figure 23), as previously reported (Lucas *et al*., 2010). Unfortunately, when N600 was assayed, the results revealed the same type of complex. The sedimentation coefficients obtained corresponded, in all the cases, to a complex formed by a molecule of protein and a molecule of DNA (Figure 23).

Since we were not able to detect differences between N293 and N600 oligomerization abilities, we conclude that DNA-dependent oligomerization is not the reason why N600 performs SSI while N293 does not. This question remains open.

However, it cannot be discarded that in the case of the N293-Rep chimera the oligomerization behavior might be conferring the protein the ability to perform SSI reaction. It could be interesting to explore if relaxases without SSI activity could gain it by adding an oligomerization domain.

The second question we addressed was: why would a conjugative relaxase have integrase activity when it is not required for conjugative DNA transfer? Our hypothesis was that, by integrating the transferred DNA into the recipient genome, this ability could allow conjugative plasmids to colonize hosts where it could not replicate. In fact, the transfer range of a plasmid is usually bigger than the replication range (Kishida *et al*., 2017; Samperio *et al*., 2021). Therefore, the integrase activity would allow the plasmid to disseminate into new hosts.

It has been shown that TrwC is able to catalyze integration of the transferred DNA not only on an *oriT* containing plasmid but also in the chromosome of the recipient cell. In addition, TrwC DNA requirements to catalyze integration are less stringent in the acceptor site. It can integrate DNA into sequences that do not match exactly the *oriT* minimal sequence, and even into two DNA sequences of human origin with a single mismatch from the minimal *oriT* sequence (Agúndez *et al*., 2012). And moreover, when translocated into human cells, it was shown that TrwC promoted integration of the covalently attached DNA into any site of the human genome, without any detectable homology to its target *oriT* sequence (González-Prieto *et al*, 2017). All these evidences prompted us to test if TrwC was able to integrate an *oriT* containing plasmid into a non-permissive host without an *oriT* copy present.

In addition to its biological implications, opening the range of conjugation recipients would be a useful tool for genetic modification purposes. Thus, we used *L. casei* as recipient cell, a Gram-positive bacterium with a high biotechnological interest. A prerequisite to test our hypothesis was to determine if the R388 system could mobilize a replicative plasmid into *L. casei*, since bacterial conjugation into Lactobacilli was not previously reported. We decided to mobilize an *oriT* containing shuttle plasmid using in parallel the conjugative systems of plasmids R388 and RP4. The latter has been previously described to mobilize DNA to other Gram-positive bacteria, although not to lactobacilli (Trieu-Cuot *et al*., 1987). To test conjugation from *E. coli* to *L. casei*, we

assayed different conditions and selective media and finally set up a conjugation protocol which allowed us to obtain transconjugants with both systems. The transconjugants were thoroughly characterized, including 16S sequencing. The negative controls run up in parallel, using donors with the same mobilizable plasmid but devoid of conjugative machinery, never rendered transconjugants, allowing us to prove that the *L.casei* colonies obtained containing the mobilizable plasmid were generated by conjugation, and not other HGT mechanisms. Thus, in this work we have proved that R388 and RP4 can mobilize DNA from *E. coli* to *L. casei.* Conjugation frequencies were different depending on the system used (around 10^{-6} transconjugants/donor for R388, and 20 times higher for RP4) (Table 7 and Figure 25). The differences in the conjugation frequencies between both systems were not surprising, as the RP4 system has been widely used to transfer DNA into not related bacteria and even into eukaryotic cells (Bates *et al*., 1998; Luzhetskyy *et al*., 2006).

These results demonstrated for the first time DNA mobilization from *E. coli* to *lactobacilli*. Furthermore, it has been the first report of DNA transfer using R388 system to a Gram-positive bacteria. The possibility to introduce DNA by conjugation into lactobacilli has great biotechnological potential. Although the laboratory strain used, *L.casei* 393, can be genetically modified using electroporation to introduce foreign DNA, many other related lactobacilli, and even other wild-type strains of *L. casei* which are relevant in the food industry, cannot be transformed. In fact, the continuation of this work by other lab members has shown that the conjugation protocol here developed can be used to transfer DNA into a wide variety of lactobacilli, including wild-type strains, species and even different genera (Samperio *et al*, 2021), such as *Lentilactobacillus parabuchneri*, an important contributor to the cheese organoleptic properties, which had never been transformed so far.

Once the conjugation protocol was established, we carried out integration assays using a mobilizable suicide plasmid (with no functional replicon in *L. casei*) instead of a shuttle plasmid. With this approach, we mobilized a plasmid piloted by TrwC or RP4_TraI into a recipient cell, where the plasmid could not replicate. Since our purpose was to determine if TrwC was promoting the integration into the recipient genome, we assayed it in parallel with RP4-TraI, a relaxase which does not promote such integration events,

in order to compare their integration efficiencies. In parallel, conjugation assays using the shuttle plasmids were performed, in order to obtain the conjugation efficiency. The integration rate was calculated as the number of integrants per number of transconjugants. We did not observe significative differences between both systems (Figure 27 and Table 8). Therefore, under the condition tested, we conclude that TrwC did not promote the integration of the transferred DNA into the *L. casei* chromosome.

We looked for minimal *oriT*-like sequences in the *L. casei* genome in order to detect potential targets for TrwC-mediated SSI (Agúndez *et al*., 2012), however, no homologies were found. In fact, when we analyzed the integrants, full *oriT* sequences were amplified, confirming that they were random integration events. Our results suggest that, despite TrwC promotes unspecific DNA integration in human cells, regardless of the presence of a target *oriT* sequence, in bacteria this phenomenon is not observed. This is not surprising, considering that the host DNA recombination and repair systems which presumably contribute to the integration reactions (César *et al*, 2007; Agúndez *et al*, 2012) are very different in the bacterial or human cells. In particular, it is well known that integration of a foreign DNA in the human genome occurs at any nicked DNA site with much higher efficiency than into homologous sequences. Future experiments could address TrwC-mediated integration of conjugatively transferred DNA into recipient non-permissive bacteria harboring *oriT*-like target sequences.

Apart from being recognized and translocated by its own T4SS, TrwC can also be recognized by the T4SS of *B. henselae* VirB/D4 and be translocated covalently bound to an *oriT*-containing plasmid into a human cell, where it is active (Fernández-González *et al*., 2011). This ability has been also demonstrated for the Mob relaxase of the cryptic plasmid pBGR1 of *B. henselae* (although a BID signal was added to detect DNA transfer level efficiently (Schröder *et al*., 2011)); Once in the recipient cells, both relaxases behaved differently, as TrwC promoted the integration of the transferred DNA into the genome of the recipient cell while Mob:BID did not (Gonzalez-Prieto *et al*., 2017). The integration pattern observed was unspecific, although one site-specific event was detected in an 8 bp homologous sequence, supporting the idea that the requirements of the acceptor site for the SSI reaction are less stringent. It is unclear how TrwC promotes this random integration. It could be due to a DNA chaperone-like activity of

the relaxase, protecting the DNA from degradation by binding to its ends, and thus increasing the chances of integration.

Our aim was to test if this integration-promoting ability was unique for TrwC. To test this hypothesis, we chose the relaxase MobA from the promiscuous conjugative plasmid RSF1010, which can promote *oriT*-specific recombination on single-stranded substrates (Meyer, 1989). MobA can be recruited by different T4SS, since the plasmid does not carry its own T4SS. Determination of the translocation of MobA-DNA through *B. henselae* ViB/D4 into human cells was an essential aspect previous to determine the integration ability into the human genome.

We constructed mobilizable plasmids containing *gfp* under the control of a eukaryotic promoter. They also encoded the MOB region for R388 or RSF1010. As negative controls, we generated plasmids with no functional relaxases. We infected human cells with *B. henselae* carrying the different plasmids. Only when the infections were performed using a relaxase-containing plasmid (MobA or TrwC), we detected DNA transfer, which was inferred by GFP expression. (Figure 30 and Table 9). This is the first report of DNA transfer led by MobA through a T4SS into human cells. Moreover, MobA relaxase transferred the DNA with higher efficiency than TrwC, both into EA.hy296 and HeLa cells.

The differences on DNA transfer efficiency between relaxases could be correlated with their recognition efficiency by the T4SS. A previous report showed that the relaxase Mob could transfer DNA to human cells by VirB/D4 T4SS with very low efficiency. However, when a recruitment secretion signal (BID signal) was fused to its C-terminal end, the DNA transfer frequency was increased almost 100-fold (Schröder *et al*., 2011). In the case of TrwC, when the conjugative coupling protein was deleted, a component believed to play a key role in the recruitment of the conjugative substrate, the efficiency of DNA to human cells decreased 10-fold (Fernandez- Gonzalez *et al*., 2011). MobA relaxase can be translocated through various T4SS, not only belonging to conjugative plasmids. MobA translocation through the T4SS of *Legionella pneumophila* alone or bound to DNA into a recipient bacteria has been described (Vogel *et al*., 1998; Luo and Isberg, 2004). Translocation through *A. tumefaciens* T4SS into plant cells of MobA has also been reported (Vergunst *et al*., 2005). Thus, the requirements for MobA

recruitment could be less stringent than for TrwC. In fact, subsequent experiments of relaxase translocation into human cells through the Dot/Icm T4SS of *L. pneumophila* and *Coxiella burnetii* showed that both TrwC and MobA could be recruited, but MobA was transferred with much higher efficiency (Guzmán-Herrador *et al*, 2017). All these data strongly support the idea that the efficiency of relaxase recruitment by the T4SS directly affects DNA transfer efficiency.

Once we were able to transfer DNA piloted by MobA into human cells through VirB/D4, we studied the integration ability of the relaxase into the human genome. We carried out the experiments with similar plasmids used for DNA transfer detection, but they also carried a hygromycin resistance cassette to select stable integrants. For these experiments, we compared MobA integration ability with TrwC (which enhanced integration) and with Mob:BID (which did not enhance integration; (Gonzalez-Prieto *et al*., 2017)). Results showed that MobA promoted DNA integration of the transferred DNA in the human genome. When the DNA transferred was piloted by MobA, the integration frequency was 10-fold higher than when Mob:BID was leading the DNA. By contrast, MobA integration frequency was almost five-fold lower than the frequency of TrwC (Figure 31 and Table 10). Therefore, we conclude that MobA relaxase can promote DNA integration of the transferred DNA in the human genome, although with less efficiency than TrwC. Thus, TrwC is not the only relaxase which can promote DNA integration of the transferred DNA into the human genome. The finding that other relaxase can be used to transfer and promote integration of DNA into a human cell suggests that the recruitment and translocation of conjugative relaxases by T4SS involved in pathogenesis could be extended to other systems.

Although it is unclear what makes a relaxase able to perform SSI, the ability to enhance integration of the transferred DNA into the recipient cell genome should reside in an intrinsic property of the relaxase, as it is the only protein entering the recipient cell covalently bound to the ssDNA transferred. These differences between relaxases could be caused by different factors, such as the catalytic activity of the different Tyr residues, or the binding affinity to its target DNA. In fact, although both relaxases are able to perform a recombination reaction, TrwC catalyzes site-specific recombination on scDNA substrates (César *et al*., 2006), while MobA was shown to catalyze site-specific

recombination between two *oriT* copies only when the substrate was single-stranded (Meyer, 1989).

DNA transfer into human cells has been reported by different laboratories, and through the T4SS of different human pathogens (Fernández-González *et al*., 2011; Schröder *et al*., 2011; Guzmán-Herrador *et al*., 2017). The DNA is transferred and integrated with significant efficiencies. The DNA was transferred by the relaxases of a natural plasmid of *B. henselae*, or promiscuous plasmids such as R388 or RSF1010 which can be transferred among most Gram-negative bacteria. Thus, the possibility exists that this DNA transfer process is occurring in nature. We could wonder if this ability to transfer and integrate DNA in the human host could be playing a biological role. Our hypothesis is that pathogens may use DNA transfer and integration for their own benefit for long-term subversion of the host cells (**Figure 48**). The DNA transferred by the relaxase may encode beneficial traits for the pathogen and its integration would allow its stable expression in the eukaryotic cells. Also, random integration could produce insertional mutagenesis, generating eventually uncontrolled growth of the cells, extending the niche for bacterial growth.

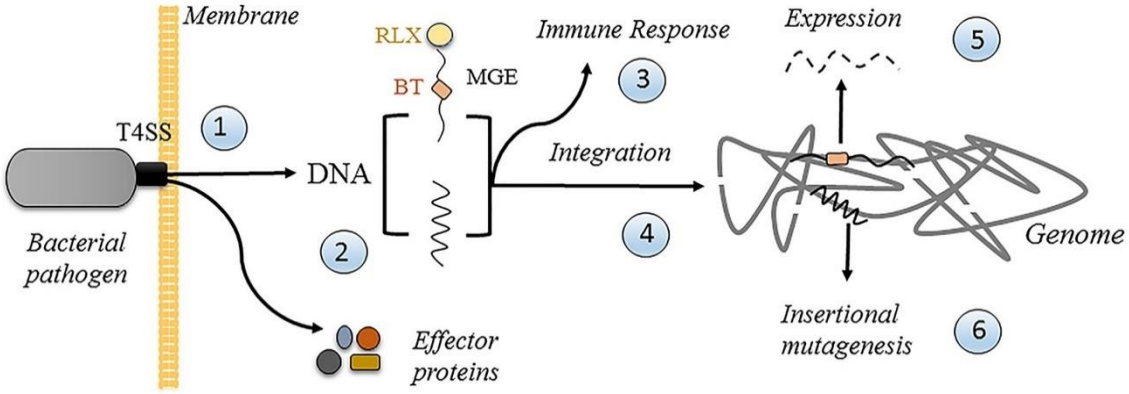


Figure 48. Possible destinations of mobilized DNA into human cells. 1) A bacterial pathogen translocated effector proteins and DNA through its T4SS to the human cell. 2) The transferred DNA could either be random DNA, or a specifically recruited mobile genetic element (MGE), in which case a dedicated transfer system would attach a relaxase to its end. 3) The DNA could induce an immune response. 4) Also, it could get integrated into the human cell genome by the host repair systems, and/or by the covalently attached conjugative relaxase. 5) The integrated DNA could be stable expressed, allowing the expression of beneficial traits for the pathogen. 6) Random integrations could produce insertional mutagenesis, generating eventually uncontrolled growth of the cells. Random DNA: jagged line. Mobile genetic element (MGE): wavy line. RLX, relaxase: small yellow sphere. BT, beneficial trait small orange box. Taken from (Guzmán-Herrador *et al*., 2017).

5.2. Relaxases as protein/DNA delivery systems for biotechnological purposes

Genomic editing of eukaryotic and prokaryotic cells allows the development of biotechnological and biomedical applications. *In vivo* access to the target cell and modification of specific sites in the genome (gene targeting) are two of the main challenges, which demand the development of new genetic modification tools.

Bacterial conjugation fulfills the requirements to address the first of these challenges. It is a promiscuous *in vivo* DNA delivery system, both into prokaryotic and eukaryotic cells, when combined with T4SS from human pathogens(Llosa et al., 2012a). In addition, the DNA enters covalently linked to a relaxase which, as previously shown, can mediate the integration of this incoming DNA into the recipient genome. In the case of TrwC, it can directly mediate SSI of the incoming DNA into the recipient bacterial genome. In human cells, it was shown that it could not perform SSI; however, it promotes random integration of the attached DNA (Gonzalez-Prieto et al., 2017). Thus, TrwC could be used as vehicle to deliver DNA into a human cell, and it could assist the action of a site-specific endonuclease to achieve gene targeting.

CRISPR-Cas systems have revolutionized the genetic modification field because of their easy use and simplicity, allowing the genetic modification of many organisms (Manghwar et al., 2019; Liu et al., 2020; Tyagi et al., 2020). CRISPR-Cas technology has been widely used in eukaryotic cells, where its use is almost unlimited (Doudna, 2020). The use of CRISPR system in bacteria has been less exploited, although it has been increased in the last years with biotechnological or biomedical purposes (see Section 1.3.2.1.). Despite its success, the technology has limitations. A critical step to achieve the genetic modification is the delivery of the endonuclease, gRNA and DNA template to the target cell. There are different methods for *in vivo* delivery, as detailed in Section 1.3.1.. The more widely used methods consist on the introduction in the target cell of the DNA which encodes the different elements of the CRISPR-Cas system. Bacterial conjugation has been used to introduce a CRISPR-Cas system into different prokaryotes using mobilizable plasmids (Citorik et al., 2014; Hamilton et al., 2019). However, the introduction of the endonuclease gene requires its expression in recipient genera not

always well characterized at this level; and overexpression of the endonuclease can lead to toxicity and off-target activity.

Cas proteins have been fused to other proteins in order to improve their activity or acquire new ones. For example, Cas9 has been fused to CtIP to increase integration by HR (Charpentier et al., 2018) or to an engineered reverse transcriptase (Anzalone et al., 2019). Our aim in this work was to fuse the Cas endonuclease to a conjugative relaxase. During conjugation, the relaxase covalently bound to the transferred DNA molecule is recognized and translocated through the T4SS into the recipient cell. Therefore, the Cas protein itself would be delivered *in vivo* through the T4SS into a recipient bacterium. Furthermore, as the transferred DNA has the only requisite of harboring an *oriT*, this DNA can encode the gRNA or the template DNA for recombination-mediated seamless genetic modifications. The integration-promoting activity of TrwC in human cells could favor its Cas-mediated SSI.

The relaxase-like protein VirD2 from *A. tumefaciens* has been previously fused to endonucleases. Rolloos and collaborators fused and I-Sce endonuclease with VirD2 and translocated it through the T4SS of *A. tumefaciens* VirB/D4 into yeast (Rolloos et al., 2015). They showed that the fusion was functional and that it enhanced the integration of the transferred DNA by homologous recombination. Recently, Ali and collaborators constructed a chimeric protein fusing VirD2 to Cas9 (Ali et al., 2020). They showed that this fusion protein increased the homology-directed repair because VirD2 could bind to the DNA template and bring it close to the DSB produced by the endonuclease. During that work, they introduced the fusion protein by bombardment of the plasmids and repair template into rice. Therefore, to our knowledge, no translocation through a T4SS of a Relaxase-Cas fusion protein has been demonstrated.

For the development of this work, we have fused the Cas endonuclease AsCas12a with the conjugative relaxase TrwC. Cas12a has demonstrated to be an alternative to Cas9 (see Section 1.3.2.1.) and it has been widely used for multiplex genome editing in prokaryotic and eukaryotic cells (Zetsche et al., 2017; Ao et al., 2018; Li et al., 2018b; Port et al., 2020). Cas12a has been shown to have less toxicity in bacteria than Cas9 (Ungerer and Pakrasi, 2016). Bai and collaborators tried to fuse Cas9 with a signal recognized by the Type III secretion system of *Pseudomonas aeruginosa.* However, the

fusion was toxic for the bacteria and they could not generate it (Bai et al., 2018). In addition, Cas12a has a smaller size than Cas9; this difference could be important for the translocation through the T4SS. As relaxase, we chose the conjugative relaxase TrwC. It has been shown to be translocated alone or with the DNA, by its own T4SS during conjugation into another bacterial cell or by heterologous T4SS involved in pathogenesis into a eukaryotic cell, where the relaxase is active (Draper et al, 2005; Alperi et al, 2013). Also, it has been fused to other polypeptides without losing its relaxase activity (Alperi et al, 2013; Agúndez et al, 2018).

The TrwC-Cas12a fusion was validated in prokaryotic and eukaryotic cells. By western blot we were able to visualize the fusion protein in both type of cells (Figure 34 and Figure 45). TrwC-Cas12a showed RNA-guided endonuclease activity in human cells (Figure 47) and in bacteria, where it killed >99% of the target cells (Figure 35). TrwC-Cas12a also showed 100% efficiency as a conjugative relaxase in complementation assays of a relaxase-deficient R388 in bacteria (Table 11). This result confirmed that TrwC was functional within the fusion protein. The relaxase is required in the recipient cell to finish the conjugation process, therefore this result suggested that the fusion protein could be translocated through the T4SS into the recipient cell. However, as we observed in the Western blot some degradation of the fusion protein, especially in the bacterial cell (Figure 34), the result could be due to the activity of the TrwC moiety devoid of the rest of the fusion protein, so we could not confirm that we were detecting the translocation and activity of the fusion protein in the recipient cell.

It has been reported that TrwC must be partially unfolded during its secretion through the T4SS, and in fact, the inclusion of an unfolding resistant domain avoids protein translocation (Trokter and Waksman, 2018). Considering the big size of TrwC-Cas12a, its translocation through the T4SS was one of the challenges of our experiment. Detecting the activity of the whole fusion protein in the recipient cell would confirm the translocation of the protein trough the T4SS channel and the recovery of Cas12a activity after translocation. For this purpose, two different approaches were performed in bacterial cells, as schematized in Figure 36.

First, we measured the induction of the SOS response in the recipient cell due to the endonuclease activity of Cas12a after TrwC-Cas12a translocation. Despite some GFP

background, probably due to the induction of the SOS response by conjugation, an increase in SOS-driven GFP expression was detected only when the combination of gRNA and translocated TrwC-Cas12a existed in the target recipient cell (Figure 38). Using this report assay we indirectly demonstrated that TrwC-Cas12a could be translocated trough the T4SS and that both moieties of the fusion were active in the recipient cell. Also, to our knowledge, this was the first report of Cas12a induction of the SOS response. This was expected, since Cas12a, as Cas9, introduces DSB which are responsible for inducing this response (Simmons et al., 2008).

As a more direct approach to detect Cas12a activity, we aimed to detect directly mutations produced by the Cas12a cleavage. It has been shown that cell death is not the only outcome for Cas cleaving bacterial chromosome, and that recombination between microhomologies around the cleavage site could produce indels as a result of DSB repair (Cui and Bikard, 2016). The indels produce mutations in the target ORF which usually lead to truncated protein products. In order to detect these mutations, we used a recipient strain encoding *sacB,* we targeted *sacB* gene with the gRNA, and we selected *sacB* mutants by their resistance to sucrose. The sucrose resistant transconjugants were only obtained when the appropriate gRNA was also expressed in the recipient. *sacB* sequence from the transconjugants resistant to sucrose was analyzed and small deletions of 1-4 nt were found right next to the expected TrwC-Cas12a cleavage site, all leading to truncated SacB products (Figure 40). The mutations obtained were similar when a plasmid encoding *cas12a* was transformed into these recipient cells, thus ruling out an effect of TrwC or the conjugation process in the generation of these mutations.

These results were a direct evidence of the site-specific endonuclease activity of TrwC-Cas12a in the recipient cell after being translocated. In this way, we obtained proof of concept for the use of bacterial conjugation to deliver *in vivo* Cas endonucleases for targeted genetic modification of recipient bacteria.

The analysis of the *sacB* mutants allowed us to observe the mutation pattern produced by Cas12a-induced DSB repair in bacteria. 5 out of 11 mutants could not be analyzed, probably due to the presence of larger deletions which prevented PCR amplification of the *sacB* region. The rest contained 1-4 nt deletions at the cleavage site. Our results contrasted with the large deletions described by (Yan et al., 2017) when

Cas12a was used to target *lacZ*, or to the ones described for Cas9, which has also been reported to produce large deletions in *E. coli* as a consequence of DSB repair (Cui and Bikard, 2016). These mutations have been associated with *recA*-dependent recombination guided by microhomologies when no DNA template is present in the cell.

The mutations found in the *sacB* sequences resembled the result of NHEJ-mediated repair of DSB observed in Cas12a-induced mutations in eukaryotic cells (2-38 nt deletions; (Zetsche et al., 2015; Kim et al., 2016; Hu et al., 2017; Xu et al., 2017; Port et al., 2020)). In prokaryotic cells, similar pattern of Cas12a-induced mutations has been found in the actinobacteria *Amycolatopsis mediterranei* (Zhou et al., 2020); in this work the authors concluded that the mutations were the consequence of the existence of a NHEJ pathway. Yan and collaborators found bigger DNA deletions (11 to 214 nt) when they edited *M. tuberculosis* with Cas12a assisted with a NHEJ strategy (Yan et al., 2020). However, no such NHEJ pathway has been described in *E. coli*, and in fact, type II CRISPR-Cas systems seldom co-occur in bacteria with NHEJ (Bernheim et al., 2017). Irrespective of the explanation underlying the mutagenesis pattern of Cas12a, its ability to introduce a few nucleotide deletions, together with its lower toxicity, could be an advantage in order to obtain knock-out mutants in bacteria.

Surprisingly, when we analyzed the *sacB* sequence of transconjugants, we found in 3 of them a mixture of sequences. This mixture appears next to the Cas12a cleavage site and continues along the rest of the sequence (Figure 41). When we analyzed the chromatograms, we realized that it was the mixture of two sequences corresponding to two different deletions in the *sacB* sequence. Specifically, it was the deletion of one or two nucleotides in the sequence of transconjugants 2 and 5, or the deletion of 2 or 3 nucleotides in the transconjugant 3. The FD3 strain only contains a copy of the *sacB* gene (confirmed by the Bikard lab; unpublished observations). Therefore, the mixture of mutations could not be the result of different cleavage and repair events in different *sacB* copies. In all the cases, the transconjugants were replicated and analyzed from single colonies, therefore a mixture of populations carrying different mutations could not be the explanation either. At the moment, we have no explanation for the frequent occurrence of this mixture of sequences after Cas12a cleavage and repair. However, it suggests that rearrangements leading to tandem copies of the target region could be

frequent, similar to the tandem arrays observed upon integration of the T-DNA in or the AAV viral genome in the plant or human genome, respectively (Krizkova and Hrouda, 1998; De Buck et al., 2009; Henckaerts et al., 2009).

In summary, with these results we have proved that bacterial conjugation can be used to deliver Cas endonuclease to a recipient bacterium by fusing it to the relaxase, and use it to generate mutations in specific genes of the recipient cell. With this approach, we eliminated the necessity of expressing the endonuclease in the recipient cell, bypassing the off-target activity and toxicity problems.

We wanted to go one step further and prove that we could use conjugation to deliver the Cas endonuclease together with the homologous template DNA, in order to accomplish seamless targeted mutations in the recipient cell. Since the relaxase can mobilize any *oriT*-containing DNA, we can insert the homologous template needed for homologous recombination in the mobilizable plasmid. We constructed a mobilizable suicide plasmid carrying a homologous recombination template of the *sacB* gene carrying a point mutation in the PAM sequence which generated a premature STOP codon, and we mobilized it in parallel by TrwC or TrwC-Cas12a into a recipient cell carrying the targeting gRNA or without it. Our goal was to obtain colonies which were sucrose-resistant and Cm-sensitive, the expected phenotype for the double-recombinants incorporating the mutation. We analyzed such colonies from the different matings (Figure 44). Significantly, even though the number of sucrose-resistant colonies obtained was not very different under the different conditions tested, we found that the ratio of edited cell was much higher when TrwC-Cas12a and the gRNA were present: almost half of the colonies had incorporated the desired mutations, while in the absence of TrwC-Cas12a or in the absence of the gRNA, this fraction dropped to less than one in 15.

With these results we show the proof of concept for using relaxase-Cas fusion proteins to deliver the active endonuclease plus the DNA template covalently bound to them to a recipient cell. The observed increase in the efficiency of targeted mutagenesis may allow the use of Relaxase-Cas systems without the need for additional recombineering systems, usually combined with CRISPR-Cas edition of bacteria using HR templates in order to increase effectivity. With these systems, 65 % of the cells

recovered were edited cells when ssDNA was using as DNA template (Jiang et al., 2013). Part of the observed increase in efficiency could be attributable to the fact that the editing template would be translocated as ssDNA, which has been demonstrated to be better template than dsDNA for editing cells (Vento et al., 2019).

Relaxase-Cas proteins could be used for targeted genetic modifications of different prokaryotes, especially wild-type strains which are difficult to transform, and poorly characterized genera for which gene expression tools are underdeveloped. Considering the ability of broad-host-range conjugative systems to reach almost all Gram-negative, and even Gram-positive bacteria, conjugation could be widely used for the targeted genetic modification of prokaryotes.

Finally, it is important to note that relaxases can also deliver DNA into human cells through the T4SS of human pathogens. Thus, this same strategy could be adapted to deliver the CRISPR-Cas toolkits *in vivo* from bacteria into human cells. Also, the delivery of Relaxase-Cas:ssDNA complexes could increase the editing efficiency by bringing the recombinogenic DNA template close to the DSB. This has been demonstrated by Ali and collaborators, by expressing a VirD2-Cas9 fusion protein in plant cells, which increased 5 fold the editing efficiency compared to Cas9 protein (Ali et al., 2020).

During this work, we have generated Relaxase-Cas protein fusions for their expression in human cells. Apart from TrwC-Cas12a fusion protein, we constructed a fusion carrying the relaxase MobA by fusing Cas12a to the C-terminus of MobA, and a Cas12a-BID fusion protein, by fusing a BID signal recognized by *B. henselae* VirB/D4 T4SS. We expressed these fusion proteins in human cells and demonstrated that Cas12a maintained its gRNA-guided site-specific endonuclease activity in all the fusions (Figure 47). As we have previously discussed, the relaxase MobA can be translocated through the T4SS of several human pathogens (*B. henselae, L. pneumophila, C. burnetii*) with higher efficiency than TrwC (Guzmán-Herrador et al., 2017). The translocation of Cas12a fusion proteins to human cells by T4SS of bacterial pathogens could be used to guide site-specific mutagenesis into human genes. In order to test if this approach is possible, it would be necessary to confirm that the T4SS VirB/D4 is able to recognize these fusion proteins, as modifications of the C-terminal of TrwC could affect its translocation

through this T4SS (Alperi et al., 2013). That is the reason why Cas12a-BID was also generated, since the C-terminal BID domain has been shown to be an efficient recruiter of fused proteins to its cognate VirB/D4 T4SS.

We have demonstrated the possibility of using conjugative relaxases as a vehicle to deliver Cas nucleases into prokaryotic cells, and we have proved their functionality in both template- and non-template driven mutagenesis. With this delivery method, we eliminated the disadvantages of Cas expression from the target cell, or the necessity to purify proteins. The next step would be the determination of Relaxase-Cas recognition and translocation through the T4SS of *B. henselae* into a human cell. This possibility would have an important biotechnological application, as we could deliver Cas12a proteins into the human cell *in vivo*. Furthermore, the addition of signals recognized by different T4SS to the Cas protein would allow its translocation to different target cells.

6. Conclusions

6. Conclusions

1. The recombinase domain of TrwC, N600, behaves as a monomer alone or in the presence of its target ssDNA.
2. There is no correlation between the site-specific integrase activity of the relaxase TrwC and its oligomerization ability.
3. We have developed a conjugation protocol to transfer DNA from *E. coli* to *L. casei* using the conjugative systems of plasmids R388 and RP4.
4. We have described for the first time conjugative DNA transfer from *E. coli* to Gram-positive bacteria using the R388 conjugative system.
5. We have shown for the first time conjugative DNA transfer from *E. coli* to *L. casei*.
6. TrwC does not promote integration of the incoming DNA into the genome of the non-permissive host *L. casei*.
7. The relaxase MobA can be translocated through the T4SS VirB/D4 of *B. henselae* to a human cell. MobA is 5-10 times more efficient than TrwC in transferring DNA to human cells.
8. MobA relaxase promotes integration of the transferred DNA into human cells. TrwC integration ratio is five-fold higher than that of MobA.
9. We have generated and validated a TrwC-Cas12a fusion protein which can be expressed in *E. coli* and we have confirmed that both moieties of the protein were active.
10. TrwC-Cas12a can be translocated through the R388 T4SS into recipient bacteria.
11. The endonuclease activity of Cas12a in the bacterial chromosome induces the SOS response.
12. Upon translocation into the recipient cell, TrwC-Cas12a shows gRNA-guided endonuclease activity leading to site-specific mutations.
13. The mutations generated by the bacterial repair pathways after TrwC-Cas12a cleavage, were similar to the ones produced by Cas12a.

14. TrwC-Cas12a can be introduced by conjugation covalently linked to a DNA molecule providing a homology cassette to generate seamless mutations in the recipient genome, leading to an 8-fold increase in the gene targeting rate.
15. We have generated Cas12a-BID, TrwC-Cas12a and MobA-Cas12a fusion proteins which can be expressed in human cells. We have confirmed Cas12a site-specific endonuclease activity in all of the fusions with an efficiency similar to Cas12a.

7. Summary in Spanish

7. Summary in Spanish

7.1. Introducción

La conjugación bacteriana es un mecanismo de transferencia horizontal de ADN entre dos células que tienen que estar en contacto físico a través de un sistema de secreción tipo IV (SST4). Bajo condiciones de laboratorio se ha descrito conjugación bacteriana a levaduras, plantas e incluso células de mamífero (Lacroix and Citovsky, 2018). Durante este proceso, las relaxasas conjugativas son las proteínas encargadas de iniciar y terminar el procesamiento del ADN que va a ser transferido. Estas proteínas reconocen y actúan sobre una secuencia específica en el ADN a movilizar, el origen de transferencia (*oriT)* **(Figura 1)** (Getino and de la Cruz, 2019).

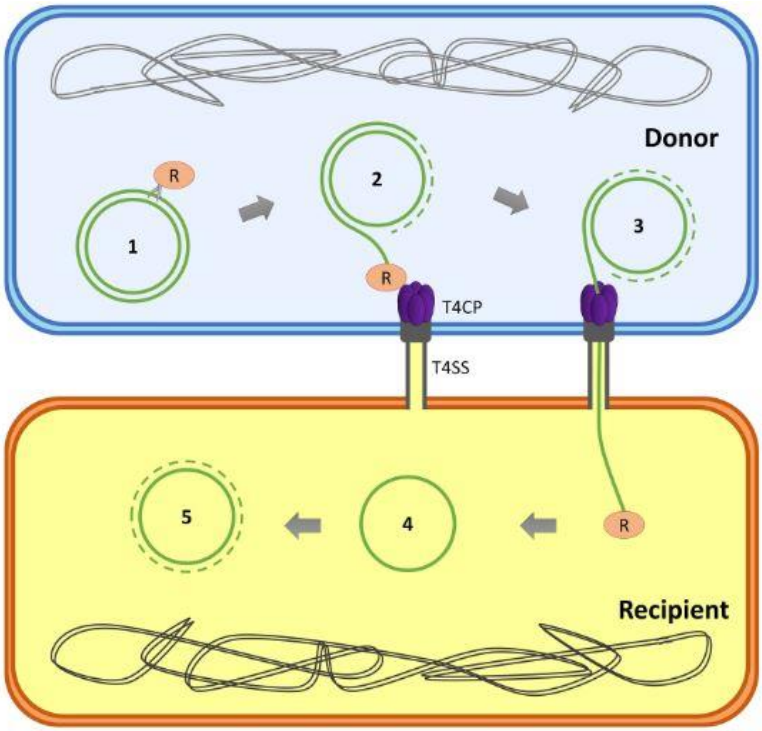


Figura 1. Esquema de la conjugación. (1) La relaxasa (R) corta el ADN plasmídico y forma intermediarios covalentes con el *oriT.*(2) El SST4 recluta al relaxosoma mediante la interacción de la proteína acopladora, mientras el ADN es replicado. (3) La relaxasa se libera de la cadena de ADN mediante un segundo corte y guía al ADN de cadena sencilla a través del SST4 ayudado por la proteína acopladora. (4) En la célula receptora, la relaxasa recirculariza la cadena de ADN. (5) El ADN de cadena sencilla transferido es replicado para dar lugar a un plásmido de cadena doble (Getino and de la Cruz, 2019).

Los sistemas de secreción tipo IV forman una familia de transportadores moleculares con una gran plasticidad en cuanto al sustrato transportado y el destino de este. Son capaces de translocar ADN, proteínas o complejos nucleoprotéicos al medio

extracelular o a una célula receptora (eucariota o procariota). Esta versatilidad hace que estén implicados en procesos muy variados, como la conjugación bacteriana o la translocación de efectores a células humanas durante la infección de patógenos bacterianos (Grohmann *et al*., 2018).

Las relaxasas conjugativas han sido clasificadas tradicionalmente dentro de la superfamilia de proteínas HUH (histidina-hidrofóbica-histidina), aunque cada vez las excepciones dentro de la familia son más habituales (Garcillán-Barcia *et al*., 2020). Además de estar involucradas en la conjugación, se ha visto que pueden llevar a cabo otras actividades, incluso en ausencia de la conjugación, como controlar el número de copias de un plásmido, la recombinación sitio-especifica o la integración sitio-específica. Actualmente se desconoce qué es lo que confiere a una relaxasa la capacidad de catalizar estas reacciones y qué posible papel biológico pueden tener estas actividades (Guzmán-Herrador and Llosa, 2019).

TrwC es la relaxasa del plásmido conjugativo R388. Aparte de su papel en conjugación, 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 a células procariotas, si existe en la célula receptora una copia del *oriT,* bien en un plásmido o en el cromosoma (Draper *et al*., 2005; Agúndez *et al*., 2012). Además, la reacción catalizada por TrwC permite cierta flexibilidad en la secuencia del *oriT* aceptor, habiéndose encontrado en el genoma de *E. coli* secuencias humanas con una elevada homología a la secuencia mínima del *oriT*, en las que TrwC es capaz de catalizar la integración (Agúndez *et al*., 2012). Además de catalizar la integración sitio-específica en bacterias, TrwC es capaz de promover la integración inespecífica del ADN movilizado en el genoma de células humanas (Gonzalez-Prieto *et al*., 2017).

Complejos nucleoprotéicos (ADN-relaxasa) han sido traslocados a través de SST4 involucrados en patogénesis a células humanas, en concreto a través del SST4 del patógeno *B. henselae*, VirB/D4. Estas relaxasas son Mob (Schröder *et al*., 2011), de un plásmido natural de *B. henselae* (pBGR1) y TrwC (Fernández-González *et al*., 2011). Los plásmidos movilizados contenían el *oriT* reconocido por cada relaxasa y el gen de la proteína verde fluorescente (GFP) bajo el control de un promotor eucariota. La transferencia de ADN se midió como porcentaje de GFP, de este modo, se detectaba la

entrada y expresión del ADN al núcleo de la célula humana. En ambos casos, los autores detectaron transferencia de ADN únicamente cuando la relaxasa se encontraba presente en el plásmido. Cuando la transferencia era llevada a cabo por Mob, se detectó un 0.02% de células GFP-positivas. En cambio, cuando la transferencia era mediada por TrwC, un 3% de las células eran positivas. Estas eficiencias se incrementaron cuando la señal BID (*Bartonella* efector protein intracelular delivery, necesaria para el reconocimiento de los efectores por parte del VirB/D4 de *B. henselae*) se añadió a la relaxasas (en el caso de Mob, la eficiencia se incrementó hasta 100 veces).

La relaxasa TrwC es activa en células humanas y promueve la integración del ADN movilizado en el cromosoma humano. González-Prieto y colaboradores estudiaron la capacidad de TrwC de integrar el ADN movilizado a través del SST4 de *B. henselae* en células humanas (Gonzalez-Prieto *et al*., 2017). Durante este trabajo analizaron en paralelo la actividad de integrar el ADN de TrwC con Mob:BID. A pesar de no observar actividad integrasa sitio-específica de TrwC, detectaron un incremento en el número de integrantes obtenidos cuando la reacción era mediada por TrwC. De este modo, mostraron que TrwC promovía la integración del ADN transferido en células humanas, aunque no de manera sitio específica, mientras que Mob no la promovía. Además, detectaron un evento de integración sitio-específico en una región que contenía 8 pb de homología con la secuencia mínima del *oriT*, dato que refuerza la idea de que los requerimientos en la secuencia aceptora son más laxos.

Los sistemas CRISPR-Cas han supuesto una revolución en las técnicas de edición genética debido a su fácil manejo y versatilidad (**Figura 2**). Sin embargo, cuentan con importantes limitaciones, como un sistema de envío *in vivo* a la célula diana. En procariotas, se utilizan principalmente como sistemas de envío técnicas de transformación o transducción. A pesar de ser efectivas, cuentan con las limitaciones propias de la técnica (hay bacterias difíciles de transformar, o la especificidad de los bacteriófagos genera la necesidad de adecuar el sistema de envío cada vez que se modifique la célula diana) (Ramachandran and Bikard, 2019). Además, el formato en el que se envía el sistema CRISPR-Cas es ADN, el cual puede producir problemas de sobreexpresión y toxicidad en la célula.

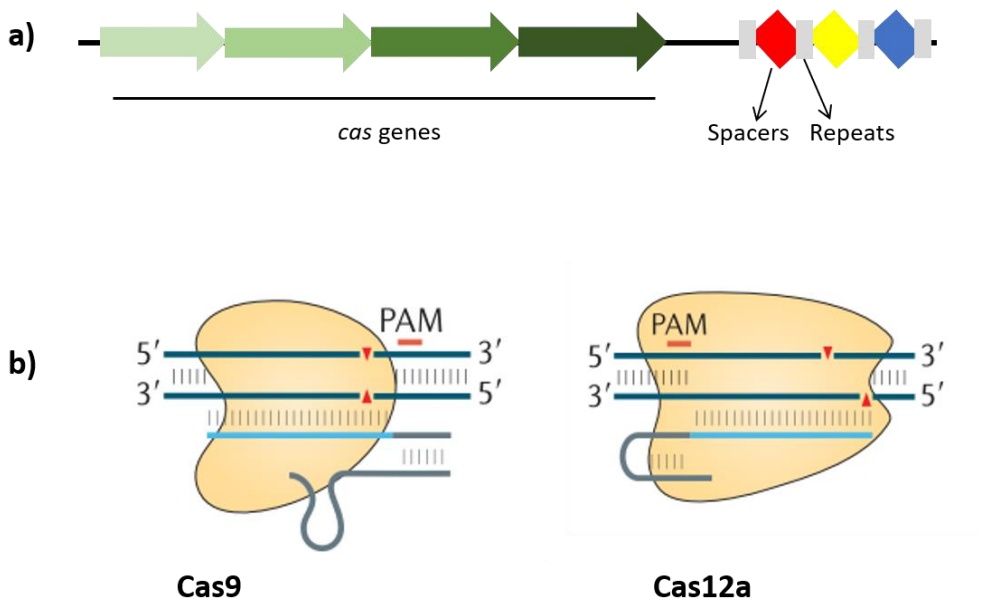


Figura 2. Esquema de los sistemas CRISPR-Cas. a) Estructura de un locus CRISPR. En azul, los genes *cas*, que codifican a las proteínas Cas. Las repeticiones y los espaciadores, que codifican para los diferentes ARNgs, están representadas como cuadrados grises y de colores, respectivamente. **b)** Complejo formado por el ADN y el sistema CRISPR-Cas9-ARNg (izquierda) y complejo formado por formado por el ADN y el sistema CRISPR-Cas12a-ARNg (derecha). La secuencia PAM (en español, Motivo Adyacente de Protoespaciador) es la secuencia reconocida por la proteína Cas en el ADN y está representada en amarillo. Los triángulos rojos señalan la posición de corte de la endonucleasa en el ADN diana. Modificada de (Jiang and Doudna, 2017) (Shmakov *et al*., 2017).

En eucariotas existen diferentes métodos de envío de los sistemas CRISPR a la célula diana. Dos de los más utilizados son los sistemas víricos como el AAV (virus adenoasociados) o los complejos formados por nanopartículas. El uso de los AAV tiene como principal limitación el tamaño del ADN que puede contener, así como el uso de ADN como formato de envío del sistema. Con respecto a los complejos formados por nanopartículas requieren la purificación previa de la proteína, con las dificultades y costos que esto conlleva, así como la inestabilidad de estos (Glass *et al*., 2018).

7.2. Objetivos

Los objetivos que nos planteamos para este trabajo fueron:

1. Estudiar y comparar la actividad integrasa de las relaxasas conjugativas. Los objetivos específicos para llevar a cabo este punto fueron:

a. Estudiar la relación entre la actividad integrasa sitio-específica y la habilidad de oligomerizar en presencia del sustrato diana de ADN de cadena sencilla de las relaxasas conjugativas, utilizando la relaxasa TrwC como modelo.
b. Estudiar el posible papel biológico de la integración mediada por las relaxasas en la colonización de huéspedes no permisivos tras la transferencia del ADN conjugativo.
c. Analizar y comparar la habilidad de diferentes relaxasas en promover la integración del ADN en células humanas.

2. Uso de relaxasas como sistemas de envío de proteínas/ADN con propósitos biotecnológicos.
    a. Utilizas la proteína de fusión TrwC-Cas12a como método de envío de Cas12a a células procariotas.
    b. Probar la actividad de TrwC-Cas12a una vez translocada a través del SST4 a una célula procariota receptora.
    c. Construir y validar las fusiones Cas12a en células humanas.

Con estas aproximaciones, nuestro objetivo es aumentar nuestro conocimiento sobre la capacidad integrasa de las relaxasas conjugativas y su implicación biológica. También pretendemos explorar su posible aplicación como vehículos de envío *in vivo* de ADN/proteínas a células receptoras difíciles de editar por otros métodos, en particular, su contribución a la edición genómica en combinación con los sistemas CRISPR-Cas.

7.3. Resultados

7.3.1. Estudio de la actividad integrasa en relaxasas conjugativas

7.3.1.1. Explorando la relación entre la actividad integrasa sitio-específica y la capacidad de oligomerizar de manera DNA-dependiente.

Actualmente se desconoce qué permite a una relaxasa catalizar la reacción de integración sitio-específica. Recientemente, Agúndez y colaboradores publicaron un trabajo que sugiere que la capacidad de oligomerización de la relaxasa puede jugar un papel importante en la catalización de la reacción de integración sitio-específica.

Muestran cómo la quimera formada por el dominio relaxasa N293 de TrwC (que se comporta como un monómero incluso cuando se une a su diana) y el dominio C-terminal de Rep68 (incluyendo la región OBD (Origin Binding Domain), necesaria para la oligomerización de Rep68), era capaz de catalizar la reacción de integración sitio-específica en bacterias, a pesar de que N293 no era capaz de realizarla por sí misma. Además, vieron que la capacidad de oligomerización de la quimera era diferente a la esperado, ya que formaba complejos oligoméricos diferentes a los formados por las proteínas parentales (Agúndez *et al*., 2018).

En este trabajo hemos estudiado la capacidad de oligomerización del dominio N600 de TrwC mediante sedimentación analítica. N600 había sido previamente descrito como el dominio mínimo capaz de catalizar la reacción de integración de forma eficiente y además había sido descrito como un monómero mediante cromatografía de gel, sin embargo no se había estudiado sus propiedades de oligomerización en presencia de su ADN diana (César *et al*., 2006; Agúndez *et al*., 2012). Para los ensayos, decidimos comparar el dominio N293 con el N600 en presencia y en ausencia de un oligonucleótido de cadena sencilla que contenía la secuencia del *oriT* 25+8. Las proteínas se purificaron siguiendo el protocolo descrito en (Boer *et al*., 2006; César *et al*., 2006). A lo largo del proceso realizamos geles de poliacrilamida SDS-PAGE para comprobar que el proceso de purificación se estaba llevando a cabo correctamente. Una vez purificados, se llevaron a cabo los experimentos de centrifugación analítica se realizaron en colaboración con el grupo del Dr. Germán Rivas en el CIB (Madrid), para determinar los coeficientes de sedimentación de los dominios, los cuales dan información acerca del estado de oligomerización de las muestras. Se realizaron mediciones con las proteínas y el oligonucleótido de forma independiente, y a continuación tras incubar proteína y oligonucleótido (en estas muestras probamos diferentes ratios de oligonucleótido y proteínas).

En el caso del dominio relaxasa N293, los resultados fueron los esperados (**Figura 3**). Cuando la proteína se analizó sola, el coeficiente de sedimentación obtenido fue de 2.1S, correspondiente a un monómero. Cuando se analizó la proteína tras ser incubada con el oligonucleótido codificante para *oriT* 25+8, los valores obtenidos se encontraban

entre 3-3.1S, valores correspondientes al complejo formado por una molécula de proteína y una molécula de ADN.

En el caso del dominio N600, el coeficiente de sedimentación obtenido fue 2.8S, confirmando su comportamiento como un monómero. Cuando se analizaron las muestras de la proteína incubada con el oligonucleótido, los valores obtenidos fueron 3.7-3.9S (Figura 3). Estos valores coinciden con la formación de complejos formados por una molécula de ADN y una molécula de proteína, es decir, N600 se comportaba como un monómero en presencia de su diana.

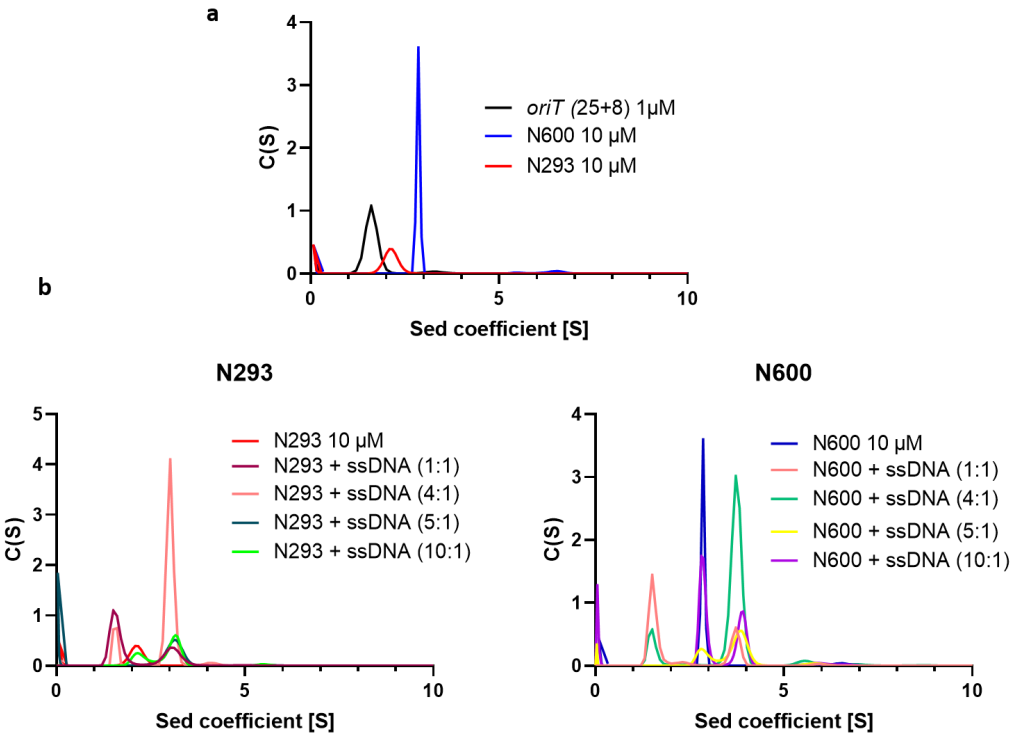


Figura 3. Análisis de la velocidad de sedimentación de los dominios N293 y N600 y del *oriT* 25+8 diana. **a)** Perfiles de sedimentación del oligonucleótido *oriT* 25+8 (1µM), N293 (10µM) y N600 (10µM). **b)** Perfiles de sedimentación de N293 (izquierda) y N600 (derecha) incubados con su sustrato específico, el oligonucleótido *oriT*(25+8). Se probaron los ratios 1:1, 4:1, 5:1 y 10:1 (proteína:oligonucleótido). (S); coeficiente de sedimentación. Los datos se obtuvieron utilizando el programa SEDFIT.

7.3.1.2. Posible papel biológico de la actividad integrasa en las relaxasas conjugativas.

Con el fin de estudiar el posible papel biológico de la actividad integrasa en las relaxasas, decidimos analizar la actividad de la relaxasa TrwC (con actividad integrasa sitio-específica en procariotas y capaz de promover la integración inespecífica en células humanas), y compararla con RP4_TraI (sin esta actividad). Nuestra hipótesis es que dicha

actividad pudiese servir para colonizar huéspedes no permisivos, donde la única forma de permanencia del plásmido tras la transferencia pasaría por su integración en el genoma huésped. Como bacteria receptora no permisiva elegimos a *L. casei*¸ una bacteria Gram-positiva perteneciente al grupo de las bacterias del ácido láctico o BAL, con un importante interés biotecnológico.

Antes de estudiar la actividad integrasa, fue necesario desarrollar un protocolo de conjugación de *E. coli* a *L. casei* utilizando ambos sistemas conjugativos, ya que no había ninguno descrito hasta el momento. Tras optimizar el protocolo, fuimos capaces de movilizar plásmidos anfibios, capaces de replicar en ambas bacterias, utilizando ambos sistemas de conjugación. Las frecuencias de conjugación (calculadas como transconjugantes/donadores) obtenidas fueron $1.17x10^{-6}$ para R388 y $2.76x10^{-5}$ para RP4. Los transconjugantes fueron analizados, confirmando que contenían el plásmido movilizado y que se trataban de *L. casei.*

A continuación, pasamos a realizar los ensayos de integración utilizando plásmidos suicidas, capaces de replicar en *E. coli* pero no en *L. casei*. Tras realizar los ensayos de integración, con ambas relaxasas obtuvimos integrantes (la frecuencia de integración con TrwC fue $1.24x10^{-8}$ mientras que para TraI fue $2.73x10^{-8}$) (**Figura 4)**.

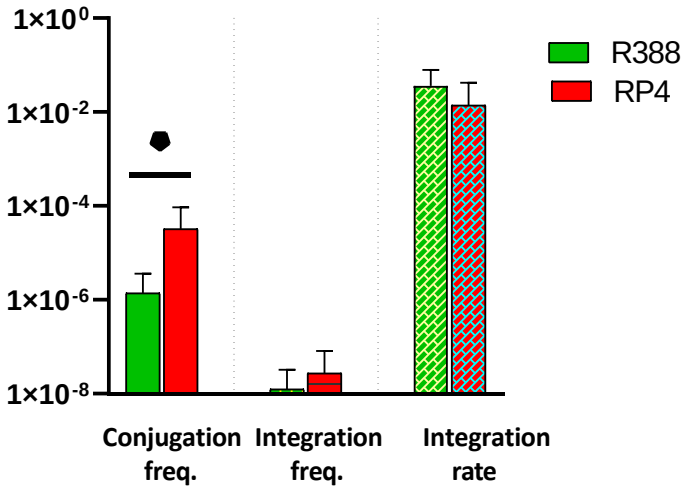


Figura 4. Ensayos de integración desde *E. coli* a *L. casei* utilizando ambos sistemas. La gráfica muestra las frecuencias de conjugación e integración calculadas como transconjugantes (o integrantes) por donador. El ratio de integración se calculó como frecuencia de integración por frecuencia de conjugación. Los datos representan la media de al menos 3 experimentos independientes. *, $p<0.05$.

Además de no observar diferencias significativas entre ambos sistemas, cuando los analizamos observamos que no se trataban de eventos de integración mediados por las relaxasas, sino que se trataba de eventos de integración al azar, ya que todos mantenían la secuencia del *oriT* completa.

7.3.1.3. Análisis de la habilidad de promover la integración del ADN en células humanas de diferentes relaxasas.

González-Prieto y colaboradores demostraron previamente que TrwC era capaz de promover la integración del ADN transferido en el genoma de la célula humana una vez translocada a través del SST4 de *B. henselae* VirB/D4 (Gonzalez-Prieto *et al*., 2017). Durante este trabajo hemos querido comprobar si la capacidad de promover la integración del ADN en células humanas está extendida a otras relaxasas o no. Para ello, elegimos la relaxasa MobA, del plásmido RSF1010. En primer lugar, comprobamos si el SST4 de *B. henselae* VirB/D4 era capaz de translocar a la relaxasa MobA unida a ADN a una célula humana. Para ello realizamos experimentos de transferencia de ADN como los mencionados en la introducción. En paralelo ensayamos plásmidos movilizados por MobA y por TrwC. En ambos casos detectamos transferencia de ADN en los tipos celulares analizados (HeLa y EAhy.926). El porcentaje de células positivas para la expresión de la GFP para MobA fue de 5.72%, en células EA.hy926 y 2% en células HeLa. En el caso de TrwC, fue de 1% en células EA.hy926 y 0.20% HeLa. De este modo, pudimos determinar que MobA también era reconocida y translocada unida al ADN a través del SST4 de *B. henselae.* Además, la transferencia de ADN era más eficiente con MobA que con TrwC.

A continuación, estudiamos la capacidad de promover la integración de ambas relaxasas en el genoma humano. Para estos ensayos añadimos también a la relaxasa Mob con la señal BID fusionada (Mob:BID), previamente descrita como relaxasa que no promueve la integración en el genoma humano (Gonzalez-Prieto *et al*., 2017). Para el estudio de la integración, a todos los plásmidos utilizados para el estudio de la transferencia de ADN se les añadió un gen de resistencia al antibiótico higromicina (excepto para Mob:BID que se utilizó neomicina), el cual permitiría seleccionar integrantes estables. Estos plásmidos se introdujeron en *B. henselae* y se llevaron a cabo

infecciones utilizando en este caso la línea celular HeLa. En paralelo con los ensayos de integración, se realizaron ensayos de transferencia de ADN.

Como era de esperar, los resultados de los ensayos de integración mostraron que Mob:BID no era capaz de promover la integración del ADN, mientras que TrwC sí lo era. Además, mostraron que MobA promovía la integración, aunque con menor eficiencia que TrwC (5 veces menos aproximadamente) (**Figura 5**).

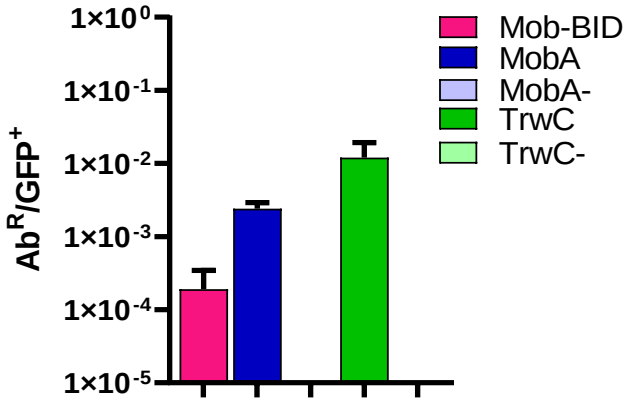


Figura 5. Ratio de integración del AND de las diferentes relaxasas. La gráfica muestra el ratio entre el número de células HeLa resistentes a higromicina y el número de células HeLa GFP positiva. Cada relaxasa se muestra con un color diferente.Ab^R: Higromicina para los ensayos con TrwC y MobA, neomicina para los ensayos con Mob:BID. Los datos son el resultado de dos experimentos independientes.

7.3.2. Uso de relaxasas como vehículos de envío de ADN/proteínas con fines biotecnológicos

La segunda parte del trabajo se basó en el estudio y desarrollo de una herramienta biotecnológica utilizando a las relaxasas como vehículos para transportar ADN o proteínas a células receptoras, ya que pueden ser translocadas a células procariotas y eucariotas a través de SST4. Para ello decidimos combinar el uso de las relaxasas conjugativas con los sistemas CRISPR-Cas, muy utilizados en edición genética. Para el desarrollo de esta parte elegimos el sistema CRISPR-Cas12a, en concreto la variante AsCas12a.

7.3.2.1. Generación y validación de la construcción TrwC-Cas12a en células procariotas.

En primer lugar, generamos la proteína de fusión TrwC-Cas12a para su correcta expresión en células procariotas. Fusionamos al extremo N-terminal de *trwC* (eliminando su codón de STOP) el gen de *cas12a,* y tras probar diferentes promotores, lo clonamos bajo el promotor de expresión P*tet*, el cual permitió una sobrexpresión de la proteína muy controlada. La estabilidad de la fusión fue validada por western blot. Los resultados mostraron que a pesar de la inestabilidad de la proteína (se observó degradación de la misma), éramos capaces de producirla en células procariotas.

A continuación, validamos la actividad de las diferentes partes de TrwC-Cas12a. La fusión mantuvo la actividad de Cas12a en procariotas generando una disminución drástica en el número de bacterias cuando plásmidos codificando TrwC-Cas12a y el ARN guía (ARNg) contra el gen diana *lacZ* fueron co-electroporados en la bacteria diana. Además, esta disminución fue observada únicamente en condiciones de inducción, confirmando el control de la expresión de la proteína de fusión.

La proteína de fusión también mostró frecuencias de conjugación similares a las obtenidas para TrwC, en la complementación de un plásmido R388 deficiente en TrwC. Con estos resultados confirmamos el mantenimiento de la actividad TrwC. Sin embargo, debido a la degradación de la proteína observada en el western blot, este resultado no pudo confirmar la translocación de la proteína de fusión completa a través del SST4 a la célula receptora.

7.3.2.2. Validación de la actividad Cas12a tras ser translocada a través del SST4.

Uno de los puntos más importantes del trabajo era demostrar que la proteína era translocada a través del SST4 y, además, que ésta mantenía su actividad en la célula receptora. Para ello, decidimos detectar la actividad de Cas12a en la célula receptora de dos formas diferentes. La **Figura 6** resume ambas aproximaciones.

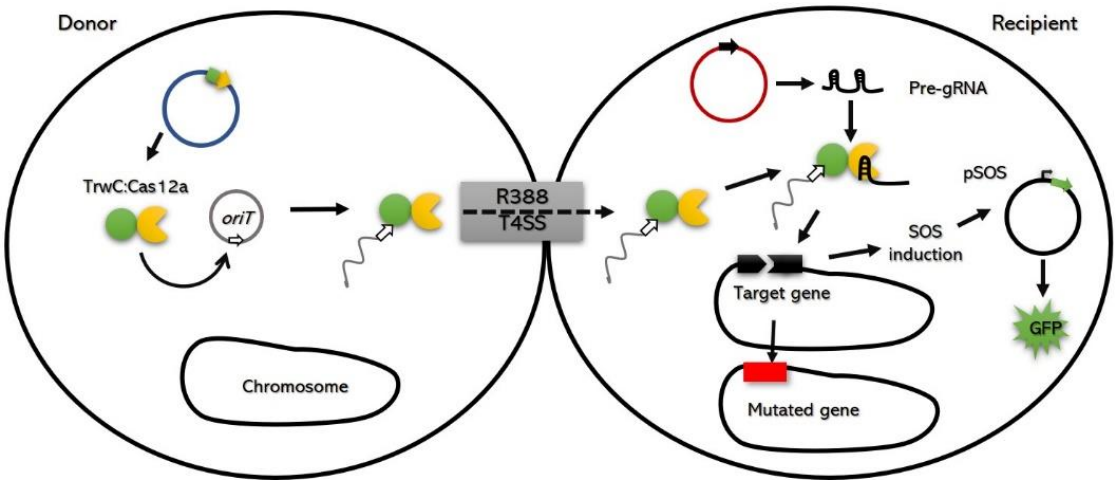


Figura 6. Representación esquematizada de los ensayos de la actividad de Cas12a en la célula receptora. En la célula donadora, el pLG24 (en azul) producirá TrwC-Cas12a. Gracias a su actividad relaxasa, la proteína de fusión cortará y se unirá covalentemente al *oriT* (flecha blanca), y el complejo será reclutado y translocado a la célula receptora a través del SST4. En la célula receptora, los plásmidos pLG15 o pLG19 (plásmido en rojo), producirán el ARNg con diana en un gen cromosómico. Debido a su actividad de endonucleasa sitio-específica, TrwC-Cas12a procesará el ARNg y ambos formarán un complejo. Este complejo será guiado al gen diana, donde TrwC-Cas12a producirá un corte de doble cadena. Este corte activará la señal SOS, la cual inducirá el promotor SOS del plásmido pZA31-sulA-GFP (en verde), y producirá GFP. La rotura de la doble cadena será reparada por las vías de reparación bacterianas, generando mutaciones en el gen (en rojo).

En primer lugar, de manera indirecta, detectando un incremento de la señal SOS bacteriana como consecuencia de la generación de roturas de la doble cadena de ADN debido al corte de Cas12a en el cromosoma bacteriano. Para ello, adaptamos los experimentos realizados por Lun Cui y David Bikard en los que demostraron que los cortes producidos por Cas9 en el cromosoma bacteriano inducían la respuesta SOS (Cui and Bikard, 2016). Durante estos experimentos se realizaron conjugaciones translocando TrwC-Cas12a a una célula receptora que contenía el plásmido pZA31-sulAGFP. Este plásmido contenía la proteína GFP bajo un promotor inducible por la respuesta SOS. De este modo, si la señal SOS se inducía, lo detectaríamos por un incremento en la expresión de GFP. Además de este plásmido reportero, las células receptoras contenían un ARNg con diana en un gen cromosómico (*lacZ*), o sin diana. La señal GFP se midió tras realizar los experimentos de conjugación y únicamente cuando TrwC-Cas12a se translocó a una célula receptora con el ARNg diana contra el gen

cromosómico *lacZ*, se detectó un incremento significativo en la expresión de la GFP. Demostrando de forma indirecta la actividad de Cas12a en la célula receptora.

Por otro lado, decidimos detectar las mutaciones producidas por los sistemas de reparación celular de la bacteria receptora tras el corte de Cas12a en el cromosoma bacteriano. Para ello, elegimos como diana el gen *sacB*. La expresión de este gen en presencia de sacarosa produce la muerte de la bacteria. Si Cas12a corta en este gen, y los sistemas de reparación celular producen una mutación en él, seríamos capaces de seleccionar estos mutantes ya que serían resistentes a sacarosa. De este modo, realizamos conjugaciones bacterianas translocando TrwC o TrwC-Cas12a a la receptora MG1655::*sacB*, la cual contenía una copia del gen *sacB* en su cromosoma. Además, esta receptora contenía un plásmido que codificaba un ARNg contra *sacB*. Únicamente cuando TrwC-Cas12a fue translocada, obtuvimos transconjugantes resistentes a sacarosa. La secuencia de *sacB* de estos transconjugantes fue analizada. De los 11 obtenidos, para 5 no obtuvimos ampliación de la región, probablemente debido a deleciones de gran tamaño en la zona amplificada. De los 6 restantes, en 5 de ellos pudimos detectar deleciones de 2-4 pb en la región de corte de Cas12a (**Figura 7**). Estos resultados confirmaron la actividad de Cas12a en la célula receptora una vez translocada a través del SST4.

Figura 7. Análisis de las mutaciones en *sacB* de las colonias resistentes a sacarosa. **a)** Alineamiento de la región de *sacB* cercana a la PAM y a los sitios diana de Cas12a-ARNg. La región de *sacB* de los transconjugantes resistentes a sacarosa (TC1-TC6) y del transformante resistente a sacarosa (TF1) fue amplificada por PCR. La secuencia *sacB* de la cepa MG1655::*sacB* se muestra en la parte superior (WT). La secuencia PAM y la secuencia espaciadora se encuentran en la parte superior. El triángulo rojo señala el sitio de corte de Cas12a. Los nucleótidos en rojo señalan las zonas donde aparecen dos secuencias de ADN. **b)** Secuencia de aminoácidos de las variaciones de *sacB* resultantes de las diferentes mutaciones. Las deleciones están indicadas a la izquierda en azul. Los aminoácidos en verde y subrayados son codificados por los nucleótidos de la PAM. Los codones de parada están señalados como *.

7.3.2.3. Incorporación de mutaciones sitio-específicas utilizando un casete de recombinación homóloga

Finalmente quisimos determinar si con nuestro sistema podíamos editar células receptoras utilizando un casete de recombinación homóloga con una mutación determinada. Para ello generamos el plásmido pLG27, un vector movilizable (contiene un *oriT*$_{w}$) y suicida (necesita la proteína Pir para replicar) con un casete de recombinación homóloga del gen *sacB*. Este casete contenía las primeras 430 pb de la región 5' del gen *sacB*, las cuales incluyen la secuencia diana del plásmido pLG19 (el cual codifica un ARNg contra *sacB*). Además, la PAM había sido mutada, generando un codón de stop prematuro y evitando ser cortada por Cas12a. A ambos lados de la PAM se construyeron un brazo izquierdo de homología con 277 pb y un brazo derecho de homología con 150

pb. Este plásmido se movilizó de forma paralela con TrwC o TrwC-Cas12a a la receptora MG1655::*sacB* la cual contenía o el pLG19 (+$sacB_{gRNA}$) o un plásmido vacío (-$sacB_{gRNA}$). De este modo, en los ensayos en los que TrwC-Cas12a y el ARNg estaban presentes, seríamos capaces de detectar un incremento en el número de células editadas (las cuales serán sacarosa resistentes).

Tras realizar los experimentos, para estudiar el número de bacterias que habían incorporado la mutación en *sacB* seleccionamos receptoras resistentes a sacarosa y las analizamos. 90 colonias para cada condición (30 colonias obtenidas en los ensayos de TrwC + ARNg, 30 de los ensayos de TrwC-Cas12a -ARNg y 30 de TrwC-Cas12a +ARNg) fueron analizadas. De las 90 colonias, analizamos qué colonias eran Cm^S, y determinamos la secuencia de *sacB* que contenían, para comprobar si habían incorporado la mutación deseada. La **Figura 8** muestra los ratios de edición, calculados dividiendo el número de colonias Cm^s que habían incorporado la mutación entre el número de colonias totales de cada ensayo.

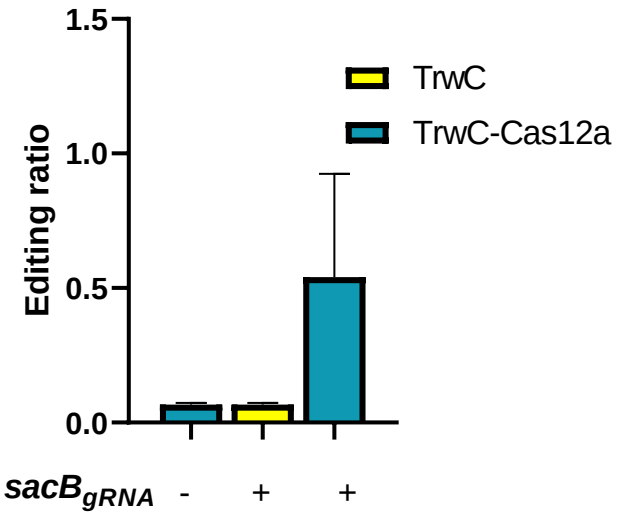


Figura 8. Ratios de edición de TrwC y TrwC-Cas12a. Los ratios de edición se calcularon como el número de colonias editadas, sensibles a cloranfenicol dividido entre el número de colonias totales. La presencia del ARNg contra *sacB* en la célula receptora está indicado como + o -. Los datos representan los resultados de dos y cuatro experimentos independientes

Observamos que cuando TrwC-Cas12a y el ARNg con diana en *sacB* estaban presentes, el porcentaje de células editadas era más de 8 veces mayor que cuando TrwC-Cas12a sin el ARNg o TrwC eran ensayadas.

7.3.2.4. Generación y validación de fusiones Cas12a en células humanas

Tras comprobar que TrwC-Cas12a podía ser translocada a través del SST4 y que la proteína mantenía su actividad, quisimos extender el uso de esta proteína a células eucariotas, ya que TrwC puede ser translocada a través de VirB/D4 de *B. henselae* a células humanas (Fernández-González *et al*., 2011; Gonzalez-Prieto *et al*., 2017). Para ello, generamos una proteína de fusión TrwC-Cas12a para su expresión en células humanas. La fusión se construyó como la diseñada en procariotas, y se comprobó su estabilidad por western blot, obteniendo una mayor estabilidad que en procariotas.

Además de la fusión TrwC-Cas12a, también generamos las proteínas de fusión MobA-Cas12a y Cas12a-BID. MobA-Cas12a se generó igual que TrwC-Cas12a pero sustituyendo la secuencia de *trwC* por la de m*obA*. Como hemos descrito anteriormente, MobA es reconocida y translocada a células humanas por VirB/D4 con mayor eficiencia que TrwC, por lo que decidimos generar esta fusión para determinar su actividad. La fusión Cas12a-BID se generó añadiendo al extremo C-terminal de *cas12a* la secuencia BID reconocida por el SST4 de *B. henselae*. Esta proteína sería reconocida por VirB/D4 como una proteína efectora y translocada a la célula eucariota.

Validamos la actividad de Cas12a en células humanas en las diferentes construcciones. Tras co-transfectar células humanas con los plásmidos que codificaban para las diferentes fusiones y los ARNg con diana en el gen *dnmt1* en células HEK293T, confirmamos que todas las fusiones eran activas en células humanas.

7.4. Discusión

Las relaxasas conjugativas son enzimas con diferentes dominios y están bien caracterizadas. A pesar de haber sido clasificadas dentro de la superfamilia de las endonucleasas HUH, están surgiendo muchas excepciones entre sus miembros. Además de estar implicadas en la conjugación, se ha descrito que las relaxasas realizan otras

reacciones, como la integración sitio-específica. Estas otras funciones han sido validadas en muchos sistemas diferentes, lo que sugiere que podrían desempeñar un papel biológico importante, y que no se trata sólo de una observación *in vitro*. El paradigma está cambiando y las relaxasas podrían no sólo ser enzimas clave para la transferencia horizontal de genes, sino también estar implicadas en otras funciones que contribuyen a la plasticidad genética de los procariotas. Actualmente se desconoce por qué hay relaxasas capaces de catalizar la reacción de integración y otras no. Tampoco se sabe qué convierte a una relaxasa en integrasa ni el posible papel biológico que puede tener esta actividad (Guzmán-Herrador and Llosa, 2019).

A lo largo de este trabajo hemos intentado dar respuesta a estas preguntas. En primer lugar, hemos intentado establecer una relación entre la capacidad de oligomerización de las relaxasas y su habilidad de catalizar la reacción de integración sitio-específica, basándonos en los resultados previos mostrados por Agúndez y colaboradores sobre la proteína quimera N293-Rep68 (Agúndez *et al*., 2018). Sin embargo, tras estudiar los coeficientes de sedimentación de los dominios N293 (sin actividad integrasa) y N600 (con actividad integrasa), en ausencia y presencia de un oligonucleótido que contenía la secuencia diana del *oriT* (25+8), no encontramos diferencias entre ambos. Los dos mostraron formar complejos compuestos por una molécula de ADN y una molécula de proteína, es decir, eran monómeros. De este modo, podemos concluir que la oligomerización, al menos en las condiciones probadas no parece tener una relación con la actividad integrasa de las relaxasas conjugativas.

A continuación, decidimos estudiar el posible papel biológico de esta actividad. Para ello, nuestra hipótesis fue que, mediante la integración del ADN transferido en el genoma receptor, los plásmidos conjugativos podrían colonizar huéspedes no permisivos en los que no pueden replicar, permitiendo establecerse en un mayor número de huéspedes. Para probar nuestra hipótesis decidimos realizar ensayos de integración comparando la relaxasa TrwC con Rp4_TraI (sin actividad integrasa), y utilizando como cepa receptora la bacteria Gram-positiva *L. casei,* incluida dentro las Bacterias del Ácido Láctico, con un importante interés biotecnológico. En primer lugar, pusimos a punto un protocolo de movilización del ADN con ambos sistemas desde *E. coli*

a *L. casei* utilizando plásmidos anfibios capaces de replicar en las dos bacterias. En ambos casos obtuvimos resultados positivos, mostrando por primera vez movilización de ADN de R388 a Gram-positivos. Además, ha sido también la primera vez que se ha mostrado movilización de ambos sistemas desde *E. coli* a *L. casei.* Este resultado es per se muy interesante, puesto que abre una vía de manipulación genética de unas bacterias con gran interés biotecnológico y biomédico. De hecho, la continuación de esta línea de trabajo por otros miembros de nuestro grupo ha mostrado la posibilidad de utilizar este protocolo para transferir ADN a especies y cepas de Lactobacillus no transformables, e incluso al patógeno humano emergente *S. epidermidis* (Samperio *et al*., 2021).

Tras establecer el protocolo de conjugación, realizamos ensayos de integración con ambos sistemas utilizando plásmidos suicidas (que no replicaban en *L. casei*). Tanto con TrwC como con TraI obtuvimos integrantes, sin embargo, cuando los analizamos se trataban de eventos de integración independientes de la relaxasa ya que mantenían las secuencias de los *oriT* completas. Estos resultados sugieren que a pesar de que TrwC promueve la integración inespecífica del ADN en células humanas, en bacterias este fenómeno no es observado. De este modo, al menos con este modelo, no hemos podido atribuir a la habilidad integrasa el papel biológico de colonizar huéspedes no permisivos.

Finalmente, hemos estudiado si la capacidad de promover la integración de las relaxasas en células humanas es una capacidad única de TrwC o está extendida entre otras relaxasas. Para ello decidimos utilizar la relaxasa MobA, del plásmido RSF1010. Esta relaxasa ha sido translocada por diferentes T4SS, por lo que nos pareció una buena candidata. En primer lugar comprobamos que MobA era capaz de transferir DNA a través de VirB/D4 a células humanas (tanto EA.hy926 como HeLa). Los resultados mostraron que era translocada y que, además, su eficiencia de transferir DNA era mayor que con TrwC. Estos datos, junto con los trabajos publicados previamente (Fernández-González *et al*., 2011; Schröder *et al*., 2011), sugieren que las eficiencias de transferencia de ADN dependen en gran medida de la eficiencia con la que la relaxasa es reconocida y translocada por el SST4. Una vez comprobado que era posible transferir ADN a una célula humana utilizando la relaxasa MobA, estudiamos su capacidad de promover la integración del ADN, comparándola con TrwC y Mob:BID. Los resultados mostraron que

MobA era capaz de promover la integración, aunque en menor medida que TrwC. De este modo, hemos mostrado que la capacidad de promover la integración del ADN transferido en células eucariotas puede expandirse a otras relaxasas además de TrwC. A pesar de que se desconoce qué capacita a una relaxasa a realizar la reacción de integración, debe residir en alguna característica intrínseca de la misma, ya que es la única proteína que pasa a la célula eucariota (que se conozca actualmente).

Por último, hemos querido desarrollar una herramienta biotecnológica teniendo en cuenta las características de TrwC. Nuestro objetivo era utilizar las relaxasas conjugativas como vehículos de envío de proteínas y ADN a células (tanto procariotas, mediante la conjugación, como a eucariotas mediante el uso de SST4 involucrados en patogénesis). Esta herramienta permitiría el envío *in vivo* de proteínas y ADN a células receptoras. El sistema CRISPR-Cas es una herramienta de edición genética revolucionaria, utilizada ampliamente en eucariotas, y cada vez más en procariotas. Sin embargo, una de las limitaciones de esta técnica es su envío a la célula deseada (Glass *et al*., 2018). De este modo, hemos combinado ambos sistemas generando la proteína de fusión TrwC-Cas12a. Hemos construido y validado el sistema en células procariotas, así como hemos asentado las bases para en un futuro poder desarrollarlo en células eucariotas.

En primer lugar, construimos una proteína de fusión TrwC-Cas12a. A pesar de ser inestable y observar productos de degradación, hemos conseguido producirla en procariotas. Además, mantiene la actividad de los parentales, tanto de TrwC como de Cas12a. Uno de los puntos clave era comprobar si la proteína era capaz de ser translocada a través del SST4 y una vez ahí, mantener su actividad. Para ello, realizamos dos ensayos diferentes de validación de la actividad de la fusión en la célula receptora. En ambos casos fuimos capaces de detectar la actividad Cas12a en la célula receptora. Además, hemos comprobado que los cortes producidos por Cas12a en la célula receptora, al igual que Cas9, produce una activación de la señal SOS. También hemos podido analizar las mutaciones generadas como consecuencia de la reparación de los cortes de Cas12a en el cromosoma bacteriano. Hemos detectado deleciones de 2-4 pares de bases en la zona descrita como sitio de corte de Cas12a. Además, hemos

comparado el tipo de mutación producida tras la reparación del corte con TrwC-Cas12a con Cas12a, obteniendo resultados similares. De este modo, hemos sido capaces de detectar la actividad de Cas12a en la célula receptora, mostrando que la proteína de fusión es translocada por completo a través del SST4 y que, además, mantiene las funciones de los parentales.

Finalmente, quisimos determinar si nuestro sistema permitía la translocación de un casete de recombinación homóloga covalentemente unido a la proteína de fusión para la incorporación de mutaciones específicas en la célula receptora. Para ello generamos un casete de recombinación homóloga con *sacB* que contenía 430 pb de la región 5' del gen *sacB*. Además, esta región contenía la diana del ARNg producido por el plásmido pLG19, pero con la PAM mutada, generando un codón de STOP prematuro y evitando e futuros cortes de Cas12a. Tras realizar los ensayos movilizando en paralelo el plásmido con el casete de recombinación con TrwC o TrwC-Cas12a, analizamos las colonias receptoras resistentes a sacarosa. Cuando analizamos la secuencia, la eficiencia de edición de TrwC-Cas12a era más de ocho veces superior a la eficiencia de edición producida por los sistemas de recombinación celular (en ausencia del sistema CRISPR-Cas). Estos datos son muy alentadores en el futuro uso de la proteína de fusión TrwC-Cas12a debido a que permite el envío simultáneo del sistema junto al molde de edición a la célula receptora.

De este modo, hemos generado una herramienta capaz de translocar a la proteína Cas *in vivo* a la célula procariota a editar, así como hemos sido capaces de modificar las células receptoras con la mutación deseada mediante la movilización de un casete de recombinación homóloga. Con este método, disminuimos ciertas limitaciones de la técnica CRISPR-Cas, ya que conseguimos introducir la propia proteína, evitando su producción en la célula diana, con los problemas de toxicidad y actividad off-target que esto conlleva. Además, el uso de relaxasas conjugativas permite la translocación simultánea del sistema junto al ADN molde deseado, necesario para llevar a cabo mutaciones dirigidas.

El uso de la conjugación como mecanismo de translocación del sistema supone una serie de ventajas importantes como es la posibilidad de enviar la proteína Cas *in vivo* a

bacterias difíciles o imposibles de transformar. La edición genética de procariotas está muy enfocada al uso de cepas silvestres, las cuales suelen acarrear limitaciones a la hora de ser transformadas. Con la conjugación, el rango de procariotas al que podemos llegar es mucho mayor, como hemos demostrado recientemente enviando ADN por conjugación a *L. parabuchneri,* una bacteria del ácido láctico, hasta ahora imposible de transformar, con un interés biológico importante (Samperio *et al*., 2021).

Una vez demostrado que el sistema es funcional en procariotas, decidimos ponerlo a punto en células humanas. En este caso, generamos no solo la fusión TrwC-Cas12a, sino que también generamos la fusión MobA-Cas12a, ya que previamente habíamos visto que era mejor transferida a células humanas que TrwC. Además, generamos la fusión Cas12a-BID. Esta fusión, sería reconocida por VirB/D4 como un efector. Todas las fusiones mantenían la actividad de la proteína Cas en células humanas. Nuestro próximo paso será determinar, como hemos hecho en procariotas, si estas proteínas son translocadas por el T4SS y mostrar su actividad en la célula humana de destino. El hecho de que hayamos demostrado que TrwC-Cas es translocada por el T4SS de R388 y recupera su actividad Cas en la bacteria receptora, junto con las evidencias existentes de la actividad de TrwC en las células humanas tras ser translocada a través del SST4 VirB/D4 (Gonzalez-Prieto *et al*., 2017), nos llevan a ser optimistas en este sentido.

El uso de este sistema en células eucariotas puede suponer un avance importante en la técnica CRISPR-Cas. El uso de relaxasas como sistema de envío (o de la señal BID en el caso de la fusión Cas12a-BID), permite la translocación de la proteína a la célula eucariota con las ventajas que todo esto conlleva, como la eliminación de intermediarios de ADN y la sobrexpresión de la proteína en la célula, que pueden dar lugar a actividades inespecíficas o problemas tóxicos.

7.5. Conclusiones

1. El dominio recombinasa de TrwC, N600 se comporta como un monómero, sólo o en presencia de ADN de cadena sencilla diana.
2. No hay una correlación entre la actividad sitio-específica de la relaxasa TrwC y su capacidad de oligomerización.

3. Hemos desarrollado un protocolo de conjugación para transferir ADN de *E. coli* a *L. casei* usando los sistemas conjugativos de los plásmidos R388 y RP4.
4. Hemos descrito por primera vez transferencia conjugativa de ADN desde *E. coli* a una bacteria Gram-positiva utilizando el sistema conjugativo R388.
5. Hemos mostrado por primera vez transferencia conjugativa de ADN desde *E. coli* a *L. casei.*
6. TrwC no promueve la integración del ADN movilizado en el genoma del huésped no permisivo *L. casei.*
7. La relaxasa MobA puede ser translocada a través del SST4 VirB/D4 de *B. henselae* a una célula humana. MobA es 5-10 veces más eficiente que TrwC transfiriendo el ADN a células humanas.
8. La relaxasa MobA promueve la integración del ADN transferido en células humanas. El ratio de integración de TrwC es 5 veces mayor que el de MobA.
9. Hemos generado y validado una proteína de fusión TrwC-Cas12a que puede ser expresada en *E. coli* y hemos confirmado que ambas partes de la proteína son activas.
10. TrwC-Cas12a puede ser translocada a través del SST4 del plásmido R388 a una bacteria receptora.
11. La actividad endonucleasa de Cas12a en el cromosoma bacteriano induce la respuesta SOS.
12. Tras ser translocada a la célula receptora, TrwC-Cas12a muestra actividad endonucleasa guiada por el ARNg, dando lugar a mutaciones sitio-específicas.
13. Las mutaciones producidas por las vías de reparación bacterianas tras el corte de TrwC-Cas12a en el ADN diana eran similares a las producidas por Cas12a.
14. TrwC-Cas12a puede ser introducida por conjugación covalentemente unida a una molécula de ADN la cual contiene un casete de recombinación homóloga para producir mutaciones sin dejar cicatriz en el genoma de la bacteria receptora, mejorando más de 8 veces el ratio de edición.
15. Hemos generado las proteínas de fusión Cas12a-BID, TrwC-Cas12a and MobA-Cas12a que pueden ser expresadas en células humanas. Hemos confirmado la actividad endonucleasa sitio-específica en todas las proteínas con una eficiencia similar a Cas12a.

8. Bibliography

8. Bibliography

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* 9, 1911. doi:10.1038/s41467-018-04252-2.

Agúndez, L., González-Prieto, C., Machón, C., and 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. doi:10.1371/journal.pone.0031047.

Agúndez, L., MacHón, C., Elvira César, C., Rosa-Garrido, M., Delgado, M. D., and Llosa, M. (2011). Nuclear targeting of a bacterial integrase that mediates site-specific recombination between bacterial and human target sequences. *Appl. Environ. Microbiol.* 77, 201–210. doi:10.1128/AEM.01371-10.

Agúndez, L., Zárate-Pérez, F., Meier, A. F., Bardelli, M., Llosa, M., Escalante, C. R., et al. (2018). Exchange of functional domains between a bacterial conjugative relaxase and the integrase of the human adeno-associated virus. *PLoS One* 13, e0200841. doi:10.1371/journal.pone.0200841.

Ali, Z., Shami, A., Sedeek, K., Kamel, R., Alhabsi, A., Tehseen, M., et al. (2020). Fusion of the Cas9 endonuclease and the VirD2 relaxase facilitates homology-directed repair for precise genome engineering in rice. *Commun. Biol.* 3, 1–13. doi:10.1038/s42003-020-0768-9.

Alperi, A., Larrea, D., Fernández-González, E., Dehio, C., Zechner, E. L., and Llosa, M. (2013). A translocation motif in relaxase trwc specifically affects recruitment by its conjugative type IV Secretion System. *J. Bacteriol.* 195, 4999–5006. doi:10.1128/JB.00367-13.

Alvarez-Martinez, C. E., and Christie, P. J. (2009). Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* 73, 775–808. doi:10.1128/MMBR.00023-09.

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. doi:10.1038/s41586-019-1711-4.

Ao, X., Yao, Y., Li, T., Yang, T.-T., Dong, X., Zheng, Z.-T., et al. (2018). A Multiplex Genome Editing Method for Escherichia coli Based on CRISPR-Cas12a. *Front. Microbiol.* 9, 2307. doi:10.3389/fmicb.2018.02307.

Aviv, G., Rahav, G., and Gal-Mor, O. (2016). Horizontal Transfer of the Salmonella enterica Serovar Infantis Resistance and Virulence Plasmid pESI to the Gut Microbiota of Warm-Blooded Hosts. *MBio* 7. doi:10.1128/mBio.01395-16 e01395-16 [pii] mBio.01395-16 [pii].

Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. 30, 1473–1475. doi:10.1093/bioinformatics/btu048.

Baharoglu, Z., Bikard, D., and Mazel, D. (2010). Conjugative DNA Transfer Induces the Bacterial SOS Response and Promotes Antibiotic Resistance Development through Integron Activation. *PLoS Genet.* 6, e1001165. doi:10.1371/journal.pgen.1001165.

Bai, F., Li, Z., Umezawa, A., Terada, N., and Jin, S. (2018). Bacterial type III secretion system as a protein delivery tool for a broad range of biomedical applications. *Biotechnol. Adv.* 36, 482–493. doi:10.1016/j.biotechadv.2018.01.016.

Bardelli, M., Zárate-Pérez, F., Agúndez, L., Linden, R. M., Escalante, C. R., and Henckaerts, E. (2016). Identification of a Functionally Relevant Adeno-Associated Virus Rep68 Oligomeric Interface. *J. Virol.* 90, 6612–6624. doi:10.1128/jvi.00356-16.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (80-.).* 315, 1709–1712. doi:10.1126/science.1138140.

Bates, S., Cashmore, A. M., and Wilkins, B. M. (1998). IncP plasmids are unusually effective in mediating conjugation of Escherichia coli and Saccharomyces cerevisiae: involvement of the tra2 mating system. *J Bacteriol* 180, 6538–6543. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9851996.

Becker, E. C., and Meyer, R. J. (2003). Relaxed specificity of the R1162 nickase: a model for evolution of a system for conjugative mobilization of plasmids. *J. Bacteriol.* 185, 3538–46. doi:10.1128/jb.185.12.3538-3546.2003.

Bernheim, A., Calvo-Villamañán, A., Basier, C., Cui, L., Rocha, E. P. C., Touchon, M., et al. (2017). Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria. *Nat. Commun.* 8, 1–9. doi:10.1038/s41467-017-02350-1.

Bikard, D., Euler, C. W., Jiang, W., Nussenzweig, P. M., Goldberg, G. W., Duportet, X., et al. (2014). Exploiting CRISPR-cas nucleases to produce sequence-specific antimicrobials. *Nat. Biotechnol.* 32, 1146–1150. doi:10.1038/nbt.3043.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., et al. (1997). The complete genome sequence of Escherichia coli K-12. *Science (80-.).* 277, 1453–1462. doi:10.1126/science.277.5331.1453.

Boer, R., Russi, S., Guasch, A., Lucas, M., Blanco, A. G., Perez-Luque, R., et al. (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*. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16540117.

Börner, R. A., Kandasamy, V., Axelsen, A. M., Nielsen, A. T., and Bosma, E. F. (2019a). Genome editing of lactic acid bacteria: opportunities for food, feed, pharma and biotech. *FEMS Microbiol. Lett.* 366. doi:10.1093/femsle/fny291.

Börner, R. A., Kandasamy, V., Axelsen, A. M., Nielsen, A. T., and Bosma, E. F. (2019b). Genome editing of lactic acid bacteria: opportunities for food , feed , pharma and biotech. *FEMS, Microbiol. Lett.* 366. doi:10.1093/femsle/fny291.

Bosma, E. F., Forster, J., and Nielsen, A. T. (2017). Lactobacilli and pediococci as versatile cell factories – Evaluation of strain properties and genetic tools. *Biotechnol. Adv.* 35, 419–442. doi:10.1016/j.biotechadv.2017.04.002.

Bradley, D. E., Taylor, D. E., and Cohen, D. R. (1980). Specification of surface mating systems among conjugative drug resistance plasmids in Escherichia coli K-12. *J Bacteriol* 143, 1466–1470. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6106013.

Byrd, D. R., and Matson, S. W. (1997). Nicking by transesterification: the reaction catalysed by a relaxase. *Mol Microbiol* 25, 1011–22.

Cabezon, E., Ripoll-Rozada, J., Pena, A., de la Cruz, F., and Arechaga, I. (2015). Towards an integrated model of bacterial conjugation. *FEMS Microbiol Rev* 39, 81–95. doi:10.1111/1574-6976.12085 1574-6976.12085 [pii].

Cabezón, E., Ripoll-Rozada, J., Peña, A., de la Cruz, F., and Arechaga, I. (2015). Towards an integrated model of bacterial conjugation. *FEMS Microbiol. Rev.* 39, 81–95. doi:10.1111/1574-6976.12085.

Carballeira, J. D., Gonzalez-Perez, B., Moncalian, G., and de la Cruz, F. (2014). A high security double lock and key mechanism in HUH relaxases controls oriT-processing for plasmid conjugation. *Nucleic Acids Res* 42, 10632–10643. doi:10.1093/nar/gku741.

César, C. E., and Llosa, M. (2007). TrwC-mediated site-specific recombination is controlled by host factors altering local DNA topology. *J Bacteriol* 189, 9037–9043. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17921309.

César, C. E., Machón, C., de la Cruz, F., and 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. doi:10.1111/j.1365-2958.2006.05437.x.

Chandler, M., de la Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G., and Ton-Hoang, B. (2013). Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol* 11, 525–538. doi:10.1038/nrmicro3067 nrmicro3067 [pii].

Charpentier, M., Khedher, A. H. Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., et al. (2018). CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* 9, 1–11. doi:10.1038/s41467-018-03475-7.

Chen, F., Alphonse, M., and Liu, Q. (2020). Strategies for nonviral nanoparticle-based delivery of CRISPR/Cas9 therapeutics. *WIREs Nanomedicine and Nanobiotechnology* 12. doi:10.1002/wnan.1609.

Christie, P. J. (2001). Type IV secretion: intercellular transfer of macromolecules by

systems ancestrally related to conjugation machines. *Mol Microbiol* 40, 294-305.

Christie, P. J., Gomez Valero, L., and Buchrieser, C. (2017). "Biological Diversity and Evolution of Type IV Secretion Systems," in (Springer, Cham), 1–30. doi:10.1007/978-3-319-75241-9_1.

Christie, P. J., and Vogel, J. P. (2000). Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol* 8, 354–360. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10920394.

Citorik, R. J., Mimee, M., and Lu, T. K. (2014). Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nat. Biotechnol.* 32, 1141–1145. doi:10.1038/nbt.3011.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (80-.).* 339, 819–823. doi:10.1126/science.1231143.

Cui, L., and Bikard, D. (2016). Consequences of Cas9 cleavage in the chromosome of Escherichia coli. *Nucleic Acids Res* 44, 4243–4251. doi:10.1093/nar/gkw223 gkw223 [pii].

Cui, L., Vigouroux, A., Rousset, F., Varet, H., Khanna, V., and Bikard, D. (2018). A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. *Nat. Commun.* 9, 1–10. doi:10.1038/s41467-018-04209-5.

Datta, N., and Hedges, R. W. (1972). Trimethoprim resistance conferred by W plasmids in Enterobacteriaceae. *J Gen Microbiol* 72, 349–355. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=4562309.

Davies, J., and Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74, 417–33. doi:10.1128/MMBR.00016-10.

De Buck, S., Podevin, N., Nolf, J., Jacobs, A., and Depicker, A. (2009). The T-DNA integration pattern in Arabidopsis transformants is highly determined by the transformed target cell. *Plant J.* 60, 134–145. doi:10.1111/j.1365-313X.2009.03942.x.

de la Cruz, F., and Davies, J. (2000). Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol* 8, 128–133. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10707066.

Dehio, C. (2005). Bartonella-host-cell interactions and vascular tumour formation. *Nat Rev Microbiol* 3, 621–631. doi:10.1038/nrmicro1209.

Dehio, C., and Tsolis, R. M. (2017). "Type IV Effector Secretion and Subversion of Host Functions by Bartonella and Brucella Species," in (Springer, Cham), 269–295. doi:10.1007/978-3-319-75241-9_11.

Demarre, G., Guerout, A. M., Matsumoto-Mashimo, C., Rowe-Magnus, D. A., Marliere, P., and Mazel, D. (2005a). 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. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15748991.

Demarre, G., Guérout, A. M., Matsumoto-Mashimo, C., Rowe-Magnus, D. A., Marlière, P., and Mazel, D. (2005b). A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPα) conjugative machineries and their cognate Escherichia coli host strains. *Res. Microbiol.* 156, 245–255. doi:10.1016/j.resmic.2004.09.007.

Dostal, L., Shao, S., and Schildbach, J. F. (2011). Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer. *Nucleic Acids Res* 39, 2658–2670. doi:10.1093/nar/gkq1137.

Doudna, J. A. (2020). The promise and challenge of therapeutic genome editing. *Nature* 578, 229–236. doi:10.1038/s41586-020-1978-5.

Draper, O., César, C. E., Machón, C., De La Cruz, F., and 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. doi:10.1073/pnas.0506081102.

Encinas, D., Pilar Garcillán-Barcia, M., Santos-Merino, M., Delaye, L., Moya, A., and De La Cruz, F. (2014). Plasmid Conjugation from Proteobacteria as Evidence for the

Origin of Xenologous Genes in Cyanobacteria. doi:10.1128/JB.01464-13.

Fagen, J. R., Collias, D., Singh, A. K., and Beisel, C. L. (2017). Advancing the design and delivery of CRISPR antimicrobials. *Curr. Opin. Biomed. Eng.* 4, 57–64. doi:10.1016/j.cobme.2017.10.001.

Fernandez-Gonzalez, E., Bakioui, S., Gomes, M. C., O'Callaghan, D., Vergunst, A. C., Sangari, F. J., et al. (2016). A Functional oriT in the Ptw Plasmid of Burkholderia cenocepacia Can Be Recognized by the R388 Relaxase TrwC. *Front Mol Biosci* 3, 16. doi:10.3389/fmolb.2016.00016.

Fernández-González, E., de Paz, H. D., Alperi, A., Agúndez, L., Faustmann, M., Sangari, F. J., 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. doi:10.1128/JB.05905-11.

Fernandez-Lopez, C., Pluta, R., Perez-Luque, R., Rodriguez-Gonzalez, L., Espinosa, M., Coll, M., et al. (2013). Functional properties and structural requirements of the plasmid pMV158-encoded MobM relaxase domain. *J Bacteriol* 195, 3000–3008. doi:10.1128/JB.02264-12JB.02264-12 [pii].

Fernández-López, R., Pilar Garcillán-Barcia, M., Revilla, C., Lázaro, M., Vielva, L., and De La Cruz, F. (2006). Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiol. Rev.* 30, 942–966. doi:10.1111/j.1574-6976.2006.00042.x.

Francia, M. V, and Clewell, D. B. (2002a). Amplification of the tetracycline resistance determinant of pAMalpha1 in Enterococcus faecalis requires a site-specific recombination event involving relaxase. *J Bacteriol* 184, 5187–5193. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12193637.

Francia, M. V, and Clewell, D. B. (2002b). 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. *Mol Microbiol* 45, 375–395. doi:3007 [pii].

Furuya, N., and Komano, T. (2003). NikAB- or NikB-dependent intracellular

recombination between tandemly repeated oriT sequences of plasmid R64 in plasmid or single-stranded phage vectors. *J Bacteriol* 185, 3871–3877. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12813081.

Gaj, T., Gersbach, C. A., and Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. doi:10.1016/j.tibtech.2013.04.004.

Garcillán-Barcia, M. P., Francia, M. V, and de la Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol Rev* 33, 657–687. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19396961.

Garcillán-Barcia, M. P., Jurado, P., González-Pérez, B., Moncalián, G., Fernández, L. A., and de la Cruz, F. (2007). Conjugative transfer can be inhibited by blocking relaxase activity within recipient cells with intrabodies. *Mol. Microbiol.* 63, 404–416. doi:10.1111/j.1365-2958.2006.05523.x.

Garcillán-Barcia, M. P., Redondo-Salvo, S., Vielva, L., and de la Cruz, F. (2020). "MOBscan: Automated Annotation of MOB Relaxases," in (Humana, New York, NY), 295–308. doi:10.1007/978-1-4939-9877-7_21.

Getino, M., and de la Cruz, F. (2019). "Natural and Artificial Strategies to Control the Conjugative Transmission of Plasmids," in *Microbial Transmission* (American Society of Microbiology), 33–64. doi:10.1128/microbiolspec.MTBP-0015-2016.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison 3rd, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343–345. doi:10.1038/nmeth.1318.

Glass, Z., Lee, M., Li, Y., and Xu, Q. (2018). Engineering the Delivery System for CRISPR-Based Genome Editing. *Trends Biotechnol.* 36, 173–185. doi:10.1016/j.tibtech.2017.11.006.

Gonzalez-Perez, B., Lucas, M., Cooke, L. A., Vyle, J. S., de la Cruz, F., and Moncalian, G. (2007). Analysis of DNA processing reactions in bacterial conjugation by using

suicide oligonucleotides. *EMBO J* 26, 3847–3857. doi:10.1038/sj.emboj.7601806.

Gonzalez-Prieto, C., Agundez, L., Linden, R. M., and Llosa, M. (2013). HUH site-specific recombinases for targeted modification of the human genome. *Trends Biotechnol* 31, 305–312. doi:10.1016/j.tibtech.2013.02.002 S0167-7799(13)00049-8 [pii].

Gonzalez-Prieto, C., Gabriel, R., Dehio, C., Schmidt, M., and Llosa, M. (2017). The Conjugative Relaxase TrwC Promotes Integration of Foreign DNA in the Human Genome. *Appl Env. Microbiol* 83. doi:10.1128/AEM.00207-17.

Grandoso, G., Avila, P., Cayón, A., Hernando, M. A., Llosa, M., and 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, 1163–72.

Grandoso, G., Llosa, M., Zabala, J. C., and 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.

Grasseschi, H. A., and Minnick, M. F. (1994). Transformation of Bartonella bacilliformis by electroporation. *Can J Microbiol* 40, 782–786. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7954111.

Grohmann, E., Christie, P. J., Waksman, G., and Backert, S. (2018). Type IV secretion in Gram-negative and Gram-positive bacteria. *Mol. Microbiol.* 107, 455–471. doi:10.1111/mmi.13896.

Guasch, A., Lucas, M., Moncalián, G., Cabezas, M., Pérez-Luque, R., Gomis-Rüth, F. X., et al. (2003). Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. *Nat Struct Biol* 10, 1002–1010. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14625590.

Guglielmini, J., Quintais, L., Garcillán-Barcia, M. P., de la Cruz, F., and Rocha, E. P. C. (2011). The Repertoire of ICE in Prokaryotes Underscores the Unity, Diversity, and Ubiquity of Conjugation. *PLoS Genet.* 7, e1002222. doi:10.1371/journal.pgen.1002222.

Guzmán-Herrador, D. L., and Llosa, M. (2019). The secret life of conjugative relaxases.

Plasmid 104, 102415. doi:10.1016/j.plasmid.2019.102415.

Guzmán-Herrador, D. L., Steiner, S., Alperi, A., González-Prieto, C., Roy, C. R., and Llosa, M. (2017). DNA delivery and genomic integration into mammalian target cells through Type IV A and B secretion systems of human pathogens. *Front. Microbiol.* 8, 1503. doi:10.3389/fmicb.2017.01503.

Hamilton, T. A., Pellegrino, G. M., Therrien, J. A., Ham, D. T., Bartlett, P. C., Karas, B. J., et al. (2019). Efficient inter-species conjugative transfer of a CRISPR nuclease for targeted bacterial killing. *Nat. Commun.* 10, 1–9. doi:10.1038/s41467-019-12448-3.

Hansen, P. A., and Lessel, E. F. (1971). Lactobacillus casei (Orla-Jensen) comb. nov. *Int. J. Syst. Bacteriol.* 21, 69–71. doi:10.1099/00207713-21-1-69.

Harms, A., and Dehio, C. (2012). Intruders below the Radar: Molecular pathogenesis of Bartonella spp. *Clin. Microbiol. Rev.* 25, 42–78. doi:10.1128/CMR.05009-11.

Henckaerts, E., Dutheil, N., Zeltner, N., Kattman, S., Kohlbrenner, E., Ward, P., 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. doi:10.1073/pnas.0806821106.

Henderson, D., and Meyer, R. (1999). The MobA-linked primase is the only replication protein of R1162 required for conjugal mobilization. *J Bacteriol* 181, 2973–2978. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10217797.

Henderson, D., and Meyer, R. J. (1996). The primase of broad-host-range plasmid R1162 is active in conjugal transfer. *J Bacteriol* 178, 6888–6894. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8955311.

Hickman, A. B., Ronning, D. R., Perez, Z. N., Kotin, R. M., and Dyda, F. (2004). The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces. *Mol. Cell* 13, 403–14. doi:10.1016/s1097-2765(04)00023-1.

Hu, X., Wang, C., Liu, Q., Fu, Y., and Wang, K. (2017). Targeted mutagenesis in rice using CRISPR-Cpf1 system. *J Genet Genomics* 44, 71–73. doi:S1673-8527(16)30191-6

[pii]10.1016/j.jgg.2016.12.001.

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. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1630899.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakatura, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* 169, 5429–5433. doi:10.1128/jb.169.12.5429-5433.1987.

Jakočiunas, T., Jensen, M. K., and Keasling, J. D. (2016). CRISPR/Cas9 advances engineering of microbial cell factories. *Metab. Eng.* 34, 44–59. doi:10.1016/j.ymben.2015.12.003.

James, J. A., Escalante, C. R., Yoon-Robarts, M., Edwards, T. A., Linden, R. M., and Aggarwal, A. K. (2003). Crystal Structure of the SF3 Helicase from Adeno-Associated Virus Type 2. *Structure* 11, 1025–1035. doi:10.1016/S0969-2126(03)00152-7.

Jiang, F., and Doudna, J. A. (2017). CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* 46, 505–529. doi:10.1146/annurev-biophys-062215-010822.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31, 233–239. doi:10.1038/nbt.2508 nbt.2508 [pii].

Jiang, Y., Qian, F., Yang, J., Liu, Y., Dong, F., Xu, C., et al. (2017). CRISPR-Cpf1 assisted genome editing of Corynebacterium glutamicum. *Nat Commun* 8, 15179. doi:10.1038/ncomms15179 ncomms15179 [pii].

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* 337, 816–821. doi:10.1126/science.1225829.

Killmann, H., Benz, R., and Braun, V. (1996). Properties of the FhuA channel in the Escherichia coli outer membrane after deletion of FhuA portions within and outside

the predicted gating loop. *J Bacteriol* 178, 6913–6920. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8955314.

Kim, D., Kim, J., Hur, J. K., Been, K. W., Yoon, S. H., and Kim, J. S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol* 34, 863–868. doi:10.1038/nbt.3609 nbt.3609 [pii].

Kirby, J. E., and Nekorchuk, D. M. (2002). Bartonella-associated endothelial proliferation depends on inhibition of apoptosis. *Proc Natl Acad Sci U S A* 99, 4656–4661. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11904386.

Kishida, K., Inoue, K., Ohtsubo, Y., Nagata, Y., and Tsuda, M. (2017). Host Range of the Conjugative Transfer System of IncP-9 Naphthalene-Catabolic Plasmid NAH7 and Characterization of Its oriT Region and Relaxase. *Appl Env. Microbiol* 83. doi:e02359-16 [pii]AEM.02359-16 [pii]10.1128/AEM.02359-16.

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., et al. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495. doi:10.1038/nature16526.

Kleinstiver, B. P., Sousa, A. A., Walton, R. T., Tak, Y. E., Hsu, J. Y., Clement, K., et al. (2019). Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* 37, 276–282. doi:10.1038/s41587-018-0011-0.

Krizkova, L., and Hrouda, M. (1998). Direct repeats of T-DNA integrated in tobacco chromosome: Characterization of junction regions. *Plant J.* 16, 673–680. doi:10.1046/j.1365-313X.1998.00330.x.

Lacroix, B., and Citovsky, V. (2018). “Beyond Agrobacterium-mediated transformation: Horizontal gene transfer from bacteria to eukaryotes,” in *Current Topics in Microbiology and Immunology* (Springer Verlag), 443–462. doi:10.1007/82_2018_82.

Lang, S., Gruber, K., Mihajlovic, S., Arnold, R., Gruber, C. J., Steinlechner, S., et al. (2010). Molecular recognition determinants for type IV secretion of diverse families of

conjugative relaxases. *Mol Microbiol* 78, 1539–1555. doi:10.1111/j.1365-2958.2010.07423.x.

Lessl, M., Balzer, D., Weyrauch, K., and Lanka, E. (1993). The mating pair formation system of plasmid RP4 defined by RSF1010 mobilization and donor-specific phage propagation. *J Bacteriol* 175, 6415–6425. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8407818.

Li, H., Yang, Y., Hong, W., Huang, M., Wu, M., and Zhao, X. (2020). Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct. Target. Ther.* 5, 1–23. doi:10.1038/s41392-019-0089-y.

Li, L., Hu, S., and Chen, X. (2018a). Non-viral delivery systems for CRISPR/Cas9-based genome editing: Challenges and opportunities. *Biomaterials* 171, 207–218. doi:10.1016/j.biomaterials.2018.04.031.

Li, L., Wei, K., Zheng, G., Liu, X., Chen, S., Jiang, W., et al. (2018b). CRISPR-Cpf1-assisted multiplex genome editing and transcriptional repression in Streptomyces. *Appl. Environ. Microbiol.* 84. doi:10.1128/AEM.00827-18.

Li, Y. G., Hu, B., and Christie, P. J. (2019). Biological and Structural Diversity of Type IV Secretion Systems. *Microbiol. Spectr.* 7. doi:10.1128/microbiolspec.psib-0012-2018.

Linden, R. M., Winocour, E., and Berns, K. I. (1996). The recombination signals for adeno-associated virus site-specific integration. *Proc. Natl. Acad. Sci. U. S. A.* 93, 7966–72. doi:10.1073/pnas.93.15.7966.

Liu, Z., Dong, H., Cui, Y., Cong, L., and Zhang, D. (2020). Application of different types of CRISPR/Cas-based systems in bacteria. *Microb. Cell Fact.* 19. doi:10.1186/s12934-020-01431-z.

Llosa, M., and Alkorta, I. (2017). "Coupling proteins in type IV secretion," in *Current Topics in Microbiology and Immunology* (Springer Verlag), 143–168. doi:10.1007/978-3-319-75241-9_6.

Llosa, M., Bolland, S., and 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., and 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., and de la Cruz, F. (1994b). Conjugation-independent, site-specific recombination at the oriT of the IncW plasmid R388 mediated by TrwC. *J Bacteriol* 176, 3210–3217. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8195075.

Llosa, M., and de la Cruz, F. (2005). Bacterial conjugation: a potential tool for genomic engineering. *Res Microbiol* 156, 1–6. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15636742.

Llosa, M., Gomis-Ruth, F. X., Coll, M., and Cruz, F. de la (2002). Bacterial conjugation: a two-step mechanism for DNA transport. *Mol. Microbiol.* 45, 1–8. doi:10.1046/j.1365-2958.2002.03014.x.

Llosa, M., Grandoso, G., and 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., and 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. Available at: http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/referer?http://www.idealibrary.com/cgi-bin/links/citation/0022-2836/264/56.

Llosa, M., Schroder, G., and Dehio, C. (2012a). New perspectives into bacterial DNA transfer to human cells. *Trends Microbiol* 20, 355–359. doi:10.1016/j.tim.2012.05.008S0966-842X(12)00100-X [pii].

Llosa, M., Schröder, G., and Dehio, C. (2012b). New perspectives into bacterial DNA transfer to human cells. *Trends Microbiol.* 20, 355–359. doi:10.1016/j.tim.2012.05.008.

Llosa, M., Zunzunegui, S., and de la Cruz, F. (2003). Conjugative coupling proteins interact with cognate and heterologous VirB10-like proteins while exhibiting specificity for cognate relaxosomes. *Proc Natl Acad Sci U S A* 100, 10465–10470. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12925737.

Lorenzo-Diaz, F., Fernandez-Lopez, C., Lurz, R., Bravo, A., and Espinosa, M. (2017). Crosstalk between vertical and horizontal gene transfer: plasmid replication control by a conjugative relaxase. *Nucleic Acids Res* 45, 7774–7785. doi:10.1093/nar/gkx4503835314 [pii].

Lucas, M., Gonzalez-Perez, B., Cabezas, M., Moncalian, G., Rivas, G., and 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. *J Biol Chem* 285, 8918–8926. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20061574.

Luo, Z. Q., and Isberg, R. R. (2004). Multiple substrates of the Legionella pneumophila Dot/Icm system identified by interbacterial protein transfer. *Proc Natl Acad Sci U S A* 101, 841–846. doi:10.1073/pnas.0304916101.

Luzhetskyy, A., Fedoryshyn, M., Gromyko, O., Ostash, B., Rebets, Y., Bechthold, A., et al. (2006). IncP plasmids are most effective in mediating conjugation between Escherichia coli and streptomycetes. *Russ. J. Genet.* 42, 476–481. doi:10.1134/S1022795406050036.

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83. doi:10.1038/s41579-019-0299-x.

Manghwar, H., Lindsey, K., Zhang, X., and Jin, S. (2019). CRISPR/Cas System: Recent Advances and Future Prospects for Genome Editing. *Trends Plant Sci.* 24, 1102–1125. doi:10.1016/j.tplants.2019.09.006.

Mansilla-Soto, J., Yoon-Robarts, M., Rice, W. J., Arya, S., Escalante, C. R., and Linden, R.

M. (2009). DNA Structure Modulates the Oligomerization Properties of the AAV Initiator Protein Rep68. *PLoS Pathog.* 5, e1000513. doi:10.1371/journal.ppat.1000513.

Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature* 526, 55–61. doi:10.1038/nature15386.

Marraffini, L. A., and Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science (80-.).* 322, 1843–1845. doi:10.1126/science.1165771.

Matson, S. W., Sampson, J. K., and Byrd, D. R. (2001). F plasmid conjugative DNA transfer: the TraI helicase activity is essential for DNA strand transfer. *J Biol Chem* 276, 2372–9.

Mazel, D., and Davies, J. (1999). Antibiotic resistance in microbes. *Cell Mol Life Sci* 56, 742–754. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11212334.

McAllister, K. N., and Sorg, J. A. (2019). CRISPR Genome Editing Systems in the Genus Clostridium : a Timely Advancement . *J. Bacteriol.* 201. doi:10.1128/jb.00219-19.

Meyer, R. (1989). Site-specific recombination at oriT of plasmid R1162 in the absence of conjugative transfer. *J Bacteriol* 171, 799–806. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2644236.

Miroux, B., and Walker, J. E. (1996). Over-production of proteins in Escherichia coli: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* 260, 289–298. doi:10.1006/jmbi.1996.0399.

Miyazaki, R., and van der Meer, J. R. (2011). A dual functional origin of transfer in the ICEclc genomic island of Pseudomonas knackmussii B13. *Mol Microbiol* 79, 743–758. doi:10.1111/j.1365-2958.2010.07484.x.

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60, 174–182. doi:10.1007/s00239-004-0046-3.

Mojica, F. J. M., Diez-Villasenor, C., Soria, E., and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* 36, 244–246. doi:10.1046/j.1365-2958.2000.01838.x.

Moncalián, G. (2000). Arquitectura y ensamblaje del relaxosoma en el plásmido conjugativo R388. *PhD thesis Santander Univ. Cantab.*

Moncalián, G., Grandoso, G., Llosa, M., and de la Cruz, F. (1997). oriT-processing and regulatory roles of TrwA protein in plasmid R388 conjugation. *J Mol Biol* 270, 188–200. Available at: http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/referer?http://www.idealibrary.com/cgi-bin/links/citation/0022-2836/270/188.

Moriguchi, K., Yamamoto, S., Tanaka, K., Kurata, N., and 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. doi:10.1371/journal.pone.0074590.

Norrander, J., Kempe, T., and Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26, 101–106. doi:10.1016/0378-1119(83)90040-9.

Padmalayam, I., Karem, K., Baumstark, B., and 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 Cell Biol* 19, 377–382. doi:10.1089/10445490050043344.

Pálffy, R., Gardlík, R., Hodosy, J., Behuliak, M., Reško, P., Radvánský, J., et al. (2006). Bacteria in gene therapy: bactofection versus alternative gene therapy. *Gene Ther.* 13, 101–105. doi:10.1038/sj.gt.3302635.

Pansegrau, W., Schroder, W., and Lanka, E. (1993a). Relaxase (TraI) of IncP alpha plasmid RP4 catalyzes a site-specific cleaving-joining reaction of single-stranded DNA. *Proc. Natl. Acad. Sci.* 90, 2925–2929. doi:10.1073/pnas.90.7.2925.

Pansegrau, W., Schroder, W., and Lanka, E. (1993b). Relaxase (TraI) of IncP alpha plasmid RP4 catalyzes a site-specific cleaving-joining reaction of single-stranded DNA. *Proc*

Natl Acad Sci U S A 90, 2925–2929. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8385350.

Parker, C., and Meyer, R. J. (2007). The R1162 relaxase/primase contains two, type IV transport signals that require the small plasmid protein MobB. *Mol. Microbiol.* 66, 252–261. doi:10.1111/j.1365-2958.2007.05925.x.

Paul, B., and Montoya, G. (2020). CRISPR-Cas12a: Functional overview and applications. *Biomed. J.* 43, 8–17. doi:10.1016/j.bj.2019.10.005.

Platt, R., Drescher, C., Park, S. K., and Phillips, G. J. (2000). Genetic system for reversible integration of DNA constructs and lacZ gene fusions into the Escherichia coli chromosome. *Plasmid* 43, 12–23. doi:10.1006/plas.1999.1433.

Port, F., Starostecka, M., and Boutros, M. (2020). Multiplexed conditional genome editing with Cas12a in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* 117, 22890–22899. doi:10.1073/pnas.2004655117.

Qiu, P., Shandilya, H., D'Alessio, J. M., O'Connor, K., Durocher, J., and Gerard, G. F. (2004). Mutation detection using Surveyor™ nuclease. *Biotechniques* 36, 702–707. doi:10.2144/04364PF01.

Ramachandran, G., and Bikard, D. (2019). Editing the microbiome the CRISPR way. *Philos. Trans. R. Soc. B Biol. Sci.* 374. doi:10.1098/rstb.2018.0103.

Ramsay, J. P., and Firth, N. (2017). Diverse mobilization strategies facilitate transfer of non-conjugative mobile genetic elements. *Curr Opin Microbiol* 38, 1–9. doi:S1369-5274(16)30168-0 [pii] 10.1016/j.mib.2017.03.003.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi:10.1038/nprot.2013.143.

Reyrat, J. M., Pelicic, V., Gicquel, B., and Rappuoli, R. (1998). Counterselectable markers: Untapped tools for bacterial genetics and pathogenesis. *Infect. Immun.* 66, 4011–4017.

Rocco, J. M., and Churchward, G. (2006). The integrase of the conjugative transposon Tn916 directs strand- and sequence-specific cleavage of the origin of conjugal

transfer, oriT, by the endonuclease Orf20. *J Bacteriol* 188, 2207–2213. doi:188/6/2207 [pii]10.1128/JB.188.6.2207-2213.2006.

Rolloos, M., Hooykaas, P. J., and van der Zaal, B. J. (2015). Enhanced targeted integration mediated by translocated I-SceI during the Agrobacterium mediated transformation of yeast. *Sci Rep* 5, 8345. doi:10.1038/srep08345srep08345 [pii].

Rui, Y., Wilson, D. R., and Green, J. J. (2019). Non-Viral Delivery To Enable Genome Editing. *Trends Biotechnol.* 37, 281–293. doi:10.1016/j.tibtech.2018.08.010.

Sadler, J. R., Tecklenburg, M., and Betz, J. L. (1980). Plasmids containing many tandem copies of a synthetic lactose operator. *Gene* 8, 279–300. doi:0378-1119(80)90005-0 [pii].

Saenz, H. L., Engel, P., Stoeckli, M. C., Lanz, C., Raddatz, G., Vayssier-Taussat, M., et al. (2007). Genomic analysis of Bartonella identifies type IV secretion systems as host adaptability factors. *Nat Genet* 39, 1469–1476. doi:10.1038/ng.2007.38.

Sambrook, J., and Russell, D. W. (2001). *Molecular cloning : a laboratory manual*. 3rd ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Samperio, S., Guzmán-Herrador, D. L., May-Cuz, R., Cruz Martín, M., Álvarez, M. A., and Llosa, M. (2021). Conjugative DNA Transfer From E. coli to Transformation-Resistant Lactobacilli. *Front. Microbiol.* 12, 213. doi:10.3389/fmicb.2021.606629.

Schmid, M. C., Schulein, R., Dehio, M., Denecker, G., Carena, I., and Dehio, C. (2004). The VirB type IV secretion system of Bartonella henselae mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells. *Mol. Microbiol.* 52, 81–92. doi:10.1111/j.1365-2958.2003.03964.x.

Schmiederer, M., and Anderson, B. (2000). Cloning, sequencing, and expression of three Bartonella henselae genes homologous to the Agrobacterium tumefaciens VirB region. *DNA Cell Biol* 19, 141–147. doi:10.1089/104454900314528.

Schröder, G., Schuelein, R., Quebatte, M., and Dehio, C. (2011). Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen Bartonella henselae. doi:10.1073/pnas.1019074108.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation

velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* 78, 1606–19. doi:10.1016/S0006-3495(00)76713-0.

Schulein, R., and Dehio, C. (2002). The VirB/VirD4 type IV secretion system of Bartonella is essential for establishing intraerythrocytic infection. *Mol Microbiol* 46, 1053–1067. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12421311.

Schulein, R., Guye, P., Rhomberg, T. A., Schmid, M. C., Schroder, G., Vergunst, A. C., et al. (2005). A bipartite signal mediates the transfer of type IV secretion substrates of Bartonella henselae into human cells. *Proc Natl Acad Sci U S A* 102, 856–861. doi:10.1073/pnas.0406796102.

Shuman, S., and Glickman, M. S. (2007). Bacterial DNA repair by non-homologous end joining. *Nat. Rev. Microbiol.* 5, 852–861. doi:10.1038/nrmicro1768.

Simmons, L. A., Foti, J. J., Cohen, S. E., and Walker, G. C. (2008). The SOS Regulatory Network. *EcoSal Plus* 3. doi:10.1128/ecosalplus.5.4.3.

Simon, R., Priefer, U., and Pühler, A. (1983). A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Bio/Technology* 1, 784–791. doi:10.1038/nbt1183-784.

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science (80-.).* 351, 84–88. doi:10.1126/science.aad5227.

Smillie, C., Garcillan-Barcia, M. P., Francia, M. V, Rocha, E. P., and de la Cruz, F. (2010). Mobility of plasmids. *Microbiol Mol Biol Rev* 74, 434–452. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20805406.

Souza, D. P., Oka, G. U., Alvarez-Martinez, C. E., Bisson-Filho, A. W., Dunger, G., Hobeika, L., et al. (2015). Bacterial killing via a type IV secretion system. *Nat Commun* 6, 6453. doi:10.1038/ncomms7453.

St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., and Shearwin, K. E. (2013). One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.* 2, 537–541. doi:10.1021/sb400021j.

Stella, S., Alcon, P., and Montoya, G. (2017). Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* 546, 559–563. doi:10.1038/nature22398nature22398 [pii].

Stemmer, M., Thumberger, T., del Sol Keyer, M., Wittbrodt, J., and Mateo, J. L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS One* 10, e0124633. doi:10.1371/journal.pone.0124633.

Suzuki, H., Yano, H., Brown, C. J., and Top, E. M. (2010). Predicting plasmid promiscuity based on genomic signature. *J. Bacteriol.* 192, 6045–6055. doi:10.1128/JB.00277-10.

Swarts, D. C., and Jinek, M. (2018). Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip. Rev. RNA* 9, e1481. doi:10.1002/wrna.1481.

Tato, I., Matilla, I., Arechaga, I., Zunzunegui, S., de la Cruz, F., and 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. *J Biol Chem* 282, 25569–25576. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17599913.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A* 76, 4350–4354. Available at: http://www.ncbi.nlm.nih.gov/pubmed/388439.

Trieu-Cuot, P., Carlier, C., Martin, P., and Courvalin, P. (1987). Plasmid transfer by conjugation from Escherichia coli to Gram-positive bacteria. *FEMS Microbiol Lett* 48, 289–294. doi:10.1111/j.1574-6968.1987.tb02558.x.

Trokter, M., and Waksman, G. (2018). Translocation through the Conjugative Type IV Secretion System Requires Unfolding of Its Protein Substrate. doi:10.1128/JB.00615-17.

Tyagi, S., Kumar, R., Das, A., Won, S. Y., and Shukla, P. (2020). CRISPR-Cas9 system: A genome-editing tool with endless possibilities. *J. Biotechnol.* 319, 36–53.

doi:10.1016/j.jbiotec.2020.05.008.

Ungerer, J., and Pakrasi, H. B. (2016). Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. *Sci Rep* 6, 39681. doi:10.1038/srep39681 srep39681 [pii].

Varsaki, A., Lucas, M., Afendra, A. S., Drainas, C., and de la Cruz, F. (2003). Genetic and biochemical characterization of MbeA, the relaxase involved in plasmid ColE1 conjugative mobilization. *Mol Microbiol* 48, 481–493. doi:3441 [pii].

Vento, J. M., Crook, N., and Beisel, C. L. (2019). Barriers to genome editing with CRISPR in bacteria. *J. Ind. Microbiol. Biotechnol.* 46, 1327–1341. doi:10.1007/s10295-019-02195-1.

Vergunst, A. C., Schrammeijer, B., Den Dulk-Ras, A., De Vlaam, C. M. T., Regensburg-Tuink, T. J. G., and Hooykaas, P. J. J. (2000). VirB/D4-dependent protein translocation from Agrobacterium into plant cells. *Science (80-.).* 290, 979–982. doi:10.1126/science.290.5493.979.

Vergunst, A. C., van Lier, M. C., den Dulk-Ras, A., Stuve, T. A., Ouwehand, A., and Hooykaas, P. J. (2005). Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of Agrobacterium. *Proc Natl Acad Sci U S A* 102, 832–837. doi:10.1073/pnas.0406241102.

Vieira, J., and Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19, 259–268. doi:10.1016/0378-1119(82)90015-4.

Vogel, J. P., Andrews, H. L., Wong, S. K., and Isberg, R. R. (1998). Conjugative transfer by the virulence system of Legionella pneumophila. *Science (80-.).* 279, 873–876. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9452389.

Voth, D. E., Broederdorf, L. J., and Graham, J. G. (2012). Bacterial Type IV Secretion Systems: Versatile Virulence Machines NIH Public Access. *Futur. Microbiol* 7, 241–257. doi:10.2217/fmb.11.150.

Wagner, A., and Dehio, C. (2019). Role of distinct type-IV-secretion systems and secreted effector sets in host adaptation by pathogenic *Bartonella* species. *Cell. Microbiol.*

21, e13004. doi:10.1111/cmi.13004.

Walker, G. C. (1996). "The SOS Response of Escherichia coli," in *Escherichia coli and Salmonella. Neidhardt, FC Washington DC American Society of Microbiology*, 1: 1400–1416.

Waller, M. C., Bober, J. R., Nair, N. U., and Beisel, C. L. (2017). Toward a genetic tool development pipeline for host-associated bacteria. *Curr. Opin. Microbiol.* 38, 156–164. doi:10.1016/j.mib.2017.05.006.

Wang, M., Gao, Z., Zhang, Y., and Pan, L. (2016). Lactic acid bacteria as mucosal delivery vehicles: a realistic therapeutic option. *Appl. Microbiol. Biotechnol.* 100, 5691–5701. doi:10.1007/s00253-016-7557-x.

Wang, P., Zhang, C., Zhu, Y., Deng, Y., Guo, S., Peng, D., et al. (2013). The resolution and regeneration of a cointegrate plasmid reveals a model for plasmid evolution mediated by conjugation and oriT site-specific recombination. *Env. Microbiol* 15, 3305–3318. doi:10.1111/1462-2920.12177.

Waters, V. L., and Guiney, D. G. (1993). Processes at the nick region link conjugation, T-DNA transfer and rolling circle replication. *Mol Microbiol* 9, 1123–1130.

Wawrzyniak, P., Płucienniczak, G., and Bartosik, D. (2017). The Different Faces of Rolling-Circle Replication and Its Multifunctional Initiator Proteins. *Front. Microbiol.* 8, 2353. doi:10.3389/fmicb.2017.02353.

Wilbie, D., Walther, J., and Mastrobattista, E. (2019). Delivery Aspects of CRISPR/Cas for in Vivo Genome Editing. *Acc. Chem. Res.* 52, 1555–1564. doi:10.1021/acs.accounts.9b00106.

Wisniewski, J. A., Traore, D. A., Bannam, T. L., Lyras, D., Whisstock, J. C., and Rood, J. I. (2016). TcpM: a novel relaxase that mediates transfer of large conjugative plasmids from Clostridium perfringens. *Mol Microbiol* 99, 884–896. doi:10.1111/mmi.13270.

Xu, C. L., Ruan, M. Z. C., Mahajan, V. B., and Tsang, S. H. (2019). Viral Delivery Systems for CRISPR. *Viruses* 11, 28. doi:10.3390/v11010028.

Xu, R., Qin, R., Li, H., Li, D., Li, L., Wei, P., et al. (2017). Generation of targeted mutant rice using a CRISPR-Cpf1 system. *Plant Biotechnol J* 15, 713–717.

doi:10.1111/pbi.12669.

Yan, M. Y., Li, S. S., Ding, X. Y., Guo, X. P., Jin, Q., and Sun, Y. C. (2020). A crispr-assisted nonhomologous end-joining strategy for efficient genome editing in Mycobacterium tuberculosis. *MBio* 11. doi:10.1128/mBio.02364-19.

Yan, M. Y., Yan, H. Q., Ren, G. X., Zhao, J. P., Guo, X. P., and Sun, Y. C. (2017). CRISPR-Cas12a-Assisted Recombineering in Bacteria. *Appl Env. Microbiol* 83. doi:e00947-17 [pii] 10.1128/AEM.00947-17 AEM.00947-17 [pii].

Yao, R., Liu, D., Jia, X., Zheng, Y., Liu, W., and Xiao, Y. (2018). CRISPR-Cas9/Cas12a biotechnology and application in bacteria. *Synth. Syst. Biotechnol.* 3, 135–149. doi:10.1016/j.synbio.2018.09.004.

Yoon, M., Smith, D. H., Ward, P., Medrano, F. J., Aggarwal, A. K., and Linden, R. M. (2001). Amino-terminal domain exchange redirects origin-specific interactions of adeno-associated virus rep78 in vitro. *J. Virol.* 75, 3230–9. doi:10.1128/JVI.75.7.3230-3239.2001.

Zarate-Perez, F., Bardelli, M., Burgner, J. W., Villamil-Jarauta, M., Das, K., Kekilli, D., et al. (2012). The Interdomain Linker of AAV-2 Rep68 Is an Integral Part of Its Oligomerization Domain: Role of a Conserved SF3 Helicase Residue in Oligomerization. *PLoS Pathog.* 8, e1002764. doi:10.1371/journal.ppat.1002764.

Zeaiter, Z., Mapelli, F., Crotti, E., and Borin, S. (2018). Methods for the genetic manipulation of marine bacteria. *Electron. J. Biotechnol.* 33, 17–28. doi:10.1016/j.ejbt.2018.03.003.

Zechner, E. L., Moncalian, G., and de la Cruz, F. (2017). Relaxases and Plasmid Transfer in Gram-Negative Bacteria. *Curr Top Microbiol Immunol* 413, 93–113. doi:10.1007/978-3-319-75241-9_4.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771. doi:10.1016/j.cell.2015.09.038 S0092-8674(15)01200-3 [pii].

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.

M., et al. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol* 35, 31–34. doi:10.1038/nbt.3737.

Zhang, H. X., Zhang, Y., and Yin, H. (2019). Genome Editing with mRNA Encoding ZFN, TALEN, and Cas9. *Mol. Ther.* 27, 735–746. doi:10.1016/j.ymthe.2019.01.014.

Zheng, J., Wittouck, S., Salvetti, E., Franz, C. M. A. P., Harris, H. M. B., Mattarelli, P., et al. (2020). A taxonomic note on the genus Lactobacillus : Description of 23 novel genera , emended description of the genus Lactobacillus Beijerinck 1901 , and union of Lactobacillaceae and Leuconostocaceae. doi:10.1099/ijsem.0.004107.

Zhou, Y., Liu, X., Wu, J., Zhao, G., and Wang, J. (2020). CRISPR-Cas12a-Assisted Genome Editing in Amycolatopsis mediterranei. *Front. Bioeng. Biotechnol.* 8, 698. doi:10.3389/fbioe.2020.00698.

9. Publications

9. Publications

9.1. Scientific publications

Larrea, D., de Paz, H. D., Matilla, I., Guzman-Herrador, D. L., Lasso, G., de la Cruz, F., et al. (2017). Substrate translocation involves specific lysine residues of the central channel of the conjugative coupling protein TrwB. *Mol Genet Genomics*. doi:10.1007/s00438-017-1331-3 10.1007/s00438-017-1331-3 [pii].

Guzmán-Herrador, D. L., Steiner, S., Alperi, A., González-Prieto, C., Roy, C. R., and Llosa, M. (2017). DNA delivery and genomic integration into mammalian target cells through Type IV A and B secretion systems of human pathogens. *Front. Microbiol.* 8, 1503. doi:10.3389/fmicb.2017.01503.

Guzmán-Herrador, D. L., and Llosa, M. (2019). The secret life of conjugative relaxases. *Plasmid* 104, 102415. doi:10.1016/j.plasmid.2019.102415.

Samperio, S., Guzmán-Herrador, D. L., May-Cuz, R., Cruz Martín, M., Álvarez, M. A., and Llosa, M. (2021). Conjugative DNA Transfer From E. coli to Transformation-Resistant Lactobacilli. *Front. Microbiol.* 12, 213. doi:10.3389/fmicb.2021.606629.

9.2. Patent

Patent in process. Patent number: P202030890

Substrate translocation involves specific lysine residues of the central channel of the conjugative coupling protein TrwB

Delfina Larrea,[a,1] Héctor D. de Paz,[a,2] Inmaculada Matilla,[a,3] Dolores L. Guzmán-Herrador,[a] Gorka Lasso,[b] Fernando de la Cruz,[a] Elena Cabezón,[a] Matxalen Llosa[a#]

[a] Departamento de Biología Molecular, Universidad de Cantabria, and Instituto de Biomedicina y Biotecnología de Cantabria, IBBTEC (Universidad de Cantabria, CSIC, SODERCAN), Santander, Spain

[b] Department of Biochemistry and Molecular Biophysics, Center for Computational Biology and Bioinformatics, Department of Systems Biology, Howard Hughes Medical Institute, Columbia University, New York, New York 10032, USA.

Running Head: Role of TrwB central channel in bacterial conjugation

Address correspondence to: llosam@unican.es

ORCID 0000-0002-4826-2240

[1] Present address: Department of Neurology, Columbia University Medical Center, New York, USA.

[2] Present address: Molecular Microbiology Department, University Hospital Sant Joan de Déu, Esplugues de Llobregat, Barcelona, Spain

[3] Present address: Cell Biology Unit, Children's Medical Research Institute, Westmead, Australia.

Abstract

Conjugative transfer of plasmid R388 requires the coupling protein TrwB for protein and DNA transport, but their molecular role in transport has not been deciphered. We investigated the role of residues protruding into the central channel of the TrwB hexamer by a mutational analysis. Mutations affecting lysine residues K275, K398 and K421, and residue S441, all facing the internal channel, affected transport of both DNA and the relaxase protein *in vivo*. The ATPase activity of the purified soluble variants was affected significantly in the presence of accessory protein TrwA or DNA, correlating with their behaviour *in vivo*. Alteration of residues located at the cytoplasmic or the inner membrane interface resulted in lower activity *in vivo* and *in vitro*, while variants affecting residues in the central region of the channel showed increased DNA and protein transfer efficiency, and higher ATPase activity, especially in the absence of TrwA. In fact, these variants could catalyse DNA transfer in the absence of TrwA under conditions in which the wild-type system was transfer-deficient. Our results suggest that protein and DNA molecules have the same molecular requirements for translocation by Type IV secretion systems, with residues at both ends of the TrwB channel controlling the opening-closing mechanism, while residues embedded in the channel would set the pace for substrate translocation (both protein and DNA) in concert with TrwA.

Keywords: Bacterial conjugation / Type IV secretion systems / DNA transport / molecular motors

Introduction

Bacterial conjugation is a highly efficient and promiscuous process of DNA transfer from donor to recipient bacteria, which contributes to horizontal dissemination of DNA in Gram-negative and Gram-positive bacteria (Thomas and Nielsen 2005). Conjugative coupling proteins are essential elements of the DNA transfer machinery (Christie 2016). These proteins form hexamers anchored to the inner membrane. Their ATPase activity is required for the transport of the DNA molecule as well as the protein which leads the DNA into the recipient cell (Llosa et al. 2003). Early and current models of conjugative DNA transfer propose that the DNA strand is pumped into the recipient travelling along the internal channel (ICH) of the hexamer (Llosa et al. 2002; Cabezon et al. 2015), but there is no conclusive evidence.

The conjugative apparatus involves different functional modules (Llosa and de la Cruz 2005; Cabezon et al. 2015): the protein-DNA complex responsible for substrate processing called relaxosome, and a Type IV secretion system (T4SS) for substrate secretion. The relaxosome is comprised of a DNA site, the origin of transfer (*oriT*), the relaxase, and additional relaxase accessory proteins and host factors. In Gram-negative bacteria, the T4SS is comprised of a core channel complex spanning the bacterial envelope, an extracellular pilus involved in cell-to-cell contact, and three cytoplasmic hexameric ATPases that supply the energy for pilus biogenesis and substrate transport. One of these ATPases is the Type IV coupling protein (T4CP), required to couple the relaxosome to the T4SS. To accomplish DNA transfer, the relaxase cleaves the DNA strand to be transferred and remains covalently bound to the T-strand; this nucleoprotein complex is recruited and translocated by the T4SS into the recipient cell, where the relaxase catalyzes recircularization of the DNA.

Type IV coupling proteins (T4CPs) are present in all conjugative systems, and in many Type IV secretion systems (T4SSs) involved in bacterial virulence (Gonzalez-Rivera et al. 2016). T4CPs are dispensable for pilus biogenesis but required for substrate translocation (Lai et al. 2000a; Lawley et al. 2002), probably playing a major role in substrate recruitment. Evolutionary and biochemical work supports the assumption that T4CPs have an independent origin and function from their cognate T4SS (Cabezon et al. 2012; Guglielmini et al. 2013; Larrea et al. 2013). T4CPs belonging to the VirD4-like protein family, e.g., F-TraD, RP4-TraG, R388-TrwB and *Agrobacterium tumefaciens*-VirD4 (de la Cruz et al. 2010), display low sequence identities (15-20%) but share conserved features, including a nucleotide-binding domain with Walker boxes A and B, which are essential for conjugation, and a transmembrane domain (TMD) for anchoring them to the inner membrane.

The soluble derivative of the R388 T4CP, TrwBΔN70, devoid of the TMD, is a DNA-dependent ATPase (Tato et al. 2005). ATPase activity is stimulated by the CTD of the relaxosomal protein TrwA (Tato et al. 2007), and by both double stranded DNA (dsDNA) and single stranded DNA (ssDNA) (Tato et al. 2005; Tato et al. 2007), but more specifically by G4 DNA structures (Matilla et al. 2010). The crystallographic structure of TrwBΔN70 (Gomis-Rüth et al. 2001; Gomis-Rüth et al. 2002) reveals a hexamer with a 6-fold symmetry and an ICH of approximately 20 Å in diameter. The ICH is composed of 180 surface residues (solvent accessible surface area >= 10$Å^2$), of which 114 (19 per monomer) correspond to charged residues; 48 positively charged and 66 negatively charged surface residues. Each monomer is composed of two main structural domains: the nucleotide-binding domain showing a RecA-like fold, and a small membrane-distal all-alpha domain. The TMD of TrwB plays an important role in TrwB structural integrity and oligomerization (Hormaeche et al. 2002; Hormaeche et al. 2004; de Paz et al. 2010; Vecino et al. 2011), subcellular localization

93 (Segura et al. 2014), and regulation of ATPase activity (Hormaeche et al. 2006; Vecino et al.
94 2010).

95 A coupling role for T4CPs is supported by early genetic data (Cabezón et al. 1997) and
96 evidence of protein-protein interactions with both the substrate and the T4SS. Interactions
97 with relaxosomal proteins were described for different conjugative T4CPs, both from Gram-
98 negative (Schröder et al. 2002; Llosa et al. 2003; Tato et al. 2007; Lu et al. 2008; Lang et al.
99 2011) and Gram-positive (Chen et al. 2008) plasmids. Moreover, an interaction with the
100 substrate has also been reported for *Helicobacter pylori* T4CP Cag β and its secreted
101 substrate CagA (Jurik et al. 2010). The carboxy-terminal domain (CTD) is the candidate
102 domain for substrate interaction, as suggested by structural and functional data in different
103 systems (Sastre et al. 1998; Lu et al. 2008; Whitaker et al. 2015; Whitaker et al. 2016).
104 T4CP:T4SS interactions have been reported with the T4SS core component VirB10 (Gilmour
105 et al. 2003; Llosa et al. 2003; Atmakuri et al. 2004). The T4CP-VirB10 interaction was shown
106 to be responsible for efficiency of DNA transfer (Llosa et al. 2003). VirB10 is proposed to act
107 as a regulator of the T4SS outer-membrane pore (Cascales and Christie 2004); thus, a mating
108 signal could be transmitted from the outside of the cell to the relaxosome via the T4CP-
109 VirB10 interaction (de Paz et al. 2010). While the interaction with the relaxosome is highly
110 specific for its cognate system, a single T4CP can interact functionally with several
111 conjugative T4SSs (Llosa et al. 2003) and even with T4SSs involved in bacterial virulence,
112 leading to DNA transfer into the human cells targeted by the pathogen (Fernández-González
113 et al. 2011; Schröder et al. 2011). Interestingly, it was recently shown that chimeric T4CP
114 could recruit the cognate substrates of their CTD to the cognate T4SS of their TMD,
115 emphasizing their self-sufficiency as substrate recruiters for T4SS (Whitaker et al. 2016).

In addition to the interactions with VirB10, the T4CP also interacts with the two cytoplasmic ATPases VirB4 and VirB11 (Atmakuri et al. 2004; Ripoll-Rozada et al. 2013). It has been proposed that interactions between the three cytoplasmic ATPases may be dynamic, representing alternative functional conformations of the T4SS (Ripoll-Rozada et al. 2013). In Gram-positive plasmids pLS20 and pCF10, interaction of the T4CP with the VirB4 homologue was reported (Bauer et al. 2011; Li et al. 2012a). Moreover, the ATPase activity of the T4CP TrwB is inhibited in the presence of an ATPase- defective mutant of TrwK, the VirB4 homolog in plasmid R388, which suggests that both proteins can interact with each other to form heterocomplexes (Pena et al. 2012). There is a striking structural homology between TrwB and the CTD of VirB4 homologues (Pena et al. 2012; Wallden et al. 2012), although these two ATPases are proposed to act at different steps of the conjugative process. VirB4 proteins seem to mediate pilin dislocation from the inner membrane, promoting pilus formation (Kerr and Christie 2010), whereas the T4CP would be required in subsequent processing steps to pump the plasmidic DNA through the channel (Llosa et al. 2002; Cabezon and de la Cruz 2006). This proposal is based on its DNA-dependent ATPase activity (Tato et al. 2005), and on the structural similarities with other RecA-like motor proteins that pump DNA between cellular foci or across membranes, such as FtsK or SpoIIIE (Cabezon et al. 2012). However, T4CPs are also essential for substrate translocation in the absence of DNA transfer (Draper et al. 2005; Jurik et al. 2010), and their ATPase activity is also required for relaxase translocation (de Paz et al. 2010).

In a previous work, we mapped functional domains of TrwB by *in vivo* analysis of a collection of TrwB variants (de Paz et al. 2010) and we identified a region, including the cytoplasmic entrance and surface of the ICH of the TrwB hexamer, involved in substrate transfer. In order to clarify the role of the ICH of the TrwB hexamer in conjugation, we have

addressed an *in vivo* and *in vitro* analysis of TrwB variants on residues at both ends and embedded into the ICH. The results obtained in this work support a model in which the ICH of the T4CP controls translocation of both DNA and protein substrates.

Materials and Methods

Bacterial strains

E. coli strains DH5α (Grant et al. 1990) and D1210 (Sadler et al. 1980) were used for cloning procedures and plasmid maintenance. For mating assays, strains D1210, DH5α or HMS174 (Campbell et al. 1978) were used as donors and recipients, as indicated. Strain C41 (Miroux and Walker 1996) was used for overexpression under the control of the T7 promoter.

Plasmid constructions

Plasmids used in this work are listed in **Table 1**. Plasmids constructions made for this work are described in **Supplementary Table S1**. Plasmids were constructed using standard methodological techniques (Sambrook and Russell 2001). The spontaneous mutations obtained in *trwB* in plasmid pDEL045 were separated by restriction cloning as detailed in Table S1. Constructs harboring *oriT+trwC* to test DNA transfer in the absence of TrwA were obtained by cloning *oriT* in place of the *trwA* gene in plasmid pET29::*trwAC* (Table S1). Two plasmids were obtained with both orientations of the *oriT* with respect to *trwC*. Since this difference could affect *trwC* levels and transfer efficiency, both were tested in the presence of the R388 *trwC* mutant pSU1458 (Llosa et al. 1994) and shown to be mobilizable and to complement *trwC* mutations with the same efficiency (data not shown). Since the *PtrwA* promoter is located at positions 272- 300 of the *oriT*, we selected the construct with *PtrwA*

promoter facing opposite orientation of *trwC*, so that it does not affect its expression (pDEL017; Table 1).

Bacterial conjugation assays under TrwB limiting conditions

Mating assays under TrwB limiting conditions were performed as previously described (de Paz et al. 2010), but using donor cells in exponential phase instead of stationary phase. Under these conditions, TrwB steady-state levels were reproducible, as judged by Western blot (data not shown). Briefly, pHP139-derived plasmids expressing mutated *trwB* under the control of the lactose promoter were used for mobilization of a plasmid containing *oriT* plus *trwA* and *trwC* (pHP138) or *oriT* plus *trwC* (pDEL017) through the T4SS of plasmid pKM101 present in plasmid pKM101Δ*mob*. This plasmid is a SmaI deletion derivative of pKM101 devoid of the whole DNA transfer region of pKM101, so it only codes for the T4SS. Matings were carried out using as donor the *lacI*^{q} strain D1210 in the absence of Isopropyl β-D-1-thiogalactopyranoside (IPTG) (repressed conditions), or under induced conditions as follows: overnight cultures of donor strains were diluted 1/20 dilution and growth for 2 hours in the presence of 0.5 mM IPTG, and the matings were performed on LB agar plates supplemented with 0.5 mM IPTG.

Relaxase transport assay

Triparental matings were performed to check for TrwC transport in the absence of DNA as described previously (de Paz et al. 2010), with modifications further explained in the text. pHP139-derived plasmids containing *trwB* (wild-type or mutants) were introduced in donor cells (D1210) that also contain plasmid pKM101Δ*mob* coding for the pKM101 T4SS and the non-mobilizable plasmid containing *PtrwA-trwA-trwC* genes (plasmid pET29::*trwAC)*. None of the three plasmids present in the donor cell contained an *oriT*, so there is no conjugative DNA transfer from the donor. Donor cells were mated with a second

strain (DH5α) harboring a *trwC*-deficient R388 derivative (plasmid pSU1445 (Llosa et al. 1994)). TrwC transport into this second strain was detected by complementation of the *trwC* mutation and subsequent mobilization of pSU1445 into a third recipient strain (HMS174). Mating assays were carried out from the *lacI*^{q} strain D1210 under repressed or induced conditions as described in the previous section.

DNA and protein transfer results are shown as the frequency of transconjugants per donor cell, and represent the mean of 3-5 independent experiments.

Western blot

The amount of TrwB protein was estimated by Western blot of total protein extracts. Overnight cultures of *E. coli* cells harboring plasmids containing *trwB* (wild-type and mutants) under TrwB liming conditions were diluted 1/20 and cells were grown to an optical density (OD) of 0.6. When indicated, 0.5 mM IPTG was added and growth continued for 1 or 3 hours post induction. Cells were collected, centrifuged, resuspended in 1/10 volume of 2 x SDS-gel loading buffer (Sambrook & Russell, 2001) and stored at -20°C. Samples were boiled for 10 min prior to electrophoresis, and equivalent amounts of total protein were loaded per well. Proteins were transferred from the gel onto nitrocellulose filters. Anti-TrwB primary antibody (de Paz et al. 2010) and peroxidase-conjugated anti-rabbit secondary antibody (SIGMA) were used at 1:5,000 and 1 :10,000 dilutions, respectively, in 1 X TBST + 1.5% of blocking agent. Detection was performed with the Supersignal kit (Pierce), and bands were analyzed on a Bio-Rad ChemiDoc apparatus.

Protein purification

TrwBΔN70 and derivatives were purified as described previously (Tato et al. 2005). Protein TrwAh (TrwA with a C-terminal His-tag) was also purified as described (Tato et al. 2007).

211 ***Protein and DNA quantification***

212 Protein concentrations were determined using BCA Protein kit (Pierce). Double
213 stranded pUC8 (Vieira and Messing 1982) DNA was purified by using Qiagen midi Kit and
214 quantified in a NanoDrop (ND-1000 Thermo) spectrophotometer.

215 ***Electrophoretic mobility shift assay (EMSA)***

216 Nonspecific binding of TrwBΔN70 to supercoiled plasmid DNA was assayed by EMSA
217 as described (Moncalian et al. 1999). 200 ng of pUC8 DNA were incubated in binding buffer
218 (50 mM PIPES-NaOH pH 6.2, 0.1 mM EDTA, 100 mM NaCl, 2 mM $MgCl_2$, Glycerol 5 % (v/v),
219 PMSF 0.001%) with increasing concentrations of TrwBΔN70 monomer (5-15 μM), or BSA as
220 a negative control, in a final volume of 10 μl for 10 min at 37ºC. The reaction mix was then
221 loaded on 0.8% agarose gel stained with Sybr Safe DNA gel Stain (INVITROGEN) and run at
222 100 V for 2.5 h. The shift of DNA was visualized in a Gel-Doc apparatus at 30 minutes
223 intervals.

224 ***Affinity chromatography***

225 Protein interactions with TrwA were assayed as described (Llosa et al. 2003). Briefly,
226 GST-TrwBΔN75 fusion proteins were partially purified from the soluble fraction obtained
227 after cell lysis by mixing with glutathione-Sepharose resin (Pharmacia) overnight at 4ºC. The
228 resin was then washed extensively with 1X PBS to remove unbound proteins. 20 μg of
229 purified TrwAh or BSA as a negative control were added in buffer A (50 mM Tris, pH 7.6; 50
230 mM NaCl; 5 mM $MgCl_2$; BSA 1 g/ml) and incubated at room temperature for 1 h. Following
231 incubation, unbound TrwAh was removed via extensive washing with 1X PBS. Afterwards,
232 the resin was incubated for 20 minutes at room temperature with 30 mM glutathione, and
233 centrifuged to remove insoluble proteins. Bound proteins (TrwBΔN75-TrwAh) were
234 collected from the supernatant. The elution process was repeated twice. Protein

concentration was quantified and 30 μg from each sample were loaded on SDS polyacrilamide gels and stained with Coomassie brilliant blue.

ATP hydrolysis assays

ATP hydrolysis was quantified by a coupled enzyme assay as described previously (Tato et al. 2007). ATPase activity of TrwBΔN70 or derivatives was analyzed in the presence/absence of 5 μM ssDNA (M13mp18 viral single-stranded DNA or ϕx174 Virion DNA, both from New England Biolabs), 10μM supercoiled dsDNA (pUC8) or 0.2 μM TrwA (tetramer). TrwA was pre-incubated for 10 min at 37°C in ATPase assay mixture. The ATPase reaction mixture contained 150 μl of 50 mM Pipes-NaOH, pH 6.2, 75 mM NaCl, 10 mM MgCl, 10% glycerol, 0.5 mM phosphoenolpyruvate, 0.25 mM NADH, 60 μg/ml pyruvate kinase, 60 μg/ml lactate dehydrogenase (Roche Applied Science or Sigma Aldrich enzymatic mix) and 5 mM ATP (CALBIOCHEM). All reactions were initiated by the addition of 0.3 μM TrwBΔN70 monomer. ATPase activity was measured indirectly by decrease in NADH absorbance at 340 nm for 10 min at 37 °C in a UV-1603 spectrophotometer (Shimadzu). The ATPase activity was calculated as nmol of ATP hydrolyzed per minute per mg of protein.

Electrostatics of TrwB

For electrostatics analysis of TrwB and variants, the hexameric biological unit of TrwB was used as the initial structure to model the wild type (wt) and mutant complexes (Gomis-Rüth et al. 2001). K275A, K389A and K421A mutants were modelled using VMD (Humphrey et al. 1996). Hydrogens were added to the complexes using VMD and the protonation state of histidines was predicted with Propka (Li et al. 2005). Vacuum minimization was carried out for 2000 steps (time-step 1fs/step) with the conjugate gradient minimization algorithm as implemented in NAMD (Phillips et al. 2005) using the CHARMM forcefield (Mackerell et al. 2004). Electrostatic potentials were computed with the finite difference Poisson-

Boltzmann (FDPB) method (Warwicker and Watson 1982), implemented in Delphi (Li et al. 2012b). Atomic charges and radii were extracted from the CHARMM forcefield (Huang and MacKerell 2013). The dielectric constant of the protein interior and the solvent were set to four and 80, respectively (Huang and MacKerell 2013). The ion exclusion parameter was set to two and the ionic strength to 145 mM. Electrostatic calculations were carried out using a lattice with 1.7 grids per Å and a series of focusing runs of increasing percentage fill (perfil) was performed from 20% to 90%. Calculations were iterated until they reached convergence, defined as the point at which the final maximum energy change is less than 10-4kTe-1. Visualization of electrostatic surfaces was carried out with PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

Results

Construction of TrwB variants

According to current models for bacterial conjugation, the hexameric form of TrwB pumps DNA out of the cell through its ICH (Cabezon et al. 2015). Thus, positively charged residues within the ICH might interact with the negatively charged DNA backbone to facilitate transfer of the conjugative substrate (TrwC-DNA) through the T4SS. A previous report mapping TrwB functional domains (de Paz et al. 2010) suggested that several lysine residues mapping in the ICH (K275, K398 and K421) are involved in the process. Some of the constructs used in this work were double mutants, leading to variants K398A R417S and K421A D425A. In order to refine the previous analysis, new plasmids were constructed coding separately for different TrwB variants. Also, a double variant K275A K398A was

constructed. **Figure 1** shows the residues altered by mutagenesis, including a representation of the expected effect on charge distribution when each lysine is replaced by an alanine. The steady-state levels of these new TrwB proteins were similar to those of the wt protein, according to Western blot analysis (**Supplementary Figure S1**).

During the construction of TrwB K275A (de Paz et al. 2010), a plasmid including two additional spontaneous mutations was obtained, rendering TrwB variant P237L K275A S441G. The S441 residue also maps to the ICH (Fig. 1), so we isolated the mutation and included this variant in our present analysis.

Effect of* trwB *mutations on DNA transfer

In a previous work, we developed a conjugation assay in which the amount of TrwB was the limiting factor for DNA transfer, improving the detection of TrwB variant phenotypes (de Paz et al. 2010). The assay is based on controlled expression of *trwB* from the lactose promoter, and transfer through the T4SS of plasmid pKM101, which can replace the R388 T4SS although with lower efficiency (Llosa et al. 2003). We have improved the assay conditions by using donor cells in exponential phase, to avoid fluctuations in the steady-state level of TrwB (not shown). The effect of *trwB* mutations on DNA transfer was tested in this system (**Table 2**, assay I). TrwB variants N271D and K275A, previously reported not to affect DNA transfer efficiency (de Paz et al. 2010), consistently showed lower conjugation frequencies under the new assays conditions, while the newly constructed variants K398A and K421A showed conjugation frequencies higher than wt (3 and 4 fold higher, respectively) , and variant K275A K398A behaved similarly to wt. The triple variant K275A P337L S441G showed a strong reduction in DNA transfer efficiency (Table 2, assay I). A separate analysis of each mutation revealed that the variant S441G was responsible for

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 August 2017
doi: 10.3389/fmicb.2017.01503

DNA Delivery and Genomic Integration into Mammalian Target Cells through Type IV A and B Secretion Systems of Human Pathogens

*Dolores L. Guzmán-Herrador[1], Samuel Steiner[2], Anabel Alperi[1], Coral González-Prieto[1†], Craig R. Roy[2] and Matxalen Llosa[1]**

[1] Departamento de Biología Molecular, Universidad de Cantabria (UC), Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC, UC-CSIC-SODERCAN), Santander, Spain, [2] Department of Microbial Pathogenesis, Boyer Center for Molecular Medicine, Yale University School of Medicine, New Haven, CT, United States

We explore the potential of bacterial secretion systems as tools for genomic modification of human cells. We previously showed that foreign DNA can be introduced into human cells through the Type IV A secretion system of the human pathogen *Bartonella henselae*. Moreover, the DNA is delivered covalently attached to the conjugative relaxase TrwC, which promotes its integration into the recipient genome. In this work, we report that this tool can be adapted to other target cells by using different relaxases and secretion systems. The promiscuous relaxase MobA from plasmid RSF1010 can be used to deliver DNA into human cells with higher efficiency than TrwC. MobA also promotes DNA integration, albeit at lower rates than TrwC. Notably, we report that DNA transfer to human cells can also take place through the Type IV secretion system of two intracellular human pathogens, *Legionella pneumophila* and *Coxiella burnetii*, which code for a distantly related Dot/Icm Type IV B secretion system. This suggests that DNA transfer could be an intrinsic ability of this family of secretion systems, expanding the range of target human cells. Further analysis of the DNA transfer process showed that recruitment of MobA by Dot/Icm was dependent on the IcmSW chaperone, which may explain the higher DNA transfer rates obtained. Finally, we observed that the presence of MobA negatively affected the intracellular replication of *C. burnetii*, suggesting an interference with Dot/Icm translocation of virulence factors.

Keywords: protein secretion, bacterial conjugation, *Legionella pneumophila*, *Coxiella burnetii*, *Bartonella henselae*, conjugative relaxase, intracellular pathogen, gene therapy

OPEN ACCESS

Edited by:
Manuel Espinosa, Centro de Investigaciones Biológicas (CSIC), Spain

Reviewed by:
Elisabeth Grohmann, Beuth University of Applied Sciences, Germany
Jose Angel Ruiz-Masó, Centro de Investigaciones Biológicas (CSIC), Spain

****Correspondence:***
Matxalen Llosa
llosam@unican.es

†Present address:
Coral González-Prieto, Department of Medicine, Division of Infectious Diseases, Massachusetts General Hospital – Department of Microbiology and Immunobiology, Harvard Medical School, Cambridge, MA, United States

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *30 June 2017*
Accepted: *26 July 2017*
Published: *22 August 2017*

Citation:
Guzmán-Herrador DL, Steiner S, Alperi A, González-Prieto C, Roy CR and Llosa M (2017) DNA Delivery and Genomic Integration into Mammalian Target Cells through Type IV A and B Secretion Systems of Human Pathogens. Front. Microbiol. 8:1503. doi: 10.3389/fmicb.2017.01503

INTRODUCTION

Bacterial Type IV secretion systems (T4SS) selectively deliver macromolecules to other cells or to the extracellular media. An outstanding feature of these secretion systems is their ability to secrete both, protein and DNA molecules, a particularity that distinguishes them from other types of secretion systems. In addition, the secreted substrates can be delivered to either prokaryotic

or eukaryotic cells. This plasticity allows T4SS to be involved in bacterial processes as diverse as horizontal DNA transfer or virulence (Christie, 2016).

Bacterial Type IV secretion systems are multiprotein complexes formed by different constitutive elements: a core complex spanning both bacterial membranes, which forms the transport conduit; a pilus-like appendage, whose function as a transport channel is still under debate; a series of cytoplasmic ATPases, which energize the transport process; and elements necessary to recruit and present the substrates to the translocation machine, including chaperones that are variable for each system (Zechner et al., 2012). Within the family of T4SS, two sub-families were described based on sequence homologies: The Type IV A-IV B secretion systems (T4ASS and T4BSS, respectively). The formers are homologous to the prototypical VirB T4SS of *Agrobacterium tumefaciens* and have been characterized extensively, both functionally and structurally (Chandran Darbari and Waksman, 2015). Members of this family form part of conjugative systems of plasmids such as R388 or RP4; others are encoded in the genomes of human pathogens such as *Bartonella henselae (Bh), Brucella melitensis* or *Helicobacter pylori* among others, and their main role is to inject virulence factors to the target human cell. Similarly, T4BSS members are encoded in conjugative plasmids such as F, and in the chromosomes of human pathogens such as *Legionella pneumophila (Lp)* and *Coxiella burnetii (Cb)*. Research on T4BSS structure and function lags behind T4ASS; however, extensive work has been done regarding the role of T4BSS-delivered effectors within human cell (Hubber and Roy, 2010; Rolando and Buchrieser, 2014; Personnic et al., 2016).

As aforementioned, a distinctive feature of T4SS is their ability to secrete DNA molecules. This is the main molecular function of T4SS belonging to the conjugative machinery of self-transmissible plasmids (Cabezon et al., 2015). In order to secrete DNA, at least two components are essential in addition to the T4SS machinery: an origin of transfer (*oriT*), which is the DNA sequence required in *cis* on a DNA molecule to be transferred, and a conjugative relaxase, which cuts the DNA strand to be transferred at the *oriT*. Many plasmids also encode for accessory nicking proteins, which assist the DNA processing by the relaxase. The DNA is transferred as a single strand covalently attached to the relaxase, which itself is the substrate of the T4SS; the nucleoprotein complex enters the recipient cell, where the relaxase catalyzes the recircularization of the transferred DNA strand (Garcillan-Barcia et al., 2007; Gonzalez-Perez et al., 2007).

Notably, some conjugative relaxases have the ability to catalyze site-specific recombination between two copies of *oriT*. This phenomenon was first described for the R388 relaxase TrwC (Llosa et al., 1994). TrwC acts as a site-specific recombinase on supercoiled substrates containing minimal target sequences (Cesar et al., 2006). This ability is shared by some, but not all, conjugative relaxases, and it is unclear why. MobA, the relaxase of the mobilizable plasmid RSF1010 (virtually identical to plasmid R1162), is able to catalyze *oriT*–*oriT* recombination on single-stranded substrates but not on supercoiled plasmid substrates (Meyer, 1989). TrwC can also catalyze the integration of the transferred DNA molecule into a target sequence present in the recipient bacterium (Draper et al., 2005); moreover, the protein can catalyze integration into DNA sequences present in the human genome that resemble its natural target, the *oriT* (Agundez et al., 2012), opening the possibility that this relaxase could work as a site-specific integrase in human cells (Gonzalez-Prieto et al., 2013). Recently, we have shown that the relaxase TrwC is active in a human cell after delivery by the T4SS of *Bartonella henselae*, where it can promote the integration of foreign DNA into the human genome, although without site-specificity (Gonzalez-Prieto et al., 2017). The integration rate of the foreign DNA introduced by TrwC was about 100 times higher compared to when it was introduced by the Mob relaxase from *Bartonella* cryptic plasmid pRGB1, or by transfection.

Gene therapy strategies combine methods to introduce DNA into specific human cell types and to promote DNA integration in the human genome for stable expression. Bacteria have previously been used as vectors for DNA delivery into mammalian cells; the process, known as bactofection, is based on the engulfment of bacteria by an eukaryotic cell, which causes bacterial lysis and DNA release (Celec and Gardlik, 2017). We have previously shown that DNA of any origin and length can be introduced into specific human cell types using *B. henselae* as a delivery agent (Fernandez-Gonzalez et al., 2011). In contrast to bactofection, in this case the DNA is secreted by the living bacterium. *B. henselae* encodes a T4ASS named VirB/D4, which translocates effector proteins to the infected human cell, contributing to its virulence (Saenz et al., 2007). We showed that the VirB/D4 T4SS is also capable of translocating relaxase-DNA complexes via a process resembling bacterial conjugation. DNA transfer was dependent on the conjugative elements required to process the DNA in the donor bacterium, which in this case were derived from the conjugative plasmid R388. No DNA transfer occurred in the absence of the relaxase TrwC, and it was severely impaired in the absence of the conjugative coupling protein TrwB. In a parallel work, Schroder et al. (2011) similarly showed DNA transfer through the *B. henselae* VirB/D4 using the Mob relaxase of a natural plasmid of *Bartonella*; in this case, it was necessary to fuse the known T4 recruiting signal (the BID domain) to the relaxase in order to attain efficient DNA transfer. This discovery had interesting biological implications, opening the possibility that pathogens naturally send DNA to their host cell, and potential biotechnological applications, constituting a new way of DNA delivery to specific human cells (Llosa et al., 2012).

In this work, we asked whether this DNA delivery system could be extended to T4SS from other human pathogens targeting different cell types. We infect cultured mammalian cell lines with *B. henselae*, *L. pneumophila*, or *C. burnetii,* all containing mobilizable plasmids with markers for eukaryotic selection and encoding different conjugative relaxases. We report that DNA can be delivered to human cells through the T4BSS of *L. pneumophila* and *C. burnetii*, which belong to a distant family of T4SS. This suggests that DNA transfer may be an intrinsic feature of T4SS. DNA transfer and integration rates depend on the relaxase used. All these elements could add to the development of useful tools for *in vivo* genetic modification

of human cells. In addition, DNA is a trackable substrate which could be used to study the T4 secretion process in the mammalian host.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

Bacterial strains used in this work are listed in **Table 1**. *Escherichia coli* (*Ec*) strains DH5α and D1210 were used for DNA manipulations. *B. henselae* strain RSE247, *L. pneumophila* serogroup 1 strain Lp01 (*hsdR*, *rpsL*; Berger and Isberg, 1993), and *C. burnetii* strain RSA439 Nine Mile phase II (NMII), or derivatives from these strains as indicated, were used for infection of cultured cells.

Escherichia coli strains were grown at 37°C in Luria-Bertani broth, supplemented with agar for growth on plates. *B. henselae* was grown on Columbia blood agar (CBA) plates at 37°C under a 5% CO_2 atmosphere. *L. pneumophila* strains were grown on charcoal yeast extract (CYE) plates [1% yeast extract, 1% *N*-(2-acetamido)-2-aminoethanesulfonic acid (ACES; pH 6.9), 3.3 mM L-cysteine, 0.33 mM $Fe(NO_3)_3$, 1.5% Bacto agar, 0.2% activated charcoal] at 37°C, supplemented with 100 μg/ml thymidine if required. *C. burnetii* was grown axenically in liquid acidified citrate cysteine medium 2 (ACCM-2) for 6 days or on ACCM-2 agarose for >8 days at 37°C, 5% CO_2, and 2.5% O_2 as previously described (Omsland et al., 2011).

For plasmid selection, antibiotics were added at the following final concentrations: ampicillin (Ap), 100 μg/ml; kanamycin monosulfate (Km), 20 μg/ml (*L. pneumophila*), 50 μg/ml (*E. coli*, *B. henselae*) or 375 μg/ml (*C. burnetii*); streptomycin (Sm), 300 μg/ml (*E. coli*) or 100 μg/ml (*B. henselae*, *L. pneumophila*); gentamicin sulfate (Gm), 10 μg/ml (*E. coli, B. henselae*) or 5 μg/ml (*L. pneumophila*); chloramphenicol (Cm), 25 μg/ml (*E. coli*) or 3 μg/ml (*C. burnetii*).

Plasmids and Plasmid Constructions

Bacterial plasmids are listed in **Table 2**. Oligonucleotides used for plasmid constructions are listed in **Table 3**. Plasmids pAA58, pLG03, pLG04, pMTX808, pMTX821, and pMTX822 were constructed by the isothermal assembly method (Gibson et al., 2009) using the HiFi assembly cloning kit (New England Biolabs). Plasmids pLG05 and pLG06 were constructed by standard restriction cloning techniques (Sambrook and Russell, 2001).

pAA58 was generated by assembling the eGFP eukaryotic expression cassette from pHP161 into the PstI sites of RSF1010K, which was itself amplified in two overlapping PCR fragments. To generate pLG03, pLG04, pLG05, and pLG06, the hygromycin resistance cassette from pMTX708 was amplified and assembled into the SgsI site of pMTX808 and pAA58, or into the ClaI site of pMTX821 and pMTX822, respectively. pMTX808 was constructed by insertion of an ampicillin resistance cassette (amplified from pJB-KAN) into the *mobA* gene of pAA58. The cassette was inserted at the unique BstZ17I site which lies at nt 320 of *mobA*, leaving unaffected the downstream *mobB* and *repB* ORFs which overlap *mobA*. pMTX821 and pMTX822 were generated by insertion of a kanamycin resistance cassette from pJB-KAN into the gentamicin resistance cassette of pHP159 and pHP181, respectively.

Plasmids were routinely introduced in all strains by electroporation. The protocol for *C. burnetii* electroporation was previously described (Newton et al., 2014); electroporation was carried out with a Bio-Rad GenePulser Xcell (settings: 1.8 kV, 500 Ω, 25 μF). To make competent *L. pneumophila* cells, bacteria were collected from 48 h-patches grown on CYE plates, resuspended in 1 ml ice-cold sterile ddH_2O, and centrifuged for 2 min in Eppendorf tubes. The washing step was repeated three times. The pellet was resuspended in 1 ml ice-cold sterile glycerol, pelleted for 5 min and resuspended in 1 ml ice-cold sterile glycerol, from which 100 μl aliquots were either frozen at −80°C or used for transformation. Electroporation was

TABLE 1 | Bacterial strains used in this work.

Name	Relevant genotype	Description/comments	Reference
Escherichia coli			
D1210	*recA hspR hsdM rpsl lacI*^{q}	Sm^R, $LacI^q$ constitutive expression	Sadler et al., 1980
DH5α T1 phage resistant	*F- φ80lacZΔM15 Δ((lacZYA-argF)U169 recA1 endA1 hsdR17(rk−, mk+) phoA supE44 λ-thi-1 gyrA96 relA1 tonA*	Nx^R, T1 phage resistant strain	Killmann et al., 1996
Bartonella henselae			
RSE247	Sm^R	Sm^R spontaneous mutant of ATCC 49882	Schmid et al., 2004
Legionella pneumophila			
Lp02	Lp01 *thyA*	Spontaneous thymidine auxotroph	Berger and Isberg, 1993
Lp03	Lp02 *dotA*	Spontaneous *dotA* mutant	Berger and Isberg, 1993
CR503	Lp01 Δ*icmS*Δ*icmW*		Coers et al., 2000
Coxiella burnetii			
RSA439	Wild type	Plaque-purified Nine Mile phase II (NMII) clone 4	Williams et al., 1981
RSA439 *dotA*::Tn	*dotA*::TnA7	Transposon insertion mutant in *dotA* (CBU_1648), Cm^R, mCherry	Newton et al., 2014
RSA439 intergenic::Tn	intergenic::TnA7	Transposon insertion mutant between *hemD* and CBU_2078, Cm^R, mCherry; shows intracellular replication comparable to wild type	Newton et al., 2014

TABLE 2 | Plasmids used in this work.

Relaxase	Other conjugative elements	Plasmid	Selection markers[1]	Description	Reference
Mob-BID	pBGR *oriT*	pRS130	Km^R Neo^R	pBGR::*mob:BID+gfp+neo*	(Schroder et al., 2011)
MobA	RSF1010 *oriT mobB mobC*	RSF1010K	Km^R	RSF1010 Δ*Sm KmR*	(Lessl et al., 1993)
MobA	RSF1010 *oriT mobB mobC*	pAA58	Km^R	RSF1010K::*egfp*	This work
MobA	RSF1010 *oriT mobB mobC*	pLG04	Km^R Hyg^R	pAA58::*hyg*	This work
TrwC	R388 *oriT trwA trwB*	pHP159	Gm^R	pBBR6::*oriT trwABC+egfp*	(Fernandez-Gonzalez et al., 2011)
TrwC	R388 *oriT trwA trwB*	pHP161	Gm^R	pBBR6::*oriT trwABC+egfp*	(Fernandez-Gonzalez et al., 2011)
TrwC	R388 *oriT trwA trwB*	pMTX821	Km^R	pHP159::*Km* Δ*Gm*	This work
TrwC	R388 *oriT trwA trwB*	pCOR31	Gm^R Neo^R	pHP159::*neo*	(Gonzalez-Prieto et al., 2017)
TrwC	R388 *oriT trwA trwB*	pLG05	Km^R Hyg^R	pMTX821::*hyg*	This work
TrwC-RalF	R388 *oriT trwA trwB*	pAA12	Gm^R	pHP159::*trwC-RalF TS*	(Alperi et al., 2013)
–	RSF1010 *oriT mobB mobC*	pMTX808	Km^R Ap^R	pAA58::*Ap MobA-*	This work
–	RSF1010 *oriT mobB mobC*	pLG03	Km^R Ap^R Hyg^R	pMTX808::*hyg*	This work
–	R388 *oriT trwA trwB*	pHP181	Gm^R	pBBR6::*oriT trwAB+egfp*	(Fernandez-Gonzalez et al., 2011)
–	R388 *oriT trwA trwB*	pMTX822	Km^R	pHP181::*Km* Δ*Gm*	This work
–	R388 *oriT trwA trwB*	pCOR35	Gm^R Neo^R	pHP181::*neo*	(Gonzalez-Prieto et al., 2017)
–	R388 *oriT trwA trwB*	pLG06	$Km^R Hyg^R$	pMTX822::*hyg*	This work
nr[2]	nr[2]	pMTX708	Ap^R Hyg^R	pTRE2hyg::*Ptac-oriT*	(Gonzalez-Prieto et al., 2017)
nr[2]	nr[2]	pJB-KAN	Km^R Ap^R	Cloning vector	(Omsland et al., 2011)

[1]R, *resistance to Ampicillin (Ap), Gentamycin (Gm), Kanamycin (Km), Hygromycin (Hyg) or Neomycin (Neo).* [2]*nr, not relevant.*

TABLE 3 | Oligonucleotides used for plasmid constructions.

Plasmid constructed (IA/RC)[1]	Oligonucleotide sequence (5′ to 3′)[2]	Amplified fragment
pLG03, pLG04 (IA)	TCCAGATGTATGCTCTTCTGCTCGGCGCGCC**TTTCGTCTCGAGGCAGTG**	Hyg^R cassette
	TGCGATGATAAGCTGTCAAACAGGCGCGCC**GTCAGTTAGGGTGTGGAAAG**	
pLG05, pLG06 (RC)	CCAAACATCGAT**GTCAGTTAGGGTGTGGAAAG**	Hyg^R cassette
	CCAAACATCGAT**CTTTCGTCTCGAGGCAGTG**	
pAA58 (IA)	**AGCTTGCCGCCGCCGCAG**	RSF1010K
	GGTCTATTGCCTCCCGGTATTCCTGT	
	CGCCCAGATCATCGACTTACAGGAATAC	
	GAGCAGAAGAGCATACATCTGGAAGC	
	GCCGCTTTCCTGGCTTTGCTTCCAGATGTATGCTCTTCTGCTCGGCGCGCC**TGTTTGACAGCTTATCATCGCAG**	eGFP cassette
	GTGCGGATGAAGTCAGCTCCACCTGCGGCGGCGGCAAGCTCCTGCAGG**CCCCGACACCCGCCAACAC**	
pMTX808 (IA)	GCACCTGACCGGTGCCGAGCGCCTGCCGTATTG**AGAAGGCCATCCTGACGGA**	Ap^R cassette
	TCGCCGCCACCGGCATGGATGGCCAGCGTA**TTACCAATGCTTAATCAGTGAG**	
pMTX821, pMTX822 (IA)	AGTATGGGCATCATTCGCACATGAA**GGCGATTCGCCGCTTTC**	Km^R cassette
	GGTGGCGGTACTTGGGTCGAT**TTATCAGAAGAACTCGTCAAG**	Km^R cassette

[1]*IA, isothermal assembly; RC, restriction cloning.* [2]*Nucleotides annealing to the PCR template are shown in bold, and restriction sites used for cloning are underlined.*

carried out adding 500 ng DNA and transferring the mixture to a cooled Bio-Rad 0.2-cm cuvette for electroshock with a Bio-Rad GenePulser Xcell set at 2.0 kV, 25 μF, and 200 Ω. After electroporation, 1 ml of AYE broth [1% yeast extract, 1% ACES pH 6.9, 3.3 mM L-cysteine, 0.33 mM $Fe(NO_3)_3$] was added, supplemented with thymidine when required, and the mixture was transferred to a 10 ml tube for incubation for 6 h at 37°C with orbital shaking. The cells were then plated on CYE supplemented with the appropriate antibiotics.

For *B. henselae,* a plate grown for 2 to 3 days was harvested with a sterile cotton swab and resuspended in 950 μl of LB. The suspension was centrifuged at 4,000 rpm for 5 min at 4°C, and the pellet was washed in 950 μl of ice-cold 10% glycerol (three times); 40 μl of these competent cells was transferred to a cooled tube, and 3 μl of DNA (300 ng/μl) was added. The mixture was incubated on ice for 15 min and transferred to a cooled Bio-Rad 0.2-cm cuvette for electroshock with a Bio-Rad Pulse controller II at 2.5 kV/cm, 25 μF, and 200 Ω. After electroporation, 1 ml of SB broth (RPMI 1640 plus L-glutamine, 42 mM HEPES, 1% sodium pyruvate, 5% heat-inactivated fetal calf serum, and 5% sheep blood lysate) was added, and the mixture was transferred to an Eppendorf tube for incubation for 3.5 h at 37°C under 5% CO_2 conditions with slow shaking. The cells were then centrifuged at 4,000 rpm for 4 min at room temperature. The pellet was resuspended in 40 μl SB broth and plated on CBA supplemented with the appropriate antibiotics.

Cell Lines and Cell Culture Conditions

The cell lines used for bacterial infections are listed in **Table 4**. EA.hy926 and HeLa cell lines were routinely grown in Dulbecco's modified Eagle medium (DMEM; Lonza or Gibco), and Chinese Hamster Ovary (CHO) cells were maintained in minimal essential medium MEMα (Gibco); both media were supplemented with 10% heat inactivated fetal bovine serum (FBS; Lonza or Sigma). Cells were incubated at 37°C under 5% CO_2.

Infections

Bartonella henselae strains containing the appropriate plasmids were grown on CBA plates for 3 to 4 days. Human cells were seeded 1 day before infection. For routine infections, cells were seeded in 6-well plates (80,000 cells per well) in 3 ml of medium. 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 10-cm tissue culture dishes seeded with 450,000 cells in 12 ml of medium. The day of infection, DMEM was replaced by M199 medium (Gibco) supplemented with 10% FBS and appropriate antibiotics to select for the *B. henselae* strains to be added. 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} of 1 corresponds to 10^9 bacteria/ml (Kirby and Nekorchuk, 2002). Bacteria were added to the human cells to get a multiplicity of infection (MOI) of 400 bacteria per host cell. The dishes or plates were incubated for 72 h at 37°C under 5% CO_2.

Coxiella burnetii strains containing the appropriate plasmids were grown for 6 days in liquid cultures. 25,000–50,000 HeLa 229 cells were seeded in DMEM 5% FBS into 24-well plates 6–8 h before they were infected at a MOI of 500, unless specified otherwise. Bacteria were quantified measuring genome equivalents (GE) as previously described (Newton et al., 2014). Infections were incubated for 96 h at 37°C under 5% CO_2. Wells for quantification of intracellular replication were washed once with PBS at approximately 15 h post infection (hpi) before the addition of fresh DMEM 5% FBS. Wells for flow cytometry experiments were not washed.

Legionella pneumophila strains containing the appropriate plasmids were harvested from a heavy patch (after 48 h growth on CYE plates), and used to infect CHO FcγRII cells, stably expressing the receptor FcγRII. This receptor allows *L. pneumophila* opsonized with anti-*Legionella* antibodies to be internalized efficiently by non-phagocytic cells (Arasaki and Roy, 2010). FcγRII cells were grown to near confluency in 24-well dishes. Bacteria were opsonized with rabbit anti-*Legionella* antibody diluted 1/1000 for 20 min at room temperature with shaking. Bacteria were then added to the cells at an estimated MOI of 10. The cells were centrifuged 5 min at 1000 rpm and incubated for 1 h, washed three times with PBS (Gibco) and incubated in fresh media for 24 h at 37°C under 5% CO_2.

TABLE 4 | Mammalian cell lines used in this work.

Name	Description	Reference
CHO FcγRII	Chinese hamster ovary cells producing the FcγRII protein	Joiner et al., 1990
EA.hy926	Fusion cell line of human umbilical vein endothelial cells (HUVEC) and adenocarcinomic human alveolar basal epithelial cells (A549)	ATCC CRL-2922
HeLa	Human epithelial cells of cervix adenocarcinoma	ATCC CCL-2
HeLa 229	Human epithelial cells of cervix adenocarcinoma	ATCC CCL-2.1

Detection of GFP Positive Cells by Flow Cytometry

At the indicated hours post infection (hpi) indicated for each bacteria, infected cells were washed with PBS, trypsinized, and analyzed by flow cytometry using a Cytomics FC500 flow cytometer (Beckman Coulter) for *B. henselae* infections, or a BD Accuri C6 flow cytometer (BD Biosciences) for *L. pneumophila* and *C. burnetii* infections. Data were analyzed using the software for each cytometer and FlowJo (Tree Star, Inc.) software. Singlet cells were gated based on SSC-H/FSC-H and GFP positive cells (detected in the FL1-H channel) were gated based on uninfected control cells. The gate was set to approximately 0.05% GFP^+ cells in the uninfected control sample.

Fluorescence Microscopy

At the indicated hpi, wells with infected cells were washed with PBS and the plates were placed directly on a Nikon Eclipse TE2000-S inverted fluorescence microscope with a 10× objective lens. Digital images were acquired with a microscope camera (Photometrics CoolSNAP EZ) controlled by SlideBook™ (Intelligent Imaging Innovations).

Detection of Stable Integrants

At 72 hpi, either 500 μg/ml G418 disulfate salt (Sigma–Aldrich) or 300 μg/ml Hygromycin B (Invitrogen), as appropriate, were added to HeLa cells infected with *B. henselae*, and selection was maintained for 4 to 5 weeks. Resistant colonies on the plates were counted.

In order to calculate the integration rate, integration experiments were always performed in parallel with infections to measure GFP positive cells by flow cytometry. The resulting percentage of GFP positive cells was extrapolated to the number of cells in the 10-cm plate used to detect integrants, and the number of resistant colonies was divided by the inferred number of GFP positive cells.

Determination of Genome Equivalents (GE)

Quantification of *C. burnetii* intracellular replication was performed as described in Newton et al. (2014). Briefly, infected HeLa cells were lysed in ddH_2O at specific time points post infection. Total genomic DNA was extracted using the Illustra Bacteria GenomicPrep Mini Spin Kit (GE Healthcare) and GE were quantified by qPCR using *dotA*-specific primers (GCGCAATACGCTCAATCACA,

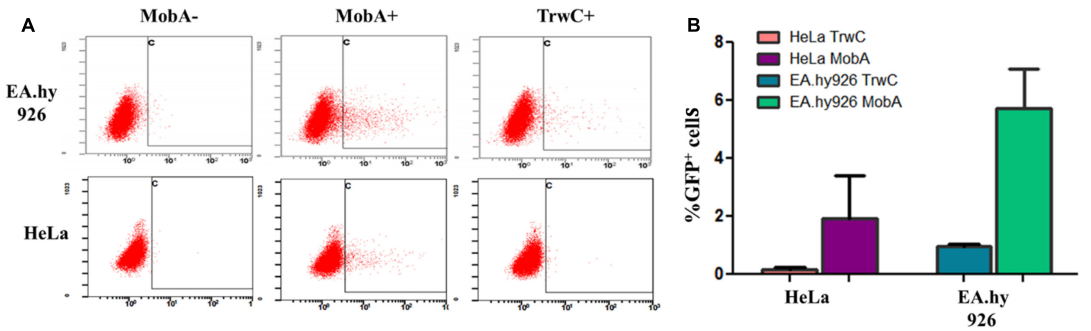


FIGURE 1 | Transient expression of transferred DNA in HeLa and EA.hy926 cells. Pools of cells obtained at 3 days post infection with *B. henselae* were analyzed by flow cytometry using uninfected cells as control. **(A)** Representative plots (cell granularity versus GFP intensity). The square marks the population considered as positive. The relaxase present in each experiment is indicated on top of the panels. **(B)** Mean values of the percentage of GFP positive cells from 2 to 4 independent experiments, after subtracting the background values of the negative controls with no relaxase. The relaxase present in each experiment is indicated.

CCATGGCCCCAATTCTCTT). The generation of this short PCR product is not affected by the presence of a transposon in the *dotA*::Tn mutant strain.

RESULTS

The conjugative relaxase TrwC can be translocated through the T4SS VirB/D4 of *B. henselae* to human cells, where it promotes the integration of the transferred DNA into the recipient genome (Gonzalez-Prieto et al., 2017). In this work, we wanted to test whether this is a unique feature of TrwC and VirB/D4, or other systems can also be combined to deliver and integrate DNA into human cells.

To test DNA transfer mediated by the relaxase MobA of the mobilizable plasmid RSF1010, we constructed a derivative carrying an eukaryotic eGFP expression cassette to detect gene expression from the human cell nucleus. An insertion of an ampicillin resistance cassette in *mobA* served as a negative control. The insertion is located in the 5′ region of the ORF, thus not affecting the expression of the ORFs *mobB* and especially *repB'*, which encodes a DNA primase required for plasmid replication. We observed that this *mobA*$^{-}$ construct had a higher copy number than the parental plasmid, as judged from the amount of DNA extracted from parallel cultures (data not shown). This phenomenon has previously been reported, and attributed to the repressor role of MobA/RepB in replication (Frey et al., 1992).

These plasmids (pAA58 and pMTX808; **Table 2**) were introduced in *B. henselae*, and the resulting strains were used to infect both EA.hy926 and HeLa human cell lines. The former is derived from HUVEC cells, which are the natural target of *B. henselae in vivo*; however, HeLa cells can also be infected by *B. henselae* with lower efficiency, and we showed that TrwC-mediated DNA transfer takes place to HeLa cells as well (Gonzalez-Prieto et al., 2017). *B. henselae* carrying plasmids coding for either MobA or TrwC, or relaxase mutants as negative controls, were used for infections. To assess transfer of the plasmid DNA to the human cells, flow cytometry was used to quantify the expression of the eGFP cassette per cell, thus allowing the determination of the percentage of GFP positive cells. The results are shown in **Figure 1** and **Table 5**, top 8 rows. We observed DNA transfer when the plasmids encoded a functional relaxase, and background levels in the absence of a relaxase. DNA transfer rates were notably higher when using MobA as the leading relaxase compared to TrwC.

In order to measure genomic integration of the transferred DNA, we constructed plasmid derivatives encoding antibiotic resistance cassettes (see **Table 2**). The plasmids containing R388 conjugative elements carried a neomycin gene; however, this was not used in these experiment because of the presence of a kanamycin resistance gene in the RSF1010K backbone, which could lead to recombination between both cassettes. Instead, a hygromycin resistance cassette was inserted. In order to avoid an effect caused by the different antibiotic selections applied, we also constructed Hygromycin-resistant derivatives encoding TrwC (**Table 2**), and we found that TrwC-mediated integration rate did not vary when the selection applied was hygromycin B or Geneticin (data not shown).

HeLa cells were used as target cells to measure DNA integration, because in contrast to EA.hy926 cells HeLa cells show enhanced survival during the 4–5 weeks of antibiotic selection required to measure resistant colonies (Gonzalez-Prieto et al., 2017). The cells were infected with *B. henselae* carrying the different plasmids. A plasmid derived from the cryptic *Bartonella* plasmid pBGR1 was also assayed for comparison, since it has been reported that its relaxase mediates DNA transfer but does not promote integration of the transferred DNA (Gonzalez-Prieto et al., 2017). After applying the antibiotic selection, resistant colonies were counted, and integration rates were calculated dividing this number by the number of GFP positive cells determined in parallel infection experiments (see Materials and Methods for details). The results (**Figure 2**) indicate that the integration rate for the MobA constructs was approximately one-log higher than in case of Mob-BID, which suggest that MobA promotes integration of the transferred DNA. It can also be observed that TrwC has a stronger effect on integration than MobA (approximately five-fold higher DNA integration).

TABLE 5 | Rates of DNA transfer to mammalian cells through T4ASS and T4BSS.

Donor bacteria (genotype)	T4SS	Transfer system	Relaxase	Infected cells	GFP$^+$ mammalian cells[1]	
					Flow cyt %	Scope
Bh RSE247 (wt)	Functional	RSF1010	MobA	EA.hy926	5.72 ± 1.37	nq[2]
Bh RSE247 (wt)	Functional	RSF1010	–	EA.hy926	0.29 ± 0.07	nq[2]
Bh RSE247 (wt)	Functional	R388	TrwC	EA.hy926	1.00 ± 0.09	nq[2]
Bh RSE247 (wt)	Functional	R388	–	EA.hy926	0.14 ± 0.19	nq[2]
Bh RSE247 (wt)	Functional	RSF1010	MobA	HeLa	2.00 ± 1.48	nq[2]
Bh RSE247 (wt)	Functional	RSF1010	–	HeLa	0.07 ± 0.05	nq[2]
Bh RSE247 (wt)	Functional	R388	TrwC	HeLa	0.20 ± 0.03	nq[2]
Bh RSE247 (wt)	Functional	R388	–	HeLa	0.04 ± 0.06	nq[2]
Lp Lp02 (wt)	Functional	RSF1010	MobA	CHO FcγRII	0.35 ± 0.12	nq[2]
Lp Lp03 *(dotA)*	No transport	RSF1010	MobA	CHO FcγRII	0.03 ± 0.05	$<5 \times 10^{-6}$
Lp Lp02 (wt)	Functional	RSF1010	–	CHO FcγRII	0.00 ± 0.00	$<5 \times 10^{-6}$
Lp CR503 (*icmS icmW)*	No chaperone	RSF1010	MobA	CHO FcγRII	0.00 ± 0.00	$<5 \times 10^{-6}$
Lp Lp02 (wt)	Functional	R388	TrwC	CHO FcγRII	0.00 ± 0.00	$<5 \times 10^{-6}$
Lp Lp02 (wt)	Functional	R388	TrwC-RalF	CHO FcγRII	0.00 ± 0.00	1×10^{-5}
Lp Lp03 *(dotA)*	No transport	R388	TrwC-RalF	CHO FcγRII	nq[2]	$<5 \times 10^{-6}$
Lp Lp02 (wt)	Functional	R388	–	CHO FcγRII	nq[2]	$<5 \times 10^{-6}$
Lp CR503 (*icmS icmW)*	No chaperone	R388	TrwC-RalF	CHO FcγRII	nq[2]	2×10^{-5}
Cb intergenic::Tn (wt)	Functional	RSF1010	MobA	HeLa	0.56 ± 0.53	nq[2]
Cb dotA::Tn	No transport	RSF1010	MobA	HeLa	0.04 ± 0.02	nq[2]
Cb intergenic::Tn (wt)	Functional	RSF1010	–	HeLa	0.10[3] ± 0.04	$<5 \times 10^{-6}$

[1]*DNA transfer is measured as the ratio of mammalian recipient cells expressing GFP. Data from flow cytometry (left column) show the percentage of GFP positive cells (mean ± SD of two to eight independent assays). Infected cells were also screened visually under the microscope (right column). Positive cells were counted and divided by the total number of cells per well (estimated as 200.000). The screen was performed at least twice for each condition.* [2]*nq, not quantified.* [3]*Due to higher background (see text for details).*

Earlier studies reported Dot/Icm-dependent conjugative DNA transfer of RSF1010 (Vogel et al., 1998), implying that MobA can mediate the translocation of an attached DNA substrate through the T4BSS Dot/Icm of *L. pneumophila*. Thus, we asked whether the Dot/Icm T4SS could also promote DNA transfer to mammalian cells upon infection by *L. pneumophila*. In addition to testing MobA-mediated transfer, we tested DNA transfer mediated by TrwC and TrwC-RalF, a fusion protein carrying the C-terminal 20 residues of the *L. pneumophila* Dot/Icm substrate RalF, that has been shown to be sufficient for translocation (Nagai et al., 2005). In contrast to the infection experiments done with *B. henselae*, for infections with *L. pneumophila* a MOI of 10 was used and DNA transfer was monitored at 24 hpi. As shown in **Figure 3A** and **Table 5**, we detected GFP positive cells after infection by a mechanism dependent on the Dot/Icm T4BSS and the relaxase MobA. Thus, we show for the first time that DNA transfer can occur through a T4BSS into mammalian cells. Using the same flow cytometry assay, we did not detect GFP positive cells above the background when the mobilizable plasmids encoded the relaxase TrwC or TrwC-RalF. However, inspection of the infected cells by fluorescence microscopy did reveal a small number of positive cells that expressed GFP uniformly and strongly after infection with *L. pneumophila* producing TrwC-RalF (**Figure 3B**). Positive cells were not observed in the negative controls or with TrwC-encoding plasmids.

The rate of DNA transfer was highly dependent on the conjugative DNA processing system used. This could be due to different relaxase recruitment efficiencies. The Dot/Icm T4BSS recruits a subset of its substrates through a chaperone complex formed by IcmS and IcmW (Cambronne and Roy, 2007). To determine if recruitment of the relaxases was dependent on this complex, a *L. pneumophila* Δ*icmS* Δ*icmW* mutant strain was used in infection experiments carrying plasmids which encode either MobA or TrwC-RalF. The results (**Table 5** and **Figure 3B**) indicate that the absence of IcmSW did not affect DNA transfer mediated by TrwC-RalF, while DNA transfer mediated by MobA was abolished in the absence of IcmSW.

The Dot/Icm T4BSS of *L. pneumophila* is closely related to that of *C. burnetii*, and several reports have shown that both can recruit the same effector proteins and cross-complement *icmSW* mutants (Zamboni et al., 2003; Zusman et al., 2003; Carey et al., 2011). Thus, we decided to test MobA-mediated DNA transfer through the Dot/Icm T4BSS of *C. burnetii*. HeLa cells were infected with *C. burnetii* strains harboring the plasmids with and without MobA at a MOI of 500, and GFP expression was investigated at 4 days post infection. The results are shown in **Table 5**, and **Figure 3C** shows representative plots. Similar to what was observed with *L. pneumophila*, GFP positive cells were only detected when the Dot/Icm T4BSS and the MobA relaxase were present.

Performing these experiments, we observed a difference in the background fluorescence intensity of HeLa cells depending

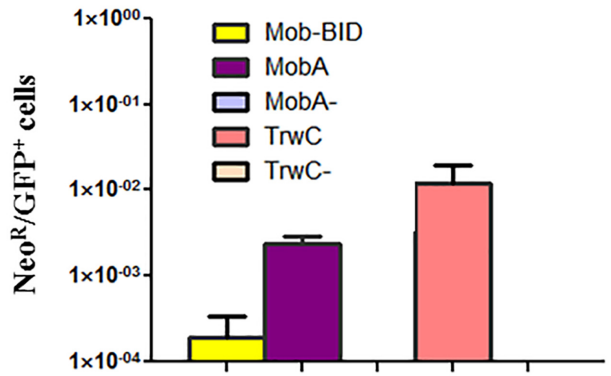


FIGURE 2 | DNA integration rate for different relaxases. HeLa cells were infected with *B. henselae* containing the plasmids encoding the indicated relaxases (or the negative controls lacking the relaxase) and at 3 days post infection, cells were either analyzed by flow cytometry (to determine transient expression of eGFP), or subjected to antibiotic selection (to determine permanent expression of the antibiotic resistance gene). The graph shows the ratio between the number of antibiotic-resistant cells and the number of GFP positive cells.

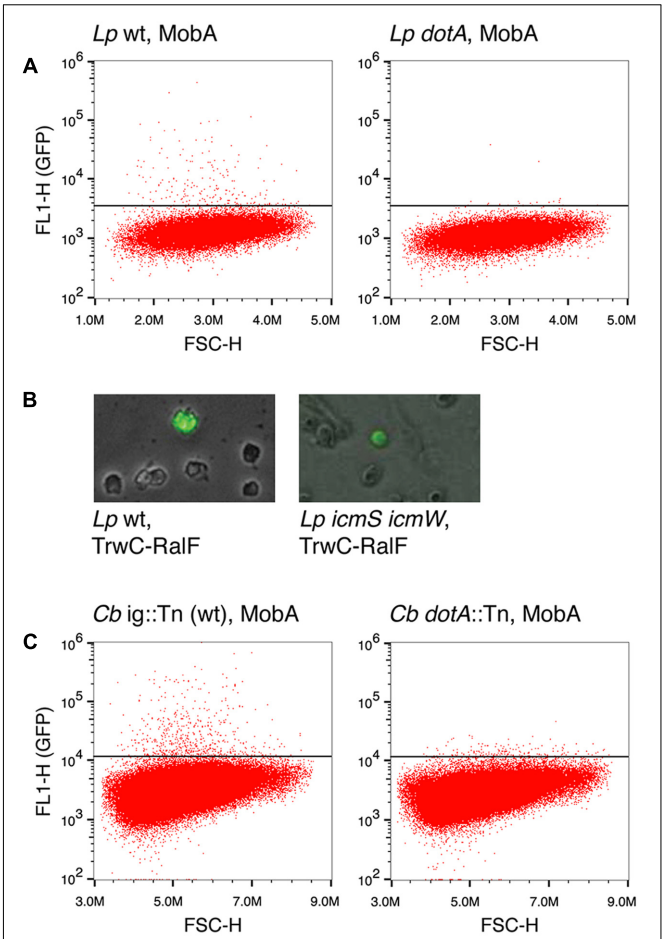


FIGURE 3 | DNA transfer through the Dot/Icm T4BSS. **(A)** Representative flow cytometry plots for quantification of GFP positive CHO FcγRII cells after infection with *L. pneumophila* strains Lp02 (wild type, left panel) or Lp03 (*dotA* mutant, right panel) harboring plasmid pAA58, which encodes MobA. **(B)** Fluorescence microscope images showing GFP positive CHO FcγRII cells infected with *L. pneumophila* strains Lp02 (wild type, left panel) or CR503 (Δ*icmS* Δ*icmW* mutant, right panel) harboring plasmid pAA12, which encodes TrwC-RalF. **(C)** Representative flow cytometry plots for quantification of GFP positive HeLa cells after infection with *C. burnetii* strains RSA439 intergenic::Tn (wild type, left panel) or RSA439 *dotA*::Tn (*dotA* mutant, right panel) harboring plasmid pAA58, which encodes MobA. ig, intergenic.

on the bacterial strain used for infection. A representative flow cytometry histogram is shown in **Figure 4A**. The background GFP fluorescence peak shifts toward a higher intensity when HeLa cells were infected with wild type *C. burnetii* or wild type *C. burnetii* harboring the plasmid with the *mobA* mutation, but not when cells were infected with wild type *C. burnetii* carrying the plasmid with the intact *mobA* gene. This higher fluorescence did not correspond to DNA transfer, since we did not detect any proper GFP positive cells by flow cytometry or using microscopy, but it contributed to a minimal raise in the background frequencies observed when infecting with a *mobA*$^-$ strain (see **Table 5**). However, the difference in background fluorescence may be attributed to a different amount of intracellular bacteria per cell. To test this hypothesis, HeLa cells were infected at a MOI of 50 and the number of intracellular *C. burnetii* was determined by measuring GE at two time points post infection. The results are shown in **Figure 4B**. A strain carrying the *mobA*-deficient plasmid replicates nearly as efficiently as a strain with no plasmid. In contrast, the same strain carrying a plasmid that encodes a functional MobA protein was severely impaired in intracellular replication. A *dotA* mutant that fails to replicate intracellularly due to the absence of a functional T4SS was used as a control in this assay.

DISCUSSION

In our previous reports, we showed that the conjugative relaxase TrwC can be translocated to human cells through the T4SS VirB/D4 of *B. henselae* (Fernandez-Gonzalez et al., 2011), and also that it promotes integration of the transferred DNA into the recipient genome (Gonzalez-Prieto et al., 2017). Whether these abilities were unique for TrwC and VirB/D4 remained to be tested. In this work, we report that different relaxases and T4SS can be used to transfer DNA to human cells and to promote DNA integration. In other words, relaxases and T4SS from various bacterial species can be combined to create tools intended to genetically modify specific human target cells in a permanent way, thus generating enormous biotechnological potential.

Firstly, we compared the ability of different relaxases to transfer DNA to mammalian cells and to promote DNA integration into the recipient genome when translocated by the same T4SS, VirB/D4. Human cells were infected with *B. henselae* carrying derivatives of the mobilizable plasmid RSF1010, encoding the relaxase MobA; with constructs containing the conjugative processing elements of the self-transferable plasmid R388, which encodes the relaxase TrwC; or with derivatives

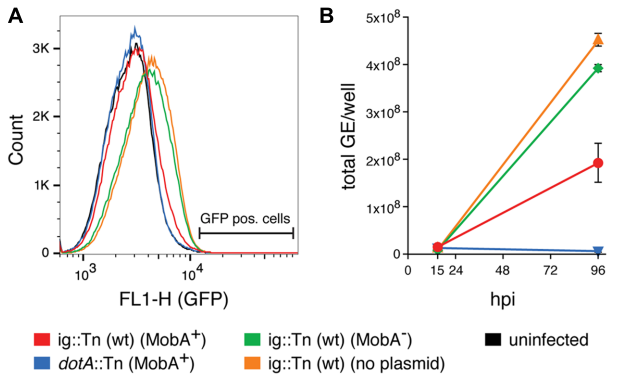


FIGURE 4 | The presence of MobA interferes with intracellular replication of *C. burnetii* in HeLa cells. **(A)** Flow cytometry histogram analysis of HeLa cells infected with the indicated *C. burnetii* strains at 4 days post infection. The location of GFP positive cells (see **Figure 3C** and Materials and Methods) is indicated. **(B)** Quantification of intracellular replication of *C. burnetii*. Total genome equivalents (GE)/well are shown at two time points post infection for infections with the indicated strains at a MOI of 50. The bottom strain legend applies for **(A,B)**. The presence or absence of plasmid-encoded MobA is indicated in brackets. ig, intergenic. hpi, hours post infection.

of *B. henselae* cryptic plasmid pBGR1, coding for the relaxase Mob fused to the BID signal for efficient recruitment by VirB/D4 (Schroder et al., 2011). When the three plasmids are compared in terms of DNA transfer and integration rates (**Figures 1**, **2**), we find that these vary significantly, with RSF1010 being the most efficiently transferred, while TrwC is the relaxase showing higher integration rates. The rate of DNA transfer is probably proportional to the efficiency with which the relaxase is recruited to the T4SS machinery; this assumption comes from previous works showing that the relaxase Mob itself could transfer DNA to human cells with barely detectable frequency, but when a recruitment secretion signal was fused to its C-terminal end, it transferred DNA to similar frequencies than TrwC (Schroder et al., 2011). In addition, in case of R388, a deletion of the conjugative coupling protein, a component believed to play a key role in the recruitment of the conjugative substrate, caused DNA transfer rates to drop 10-fold (Fernandez-Gonzalez et al., 2011). The relaxase MobA belongs to a mobilizable plasmid which hijacks the T4SS of co-residing conjugative plasmids, so it can be translocated through various T4SS; thus, it is plausible that the requirements for MobA recruitment are less stringent. In fact, the C-terminal 48 residues of MobA were shown to direct translocation of a Cre fusion through the VirB T4SS of *A. tumefaciens* into plant cells (Vergunst et al., 2005). Now, we show that MobA can also be translocated through a T4ASS into mammalian cells.

The ability to enhance integration of the transferred DNA into the recipient cell genome must reside in an intrinsic property of the relaxase, which is the only protein entering the recipient cell covalently attached to the transferred DNA strand. We report here that the promiscuous relaxase MobA also promotes DNA integration, resulting in resistant colonies with about 10-fold higher frequency than Mob-BID, which does not promote integration above background levels obtained by DNA transfection (Gonzalez-Prieto et al., 2017), but roughly five-fold lower frequency than TrwC. These differences observed among relaxases could be due to differential nuclear targeting, catalytic activity, or binding affinity to its target, which could protect the DNA ends, thus favoring integration by host-mediated mechanisms, as previously suggested (Gonzalez-Prieto et al., 2017). Subcellular localization of TrwC and MobA in human cells showed no preferential nuclear localization for either relaxase (Silby et al., 2007; Agundez et al., 2011). It is noteworthy that TrwC catalyzes site-specific recombination on supercoiled DNA substrates (Cesar et al., 2006), while MobA was shown to catalyze site-specific recombination between two *oriT* copies when the substrate was single-stranded (Meyer, 1989), and other relaxases do not catalyze this reaction at all. Although the integration pattern in the human genome is random (Gonzalez-Prieto et al., 2017), site-specific recombination ability could play a role in strand-transfer reactions when the nucleoprotein complex is directed to a nicked DNA strand by the host repair machinery.

MobA can be translocated by the T4BSS of *L. pneumophila*, alone or bound to DNA, into recipient bacteria (Vogel et al., 1998; Luo and Isberg, 2004). These results prompted us to test its translocation by T4BSS into mammalian cells. Our results (**Figure 3**) show for the first time that DNA transfer to human cells can also be accomplished through the Dot/Icm T4BSS of *L. pneumophila* and *C. burnetii*, only remotely related to T4ASS. Thus, it is reasonable to assume that DNA translocation may be an intrinsic ability of T4SS. An important difference between both Dot/Icm systems is the temporal pattern of secretion: while *L. pneumophila* has been shown to secrete effectors as internalization into host cells is initiated (Nagai et al., 2005) in case of *C. burnetii* effector translocation is initiated when the pathogen has reached an acidified lysosomal compartment (Newton et al., 2013); thus, DNA transfer in *C. burnetii* must occur from within the *Coxiella*-containing vacuole.

DNA transfer was dependent on the presence of the Dot/Icm T4SS and a functional relaxase, as expected for a *bona fide* conjugation-like DNA transfer process. The wide differences in DNA transfer rates depending on the relaxase (MobA, TrwC, or TrwC-RalF, including the translocation signal of the natural T4SS substrate RalF) and on the presence/absence of the chaperones IcmSW (see **Table 5**) support the concept that relaxase recruitment is the main driver of DNA transfer.

During the course of performing *C. burnetii* infection experiments, we noticed an inhibition of *C. burnetii* intracellular replication caused by the presence of RSF1010 derivatives carrying a functional MobA relaxase while isogenic strains with a *mobA* mutation did not affect growth (**Figure 4**). Similarly, RSF1010 conjugation was shown to inhibit intracellular replication and virulence of *L. pneumophila* (Segal and Shuman, 1998), probably by MobA interference with effector secretion by Dot/Icm. This result should be taken into account when using vectors based on RSF1010, which are the more commonly used by both *L. pneumophila* and *C. burnetii*.

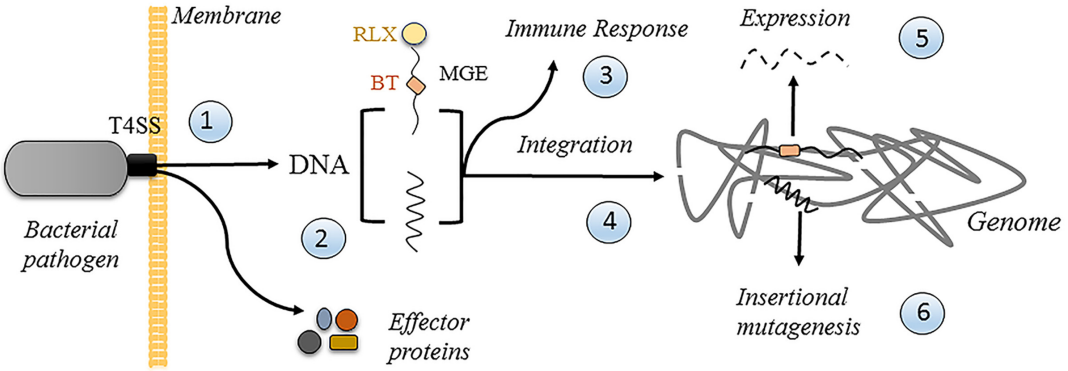


FIGURE 5 | Possible fates of translocated DNA in a human cell. See text for details. Random DNA is represented by the jagged line. Mobile genetic element (MGE) is represented by the wavy line. RLX, relaxase (small yellow sphere). BT, beneficial trait (small orange box).

Finally, an attractive question that remains open is the possible biological role, if any, of DNA transfer to mammalian cells by bacterial pathogens harboring a T4SS. Is the DNA transfer ability an evolutionary remnant of the conjugative T4SS from which the T4SS involved in virulence probably have evolved? Or is it an ability which the pathogens have evolved to use to their own benefit, in the same way as *A. tumefaciens* uses it to subvert its eukaryotic host cell?

In support of the first possibility, it is relevant to point out that in spite of many attempts, no T4 protein, protein domain or amino acid residue has been identified to date, which is specifically involved in DNA transfer. All analyzed mutants in T4 components, even in the conjugative coupling protein ATPase, affected DNA and protein translocation to the same extent, leading to the suggestion that relaxase and DNA translocation may have the same molecular requirements (de Paz et al., 2010; Larrea et al., 2017). Thus, the ability to transfer DNA could not be lost in a T4SS even if it evolved to only secrete proteins. However, the potential of DNA transfer for long-term subversion of the host cells makes it attractive to think that pathogens may utilize such a process for their own profit. **Figure 5** illustrates the possible fates of secreted DNA in a human cell. A pathogen translocates effector proteins and DNA through its T4SS once in contact with the membrane (1 in **Figure 5**), whether it is from within a vacuolar compartment, as in case of *C. burnetii*, or from the outside. The secreted DNA could either be random DNA, as proposed for *H. pylori* (Varga et al., 2016), or a specifically recruited mobile genetic element (MGE), in which case a dedicated transfer system would attach a relaxase to its end (2). The cytoplasmic DNA could elicit an immune response (3), as proposed for *H. pylori* (Varga et al., 2016), which could be used by the pathogen for its own benefit. DNA could also get integrated into the host cell genome (4) by the host repair/recombination systems, and/or by the covalently attached conjugative relaxase. Integration will lead to the stable expression of the encoded information (5), including any beneficial traits that the pathogen may have evolved to encode in MGE for that purpose. Finally, random integration has an inherent risk of insertional mutagenesis (6), which could lead to increased growth of the host cell, thereby promoting the extension of the niche of the pathogen.

In this context, it has to be stressed that human pathogens contain many poorly characterized MGE, which could be substrates for DNA transfer (in addition to the possibility of sporadic transfer of visiting promiscuous plasmids, such as RSF1010). As examples from the pathogens used in this study, the pBRG1 cryptic plasmid of *B. henselae* can be recruited by VirB/D4 and translocated to human cells (Schroder et al., 2011); conjugative transfer of chromosomal DNA has been reported for *L. pneumophila* (Miyamoto et al., 2003), and its genome includes several genomic islands; and notably, a cryptic plasmid in *C. burnetii* is enriched in important effector genes (Voth et al., 2011); it is tempting to speculate that this plasmid may be transferred to the host cell.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception and design of the work, data acquisition and/or analysis. All authors contributed to drafting, revising, and final approval of the work. All authors agree to be accountable for all aspects of the work.

FUNDING

This work was supported by grant BIO2013-46414-P from the Spanish Ministry of Economy and Competitiveness to ML, and NIH grants AI041699 and AI114760 to CRR. DLG was supported by a predoctoral fellowship from the University of Cantabria (Spain). SS was supported by an Advanced Postdoc Mobility fellowship from the Swiss National Science Foundation (SNSF).

ACKNOWLEDGMENT

ML wishes to thank the Roy lab, and especially David Chetrit and Stephanie Shames, for their support with *Legionella* lab protocols.

REFERENCES

Agundez, L., Gonzalez-Prieto, C., Machon, C., and 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. doi: 10.1371/journal.pone.0031047

Agundez, L., Machon, C., Cesar, C. E., Rosa-Garrido, M., Delgado, M. D., and Llosa, M. (2011). Nuclear targeting of a bacterial integrase that mediates site-specific recombination between bacterial and human target sequences. *Appl. Environ. Microbiol.* 77, 201–210. doi: 10.1128/AEM.01371-10

Alperi, A., Larrea, D., Fernandez-Gonzalez, E., Dehio, C., Zechner, E. L., and Llosa, M. (2013). A translocation motif in relaxase TrwC specifically affects recruitment by its conjugative type IV secretion system. *J. Bacteriol.* 195, 4999–5006. doi: 10.1128/JB.00367-13

Arasaki, K., and Roy, C. R. (2010). *Legionella pneumophila* promotes functional interactions between plasma membrane syntaxins and Sec22b. *Traffic* 11, 587–600. doi: 10.1111/j.1600-0854.2010.01050.x

Berger, K. H., and Isberg, R. R. (1993). Two distinct defects in intracellular growth complemented by a single genetic locus in *Legionella pneumophila*. *Mol. Microbiol.* 7, 7–19. doi: 10.1111/j.1365-2958.1993.tb01092.x

Cabezon, E., Ripoll-Rozada, J., Pena, A., de la Cruz, F., and Arechaga, I. (2015). Towards an integrated model of bacterial conjugation. *FEMS Microbiol. Rev.* 39, 81–95. doi: 10.1111/1574-6976.12085

Cambronne, E. D., and Roy, C. R. (2007). The *Legionella pneumophila* IcmSW complex interacts with multiple Dot/Icm effectors to facilitate type IV translocation. *PLoS Pathog.* 3:e188. doi: 10.1371/journal.ppat.0030188

Carey, K. L., Newton, H. J., Luhrmann, A., and Roy, C. R. (2011). The *Coxiella burnetii* Dot/Icm system delivers a unique repertoire of type IV effectors into host cells and is required for intracellular replication. *PLoS Pathog.* 7:e1002056. doi: 10.1371/journal.ppat.1002056

Celec, P., and Gardlik, R. (2017). Gene therapy using bacterial vectors. *Front. Biosci.* 22:81–95. doi: 10.2741/4473

Cesar, C. E., Machon, C., de la Cruz, F., and 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. doi: 10.1111/j.1365-2958.2006.05437.x

Chandran Darbari, V., and Waksman, G. (2015). Structural biology of bacterial type IV secretion systems. *Annu. Rev. Biochem.* 84, 603–629. doi: 10.1146/annurev-biochem-062911-102821

Christie, P. J. (2016). The mosaic type IV secretion systems. *EcoSal Plus* 7, doi: 10.1128/ecosalplus.ESP-0020-2015

Coers, J., Kagan, J. C., Matthews, M., Nagai, H., Zuckman, D. M., and Roy, C. R. (2000). Identification of Icm protein complexes that play distinct roles in the biogenesis of an organelle permissive for *Legionella pneumophila* intracellular growth. *Mol. Microbiol.* 38, 719–736. doi: 10.1046/j.1365-2958.2000.02176.x

de Paz, H. D., Larrea, D., Zunzunegui, S., Dehio, C., de la Cruz, F., and Llosa, M. (2010). Functional dissection of the conjugative coupling protein TrwB. *J. Bacteriol.* 192, 2655–2669. doi: 10.1128/JB.01692-09

Draper, O., Cesar, C. E., Machon, C., de la Cruz, F., and 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. doi: 10.1073/pnas.0506081102

Fernandez-Gonzalez, E., de Paz, H. D., Alperi, A., Agundez, L., Faustmann, M., Sangari, F. J., 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. doi: 10.1128/JB.05905-11

Frey, J., Bagdasarian, M. M., and Bagdasarian, M. (1992). Replication and copy number control of the broad-host-range plasmid RSF1010. *Gene* 113, 101–106. doi: 10.1016/0378-1119(92)90675-F

Garcillan-Barcia, M. P., Jurado, P., Gonzalez-Perez, B., Moncalian, G., Fernandez, L. A., and de la Cruz, F. (2007). Conjugative transfer can be inhibited by blocking relaxase activity within recipient cells with intrabodies. *Mol. Microbiol.* 63, 404–416. doi: 10.1111/j.1365-2958.2006.05523.x

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. III, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318

Gonzalez-Perez, B., Lucas, M., Cooke, L. A., Vyle, J. S., de la Cruz, F., and Moncalian, G. (2007). Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. *EMBO J.* 26, 3847–3857. doi: 10.1038/sj.emboj.7601806

Gonzalez-Prieto, C., Agundez, L., Linden, R. M., and Llosa, M. (2013). HUH site-specific recombinases for targeted modification of the human genome. *Trends Biotechnol.* 31, 305–312. doi: 10.1016/j.tibtech.2013.02.002

Gonzalez-Prieto, C., Gabriel, R., Dehio, C., Schmidt, M., and Llosa, M. (2017). The conjugative relaxase TrwC promotes integration of foreign DNA in the human genome. *Appl. Environ. Microbiol.* 83:e00207-17. doi: 10.1128/AEM.00207-17

Hubber, A., and Roy, C. R. (2010). Modulation of host cell function by *Legionella pneumophila* type IV effectors. *Annu. Rev. Cell Dev. Biol.* 26, 261–283. doi: 10.1146/annurev-cellbio-100109-104034

Joiner, K. A., Fuhrman, S. A., Miettinen, H. M., Kasper, L. H., and Mellman, I. (1990). *Toxoplasma gondii*: fusion competence of parasitophorous vacuoles in Fc receptor-transfected fibroblasts. *Science* 249, 641–646. doi: 10.1126/science.2200126

Killmann, H., Benz, R., and Braun, V. (1996). Properties of the FhuA channel in the *Escherichia coli* outer membrane after deletion of FhuA portions within and outside the predicted gating loop. *J. Bacteriol.* 178, 6913–6920. doi: 10.1128/jb.178.23.6913-6920.1996

Kirby, J. E., and Nekorchuk, D. M. (2002). *Bartonella*-associated endothelial proliferation depends on inhibition of apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4656–4661. doi: 10.1073/pnas.072292699

Larrea, D., de Paz, H. D., Matilla, I., Guzman-Herrador, D. L., Lasso, G., de la Cruz, F., et al. (2017). Substrate translocation involves specific lysine residues of the central channel of the conjugative coupling protein TrwB. *Mol. Genet. Genomics* doi: 10.1007/s00438-017-1331-3 [Epub ahead of print].

Lessl, M., Balzer, D., Weyrauch, K., and Lanka, E. (1993). The mating pair formation system of plasmid RP4 defined by RSF1010 mobilization and donor-specific phage propagation. *J. Bacteriol.* 175, 6415–6425. doi: 10.1128/jb.175.20.6415-6425.1993

Llosa, M., Bolland, S., Grandoso, G., and de la Cruz, F. (1994). Conjugation-independent, site-specific recombination at the oriT of the IncW plasmid R388 mediated by TrwC. *J. Bacteriol.* 176, 3210–3217. doi: 10.1128/jb.176.11.3210-3217.1994

Llosa, M., Schroder, G., and Dehio, C. (2012). New perspectives into bacterial DNA transfer to human cells. *Trends Microbiol.* 20, 355–359. doi: 10.1016/j.tim.2012.05.008

Luo, Z. Q., and Isberg, R. R. (2004). Multiple substrates of the *Legionella pneumophila* Dot/Icm system identified by interbacterial protein transfer. *Proc. Natl. Acad. Sci. U.S.A.* 101, 841–846. doi: 10.1073/pnas.0304916101

Meyer, R. (1989). Site-specific recombination at oriT of plasmid R1162 in the absence of conjugative transfer. *J. Bacteriol.* 171, 799–806. doi: 10.1128/jb.171.2.799-806.1989

Miyamoto, H., Yoshida, S., Taniguchi, H., and Shuman, H. A. (2003). Virulence conversion of *Legionella pneumophila* by conjugal transfer of chromosomal DNA. *J. Bacteriol.* 185, 6712–6718. doi: 10.1128/JB.185.22.6712-6718.2003

Nagai, H., Cambronne, E. D., Kagan, J. C., Amor, J. C., Kahn, R. A., and Roy, C. R. (2005). A C-terminal translocation signal required for Dot/Icm-dependent delivery of the *Legionella* RalF protein to host cells. *Proc. Natl. Acad. Sci. U.S.A.* 102, 826–831. doi: 10.1073/pnas.0406239101

Newton, H. J., Kohler, L. J., McDonough, J. A., Temoche-Diaz, M., Crabill, E., Hartland, E. L., et al. (2014). A screen of *Coxiella burnetii* mutants reveals important roles for Dot/Icm effectors and host autophagy in vacuole biogenesis. *PLoS Pathog.* 10:e1004286. doi: 10.1371/journal.ppat.1004286

Newton, H. J., McDonough, J. A., and Roy, C. R. (2013). Effector protein translocation by the *Coxiella burnetii* Dot/Icm type IV secretion system requires endocytic maturation of the pathogen-occupied vacuole. *PLoS ONE* 8:e54566. doi: 10.1371/journal.pone.0054566

Omsland, A., Beare, P. A., Hill, J., Cockrell, D. C., Howe, D., Hansen, B., et al. (2011). Isolation from animal tissue and genetic transformation of *Coxiella*

burnetii are facilitated by an improved axenic growth medium. *Appl. Environ. Microbiol.* 77, 3720–3725. doi: 10.1128/AEM.02826-10

Personnic, N., Barlocher, K., Finsel, I., and Hilbi, H. (2016). Subversion of retrograde trafficking by translocated pathogen effectors. *Trends Microbiol.* 24, 450–462. doi: 10.1016/j.tim.2016.02.003

Rolando, M., and Buchrieser, C. (2014). *Legionella pneumophila* type IV effectors hijack the transcription and translation machinery of the host cell. *Trends Cell Biol.* 24, 771–778. doi: 10.1016/j.tcb.2014.06.002

Sadler, J. R., Tecklenburg, M., and Betz, J. L. (1980). Plasmids containing many tandem copies of a synthetic lactose operator. *Gene* 8, 279–300. doi: 10.1016/0378-1119(80)90005-0

Saenz, H. L., Engel, P., Stoeckli, M. C., Lanz, C., Raddatz, G., Vayssier-Taussat, M., et al. (2007). Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors. *Nat. Genet.* 39, 1469–1476. doi: 10.1038/ng.2007.38

Sambrook, J., and Russell, D. W. (2001). *Molecular Cloning : A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schmid, M. C., Schulein, R., Dehio, M., Denecker, G., Carena, I., and Dehio, C. (2004). The VirB type IV secretion system of *Bartonella henselae* mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells. *Mol. Microbiol.* 52, 81–92. doi: 10.1111/j.1365-2958.2003.03964.x

Schroder, G., Schuelein, R., Quebatte, M., and Dehio, C. (2011). Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen *Bartonella henselae*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14643–14648. doi: 10.1073/pnas.1019074108

Segal, G., and Shuman, H. A. (1998). Intracellular multiplication and human macrophage killing by *Legionella pneumophila* are inhibited by conjugal components of IncQ plasmid RSF1010. *Mol. Microbiol.* 30, 197–208. doi: 10.1046/j.1365-2958.1998.01054.x

Silby, M. W., Ferguson, G. C., Billington, C., and Heinemann, J. A. (2007). Localization of the plasmid-encoded proteins TraI and MobA in eukaryotic cells. *Plasmid* 57, 118–130. doi: 10.1016/j.plasmid.2006.08.006

Varga, M. G., Shaffer, C. L., Sierra, J. C., Suarez, G., Piazuelo, M. B., Whitaker, M. E., et al. (2016). Pathogenic *Helicobacter pylori* strains translocate DNA and activate TLR9 via the cancer-associated cag type IV secretion system. *Oncogene* 35, 6262–6269. doi: 10.1038/onc.2016.158

Vergunst, A. C., van Lier, M. C., den Dulk-Ras, A., Stuve, T. A., Ouwehand, A., and Hooykaas, P. J. (2005). Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 832–837. doi: 10.1073/pnas.0406241102

Vogel, J. P., Andrews, H. L., Wong, S. K., and Isberg, R. R. (1998). Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science* 279, 873–876. doi: 10.1126/science.279.5352.873

Voth, D. E., Beare, P. A., Howe, D., Sharma, U. M., Samoilis, G., Cockrell, D. C., et al. (2011). The *Coxiella burnetii* cryptic plasmid is enriched in genes encoding type IV secretion system substrates. *J. Bacteriol.* 193, 1493–1503. doi: 10.1128/JB.01359-10

Williams, J. C., Peacock, M. G., and McCaul, T. F. (1981). Immunological and biological characterization of *Coxiella burnetii*, phases I and II, separated from host components. *Infect. Immun.* 32, 840–851.

Zamboni, D. S., McGrath, S., Rabinovitch, M., and Roy, C. R. (2003). *Coxiella burnetii* express type IV secretion system proteins that function similarly to components of the *Legionella pneumophila* Dot/Icm system. *Mol. Microbiol.* 49, 965–976. doi: 10.1046/j.1365-2958.2003.03626.x

Zechner, E. L., Lang, S., and Schildbach, J. F. (2012). Assembly and mechanisms of bacterial type IV secretion machines. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1073–1087. doi: 10.1098/rstb.2011.0207

Zusman, T., Yerushalmi, G., and Segal, G. (2003). Functional similarities between the icm/dot pathogenesis systems of *Coxiella burnetii* and *Legionella pneumophila*. *Infect. Immun.* 71, 3714–3723. doi: 10.1128/IAI.71.7.3714-3723.2003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer JRM and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Guzmán-Herrador, Steiner, Alperi, González-Prieto, Roy and Llosa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Plasmid 104 (2019) 102415

Contents lists available at ScienceDirect

Plasmid

journal homepage: www.elsevier.com/locate/yplas

Review

The secret life of conjugative relaxases

Dolores Lucía Guzmán-Herrador, Matxalen Llosa*

Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Universidad de Cantabria-CSIC-SODERCAN, Santander, Spain

ARTICLE INFO

Keywords:
Bacterial conjugation
Conjugative relaxase
Site-specific endonuclease
Genetic plasticity
Site-specific recombination
Rolling circle replication

ABSTRACT

Conjugative relaxases are well-characterized proteins responsible for the site- and strand-specific endonucleolytic cleavage and strand transfer reactions taking place at the start and end of the conjugative DNA transfer process. Most of the relaxases characterized biochemically and structurally belong to the HUH family of endonucleases. However, an increasing number of new families of relaxases are revealing a variety of protein folds and catalytic alternatives to accomplish conjugative DNA processing. Relaxases show high specificity for their cognate target DNA sequences, but several recent reports underscore the importance of their activity on secondary targets, leading to widespread mobilization of plasmids containing an *oriT*-like sequence. Some relaxases perform other functions associated with their nicking and strand transfer ability, such as catalyzing site-specific recombination or initiation of plasmid replication. They perform these roles in the absence of conjugation, and the validation of these functions in several systems strongly suggest that they are not mere artifactual laboratory observations. Other unexpected roles recently assigned to relaxases include controlling plasmid copy number and promoting retrotransposition. Their capacity to mediate promiscuous mobilization and genetic reorganizations can be exploited for a number of imaginative biotechnological applications. Overall, there is increasing evidence that conjugative relaxases are not only key enzymes for horizontal gene transfer, but may have been adapted to perform other roles which contribute to prokaryotic genetic plasticity. Relaxed target specificity may be key to this versatility.

1. Introduction

Prokaryotes have successfully colonized the world thanks to their genetic plasticity. Horizontal gene transfer (HGT) is the main driver of this plasticity, and bacterial conjugation is one of the major HGT mechanisms, being responsible for the transfer of mobile genetic elements (MGE) and chromosomal DNA in both Gram-negative and positive bacteria. Evidences both from natural sources and experimental settings prove that conjugation can be a very promiscuous process, capable of mediating HGT between Gram-negative and positive bacteria, and even between prokaryotic and eukaryotic cells (Lacroix and Citovsky, 2018).

Bacterial conjugation is broadly defined as the transfer of DNA from one donor bacterium to one recipient bacteria which need to be in physical contact. This definition includes a set of processes with little in common, such as the Type VII-dependent transfer of chromosomal segments in mycobacteria (Gray and Derbyshire, 2018), or the transfer of double-stranded DNA in a Type IV-independent manner in *Streptomyces* and other actinobacteria (Thoma and Muth, 2016). In this review, we will refer only to conjugative transfer of single stranded DNA (ssDNA) through a Type IV secretion system (T4SS) in Gram-positive and –negative bacteria, which requires the action of a conjugative relaxase. Most of our knowledge has come from the study of conjugative and mobilizable plasmids, although in recent years it has become apparent that this mechanism is as frequent in plasmids as in Integrative and Conjugative Elements (ICEs), and both kind of elements share similar conjugative systems (Guglielmini et al., 2011; Carraro and Burrus, 2014). The conjugative DNA transfer process can be outlined as follows: in the donor cell, the DNA strand to be transferred is cleaved at the origin of transfer (*oriT*) by a site-specific endonuclease known as the relaxase, which makes a covalent bond with the nicked strand; this nucleoprotein complex is transferred through a T4SS into the recipient cell, where the relaxase actively catalyzes the strand transfer reaction, leading to the end of the transfer process. This mechanism has been validated in different conjugative systems (Cabezon et al., 2015).

Conjugative relaxases are key enzymes in conjugative ssDNA transfer processes. They are characterized by their site- and strand-specific endonuclease activity. Initial characterization of relaxases from several different conjugative systems described them as proteins highly selective for their target DNA and which catalyzed transesterification reactions through a covalent adduct between the cut DNA and a catalytic Tyr residue. In support for this uniformity, the first solved crystal structures of several relaxases indicated that they all belonged to the

* Corresponding author.
E-mail address: llosam@unican.es (M. Llosa).

https://doi.org/10.1016/j.plasmid.2019.102415
Received 28 February 2019; Received in revised form 17 April 2019; Accepted 2 May 2019
Available online 17 May 2019
0147-619X/ © 2019 Elsevier Inc. All rights reserved.

HUH superfamily of site-specific single-stranded endonucleases. However, exceptions have become so numerous that the paradigm needs to be revisited. There are relaxases lying outside of the HUH superfamily; relaxases that do not use a catalytic Tyr; and relaxases which might not even make a covalent complex with the DNA. In particular, a growing number of recent reports show the ability of relaxases to act, with lower efficiency, on sequences other than their cognate targets, with intriguing biological consequences. The purpose of this review is to revisit the concept of conjugative relaxases, emphasizing the diversity rather than the unity, and questioning their target specificity to accomplish conjugative ssDNA as their only biological role.

2. The growing family of conjugative relaxases

The name "relaxase" honors the pioneering work by Clewell and Helinski, who discovered the "relaxation complexes" formed by mobilizable plasmid ColE1, which, when isolated as a protein-DNA complex, underwent conversion from supercoiled to open circular form in the presence of denaturing agents (Clewell and Helinski, 1969). The authors soon discovered the strand specificity of the relaxation event (Clewell and Helinski, 1970). Discovery of the proteins responsible for this relaxation had to wait for almost two decades (Traxler and Minkley Jr., 1988). Biochemical characterization of the covalent interaction between the relaxase and its cognate *nic* site was first reported for the TraI relaxase of IncP plasmid RP4 (Pansegrau et al., 1990), and similar features were soon found for the relaxases of other conjugative and mobilizable plasmids (Bhattacharjee and Meyer, 1991; Reygers et al., 1991; Matson and Morton, 1991; Scherzinger et al., 1993). Relaxases were then related through a set of three conserved motifs to other ssDNA endonucleases involved in DNA replication and transposition (Ilyina and Koonin, 1992; Mendiola and de la Cruz, 1992), which defined the HUH superfamily of site-specific ssDNA endonucleases. The HUH signature motifs were also found in relaxases from Gram-positive bacteria (Guzman and Espinosa, 1997), leading to a proposal for a universal relaxase mode of action (Byrd and Matson, 1997). Motif I contains the catalytic Tyr residue, which forms the covalent complex with the nicked DNA, while the HUH motif III, characterized by a set of three His residues, is important for coordination of the metal cation required for endonuclease activity.

There was an increasing need for relaxase classification, which led to several studies analyzing their taxonomy. Table 1 summarizes current relaxase classification and their main biochemical and biological features. It is important to note that relaxases were phylogenetically analyzed according to their N-terminal 300 residues, which contain the catalytic domain; many relaxases harbor different C-terminal domains, which often play additional roles in the DNA transfer process. Known relaxases were grouped in six families by Garcillan-Barcia et al. (2009), although the authors already proposed the existence of new families coming from uncharacterized transfer systems, where no relaxase homologue was apparent. The vast majority of relaxases possess conserved HUH motifs. This relationship among HUH relaxases would be confirmed by the resolution of the 3D structure of different members of the superfamily, which showed the conservation of the HUH catalytic fold (reviewed by Chandler et al. (2013)). Despite this conservation, some variants were reported: the characteristic 3-His motif III was replaced by a HEN motif in relaxase MbeA of mobilizable plasmid ColE1 (Varsaki et al., 2003), and the third His is not conserved in a subset of MOB_V relaxases (Garcillan-Barcia et al., 2009). With respect to motif I, the MOB_F family harbors several conserved Tyr residues, although the number and function of catalytic Tyr varies in each relaxase (Grandoso et al., 2000; Street et al., 2003; Nash et al., 2011). A recent review summarizes the detailed knowledge that we have acquired on these canonical relaxases (Zechner et al., 2017).

However, increasing knowledge of relaxases belonging to different families challenged this paradigm. Early works on relaxases of the MOB_V family were unable to assign a catalyitic Tyr residue, in spite of their conservation of the HUH motifs (Guzman and Espinosa, 1997; Antoine and Locht, 1992), and elucidation of the 3D structure revealed that these relaxases use a His residue instead of Tyr to make the nucleophilic attack and covalent complex (Pluta et al., 2017). Another significant divergence was reported for the relaxase MobC of mobilizable plasmid CloDF13, the prototype of the MOB_C family, which showed no homology to HUH relaxases; interestingly, the nicked *oriT* DNA did not have any blocked end, suggesting that covalent complexes were not formed (Núñez and de la Cruz, 2001). Modelling of the 3D structure of another relaxase of the MOB_C family, TraX of plasmid pAD1 from *Enterococcus faecalis*, suggested a structure unrelated to the HUH fold, instead resembling restriction endonucleases. In spite of these structural differences, a Tyr residue was essential for the cleavage reaction, and a Tyr-mediated covalent adduct was proposed, although never detected (Francia et al., 2013). There are other relaxase families, less characterized, which do not include the HUH motifs. The best characterized examples are relaxase TraI of *Neisseria gonorrhoeae* GGI (Salgado-Pabon et al., 2007), representative of the MOB_H family (Garcillan-Barcia et al., 2009); Orf20 of conjugative transposon Tn*916* (Rocco and Churchward, 2006), representing family MOB_T (Guglielmini et al., 2011); and relaxase TcpM of the *Clostridium perfringens* conjugative plasmid pCW3 (Wisniewski et al., 2016), which has not been assigned to any MOB family. Although structural information is still lacking, these proteins do not resemble the previously characterized relaxases, and rather show similarity, or conservation of

Table 1
Current classification and main features of conjugative relaxases (see text for details).

MOB Family[a]	F	P	Q	V	C	H	T	TcpM[b]
Prototype relaxase	R388-TrwC	RP4-TraI	RSF1010-MobA	pMV158-MobM	pAD1-TraX	GGI-TraI	Tn916-Orf20	pCW3-TcpM
3D Fold[c]	HUH	HUH,HEN	HUH	HUH	RE	HD	Rep-trans	Y-rec
Catalytic residue	Tyr x2	Tyr	Tyr	His	Tyr	Tyr		Tyr
Covalent complex[d]	Yes	Yes	Yes	Yes	No	Yes?		
2nd Function[e]	Pre	Pre	Pre*	Pre	Pre	Pre		
	Mob	Mob	Mob	Mob		Rep	Rep	
		rTn		Cop				

[a] As defined by Garcillan-Barcia et al. (2009) and Guglielmini et al. (2011).
[b] This relaxase was described after the MOB classification was reported, and does not fit into any of the defined families.
[c] Structural family based on the presence of signature motifs or 3D structure (in bold): HUH, HUH superfamily; HEN, HUH superfamily with variant HEN motifs; RE, restriction endonuclease; HD, HD hydrolase; Rep-trans, RCR initiation proteins; Y-rec, Tyrosine recombinase.
[d] Yes, experimentally detected relaxase-DNA covalent complex. Yes?, indirect evidence suggesting protection of the 5`end of the T-DNA. No, searched but not detected. Blank, no information.
[e] Reported biological function other than conjugative self-transfer: Mob, in *trans* activity on heterologous oriT sequences; Pre, Plasmid Recombination Enzyme (Pre*, only on single-stranded substrates); Rep, initiator of plasmid replication; Cop, regulation of plasmid copy number; rTn, enhancer of retrotransposition.

motifs, which relate them to HD hydrolases, Rep-trans proteins involved in RCR, and Tyr-Recombinases, respectively, highlighting the still underexplored diversity among conjugative relaxases. No covalent complexes have been reported for these divergent protein families, but it is not clear if this issue has been experimentally addressed. It must be taken into account that the covalent complex can be difficult to detect, as happened in the case of the filamentous phage fd, or the RepB replicase in plasmid pMV158, which required elaborated approaches to determine the existence of the covalent adduct (Moscoso et al., 1997; Asano et al., 1999). The absence of a covalent complex with the relaxase would imply a substantial change in the current model for conjugative ssDNA transfer, which is based on the transfer of the nucleoprotein complex into the recipient cell, where the relaxase is required to terminate the transfer reaction. Surely, a deeper characterization of these novel families will determine if there is a covalent adduct, which requires a different methodology to be detected, or if ssDNA transfer by conjugation can be radically different in systems involving non-HUH relaxases.

Exploration of bacterial clades traditionally underrepresented has revealed new relaxase families, which await further study. Initial characterization of the relaxase RelLS20 from the *Bacillus subtilis* plasmid pLS20 showed the presence of HUH motifs and a catalytic Tyr residue, but no homology to previously defined relaxases. Interestingly, the authors found more than 800 genes in Firmicutes showing homology to this protein, which suggests RelLS20 is the prototype of a new family of relaxases restricted to this family of Gram-positives (Ramachandran et al., 2017). Also, an extensive analysis of 124 genomes from 27 species of *Streptococcus* revealed 144 Integrative Mobilizable Elements, of which 118 harbored relaxases related to RCR Rep proteins, belonging to four totally new families, or to MOB_T (Coluzzi et al., 2017). In short, the diversity of relaxases has just begun to be revealed.

3. Target specificity

Conjugative relaxases specifically bind to a target sequence in the *oriT*, and introduce a site-specific nick in the DNA strand to be transferred (*nic* site). The specificity of a relaxase for its target sequences was biochemically characterized initially for the MOB_P relaxase TraI of the IncP plasmid RP4, using *in vitro* assays with labelled oligonucleotides (Pansegrau et al., 1993). It was also determined that tight substrate binding and catalytic activity were independent (Pansegrau and Lanka, 1996). Similar experiments rendered equivalent results in the paradigmatic MOB_F relaxases R388-TrwC and F-TraI (Zechner et al., 2017). The elucidation of their 3D structures allowed fine mapping of the interactions with the DNA, leading to a detailed knowledge of the relevant protein residues as well as the *oriT* nucleotides important for the interaction. The relaxases bind to an inverted repeat near the *nic* site. The DNA requirements for specificity lie both in the DNA binding domain and in the cleaved site (Zechner et al., 2017). The detailed structural and biochemical information showed that specificity relied on just a few protein-DNA interactions, thus suggesting that specificity might be altered by rational design. In fact, specificity swapping was obtained by changing only 4 bp of the *oriT*s of the staphylococcal mobilizable plasmids pC221 and pC223 (Caryl and Thomas, 2006), or two residues of the relaxases of plasmids F and R100 (Harley and Schildbach, 2003). Moreover, González-Pérez *et al* (Gonzalez-Perez et al., 2009) showed proof of principle that variant relaxases can be obtained that recognize the desired change in the target DNA.

Concerning the relaxases belonging to other families, the situation varies significantly. In the case of the MOB_C relaxases, binding occurs specifically at a set of direct repeats located more than 70 bp away from the *nic* site (Francia et al., 2013). Two types of relaxases seem to be unable to introduce the site-specific nick by themselves. The MOB_T relaxase Orf20 of Tn*916*, showed *in vitro* non-specific endonuclease activity, but sequence- and strand- specific cleavage was conferred by the Tyr recombinase responsible for integration/excision of the conjugative transposon (Rocco and Churchward, 2006). In the case of the TcpM relaxase of plasmid pCW3, which itself resembles Tyr recombinases, binding was specific for its *oriT* site, but DNA cleavage specificity could not be proven *in vitro,* suggesting other still unknown factors must confer specificity to this atypical relaxase (Wisniewski et al., 2016). It is interesting to note that a set of MOB_T relaxases recently described in streptococci have associated genes homologous to TcpA, the coupling protein associated with relaxase TcpM (Coluzzi et al., 2017),which suggests that these two types of relaxases sharing non-specific endonuclease activity may share other evolutionary relationships on their respective transfer systems.

With few exceptions (Perez-Mendoza et al., 2006; Fernandez-Gonzalez et al., 2016), relaxases are shown to work in *trans* as efficiently as in *cis*. Thus, specificity can easily be checked *in vivo* by testing conjugal mobilization of DNA molecules containing different *oriTs*. Many reports confirmed that relaxases could mobilize plasmids containing their *oriT* site but not others, even if highly homologous. This was the case, for instance, for the related IncF plasmids F and R100 (Harley and Schildbach, 2003), the enterococcal plasmids pAD1 and pAM373 (Francia and Clewell, 2002a; Francia and Clewell, 2002b), or mobilizable plasmids pC221 and pC223 (Caryl et al., 2004). It is important in this context to distinguish between binding/cleavage assays on oligonucleotides, and assays using supercoiled substrates with full *oriTs*. While the former address specifically the intrinsic binding/cleavage specificity of the relaxases, the latter mimic the *in vivo* process by including binding sites for accessory proteins, which are required to form the relaxosome, contributing to the extrusion of the binding site and exposure of the target as a single stranded region amenable to relaxase function (Cabezon et al., 2015). This role may also contribute in a decisive manner to plasmid specificity, such as in the case of the related IncP plasmids RP4 and R751, where the relaxases can be exchanged, but auxiliary factors could not, determining the *in vivo* specificity (Pansegrau et al., 1988). Another example is the staphylococcal pWBG749 family of conjugative plasmids, where the SmpO accessory protein determines *oriT* specificity (O'Brien et al., 2015). In summary, most relaxases bind *in vitro* with high specificity to their target sequences, which is a prerequisite for conjugal transmission. *In vivo*, specificity involves a set of protein-protein and protein-DNA interactions among the relaxase, accessory protein/s, and the *oriT* site.

In spite of the specificity for their cognate targets, lower efficiency recognition of heterologous sequences has been reported for members of all families of HUH relaxases. For instance, the MOB_F relaxases TraC of plasmids NAH7 and pWW0 could mobilize plasmids containing either *oriT*; in this case, the full *oriT* fragments shared only 63% identity, but the regions around the *nic* site were identical (Kishida et al., 2017). Relaxase MobM from plasmid pMV158 was shown to relax *in vitro* other mobilizable plasmids from Gram-positive organisms, whose *oriT*s shared 67–100% homology with the pMV158 minimal *oriT* (Fernandez-Lopez et al., 2013). Interestingly, not all relaxases are equally stringent on their DNA sequence requirements. The relaxases of the mobilizable plasmids pSC101 and R1162 (virtually identical to RSF1010), which recognize highly homologous *oriT* sequences, nonetheless had different stringencies: while the relaxase of pSC101 could not mobilize RSF1010, MobA of RSF1010 could also act on the pSC101 *oriT* (Jandle and Meyer, 2006). The authors found that MobA could even initiate transfer from chromosomal sites, and discussed the implication of this promiscuity for horizontal gene transfer by this broad host range plasmid. A similar situation was reported in two other plasmids, which are totally unrelated except in their transfer regions: the enterococcal plasmid pCF10 and plasmid pRS01 from *Lactococcus lactis*. PcfG, the relaxase of plasmid pCF10, could mobilize plasmids containing the heterologous *oriT*, while the relaxase LtrB of pRS01 was specific (both *in vitro* and *in vivo*) for its own *oriT* (Chen et al., 2007). More surprisingly, the relaxase TrwC of plasmid R388 was shown to mobilize plasmids containing the *oriT* region of the Ptw plasmid of *Burkholderia cenocepacia*; while the

relaxases of both plasmids are closely related, there is no significant homology among the *oriT* regions. The PtwC relaxase could not complement TrwC for mobilization of R388-*oriT* containing plasmids, although this could also be caused by a *cis*-acting preference (Fernandez-Gonzalez et al., 2016).

The ability of some relaxases to cross-react on the *oriT* sequences targeted by other relaxases illustrates the biological relevance that their relaxed specificity may have for promiscuous horizontal gene transfer. This *trans*-mobilization phenomenon is more frequent than previously thought. Different strategies exist for achieving horizontal transfer by hitchhiking on the transfer machinery of co-resident plasmids (recently reviewed by Ramsay and Firth (2017)). Mobilizable plasmids could be classified in the classical "ready-to-go" plasmids, which encode for their relaxase (and even for their own coupling protein, in the case of CloDF13 (Núñez and de la Cruz, 2001)), and "orphan" plasmids which rely solely on *oriT*-like sequences (sometimes encoding also for accessory proteins) to be mobilized by the relaxases present in a co-resident plasmid. The latter are the outmost expression of this plasmid piracy, and represent the natural manifestation of a well-known laboratory fact: the *oriT* site is the only element of the conjugative machinery required in *cis*, and thus, any DNA molecule containing *oriT* can be mobilized if the appropiate transfer machinery is provided in *trans*. In staphylococci, a diverse range of such *oriT*-containing plasmids lacking any transfer gene, which have been associated with the spread of antibiotic resistance determinants, have been shown to be mobilizable by co-resident conjugative plasmids (O'Brien et al., 2015; Pollet et al., 2016). Another illustrative example of the power of this kind of low-cost mobilization can be found in the *Escherichia coli* plasmid pBuzz, less than 2 kb in size, which relies on the conjugative machinery of a helper plasmid (Moran and Hall, 2019). These recent reports also searched for other potential *oriT*-containing plasmids and found many candidates, indicating that this is probably just the first glimpse of a widespread phenomenon.

In this new scenario, relaxases are not only responsible for the selfish transfer of the DNA molecule which encodes them, but also for *in trans* mobilization of opportunistic plasmids containing short sequences which resemble their targets. Harboring an *oriT*-like sequence could be a low-cost strategy for horizontal mobility, which relies on the presence of co-resident plasmids, but bypasses the added burden of maintaining dedicated transfer regions in their DNA. It is possible that many plasmids classified as non-mobile due to the absence of putative relaxases (Smillie et al., 2010), may in fact be orphan mobilizable plasmids (Ramsay and Firth, 2017). *oriTs* alone can be more difficult to spot than when accompanied by relaxases or other conjugative functions. However, now that some reports have elaborated bioinformatics methods of detecting *oriTs* based on sequence homologies and on structural features (Zrimec and Lapanje, 2018; Li et al., 2018), it can be anticipated that many more orphan mobilizable plasmids will be described.

4. Moonlighting relaxases

Conjugative relaxases are classified as such based on their role in conjugative DNA transfer. Often, these enzymes are multi-domain proteins harboring other functional domains involved in the DNA transfer process. This is a frequent situation in the HUH relaxases, probably reflecting the modular evolution of this protein superfamily (Chandler et al., 2013; Agundez et al., 2018). The covalently attached domains provide functions which either are essential or contribute to the efficiency of the conjugative transfer process, such as oligomerization, DNA binding, or the DNA helicase domain linked to the MOB_F family of relaxases (Zechner et al., 2017). Even the primase domain linked to the RSF1010 relaxase MobA, which is required for plasmid replication, was shown to increase the efficiency of conjugative DNA transfer, probably reflecting an adaptation of this broad host range plasmid to carry its own priming system to the recipient cell (Henderson and Meyer, 1999; Henderson and Meyer, 1996). In many other occasions, however, relaxases behave as moonlighting proteins, performing additional functions independently of conjugation.

The ability of some conjugative relaxases to promote RecA-independent, site-specific recombination between two *oriT* copies was reported even before the characterization of these proteins as relaxases (Gennaro et al., 1987). *oriT*-specific recombination is dependent on the relaxase and occurs in the absence of the rest of the transfer machinery (Llosa et al., 1994). Recombination can be intra- or inter-molecular, and relaxases can even catalyze the integration of the transferred DNA strand into a resident *oriT* copy in the recipient (Draper et al., 2005). This site-specific recombinase/integrase ability has been reported for many relaxases, both from Gram-positive and –negative systems, belonging to different MOB families (reviewed by Wawrzyniak et al. (2017)), but it is not an inherent characteristic of relaxases; at this point it is unknown which factor(s) allow a relaxase to act as a site-specific recombinase. Probably, relaxases act only on single-stranded *oriT* copies, which can be generated by the action of accessory factors (Furuya and Komano, 2003), or during the plasmid replication process, and completion of the reaction is mediated by the host-encoded replication/repair machinery (Cesar et al., 2006). The *oriT* sequence itself also plays an important role, since the MOB_H relaxase of ICE*clc* catalyzes recombination only on one of the two *oriTs* present in this ICE, while it can act on both *oriT1* and *oriT2* for conjugal DNA transfer (Miyazaki and van der Meer, 2011). DNA sequence requirements at the different *oriT* copies involved in the recombination reaction suggested that recombination events mimicked the initiation and termination steps of conjugative DNA transfer (Cesar et al., 2006; Barlett et al., 1990). In accordance with this idea, the target DNA requirements for integration of a relaxase-bound DNA strand are less stringent (Agundez et al., 2012). In both conjugal DNA transfer and site-specific integration, tight controls restrict the initiation of the reaction, but once the covalent nucleoprotein complex is formed, the process can be finished with lower efficiency on DNA sequences differing from that of the cognate *oriT*. In this way, the cell ensures that the energy consumed to start the process will not be wasted vainly.

The biological function most obviously related to conjugative DNA transfer would be plasmid replication. Replication and conjugation are two faces of the same phenomenon: plasmid dissemination, either vertical or horizontal, respectively. In fact, early reports suggested that plasmids coordinate the decision-making process to decide whether to promote horizontal or vertical replication, depending on environmental circumstances (Jagura-Burdzy and Thomas, 1994). The aforementioned primase domain linked to the conjugative relaxase MobA and involved in both plasmid replication and transfer would be another example of the close interrelationship between both processes. As already mentioned, most relaxases are evolutionarily related to RCR replicases: HUH Mob relaxases with HUH Rep proteins, and MOB_T relaxases with Rep-trans proteins. In the last decade, different reports have highlighted the fact that both kind of proteins are functionally exchangeable to a certain extent (reviewed by Wawrzyniak et al. (2017)). Several HUH Rep proteins have been reported to initiate conjugal DNA transfer of their own replicons by cleaving the DNA at the nick *dso*, which then serves as an *oriT*. Conversely, ICE relaxases belonging to the MOB_T and MOB_H families were shown to initiate both conjugal transfer and vegetative replication of the ICE, which were considered, until then, unable to replicate autonomously.

A recent report constitutes an interesting addition to the catalogue of functions that conjugative relaxases can play, independent of conjugal DNA transfer. The relaxase MobM of the RCR plasmid pMV158 was found to participate in regulation of plasmid copy number by transcriptional repression of the antisense RNA, thus increasing the number of plasmid molecules ready to be transmitted, whether it is horizontally or vertically (Lorenzo-Diaz et al., 2017). Probably, the most unexpected function reported for a conjugative relaxase is the ability of LtrB, the relaxase of plasmid pRS01, to stimulate both the frequency and diversity of retrotransposition of a mobile group II

intron, which resides precisely within the relaxase gene itself. LtrB was found to have weak off-target activity in addition to its *oriT*-specific cleavage activity; this introduction of spurious nicks would stimulate the frequency and density of intron mobility events (Novikova et al., 2014). In this way, intron mobility is promoted when the conjugative relaxase is active, *i.e.* during the conjugative process, thus stimulating the dissemination of the retrotransposon in donor and recipient cells.

5. Biotechnological applications

The specificity of conjugative relaxases for their target sequences can be exploited for biotechnological purposes. The biological autonomy of promiscuous transfer systems provides an excellent source of basic building blocks for synthetic biology (Martinez-Garcia et al., 2015), and the use of relaxases and their target sequences for plasmid mobilization would be the most obvious example. The increasing collection of characterized relaxase/target DNA pairs allows for the generation of different plasmid combinations, which have been proposed also as computing wires in synthetic biological circuits for digital cell-to-cell communication (Goni-Moreno et al., 2013). Relaxases can also be used for the sequence-specific modification of DNA-based nanostructures. Due to their covalent binding to specific single-stranded oligonucleotides, different target DNAs can serve as specific loading sites for their cognate relaxase. Proof of principle was obtained using the relaxases of plasmids R388, pKM101, RSF1010 and R100, and showing that each of them bound specifically to the oligonucleotide containing its target sequence, on two different types of DNA origami structures (Sagredo et al., 2016a). The specificity of relaxases can be changed by rational design, as previously mentioned (Gonzalez-Perez et al., 2009), and new substrates can be constructed by playing with the *oriT* elements which define binding specificity, rendering a wider catalogue of possible substrates to construct the nanostructures (Sagredo et al., 2016b). Thus, relaxases constitute a potential new class of sequence-selective protein linkers for DNA nanotechnology, which can be used for the modification of DNA nanostructures *in vivo* and for biological generation of DNA–protein hybrid nanostructures. In addition, relaxases are in general very permissive to fusions with other proteins of choice, maybe reflecting their own evolution (Agundez et al., 2018), so they could be used as anchors for other relevant functional proteins.

The ability of some relaxases to catalyze site-specific recombination fits into many biotechnological applications, and it could be of special interest in microorganisms where there is a lack of genetic tools. A relaxase-based recombination system has been used in *Streptomyces coelicolor* to amplify gene clusters for antibiotic production, improving the yield (Murakami et al., 2011). In another example, a site-specific recombination system was applied in *Bacillus* to obtain unmarked genetic manipulation by flanking the desired region with relaxase target sites (Wang et al., 2016). On the other hand, relaxed specificity could be useful in order to catalyze site-specific recombination or integration into a wide variety of DNA targets. As discussed above, the DNA specificity is very high at the start of the process, but less stringent on the second target to complete the reaction. This allows for strict choice of the DNA to be delivered, while having better options of finding the appropriate target in any given recipient genome (Agundez et al., 2012).

As biotechnological tools, relaxases have the added bonus of being part of a horizontal DNA transfer system, and so they can be delivered *in vivo*, covalently linked to any DNA molecule of choice, into any cell capable of acting as a recipient in conjugation. This includes virtually any prokaryotic cell, and even eukaryotic cells (Lacroix and Citovsky, 2018). The use of T4SS targeting eukaryotic cells to deliver relaxase-DNA complexes into human cells has proven as an efficient alternative to conjugation (Kunik et al., 2001; Llosa et al., 2012). Adding the appropriate secretion signal, different relaxases can be translocated through T4SS hosted by bacteria which target different human cell types (Guzmán-Herrador Dolores et al., 2017).

The possibility of sending site-specific recombinases covalently linked to a foreign DNA molecule into specific human cells is a promising genetic tool (Gonzalez-Prieto et al., 2013). Attempts have been made to use relaxases for genomic modification in eukaryotic cells. However, the site specificity of the integration event is challenged by the overwhelming efficiency of so-called illegitimate recombination processes in the eukaryotic cell. Integration of DNA into the genome of plant cells is routinely accomplished using the conjugation-like system of the Ti plasmid of *Agrobacterium tumefaciens*, which has been the major tool for plant genomic modification for decades (Guo et al., 2019). A T-DNA strand covalently linked to the relaxase-like protein VirD2 reaches the nucleus thanks to the nuclear localization signals present in VirD2, and DNA is integrated in a non-specific manner. This integration process is mediated by the DNA polymerase theta (van Kregten et al., 2016), which promotes microhomology-mediated end joining. The fusion of a site-specific nuclease to VirD2 increased the specificity of the integration events in yeast cells (Rolloos et al., 2015). Conjugative relaxase TrwC was used to deliver DNA into human cells through the T4SS of bacterial pathogens. Analysis of integration events indicated that the vast majority of integration events were not sequence-specific, but interestingly, the integration rate was up to 100-fold higher than when foreign DNA was introduced by transfection or by another relaxase with no reported recombinase activity (Gonzalez-Prieto et al., 2017). TrwC-DNA complexes may account for this improvement in integration efficiency due to a protecting role of the DNA ends in the human cell, and/or the lack of specificity for the final target sequence to complete the site-specific integration reaction. This ability to promote integration could be combined with a site-specific endonuclease, as shown for VirD2, in order to accomplish *in vivo* delivery and site-specific integration of foreign DNA in the human genome.

6. Biological implications

From a biological perspective, the high specificity of conjugative relaxases for their target sequences ensures that they transfer their own encoding DNA, as expected in a selfish DNA world. However, it becomes evident that relaxases are also involved in mobilization of other DNA molecules present in the same host, acting *in trans* on non-cognate targets. This phenomenon is probably much more widespread than currently thought, and it could happen that the contribution of relaxases to HGT is quantitatively higher by mobilizing orphan plasmids than its own replicon. Probably, these secondary targets have been evolutionary maintained as part of the many HGT strategies in prokaryotes.

The growing evidence of the ability of relaxases to perform functional roles independent of conjugative DNA transfer is also biologically significant. Their involvement in replication and recombination processes are not mere laboratory artifacts, since they have been validated in many instances, in unrelated systems, and with efficiencies well above biological noise. Relaxases acting as replication initiators highlight the common evolutionary origin and biological interplay between conjugation and replication (Lorenzo-Diaz et al., 2014; Waters and Guiney, 1993). The contribution of relaxases to the replication of an ICE is also a contribution to HGT, since this replication is essential to ensure that daughter cells inherit an excised form of the ICE. Site-specific recombination processes are important in plasmid evolution, creating replicons with mosaic structure and novel properties; the contribution of relaxase-mediated recombination events in plasmid evolution has been experimentally tested (Wang et al., 2013). A site-specific recombination event involving a relaxase was found to be responsible for the amplification of an antibiotic-resistance determinant in *Enterococcus faecalis* (Francia and Clewell, 2002b). Other possible biological advantages of *oriT*-specific recombination events may be envisioned, such as dimer resolution, or formation of cointegrates to favor conduction by a helper plasmid. The ability to catalyze site-specific integration into target sequences present in the recipient genome constitutes an

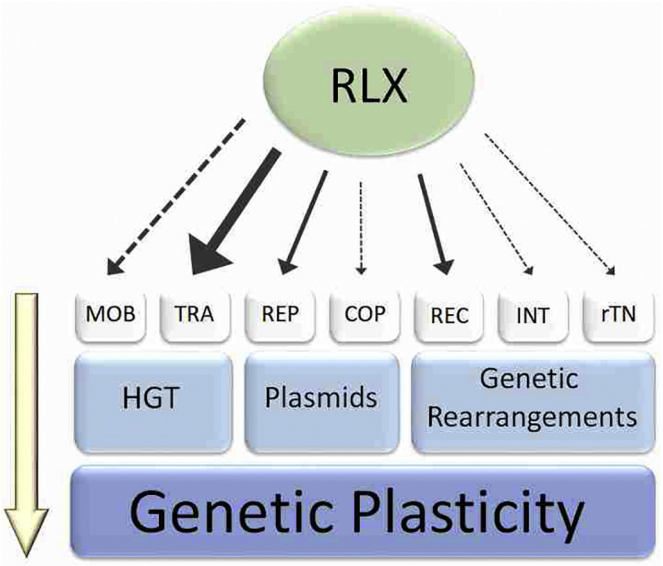


Fig. 1. Schematic of functional diversity and biological relevance of relaxases. The arrows point to the different biological functions reported for conjugative relaxases. The thickness of the arrow is indicative of the dedication of relaxases to this function. Solid arrows represent functions based on specificity of the relaxases for their target; dotted arrows represent functions derived from their activity on non-cognate targets or off-target. RLX, Relaxases; MOB, Mobilization; TRA, self-transfer; REP, Replication; COP, Copy number; REC, site-specific recombination; INT, site-specific integration; rTN, Retrotransposition; HGT, Horizontal gene transfer. The vertical arrow indicates the direction of the contribution of each layer to the following.

additional mechanism to mediate chromosomal integration of conjugative plasmids transferred into non-permissive hosts. The plasmids transfer range is usually broader than replication range (Kishida et al., 2017), so a system facilitating integration in the chromosome will contribute to the colonization of new hosts, especially if the specificity for the integration target is more relaxed, as shown for the relaxase TrwC (Agundez et al., 2012; Gonzalez-Prieto et al., 2017).

Fig. 1 highlights the different biological functions attributed to conjugative relaxases. In summary, their secondary target, off-target and moonlighting activities all contribute in the end to increasing the genomic plasticity of prokaryotes, whether it is by directing horizontal transfer of self- or non-self DNA molecules, by contributing to plasmid stabilization through replication or increasing copy number, or by enhancing genetic rearrangements through recombination reactions, or promoting retro-transposition. Conjugative relaxases are considered as key contributors to the prokaryotic horizontal gene pool, but they may play other roles in prokaryotic evolution.

Acknowledgements

We are grateful to Mapi Garcillán-Barcia for helpful suggestions. Work in our lab is supported by grants BIO2017-87190-R from the MINECO (Spanish Ministry of Economy and Innovation), and IDEAS211LLOS from the AECC (Spanish Association Against Cancer) to ML. DLG-H is a recipient of a predoctoral appointment from the University of Cantabria.

References

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 (1), e31047.
Agundez, L., Zarate-Perez, F., Meier, A.F., Bardelli, M., Llosa, M., Escalante, C.R., et al., 2018. Exchange of functional domains between a bacterial conjugative relaxase and the integrase of the human adeno-associated virus. PLoS One 13 (7), e0200841.
Antoine, R., Locht, C., 1992 Jul. Isolation and molecular characterization of a novel broad-host-range plasmid from Bordetella bronchiseptica with sequence similarities to plasmids from gram-positive organisms. Mol. Microbiol. 6 (13), 1785–1799.
Asano, S., Higashitani, A., Horiuchi, K., 1999 Apr 15. Filamentous phage replication initiator protein gpII forms a covalent complex with the 5′ end of the nick it introduced. Nucleic Acids Res. 27 (8), 1882–1889.
Barlett, M.M., Erickson, M.J., Meyer, R.J., 1990 Jun 25. Recombination between directly repeated origins of conjugative transfer cloned in M13 bacteriophage DNA models ligation of the transferred plasmid strand. Nucleic Acids Res. 18 (12), 3579–3586.
Bhattacharjee, M.K., Meyer, R.J., 1991 Mar 11. A segment of a plasmid gene required for conjugal transfer encodes a site-specific, single-strand DNA endonuclease and ligase. Nucleic Acids Res. 19 (5), 1129–1137.
Byrd, D.R., Matson, S.W., 1997. Nicking by transesterification: the reaction catalysed by a relaxase. Mol. Microbiol. 25 (6), 1011–1022.
Cabezon, E., Ripoll-Rozada, J., Pena, A., de la Cruz, F., Arechaga, I., 2015 Jan. Towards an integrated model of bacterial conjugation. FEMS Microbiol. Rev. 39 (1), 81–95.
Carraro, N., Burrus, V., 2014. Biology of Three ICE Families: SXT/R391, ICEBs1, and ICESt1/ICESt3. Microbiol. Spectr. 2 (6) Dec.
Caryl, J.A., Thomas, C.D., 2006 Jun. Investigating the basis of substrate recognition in the pC221 relaxosome. Mol. Microbiol. 60 (5), 1302–1318.
Caryl, J.A., Smith, M.C., Thomas, C.D., 2004 Jun. Reconstitution of a staphylococcal plasmid-protein relaxation complex in vitro. J. Bacteriol. 186 (11), 3374–3383.
Cesar, C.E., Machon, C., de la Cruz, F., Llosa, M., 2006 Nov. A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences. Mol. Microbiol. 62 (4), 984–996.
Chandler, M., de la Cruz, F., Dyda, F., Hickman, A.B., Moncalian, G., Ton-Hoang, B., 2013 Aug. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nat. Rev. Microbiol. 11 (8), 525–538.
Chen, Y., Staddon, J.H., Dunny, G.M., 2007 Mar. Specificity determinants of conjugative DNA processing in the enterococcus faecalis plasmid pCF10 and the Lactococcus lactis plasmid pRS01. Mol. Microbiol. 63 (5), 1549–1564.
Clewell, D.B., Helinski, D.R., 1969 Apr. Supercoiled circular DNA-protein complex in Escherichia coli: purification and induced conversion to an open circular DNA form. Proc. Natl. Acad. Sci. U. S. A. 62 (4), 1159–1166.
Clewell, D.B., Helinski, D.R., 1970 Oct 27. Properties of a supercoiled deoxyribonucleic acid-protein relaxation complex and strand specificity of the relaxation event. Biochemistry. 9 (22), 4428–4440.
Coluzzi, C., Guedon, G., Devignes, M.D., Ambroset, C., Loux, V., Lacroix, T., et al., 2017. A glimpse into the world of integrative and mobilizable elements in streptococci reveals an unexpected diversity and novel families of mobilization proteins. Front. Microbiol. 8, 443.
Draper, O., Cesar, C.E., Machon, C., de la Cruz, F., Llosa, M., 2005 Nov 8. Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. Proc. Natl. Acad. Sci. U. S. A. 102 (45), 16385–16390.
Fernandez-Gonzalez, E., Bakioui, S., Gomes, M.C., O'Callaghan, D., Vergunst, A.C., Sangari, F.J., et al., 2016. A functional oriT in the Ptw plasmid of burkholderia cenocepacia can be recognized by the R388 relaxase TrwC. Front. Mol. Biosci. 3, 16.
Fernandez-Lopez, C., Pluta, R., Perez-Luque, R., Rodriguez-Gonzalez, L., Espinosa, M., Coll, M., et al., 2013 Jul. Functional properties and structural requirements of the plasmid pMV158-encoded MobM relaxase domain. J. Bacteriol. 195 (13), 3000–3008.
Francia, M.V., Clewell, D.B., 2002 Jula. 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. Mol. Microbiol. 45 (2), 375–395.
Francia, M.V., Clewell, D.B., 2002 Octb. Amplification of the tetracycline resistance determinant of pAMalpha1 in enterococcus faecalis requires a site-specific recombination event involving relaxase. J. Bacteriol. 184 (18), 5187–5193.
Francia, M.V., Clewell, D.B., de la Cruz, F., Moncalian, G., 2013 Aug 13. Catalytic domain of plasmid pAD1 relaxase TraX defines a group of relaxases related to restriction endonucleases. Proc. Natl. Acad. Sci. U. S. A. 110 (33), 13606–13611.
Furuya, N., Komano, T., 2003 Jul. NikAB- or NikB-dependent intracellular recombination between tandemly repeated oriT sequences of plasmid R64 in plasmid or single-stranded phage vectors. J. Bacteriol. 185 (13), 3871–3877.
Garcillan-Barcia, M.P., Francia, M.V., de la Cruz, F., 2009 May. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol. Rev. 33 (3), 657–687.
Gennaro, M.L., Kornblum, J., Novick, R.P., 1987 Jun. A site-specific recombination function in Staphylococcus aureus plasmids. J. Bacteriol. 169 (6), 2601–2610.
Goni-Moreno, A., Amos, M., de la Cruz, F., 2013. Multicellular computing using conjugation for wiring. PLoS One 8 (6), e65986.
Gonzalez-Perez, B., Carballeira, J.D., Moncalian, G., de la Cruz, F., 2009 Apr. Changing the recognition site of a conjugative relaxase by rational design. Biotechnol. J. 4 (4), 554–557.
Gonzalez-Prieto, C., Agundez, L., Linden, R.M., Llosa, M., 2013 May. HUH site-specific recombinases for targeted modification of the human genome. Trends Biotechnol. 31 (5), 305–312.
Gonzalez-Prieto, C., Gabriel, R., Dehio, C., Schmidt, M., Llosa, M., 2017 Jun. The conjugative relaxase TrwC promotes integration of foreign DNA in the human genome. Appl. Environ. Microbiol. 15, 83(12).
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 (5), 1163–1172.
Gray, T.A., Derbyshire, K.M., 2018 Jun. Blending genomes: distributive conjugal transfer in mycobacteria, a sexier form of HGT. Mol. Microbiol. 108 (6), 601–613.
Guglielmini, J., Quintais, L., Garcillan-Barcia, M.P., de la Cruz, F., Rocha, E.P., 2011 Aug. The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. PLoS Genet. 7 (8), e1002222.

Guo, M., Ye, J., Gao, D., Xu, N., Yang, J., 2019. Agrobacterium-mediated horizontal gene transfer: mechanism, biotechnological application, potential risk and forestalling strategy. Biotechnol. Adv. 37 (1), 259–270 Jan - Feb.
Guzman, L.M., Espinosa, M., 1997 Mar 7. The mobilization protein, MobM, of the streptococcal plasmid pMV158 specifically cleaves supercoiled DNA at the plasmid oriT. J. Mol. Biol. 266 (4), 688–702.
Guzmán-Herrador Dolores, L.S.S., Alperi, Anabel, González-Prieto, Coral, Roy, Craig R., Llosa, Matxalen, 2017. DNA delivery and genomic integration into mammalian target cells through type IV A and B secretion systems of human pathogens. Front. Microbiol. 8, 1503.
Harley, M.J., Schildbach, J.F., 2003 Sep 30. Swapping single-stranded DNA sequence specificities of relaxases from conjugative plasmids F and R100. Proc. Natl. Acad. Sci. U. S. A. 100 (20), 11243–11248.
Henderson, D., Meyer, R.J., 1996 Dec. The primase of broad-host-range plasmid R1162 is active in conjugal transfer. J. Bacteriol. 178 (23), 6888–6894.
Henderson, D., Meyer, R., 1999 May. The MobA-linked primase is the only replication protein of R1162 required for conjugal mobilization. J. Bacteriol. 181 (9), 2973–2978.
Ilyina, T.V., Koonin, E.V., 1992 Jul 11. 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 (13), 3279–3285.
Jagura-Burdzy, G., Thomas, C.M., 1994 Oct 25. KorA protein of promiscuous plasmid RK2 controls a transcriptional switch between divergent operons for plasmid replication and conjugative transfer. Proc. Natl. Acad. Sci. U. S. A. 91 (22), 10571–10575.
Jandle, S., Meyer, R., 2006 Jan. Stringent and relaxed recognition of oriT by related systems for plasmid mobilization: implications for horizontal gene transfer. J. Bacteriol. 188 (2), 499–506.
Kishida, K., Inoue, K., Ohtsubo, Y., Nagata, Y., Tsuda, M., 2017 Jan. Host range of the conjugative transfer system of IncP-9 naphthalene-catabolic plasmid NAH7 and characterization of its oriT region and relaxase. Appl. Environ. Microbiol. 1, 83(1).
Kunik, T., Tzfira, T., Kapulnik, Y., Gafni, Y., Dingwall, C., Citovsky, V., 2001 Feb 13. Genetic transformation of HeLa cells by agrobacterium. Proc. Natl. Acad. Sci. U. S. A. 98 (4), 1871–1876.
Lacroix, B., Citovsky, V., 2018. Beyond agrobacterium-mediated transformation: horizontal gene transfer from bacteria to eukaryotes. Curr. Top. Microbiol. Immunol. 418, 443–462.
Li, X., Xie, Y., Liu, M., Tai, C., Sun, J., Deng, Z., et al., 2018 Jul 2. oriTfinder: a web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. Nucleic Acids Res. 46 (W1) W229-W34.
Llosa, M., Bolland, S., Grandoso, G., de la Cruz, F., 1994 Jun. Conjugation-independent, site-specific recombination at the oriT of the IncW plasmid R388 mediated by TrwC. J. Bacteriol. 176 (11), 3210–3217.
Llosa, M., Schroder, G., Dehio, C., 2012 Aug. New perspectives into bacterial DNA transfer to human cells. Trends Microbiol. 20 (8), 355–359.
Lorenzo-Diaz, F., Fernandez-Lopez, C., Garcillan-Barcia, M.P., Espinosa, M., 2014 Jul. Bringing them together: plasmid pMV158 rolling circle replication and conjugation under an evolutionary perspective. Plasmid. 74, 15–31.
Lorenzo-Diaz, F., Fernandez-Lopez, C., Lurz, R., Bravo, A., Espinosa, M., 2017 Jul 27. Crosstalk between vertical and horizontal gene transfer: plasmid replication control by a conjugative relaxase. Nucleic Acids Res. 45 (13), 7774–7785.
Martinez-Garcia, E., Benedetti, I., Hueso, A., De Lorenzo, V., 2015. Mining environmental plasmids for synthetic biology parts and devices. Microbiol. Spectr. 3 (1) Feb. PLAS-0033-2014.
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. J. Biol. Chem. 266 (24), 16232–16237.
Mendiola, M.V., de la Cruz, F., 1992 Jul 11. IS91 transposase IS related to the rolling-circle-type replication proteins of the pUB110 family of plasmids. Nucleic Acids Res. 20 (13), 3521.
Miyazaki, R., van der Meer, J.R., 2011 Feb. A dual functional origin of transfer in the ICEclc genomic island of Pseudomonas knackmussii B13. Mol. Microbiol. 79 (3), 743–758.
Moran, R.A., Hall, R.M., 2019 Jan. pBuzz: a cryptic rolling-circle plasmid from a commensal Escherichia coli has two inversely oriented oriTs and is mobilised by a B/O plasmid. Plasmid. 101, 10–19.
Moscoso, M., Eritja, R., Espinosa, M., 1997 May 23. Initiation of replication of plasmid pMV158: mechanisms of DNA strand-transfer reactions mediated by the initiator RepB protein. J. Mol. Biol. 268 (5), 840–856.
Murakami, T., Burian, J., Yanai, K., Bibb, M.J., Thompson, C.J., 2011 Sep 20. A system for the targeted amplification of bacterial gene clusters multiplies antibiotic yield in Streptomyces coelicolor. Proc. Natl. Acad. Sci. U. S. A. 108 (38), 16020–16025.
Nash, R.P., Niblock, F.C., Redinbo, M.R., 2011 Apr 20. Tyrosine partners coordinate DNA nicking by the Salmonella typhimurium plasmid pCU1 relaxase enzyme. FEBS Lett. 585 (8), 1216–1222.
Novikova, O., Smith, D., Hahn, I., Beauregard, A., Belfort, M., 2014 Dec. Interaction between conjugative and retrotransposable elements in horizontal gene transfer. PLoS Genet. 10 (12), e1004853.
Núñez, B., de la Cruz, F., 2001. Two atypical mobilization proteins are involved in plasmid CloDF13 relaxation. Mol. Microbiol. 39 (4), 1088–1099.
O'Brien, F.G., Yui Eto, K., Murphy, R.J., Fairhurst, H.M., Coombs, G.W., Grubb, W.B., et al., 2015 Sep 18. Origin-of-transfer sequences facilitate mobilisation of non-conjugative antimicrobial-resistance plasmids in Staphylococcus aureus. Nucleic Acids Res. 43 (16), 7971–7983.
Pansegrau, W., Lanka, E., 1996 May 31. Mechanisms of initiation and termination reactions in conjugative DNA processing. Independence of tight substrate binding and catalytic activity of relaxase (TraI) of IncPalpha plasmid RP4. J. Biol. Chem. 271 (22), 13068–13076.
Pansegrau, W., Ziegelin, G., Lanka, E., 1988 Dec 20. The origin of conjugative IncP plasmid transfer: interaction with plasmid-encoded products and the nucleotide sequence at the relaxation site. Biochim. Biophys. Acta 951 (2–3), 365–374.
Pansegrau, W., Ziegelin, G., Lanka, E., 1990 Jun 25. Covalent association of the traI gene product of plasmid RP4 with the 5′-terminal nucleotide at the relaxation nick site. J. Biol. Chem. 265 (18), 10637–10644.
Pansegrau, W., Schroder, W., Lanka, E., 1993 Apr 1. Relaxase (TraI) of IncP alpha plasmid RP4 catalyzes a site-specific cleaving-joining reaction of single-stranded DNA. Proc. Natl. Acad. Sci. U. S. A. 90 (7), 2925–2929.
Perez-Mendoza, D., Lucas, M., Munoz, S., Herrera-Cervera, J.A., Olivares, J., de la Cruz, F., et al., 2006 Nov. The relaxase of the rhizobium etli symbiotic plasmid shows nic site cis-acting preference. J. Bacteriol. 188 (21), 7488–7499.
Pluta, R., Boer, D.R., Lorenzo-Diaz, F., Russi, S., Gomez, H., Fernandez-Lopez, C., et al., 2017 Aug 8. Structural basis of a histidine-DNA nicking/joining mechanism for gene transfer and promiscuous spread of antibiotic resistance. Proc. Natl. Acad. Sci. U. S. A. 114 (32) E6526-E35.
Pollet, R.M., Ingle, J.D., Hymes, J.P., Eakes, T.C., Eto, K.Y., Kwong, S.M., et al., 2016 Jan 4. Processing of nonconjugative resistance plasmids by conjugation nicking enzyme of staphylococci. J. Bacteriol. 198 (6), 888–897.
Ramachandran, G., Miguel-Arribas, A., Abia, D., Singh, P.K., Crespo, I., Gago-Cordoba, C., et al., 2017 Feb. Discovery of a new family of relaxases in Firmicutes bacteria. PLoS Genet. 13 (2), e1006586.
Ramsay, J.P., Firth, N., 2017 Aug. Diverse mobilization strategies facilitate transfer of non-conjugative mobile genetic elements. Curr. Opin. Microbiol. 38, 1–9.
Reygers, U., Wessel, R., Muller, H., Hoffmann-Berling, H., 1991 Sep. Endonuclease activity of Escherichia coli DNA helicase I directed against the transfer origin of the F factor. EMBO J. 10 (9), 2689–2694.
Rocco, J.M., Churchward, G., 2006 Mar. The integrase of the conjugative transposon Tn916 directs strand- and sequence-specific cleavage of the origin of conjugal transfer, oriT, by the endonuclease Orf20. J. Bacteriol. 188 (6), 2207–2213.
Rolloos, M., Hooykaas, P.J., van der Zaal, B.J., 2015 Feb 9. Enhanced targeted integration mediated by translocated I-SceI during the agrobacterium mediated transformation of yeast. Sci. Rep. 5, 8345.
Sagredo, S., Pirzer, T., Aghebat Rafat, A., Goetzfried, M.A., Moncalian, G., Simmel, F.C., et al., 2016 Mr 18a. Orthogonal protein assembly on DNA nanostructures using Relaxases. Angew. Chem. Int. Ed. Eng. 55 (13), 4348–4352.
Sagredo, S., de la Cruz, F., Moncalian, G., 2016b. Design of Novel Relaxase Substrates Based on rolling circle replicases for bioconjugation to DNA nanostructures. PLoS One 11 (3), e0152666.
Salgado-Pabon, W., Jain, S., Turner, N., van der Does, C., Dillard, J.P., 2007 Nov. A novel relaxase homologue is involved in chromosomal DNA processing for type IV secretion in Neisseria gonorrhoeae. Mol. Microbiol. 66 (4), 930–947.
Scherzinger, E., Kruft, V., Otto, S., 1993 Nov 1. Purification of the large mobilization protein of plasmid RSF1010 and characterization of its site-specific DNA-cleaving/DNA-joining activity. Eur. J. Biochem. 217 (3), 929–938.
Smillie, C., Garcillan-Barcia, M.P., Francia, M.V., Rocha, E.P., de la Cruz, F., 2010 Sep. Mobility of plasmids. Microbiol. Mol. Biol. Rev. 74 (3), 434–452.
Street, L.M., Harley, M.J., Stern, J.C., Larkin, C., Williams, S.L., Miller, D.L., et al., 2003 Mar 21. Subdomain organization and catalytic residues of the F factor TraI relaxase domain. Biochim. Biophys. Acta 1646 (1–2), 86–99.
Thoma, L., Muth, G., 2016. Conjugative DNA-transfer in Streptomyces, a mycelial organism. Plasmid (87–88), 1–9 Sep - Nov.
Traxler, B.A., Minkley Jr., E.G., 1988 Nov 5. Evidence that DNA helicase I and oriT site-specific nicking are both functions of the F TraI protein. J. Mol. Biol. 204 (1), 205–209.
van Kregten, M., de Pater, S., Romeijn, R., van Schendel, R., Hooykaas, P.J., Tijsterman, M., 2016 Oct 31. T-DNA integration in plants results from polymerase-theta-mediated DNA repair. Nat. Plants. 2 (11), 16164.
Varsaki, A., Lucas, M., Afendra, A.S., Drainas, C., de la Cruz, F., 2003 Apr. Genetic and biochemical characterization of MbeA, the relaxase involved in plasmid ColE1 conjugative mobilization. Mol. Microbiol. 48 (2), 481–493.
Wang, P., Zhang, C., Zhu, Y., Deng, Y., Guo, S., Peng, D., et al., 2013 Dec. The resolution and regeneration of a cointegrate plasmid reveals a model for plasmid evolution mediated by conjugation and oriT site-specific recombination. Environ. Microbiol. 15 (12), 3305–3318.
Wang, P., Zhu, Y., Zhang, Y., Zhang, C., Xu, J., Deng, Y., et al., 2016 Jun 10. Mob/oriT, a mobilizable site-specific recombination system for unmarked genetic manipulation in Bacillus thuringiensis and Bacillus cereus. Microb. Cell Factories 15 (1), 108.
Waters, V.L., Guiney, D.G., 1993. Processes at the nick region link conjugation, T-DNA transfer and rolling circle replication. Mol. Microbiol. 9 (6), 1123–1130.
Wawrzyniak, P., Plucienniczak, G., Bartosik, D., 2017. The different faces of rolling-circle replication and its multifunctional initiator proteins. Front. Microbiol. 8, 2353.
Wisniewski, J.A., Traore, D.A., Bannam, T.L., Lyras, D., Whisstock, J.C., Rood, J.I., 2016 Mar. TcpM: a novel relaxase that mediates transfer of large conjugative plasmids from Clostridium perfringens. Mol. Microbiol. 99 (5), 884–896.
Zechner, E.L., Moncalian, G., de la Cruz, F., 2017. Relaxases and plasmid transfer in gram-negative Bacteria. Curr. Top. Microbiol. Immunol. 413, 93–113.
Zrimec, J., Lapanje, A., 2018 Jan 29. DNA structure at the plasmid origin-of-transfer indicates its potential transfer range. Sci. Rep. 8 (1), 1820.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 February 2021
doi: 10.3389/fmicb.2021.606629

Conjugative DNA Transfer From *E. coli* to Transformation-Resistant Lactobacilli

Sara Samperio[1], Dolores L. Guzmán-Herrador[1], Rigoberto May-Cuz[1], Maria Cruz Martín[2†], Miguel A. Álvarez[2] and Matxalen Llosa[1*]*

[1] *Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Universidad de Cantabria-CSIC-SODERCAN, Santander, Spain,* [2] *Dairy Research Institute (IPLA-CSIC), Villaviciosa, Spain*

Lactic acid bacteria (LAB) belonging to the genus classically known as *Lactobacillus,* recently split into 25 different genera, include many relevant species for the food industry. The well-known properties of lactobacilli as probiotics make them an attractive model also for vaccines and therapeutic proteins delivery in humans. However, scarce tools are available to accomplish genetic modification of these organisms, and most are only suitable for laboratory strains. Here, we test bacterial conjugation as a new tool to introduce genetic modifications into many biotechnologically relevant laboratory and wild type lactobacilli. Using mobilizable shuttle plasmids from a donor *Escherichia coli* carrying either RP4 or R388 conjugative systems, we were able to get transconjugants to all tested *Lactocaseibacillus casei* strains, including many natural isolates, and to several other genera, including *Lentilactobacillus parabuchneri,* for which no transformation protocol has been reported. Transconjugants were confirmed by the presence of the *oriT* and 16S rRNA gene sequencing. Serendipitously, we also found transconjugants into researcher-contaminant *Staphylococcus epidermidis*. Conjugative DNA transfer from *E. coli* to *S. aureus* was previously described, but at very low frequencies. We have purified this recipient strain and used it in standard conjugation assays, confirming that both R388 and RP4 conjugative systems mediate mobilization of plasmids into *S. epidermidis*. This protocol could be assayed to introduce DNA into other Gram-positive microorganisms which are resistant to transformation.

Keywords: bacterial conjugation, lactic acid bacteria, *Lactobacillus*, *Staphylococcus epidermidis*, plasmid RP4, plasmid R388

OPEN ACCESS

Edited by:
Eva M. Top,
University of Idaho, United States

Reviewed by:
Elisabeth Grohmann,
Beuth Hochschule für Technik Berlin, Germany
Sarah Lebeer,
University of Antwerp, Belgium

***Correspondence:**
Matxalen Llosa
llosam@unican.es
Miguel A. Álvarez
maag@ipla.csic.es

[†]*Deceased in April 2020*

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *15 September 2020*
Accepted: *22 January 2021*
Published: *11 February 2021*

Citation:
Samperio S, Guzmán-Herrador DL, May-Cuz R, Martín MC, Álvarez MA and Llosa M (2021) Conjugative DNA Transfer From E. coli to Transformation-Resistant Lactobacilli. Front. Microbiol. 12:606629. doi: 10.3389/fmicb.2021.606629

INTRODUCTION

Lactic acid bacteria (LAB) are a heterogenic group of Gram-positive bacteria with the capacity of producing lactic acid as the main product of their sugar metabolism. Consequently, LAB are an essential microbial group in the food industry due to their use as starters in the elaboration of a great variety of fermented food and drinks, being responsible for their organoleptic properties and acting as natural preservatives (Smit and Smit, 2005; Jany and Barbier, 2008; Börner et al., 2019). Due to their extensive use during the last centuries in the food industry, some species are Generally Regarded As Safe (GRAS) by the Food and Drugs Administration (FDA) and have the status of Qualified Presumption of Safety (QPS) by the European Food Safety Authority (EFSA).

The genus *Lactobacillus*, in addition, contains some strains that are well-known probiotics. Up to now, the genus *Lactobacillus* was exceptionally large and diverse, as it comprised 261 species very different at the phenotypic, ecological and genotypic level. Recently, the genus has been revisited and a new classification into 25 genera has been established (Zheng et al., 2020), which helps to reflect the great biodiversity among the species that previously were grouped as *Lactobacillus*. In this work, the term "lactobacilli" will remain used to designate all organisms previously classified as *Lactobacillus* up to 2020.

The use of lactobacilli has earned interest in human and animal biomedical applications (Wells and Mercenier, 2008; Cano-Garrido et al., 2015; Wang et al., 2016). They are crucial members of the microbiota of human mucosal surfaces, where they are involved in homeostasis processes, providing protection against pathogenic bacteria and stimulating the immune system (Isolauri and Ouwehand, 2004; Bernardeau et al., 2008). Lactobacilli have been proposed as ideal live vectors for the *in situ* production of therapeutic agents in the oral, nasal, genital and intestinal mucosae (Cano-Garrido et al., 2015; Wang et al., 2016; Rio et al., 2019), due to their tolerance to temperature, low pH, bile salts, or high alcohol concentrations (Bosma et al., 2017). So far, lactobacilli have been used as adjuvants or prophylactic agents against many different diseases (Reid, 2017; Mays and Nair, 2018) as well as in a range of animal husbandries (Syngai et al., 2016). Furthermore, their use in therapeutics for prevention and diagnosis (Mays and Nair, 2018) is gaining attention. However, the extended use of lactobacilli in industrial and biomedical applications is limited, since genetic tools are still underdeveloped, especially for wild-type strains (Bosma et al., 2017).

LAB were a pioneer group studied for development of genetic tools (de Vos, 2011), but these efforts were mainly focused on obtaining food-grade microorganisms rather than optimizing mutagenesis procedures (Derkx et al., 2014; Bachmann et al., 2015; Johansen, 2018; Vida and Berlec, 2020). There are several targeted genome editing methods currently available for LAB (Martin et al., 2000; Bosma et al., 2017; Hatti-Kaul et al., 2018). First studies focused on *Lactococcus lactis* and *Lactiplantibacillus plantarum*, due to their importance as starter cultures and probiotics, respectively, but several other LAB species have been found to be susceptible to genetic modification albeit with significantly lower efficiencies (de Vos, 2011; Bosma et al., 2017). The first step to accomplish targeted genetic modification is the introduction of DNA, which can be challenging in Gram-positive bacteria due to the thick peptidoglycan layer in their cell wall. The most widely used method is electroporation. Although a wide range of LAB species have been successfully transformed using generalized electroporation protocols, efficiencies varied strongly among strains and protocols need to be optimized (Landete et al., 2014; Bosma et al., 2017). In particular, transformation of lactobacilli wild-type strains has proven difficult or even not feasible. Thus the importance of exploring new approaches for efficient DNA introduction in these important LAB.

Bacterial conjugation is an efficient mechanism of horizontal gene transfer (HGT) of DNA from a donor bacterium to a recipient one which requires physical contact between them. Bacterial conjugation confers a high genomic plasticity to the prokaryotic world (de la Cruz and Davies, 2000), being the most important means of spreading resistance and virulence factors among bacteria. The conjugative machinery is composed by a Type IV secretion system (T4SS), which constitutes the physical channel for secretion of the DNA, and a number of proteins which recognize and process the DNA to be transferred (Cabezón et al., 2015). Among them, a key enzyme is the relaxase, which attaches covalently to the DNA and pilots it into the recipient cell (Guzmán-Herrador and Llosa, 2019). The DNA to be transferred must have a sequence of 100–400 bp known as origin of transfer (*oriT*), which is recognized by the relaxase, where it binds and cleaves the strand to be transferred.

Conjugation has been described in both Gram-negative and -positive bacteria, and even between both bacterial groups. Conjugative transfer using the RP4 transfer system from *E. coli* to several Gram-positive bacteria was described long ago (Trieu-Cuot et al., 1987). Dominguez and Sullivan (2013) describe a robust conjugation protocol that can be used in the transfer of genetic material from *E. coli* to several *Bifidobacterium* species. Although conjugative plasmids and transposons are very common in LAB, the details of conjugative mechanisms are still under research (Kullen and Klaenhammer, 2000; Bron et al., 2019). It has also been described that conjugative transfer happens *in vivo* in our microbiota, including LAB species as recipients (Aviv et al., 2016). Conjugative DNA transfer from lactobacilli has been described in a few instances to other LAB, such as *Enterococcus* and *Lactococcus* (Gevers et al., 2003). However, up to date, transfer of DNA into lactobacilli by conjugation has not been reported.

The development of genetic modification tools for Gram-positive bacteria has also focused on species with clinical relevance. An example of this is illustrated by the genus *Staphylococcus*, which includes many relevant strains for human health, and is also reluctant to genetic manipulation. Staphylococci are one of the main causative agents of severe nosocomial infections which require prolonged hospitalizations (Becker et al., 2014). The majority of genetic tools have been developed for *Staphylococcus aureus*; in other staphylococci, genetic modification is often halted by the absence of efficient transformable protocols. This is the case of coagulase-negative staphylococci (Becker et al., 2014) which include species with increasing interest in human health, such as the emerging human pathogen *Staphylococcus epidermidis* (Otto, 2009). Up to now, staphylococcal species are transformed via electroporation or, less frequently, by protoplast transformation (Götz and Schumacher, 1987; Augustin and Götz, 1990; Löfblom et al., 2007). However, the restriction-modification systems present in *S. aureus* truncate the uptake of foreign DNA (Waldron and Lindsay, 2006; Xu et al., 2011; Monk and Foster, 2012; Monk et al., 2012). The use of *E. coli* strains lacking *dcm* for production of unmethylated DNA allowed electroporation of particular strains (Monk et al., 2012; Costa et al., 2017), but it requires large amounts of DNA and is limited to specific isolates. Transformation is especially inefficient for *S. epidermidis* strains (Costa et al., 2017). Bacterial conjugation from *E. coli* to *S. aureus* was initially reported, albeit

at low frequency (Trieu-Cuot et al., 1987), and no follow-up works are available.

The limitations of current electroporation protocols for the introduction of DNA, especially in wild-type strains, prompted us to assay conjugation as an alternative to transfer DNA into lactobacilli. Furthermore, conjugation is considered a natural mechanisms and therefore is a more accepted approach than electro-transformation (Pedersen et al., 2005). To this end, we have optimized a conjugation protocol from *E. coli* to lactobacilli using the promiscuous conjugative plasmids R388 and RP4. Using this protocol, we obtained transconjugants in a number of genera and species, including many wild-type strains. Serendipitously, we found that this conjugation protocol also mediates conjugal transfer from *E. coli* to *S. epidermidis*, a researcher-contaminant bacterium which normally colonizes the human skin. This conjugation protocol could be a useful approach to genetically modify other Gram-positive microorganisms which are resistant to electroporation.

MATERIALS AND METHODS

Bacterial Strains and Culture Conditions

The bacterial strains used in this study are listed in **Table 1**. *Escherichia* coli strains were grown at 37°C in LB media and when necessary supplemented with 100 μg/ml ampicillin (Ap), or 50 μg/ml kanamycin (Km). Lactobacilli and *S. epidermidis* were grown at 37°C without aeration in MRS medium (Oxoid, Basingstoke, Hampshire, England) or on solid MRS plates supplemented with 2% agar, supplemented with 5 μg/ml erythromycin (Em) when indicated.

The *S. epidermidis* strain used in this work was isolated from the researcher's skin. The hands were placed on MRS-agar plates without antibiotics, which were then incubated at 37°C. The colonies grown were replicated on MRS-agar with and without Em 5 μg/ml. Em-sensitive colonies were selected, their 16S rRNA gene sequence was amplified by PCR, and DNA sequence was determined (STABVIDA) to confirm they were *S. epidermidis*.

DNA Manipulation

In order to extract genomic DNA from lactobacilli and *S. epidermidis*, a colony from an MRS-agar plate is punctured and resuspended in 50 μl of TE buffer. 50 μl of chloroform are added and mixed thoroughly until the mixture is homogeneous. The mixture is then centrifuged 10 min at 4°C and three phases appear. The top phase containing the genomic DNA is collected carefully and used directly for PCR analysis.

Plasmid DNA was isolated from *E. coli* with the GenElute Plasmid Miniprep Kit (Sigma). From lactobacilli and *S. epidermidis*, the protocol of Anderson and Mckay (1983) was followed with modifications to lyse the cells previous to plasmid DNA purification with the GenElute Plasmid Miniprep Kit, as follows. *Lactobacillus* and *S. epidermidis* strains were grown overnight in MRS supplemented with Em (5 μg/ml) for plasmid selection. Two milliliter cultures were centrifuged 10 min at 14,000 rpm. The pellet was resuspended in STE (sucrose 10.3%, Tris HCl 25 mM pH8, EDTA 10 mM) and centrifuged again 10 min at the same speed. The pellet was frozen at −80°C for 15 min. Then, the pellet was resuspended in 200 μl of lysis buffer (sucrose 20%, Tris HCl 10 mM pH8, EDTA 10 mM, NaCl 10 mM) with lysozyme (30 mg/ml), 2 μl of RNAse (25 mg/ml) and 20 μl of proteinase K (20 mg/ml).

TABLE 1 | Bacterial strains used in the present study.

Bacteria	Relevant properties[a]	Reference or source
Escherichia coli DH5α	Nx^R; *F- endA1 hsdR17 supE44 thi-1 recA1 gyrA96 relA1 Δ(argF-lacZYA) U 169Φ80dlac ΔM15*	Grant et al., 1990
E. coli D1210	Sm^R; *recA hspR hsdM rpsl lacI*^{q}	Sadler et al., 1980
E. coli S17.1	Sm^R; (F-)RP4-2-Tc:Mu *aph*:Tn*7 recA*	Simon et al., 1983
Furfurilactobacillus rossiae D87	Isolated from bread dough	Del Rio et al., 2018
Lacticaseibacillus casei 393	Laboratory strain	Hansen and Lessel, 1971
L. casei 12003	Isolated from wheat dough—white bread	Alvarez-Sieiro et al., 2016
L. casei 12032	Isolated from wheat dough—Pasta	Alvarez-Sieiro et al., 2016
L. casei 12042	Isolated from wheat dough—white bread	Alvarez-Sieiro et al., 2016
L. casei 13b	Dairy-derived—zamorano	Herrero-Fresno et al., 2012
L. casei 41b	Dairy-derived—zamorano	Herrero-Fresno et al., 2012
L. casei 5b	Dairy-derived—zamorano	Herrero-Fresno et al., 2012
L. casei 61b	Dairy-derived—cabrales	Herrero-Fresno et al., 2012
L. casei E2	Dairy-derived—emmental	Herrero-Fresno et al., 2012
Lacticaseibacillus paracasei 1D-CCC76	Isolated from cheese	IPLA collection
Lactiplantibacillus plantarum IPLA88	Isolated from bread dough	Laredo et al., 2013
Lactobacillus crispatus HFS47	Isolated from human feces	IPLA collection
Lactobacillus gasseri HFS29	Isolated from human feces	IPLA collection
Latilactobacillus curvatus 1b-VPZ3	Isolated from cheese	IPLA collection
Lentilactobacillus buchneri 1D-VPC30	Isolated from cheese	IPLA collection
Lentilactobacillus parabuchneri 11122	Dairy-derived—emmental	Diaz et al., 2016
Levilactobacillus brevis 1D-VCC39	Isolated from cheese	IPLA collection
Ligilactobacillus ruminis HFS44	Isolated from human feces	IPLA collection
Limosilactobacillus reuteri IPLA11078	Isolated from cheese	IPLA collection
Limosilactobacillus vaginalis IPLA11050	Isolated from cheese	Diaz et al., 2015a
Loigolactobacillus coryniformis MZ25	Isolated from cheese	IPLA collection
Staphylococcus epidermidis	Human skin—spontaneous isolate	This work

[a] Nx^R, *nalidixic acid resistance;* Sm^R, *streptomycin resistance.*

The sample was homogenized by vortexing, and incubated at 55°C during 30 min. Then, the lysates were applied to GenElute Plasmid Miniprep Kit (Sigma-Aldrich) to purify the plasmid DNA following the manufacturer's protocol.

DNA and PCR products were visualized by agarose gel electrophoresis stained with SYBR Safe (Invitrogen) and visualized with a Gel Doc2000 UV system, and images were analyzed with Quantity One software (BioRad). HyperLader I (Biolabs) was used as a molecular weight marker. DNA was quantified using a Nano-Drop Spectrophotometer ND-1000. GenElute PCR Clean-Up Kit (Sigma) was used for purification of PCR products, and GenElute Agarose Spin Columns (Sigma) were used for DNA purification from agarose gels.

Plasmids and Plasmid Constructions

Plasmids used in this work and their relevant properties are listed in **Table 2**. Plasmid constructions were done by standard recombinant DNA techniques. Plasmid pEM110 was digested with the enzymes *Cla*I and *Sma*I (Thermo Fisher Scientific). The *oriT* sequences of plasmids RP4 and R388 were PCR-amplified from plasmids pLA31 and pLA32, respectively, using the oligonucleotides shown in **Table 3** and high fidelity DNA polymerase PCRBIO HiFi (PCRBIOSYSTEMS). PCR fragments were digested with the same enzymes and ligated with the vector. Ligations were electroporated into *E. coli* (see below). The DNA sequence of the inserts was determined (STABVIDA) to verify the correct assembly of the new plasmids.

E. coli Electroporation

For preparation of electrocompetent cells, bacteria were grown to OD600 = 0.5–0.6, and pelleted by centrifugation at 4°C. Two series of washes and centrifugations (6,000 rpm on a Beckman JA-10 rotor) of 1vol milliQ water and a final wash in 1/50 volume 10% glycerol at 4°C were applied. Cells were resuspended in 1/500 vol 10% glycerol and aliquoted in 50 μl samples. Aliquots were frozen on dry ice and kept at −70°C 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, 25 μF and 200 Ω) in a MicroPulser™ (BioRad). Electroporated cells were added to 1 ml LB and incubated with shaking at 37°C to allow antibiotic expression. After incubation cells were plated on antibiotic containing media.

TABLE 2 | Plasmids used in the present study.

Plasmid	Relevant properties[a]	Reference or source
pCOR48	pEM110- based, shuttle vector *E. coli –Lactobacillus*, Ap^R Em^R; R388 *oriT*	This work
pCOR49	pEM110-borned, shuttle vector *E. coli –Lactobacillus,* Ap^R Em^R; RP4 *oriT*	This work
pEM110	P8014-2 *oriV* (*L. plantarum*), pBR322 *oriV* (*E. coli*), Em^R	Martín et al., 2004
pLA31	pSU36:RP4 *oriT*	Agúndez et al., 2012
pLA32	pSU36:R388 *oriT*	
pSU711	Km^R; R388 Δ*oriT*	Demarre et al., 2004

[a] *Ap^R, ampicillin resistance; Cm^R, chloramphenicol resistance; Em^R, erythromicyn resistance; Km^R, kanamycin resistance.*

TABLE 3 | Primers used for PCR.

Purpose of PCR	Sequence
***oriT* for cloning**	
*Cla*I-*oriT* R388-1	5′-CCGACTATCGATTCTCATTTTCTGCATCATGGTAG-3′
oriT R388 401-*Sma*I	5′-AGCTATCCCGGGCCGCCTCGTCCTCCAAAA-3′
*Cla*I-*oriT RP4-1*	5′-CCGACTATCGATCCGCTTGCCCTCATCTG-3′
oriT RP4-*Sma*I	5′-AGCTTTCCCGGGCGCTTTTCCGCTGCATAA-3′
Transconjugant confirmation	
oriT R388 F	5′-CCAAGTCTACTCATTTTCTGCATCATTGT-3′
oriT R388 R	5′-CCAAGTCTACCTCTCCCGTAGTGTTACT-3′
*Cla*I-*oriT RP4-1*	5′-CCGACTATCGATCCGCTTGCCCTCATCTG-3′
oriT RP4-*Sma*I	5′-AGCTTTCCCGGGCGCTTTTCCGCTGCATAA-3′
16S rRNA determination	
16S 1492R	5′-TACGGYTACCTTGTTACGACTT-3′
16S 27F	5′-AGAGTTTGATYMTGGCTCAG-3′

Underlined sequences represent the restriction sites.

Bacterial Conjugation

Standard conjugation assays in *E. coli* were performed as described in Grandoso et al. (2000). The conjugation protocol from *E. coli* to lactobacilli was optimized starting from the one previously described. Once the new protocol was established, conjugation to lactobacilli was performed as follows: First, donor *E. coli* strains were grown on LB supplemented with antibiotics overnight. Recipient lactobacilli were grown on MRS without antibiotics. The matings were performed on solid media by mixing the same volume of donor and recipient strains (100 μl of the overnight cultures) after washing with BHI media (Oxoid, Basingstoke, Hampshire, England). The bacterial mixture was then washed with BHI again, resuspended in 20 μL of BHI and transferred to a conjugation filter (0.22 μm nitrocellulose, Millipore) on a BHI-agar plate. The mating mixtures were incubated at 37°C for 24 h. Then, the filter was resuspended on BHI and appropriate dilutions were made and plated on selective media for donors, recipients and transconjugants. Donor *E. coli* were plated on LB agar with antibiotics for strain and plasmids selection. Recipient strains were plated on MRS. Transconjugants were plated on MRS with Em 5 μg/ml. The frequency of conjugation is expressed as the number of transconjugants per donor cell. Conjugation from *E. coli* into *Staphylococcus epidermidis* strains was performed as explained above for lactobacilli. All the manipulations of these conjugations were performed on a Faster BH-EN 2004 Class II Microbiological Safety Cabinet and using filter tips.

Analysis of Transconjugants

Transconjugants were analyzed directly from the plate for the presence of *oriT*. PCR reactions included an extra boiling step at the beginning to break the cells. PCRs were performed using DNA polymerase KapaTaq (KapaBiosystems) and primers

indicated in **Table 3**. PCR products were run on agarose gels to observe the expected amplification bands.

Several transconjugants from each conjugation assay were analyzed to confirm the lactobacilli or *S. epidermidis* species by PCR-amplification of the 16S rRNA gene, using the universal primer pair 27F and 1492R (**Table 3**; Lane, 1991), and determination of the DNA sequence from the amplicon, as explained in Diaz et al. (2016).

In order to confirm the presence of the autonomous shuttle plasmid in *S. epidermidis* transconjugants, plasmid DNA was extracted from both the transconjugants and the strain with no plasmid, and from *L. casei* with and without plasmid as a control. Plasmid DNA was visualized on agarose gels. Subsequently, this plasmid DNA was electroporated into *E. coli*, plasmid DNA extracted again from the transformants, and analyzed by restriction digestion to test its integrity.

RESULTS

Bacterial Conjugation From *E. coli* to *Lacticaseibacillus casei*

In order to set up a protocol for conjugative DNA transfer into lactobacilli from laboratory *E. coli* strains, we adapted the protocol routinely used for conjugative DNA transfer among Gram-negative bacteria on solid media (Grandoso et al., 2000). We tested two well characterized conjugative systems; those of plasmids R388 and RP4, which have been previously shown to mediate conjugative DNA transfer into a broad range of recipient cells (see section "Introduction").

For a DNA molecule to be transferred by conjugation, the only element required in *cis* is the *oriT*. We constructed mobilizable shuttle vectors carrying replication and antibiotic resistance genes for selection in *E. coli* and *Lactobacillus*, plus the *oriT* of either R388 or RP4 (pCOR48 and pCOR49; **Table 2**). The rest of the conjugative machinery was provided in *trans*, either using *E. coli* S17.1 strain as a donor, which has the conjugative machinery of RP4 integrated into the bacterial chromosome, or using a non-mobilizable helper plasmid which provides the R388 conjugative system (pSU711; **Table 2**). These plasmids were tested in conjugation between *E. coli* strains to verify their functionality (**Table 4**, top rows). As negative controls, we used donors harboring the mobilizable plasmids but devoid of the rest of the conjugative machinery.

In order to optimize a new protocol for conjugation from *E. coli* to lactobacilli, we chose the laboratory strain *L. casei* 393, which is easy to grow, manipulate and transform (Chassy and Flickinger, 1987). Different conditions were tested, such as the mating time, donor/recipient cell ratio, growth phases in the bacterial cultures, and culture media in the conjugation plate. After several trials, a functional protocol for conjugation between *E. coli* and *L. casei* was established. The protocol is detailed in section "Materials and Methods." In summary, overnight cultures of both donor and recipient bacteria were mixed on BHI medium, where both donor and recipient cells can thrive, while LB and MRS allow growth only of *E. coli* and *L. casei*, respectively, allowing counter selection of donors or recipients. The mating mixtures were incubated on solid media for 24 h. Conjugation frequencies obtained are shown in **Table 4** (lower rows). Transconjugants were obtained using both R388 and RP4 conjugative systems, with frequencies only 1−log lower than between *E. coli* strains (2.1×10^{-4} vs. 2.5×10^{-3} transconjugants/donor for R388, and 1.8×10^{-3} vs. 1.4×10^{-2} for RP4).

Several transconjugants were selected for further analysis. Total DNA was extracted and used as a template for two PCR amplifications: (i) the 16S rRNA gene region, which was used for DNA sequence determination and confirmation that they were *L. casei*; and (ii), the corresponding *oriT*. An example of this analysis is shown in **Figure 1**. It can be observed that neither *oriT* is amplified from gDNA of *L. casei*, while the expected band for each *oriT* is present in DNA from the transconjugants. All these results confirm that the transconjugants obtained were *bona-fide L. casei* colonies which had received the pCOR shuttle plasmid by conjugation.

The ability to transform lactobacilli by electroporation varies widely depending not only on the genus or species, but also on the strains. Some of the more interesting strains for biotechnological purposes are wild-type isolates, which typically show lower transformation rates than laboratory strains. We tested different strains of *L. casei* as conjugation recipients, isolated from natural environments. The results (**Table 5**) showed that we obtained transconjugants into all strains using RP4 conjugative system, whereas R388 conjugative system provided transconjugants in a subset of strains only. Transconjugants were confirmed by

TABLE 4 | Conjugation frequency of mobilizable shuttle plasmids into *E. coli* and *Lacticaseibacillus casei 393*.

Recipient	Donor strain	Conjugative system	Conjugation frequency	Stand. Dev.	n
E. coli DH5α	**D1210 + pSU711 + pCOR48**	R388+	**2.5×10^{-3}**	$\pm 1.9 \times 10^{-3}$	3
E. coli DH5α	**D1210 + pCOR48**	R388−	$<2.6 \times 10^{-7}$	$\pm 1.5 \times 10^{-7}$	3
E. coli DH5α	**S17.1 + pCOR49**	RP4+	**1.4×10^{-2}**	$\pm 1.3 \times 10^{-2}$	3
E. coli DH5α	**D1210 + pCOR49**	RP4−	$<4.6 \times 10^{-6}$	$\pm 4.0 \times 10^{-6}$	3
L. casei 393	**D1210 + pSU711 + pCOR48**	R388+	**2.1×10^{-4}**	$\pm 6.1 \times 10^{-4}$	18
L. casei 393	**D1210 + pCOR48**	R388−	$<2.0 \times 10^{-6}$	$\pm 2.3 \times 10^{-6}$	18
L. casei 393	**S17.1 + pCOR49**	RP4+	**1.8×10^{-3}**	$\pm 2.0 \times 10^{-3}$	8
L. casei 393	**D1210 + pCOR49**	RP4−	$<3.0 \times 10^{-6}$	$\pm 3.1 \times 10^{-6}$	8

Positive results are highlighted in bold.

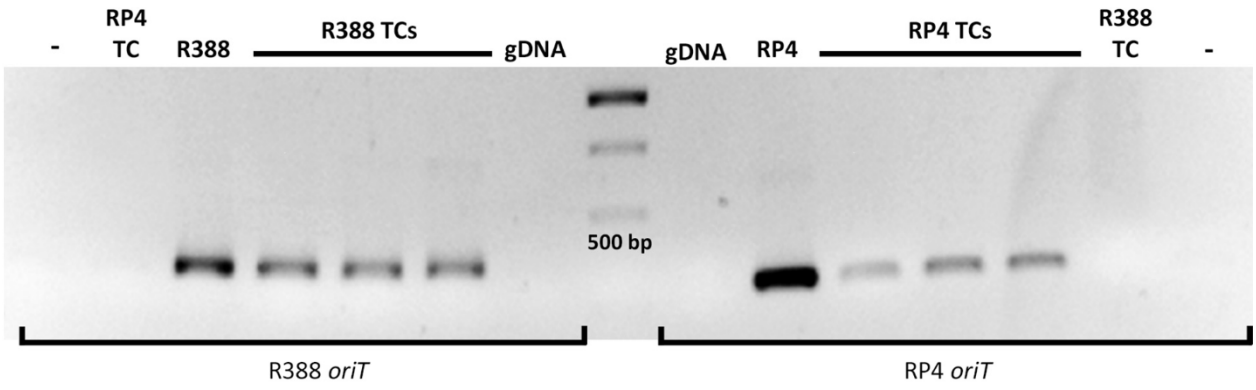


FIGURE 1 | PCR analysis of *L. casei* 393 transconjugants for the amplification of R388 and RP4 *oriT*. The *oriT* amplified by PCR is indicated at the bottom (oligonucleotides indicated in **Table 3**). The template is indicated on top of the gel, and it was either purified DNA (R388 and RP4 as positive controls, and the recipient *L. casei* gDNA as negative control) or transconjugant (TC) colonies directly used for the PCR reactions. The expected plasmid in the transconjugants (R388 or RP4) is indicated. –, control reaction with no template DNA. The central lane is a molecular weight marker (GeneRuler 1 kb DNA ladder). The 500 bp band is indicated. Expected fragment size: 330 bp for R388 *oriT* and 280 bp for RP4 *oriT*.

PCR amplification of the *oriT*. We confirmed that conjugation frequencies varied significantly among strains of *L. casei,* and frequencies were in all cases lower than that of the laboratory strain: in the case of R388, frequencies ranged around 10^{-7}–10^{-5} transconjugants/donor (compared to 10^{-4} for *L. casei* 393), and in the case of RP4, we obtained between 10^{-7} and 10^{-4} transconjugants per donor (compared to 10^{-3} for the laboratory strain).

Conjugation From *E. coli* to Other Lactobacilli

The next step was to test conjugation to other wild-type lactobacilli, some of which are reluctant to genetic transformation by electroporation. Conjugation was performed using the same donor strains harboring R388 and RP4 conjugative systems as shown in **Table 4**, and using as positive control for conjugation *L. casei* 393 as a recipient. The conjugation frequencies obtained are shown in **Table 6**. Transconjugants were obtained for *Lactiplantibacillus plantarum, Lentilactobacillus buchneri, Lentilactobacillus parabuchneri, Levilactobacillus brevis,* and *Limosilactobacillus vaginalis* when using the RP4 conjugative system. No transconjugants were obtained when using the R388 conjugative system. The frequencies obtained were in all cases significantly lower than into the laboratory strain *L. casei* 393, ranging around 10^{-5}–10^{-6} transconjugants per donor (vs. 10^{-3} for the laboratory strain).

Transconjugants were confirmed by the presence of the corresponding *oriT* (**Figure 2**). Their 16S rRNA gene sequence

TABLE 5 | Conjugation frequencies from *E. coli* to different *L. casei* strains using R388 and RP4 systems[a].

Recipient strain	Conjugation frequencies			
	R388		RP4	
	+	–	+	–
L. casei 393	**2.1×10^{-4}**	$<2.0 \times 10^{-6}$	**1.8×10^{-3}**	$<3.0 \times 10^{-6}$
L. casei 5 b	**3.5×10^{-6}**	$<3.6 \times 10^{-6}$	**1.9×10^{-4}**	$<2.5 \times 10^{-6}$
L. casei 13 b	**2.8×10^{-6}**	$<8.4 \times 10^{-6}$	**3.9×10^{-5}**	$<6.0 \times 10^{-6}$
L. casei E2	**1.5×10^{-7}**	$<6.6 \times 10^{-7}$	**1.6×10^{-4}**	$<8.1 \times 10^{-7}$
L. casei 41 b	$< 1.3 \times 10^{-6}$	$<2.9 \times 10^{-6}$	**1.1×10^{-5}**	$<3.0 \times 10^{-6}$
L. casei 61 b	**5.0×10^{-5}**	$<5.8 \times 10^{-6}$	**3.8×10^{-5}**	$<1.2 \times 10^{-5}$
L. casei 12003	**2.4×10^{-5}**	$<1.0 \times 10^{-7}$	**1.1×10^{-4}**	$<1.0 \times 10^{-7}$
L. casei 12032	$< 1.6 \times 10^{-6}$	$<1.1 \times 10^{-7}$	**7.8×10^{-6}**	$<1.0 \times 10^{-7}$
L. casei 12042	$< 3.9 \times 10^{-8}$	$<2.1 \times 10^{-8}$	**5.6×10^{-7}**	$<3.6 \times 10^{-7}$

[a]*Donor strains as shown in* ***Table 4****. Data shown are the mean of 2 independent experiments. Positive results are highlighted in bold.*

TABLE 6 | Conjugation frequencies from *E. coli* to different lactobacilli using R388 and RP4 systems[a].

Recipient lactobacilli	Conjugative system			
	R388		RP4	
	+	–	+	–
L. casei 393	**2.1×10^{-4}**	$< 2.0 \times 10^{-6}$	**1.8×10^{-3}**	$< 3.0 \times 10^{-6}$
L. curvatus	$<2.2 \times 10^{-9}$	$<5.7 \times 10^{-9}$	$<1.0 \times 10^{-7}$	$<1.0 \times 10^{-8}$
L. buchneri	$<4.8 \times 10^{-8}$	$<8.7 \times 10^{-8}$	**6.6×10^{-7}**	$<1.1 \times 10^{-7}$
L. brevis	$<6.8 \times 10^{-7}$	$<2.1 \times 10^{-7}$	**6.0×10^{-6}**	$<1.3 \times 10^{-6}$
L. paracasei	$<3.2 \times 10^{-7}$	$<1.4 \times 10^{-7}$	$<3.1 \times 10^{-6}$	$<1.4 \times 10^{-6}$
L. coryniformis	$<8.2 \times 10^{-7}$	$<1.7 \times 10^{-7}$	$<1.8 \times 10^{-6}$	$<1.6 \times 10^{-7}$
L. parabuchneri	$<3.9 \times 10^{-7}$	$<9.7 \times 10^{-8}$	**2.0×10^{-6}**	$< 2.0 \times 10^{-6}$
L. reuteri	$<1.6 \times 10^{-7}$	$<9.1 \times 10^{-9}$	$<1.4 \times 10^{-7}$	$<6.8 \times 10^{-9}$
L. vaginalis	$<5.7 \times 10^{-7}$	$<9.5 \times 10^{-7}$	**1.1×10^{-6}**	$<1.47 \times 10^{-7}$
L. rossiae	$<1.4 \times 10^{-6}$	$<2.0 \times 10^{-6}$	$<5.0 \times 10^{-6}$	$<1.6 \times 10^{-6}$
L. plantarum	$<2.6 \times 10^{-6}$	$<2.3 \times 10^{-8}$	**1.1×10^{-5}**	$<9.0 \times 10^{-9}$
L. crispatus	$<1.0 \times 10^{-7}$	$<1.8 \times 10^{-7}$	$<2.0 \times 10^{-7}$	$<2.7 \times 10^{-7}$

[a]*Donor strains as shown in* ***Table 4****. Data shown are the mean of 2 independent experiments. Positive results are highlighted in bold.*

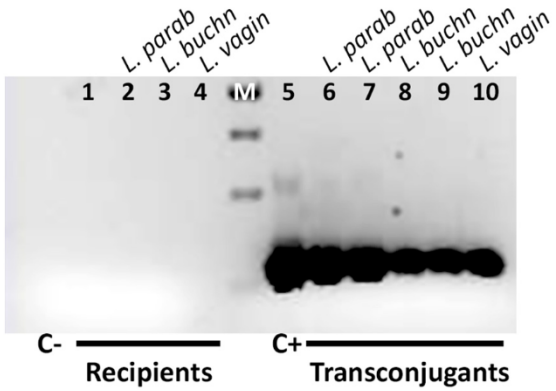


FIGURE 2 | Amplification of RP4 *oriT* from transconjugants. Primers used are shown in **Table 3**. The template DNAs are: Lane 1: no DNA (C–, negative control). Lanes 2–4: gDNAs from recipient *L. parabuchneri, L. buchneri* and *L. vaginalis.* M, molecular marker (GeneRuler 1 kb DNA ladder). The size of the bands is 1,000, 750, 500, and 250 bp from top to bottom. The expected size of the amplicon is 280 bp. To the right, Lane 5: pCOR49 plasmid DNA (C+, positive control). Lanes 6 and 7: DNA extracts from *L. parabuchneri* transconjugants. Lanes 8 and 9: DNA extracts from *L. buchneri* transconjugants. Lane 10: gDNA extracts from *L. vaginalis* transconjugant.

was amplified with primers shown in **Table 3** and the DNA sequence determined, confirming in all cases the expected genera.

Conjugation From *E. coli* to *S. epidermidis*

Serendipitously, we found a high number of putative transconjugants in one of the matings described above using the R388 conjugative system, which did not match the lactobacilli phenotype, although they did show amplification of the R388 *oriT*. Upon sequencing of the 16S rRNA gene, we found out these colonies corresponded to *Staphylococcus epidermidis*, a common isolate in human epidermis, and thus probably originated from a contamination from the researcher skin. Since this fact suggested that conjugation from *E. coli* occurred to other Gram-positives, and *S. epidermidis* itself is a recalcitrant organism of high biomedical interest, we decided to confirm and quantify this phenomenon. To confirm and quantify this finding, as well as to rule out that the observed phenotype was restricted to a particular isolate, *S. epidermidis* was isolated placing on MRS plates the hands of two other researchers, from the same and different laboratories. Colonies resembling staphylococci were obtained, cultured on MRS, checked for their sensitivity to erythromycin by replica-plating, and confirmed as *S. epidermidis* by 16S rRNA gene sequencing. A PCR for the R388 and RP4 *oriT* was performed on total DNA to verify that there was no amplification from the strains (not shown). These new isolates of *S. epidermidis* (isolates 1 and 2) were used as a recipient strains in conjugation assays from *E. coli*.

Conjugation frequencies obtained are summarized in **Table 7**. We obtained *S. epidermidis* transconjugants for both recipient strains, using both R388 and RP4 conjugative systems, confirming that plasmids can be mobilized by conjugation from *E. coli* to *S. epidermidis*. It is interesting to note that in this case, the R388 system was similar or even more efficient in rendering transconjugants than RP4. Transconjugants were confirmed by PCR for amplification of the corresponding *oriT*, and their 16S rRNA gene sequence was amplified by PCR and the DNA sequence determined, to verify that they were *S. epidermidis*.

TABLE 7 | Conjugation frequencies from *E. coli* to *S. epidermidis* using conjugative systems R388 and RP4[a].

	Conjugation frequencies			
	R388		RP4	
Recipient bacteria	+	−	+	−
S. epidermidis isolate 1	**2.5×10^{-6}**	$<1.4 \times 10^{-6}$	**8.1×10^{-8}**	$<1.1 \times 10^{-8}$
S. epidermidis isolate 2	**3.5×10^{-7}**	$<7.8 \times 10^{-7}$	**1.7×10^{-7}**	$<2.1 \times 10^{-7}$

[a]*Donor strains as shown in* ***Table 4****. Data shown are the mean of 3–4 independent experiments. Positive results are highlighted in bold.*

The mobilizable shuttle plasmids used for our mating assays (pCOR48 and pCOR49; **Table 2**) carry origins of replication for *E. coli* and *Lactobacillus*, but it is not known if the plasmid can replicate in staphylococci. In order to determine if the *S. epidermidis* transconjugants harbored the shuttle plasmid as an episome, or they were the result of integration of the plasmid into the *S. epidermidis* chromosome, plasmid DNA was extracted from several transconjugants. In parallel, plasmid DNA was extracted from the strain without plasmids as a negative control, and from *L. casei 393* transconjugants obtained in previous conjugation assays, as positive controls. The DNA samples were run on agarose gels, where the plasmid DNA was visible in the *S. epidermidis* transconjugants (not shown). For further confirmation, these plasmid DNA samples were used to electroporate *E. coli*. Ampicillin-resistant colonies were obtained and their plasmid DNA was extracted and analyzed by restriction digestion in parallel with the original plasmid DNA, to confirm the presence of the pCOR shuttle plasmid (**Figure 3**). It can be observed that the plasmid recovered from the *E. coli* cells transformed with plasmid DNA extracted from the *L. casei* or *S. epidermidis* transconjugants maintains the same restriction pattern as the original shuttle plasmid present in the donor *E. coli* strain. Thus, the shuttle plasmid is able to replicate in *S. epidermidis*.

DISCUSSION

Targeted genetic modification of bacteria with biotechnological and biomedical potential is a prerequisite for most processes of genetic improvement, whether it is to introduce a plasmid contributing to increase production of the desired substance, or to introduce scarless genetic modifications in strains to be applied to human consumption or for medical uses. These processes all have a first requisite consisting on the introduction of DNA into the target cell. A number of protocols exist using bacterial transformation, conjugation,

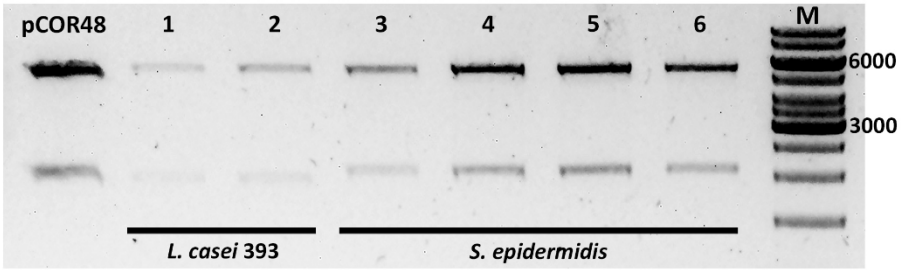


FIGURE 3 | Analysis of plasmids extracted from *E. coli* previously transformed with plasmid DNA from transconjugants. Plasmid DNA samples were digested with enzyme *Eco*RI (Fast Digest, Thermo Fisher Scientific), purified, and run on an agarose gel. pCOR48, plasmid DNA (positive control). The transconjugants from which the electroporated DNA was extracted were *L. casei* 393 (pCOR48) (1–2), and *S. epidermidis* (pCOR48) from two different colonies (3–4 and 5–6). M, molecular weight marker (GeneRuler 1 kb DNA ladder). The 6,000 and 3,000 bp are indicated. Expected fragment sizes: 5,436 and 2,000 bp.

phage transduction, or even protoplast fusions, which are available for most laboratory strains. However, the need for novel substances to use as antimicrobials, food additives, probiotics, or therapeutic substances, has propelled the search for wild-type strains providing the desired properties, which require subsequent optimization steps. Introducing foreign DNA into these microorganisms often proves challenging and even impossible. Among LAB, electroporation is the most widely used method, due to its simplicity, efficiency and wide applicability; however, efficiencies vary strongly among species and even strains, and protocols need to be optimized for each of them (Wang et al., 2020). In particular, transformation of wild-type strains has proven difficult or even not feasible. This is the case for many lactobacilli which, in addition of including some of the most relevant species in the food industry, has an increasingly important biomedical interest, due to both its potential as human live delivery vector and to the existence of emerging human pathogens. Thus, there is an open niche for new DNA introduction protocols.

Bacterial conjugation is a naturally efficient and promiscuous mechanism of horizontal gene transfer, which operates among all main bacterial types. Conjugative DNA transfer from *E. coli* to several LAB has been reported (see section "Introduction"), but to our knowledge, there are no reports of conjugation from *E. coli* into lactobacilli. In this work, we prove that it is possible to introduce DNA by conjugation into lactobacilli from *E. coli*, not only to the model laboratory strain *L. casei* 393, but to a number of other genera, species and natural isolates, typically reluctant to transformation. In particular, we successfully obtained transconjugants in *L. plantarum, L. buchneri*, *L. parabuchneri, L. brevis, and L. vaginalis* (**Table 6**). Since no transconjugants were obtained in the negative controls lacking the conjugative machinery, we conclude that DNA transfer is happening through conjugation, and not through other mechanisms such as natural transformation, nanotubes, or extracellular vesicles. Dedicated electroporation protocols have been published for each of these species, reflecting the inherent difficulty of transforming them (Stephenson et al., 2011; Spath et al., 2012; Zhang et al., 2012). In the case of *L. parabuchneri*, to our knowledge there are no reports of transformation, which makes this result especially significative. *L. parabuchneri* is a member of cheese flora, contributing to its organoleptic properties and ripening process (Fröhlich-Wyder et al., 2013). Moreover, some species have been characterized as potential probiotics (Agostini et al., 2018). On the other hand, some strains of *L. parabuchneri* are mainly responsible for the undesirable accumulation of the biogenic amine histamine in cheese (Diaz et al., 2015b). Thus, the ability to manipulate genetically this species has high scientific interest, as well as both biotechnological and biomedical potential.

The efficiency of conjugation into the model laboratory strain *L. casei* 393 was around 10^{-3} or 10^{-4} transconjugants per donor, depending on the conjugative system used (RP4 or R388), which is higher than in early reports of conjugative transfer between distantly related bacteria (Trieu-Cuot et al., 1987). Comparable rates were obtained in conjugation experiments from *E. coli* to *Bifidobacterium* (Dominguez and Sullivan, 2013) where differences between strains were apparent. The efficiency of the different lactobacilli species as recipients also varied widely, and was always lower than that of the laboratory strain (**Table 6**). Some species were not transformed. There is no taxonomic explanation for this difference: according to the recent reclassification of the genus *Lactobacillus* (Zheng et al., 2020), *L. casei* shares the genus with its closest relative *L. paracasei*, for which we obtained no transconjugants. In fact, a survey of different *L. casei* natural isolates (**Table 5**) showed also ample variation within the species. The difference in conjugation frequencies could have multiple causes, such as the existence of different restriction-modification systems, and very likely, the presence of other plasmids in the wild-type strains; further studies would be necessary to determine the factors interfering with conjugation, which could lead to increased efficiencies and a wider range of potential recipients. Moreover, as for some species the frequency of conjugation obtained in some cases is just at the limit of detection of the mating assays, we think that an optimization of the protocol will probably extend the range of recipient species.

Serendipitously, we found that our conjugation assay also mediated DNA transfer into a researcher-contaminant *S. epidermidis*, so we isolated this species from different researchers and quantified the DNA transfer, confirming the presence of transconjugants. We have also confirmed the episomal presence of the shuttle plasmid in the transconjugants,

meaning that one of the origins of replication present in this plasmid is functional in staphylococci. The mobilizable shuttle plasmids pCOR48 and pCOR49 contain the pBR322 origin of replication, which is functional in *E. coli* but not in Gram-positive organisms, and the replicon of P8014-2, a plasmid isolated from *L. plantarum* (Leer et al., 1992). There are a number of broad-host-range plasmids of Gram-positive bacteria which can replicate in both *Lactobacillus* and *Staphylococcus* spp. (Jain and Srivastava, 2013). Our results indicate that the *Lactobacillus* plasmid P8014-2 replicon is also functional in both lactobacilli and staphylococci. In fact, this replicon includes a sequence at position 2051 (5′-TTCTTATCTTGATA-3′) which is identical to the plus origin of replication of plasmid pC194 (Gros et al., 1987), capable of replication in *S. aureus* and *Bacillus subtilis* (Horinouchi and Weisblum, 1982).

Our finding that bacterial conjugation can be used to introduce DNA into *S. epidermidis* is significant. As stated in the Introduction, staphylococci are difficult to transform, and conjugative DNA transfer from *E. coli* has only been reported for *S. aureus* (Trieu-Cuot et al., 1987) using an IncP plasmid. *S. epidermidis* is a component of the human microbiota and also an emerging pathogen (Otto, 2009), leading to an increasing interest in its genetic manipulation. Up to now, few reports have addressed electroporation and transduction, respectively, of specific *S. epidermidis* strains (Monk et al., 2012; Winstel et al., 2015; Costa et al., 2017). The strategies used to increase the transformation efficiency of *S. aureus* have little efficiency on *S. epidermidis* (Monk et al., 2012; Costa et al., 2017). Thus, adding bacterial conjugation to this scarce toolbox will undoubtedly facilitate the generation of genetically modified strains. Conjugation into the laboratory strain *L. casei* 393 worked efficiently using both RP4 and R388 conjugative systems, although the efficiency was higher with the RP4 system. This result was expected, since the RP4 transfer system has been widely used to transfer DNA into distantly related bacteria and even eukaryotic cells, due to its intrinsic promiscuity (Bates et al., 1998; Luzhetskyy et al., 2006). In contrast, this is, to our knowledge, the first report of conjugative transfer to any Gram-positive bacteria mediated by R388. Moreover, conjugation into *S. epidermidis* was more efficient using R388 than RP4 conjugative system (**Table 7**). These results underscore the importance of assaying different conjugative systems, and point to the R388 conjugative system as a suitable candidate to explore other recalcitrant microorganisms as recipients of bacterial conjugation assays.

With this work, we show that a single conjugation protocol allows the introduction of foreign DNA into many different genera, species, and wild-type strains. The result obtained accidentally with *S. epidermidis* suggests that the range of Gram-positive bacteria which can act as recipients of conjugative DNA transfer from *E. coli* may be wider than suspected. Using *E. coli* as a donor laboratory strain implies access to almost unlimited genetic tools to generate the desired DNA to be transferred. Bacterial conjugation is a simple assay, which allows the transfer of DNA molecules of any size, even whole genomes (Isaacs et al., 2011). In addition, bacterial conjugation is considered a natural process, as opposed to electroporation; conjugation has been exploited to introduce natural plasmids into LAB strains, which can be considered non-genetically modified when this technology is used instead of electroporation (Pedersen et al., 2005; Bron et al., 2019). These features are relevant for the genetic manipulation of LAB, for their use in food fermentation as probiotics or as live vector for mucosal delivery of therapeutic proteins (Wells and Mercenier, 2008).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

MÁ and ML conceived the work. SS, DG-H, RM-C, and MM performed the experiments. SS, DG-H, RM-C, MÁ, and ML analyzed the results. SS, MÁ, and ML wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

Work in ML lab was supported by the grant BIO2017-87190-R from the Spanish Ministry of Science and Innovation. Work in MÁ lab was funded by the Spanish State Research Agency (AEI) and the European Regional Development Fund (FEDER) (AGL2016-78708-R, AEI/FEDER, EU). DG-H was a recipient of a predoctoral appointment from the University of Cantabria. RM-C received an Erasmus+ traineeship grant.

ACKNOWLEDGMENTS

We are grateful to Matilde Cabezas for technical help, and to Mapi Garcillán for helpful suggestions.

DEDICATION

This article is dedicated to the memory of María Cruz Martín.

REFERENCES

Agostini, C., Eckert, C., Vincenzi, A., Machado, B. L., Jordon, B. C., Kipper, J. P., et al. (2018). Characterization of technological and probiotic properties of indigenous *Lactobacillus* spp. from south Brazil. *3 Biotech.* 8:451. doi: 10.1007/s13205-018-1469-7

Agúndez, L., González-Prieto, C., Machón, C., and 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. doi: 10.1371/journal.pone.0031047

Alvarez-Sieiro, P., Montalbán-López, M., Mu, D., and Kuipers, O. P. (2016). Bacteriocins of lactic acid bacteria: extending the family. *Appl. Microbiol. Biotechnol.* 100, 2939–2951. doi: 10.1007/s00253-016-7343-9

Anderson, D. G., and Mckay, L. L. (1983). Simple and rapid method for isolating large plasmid DNA from lactic streptococcit. *Appl. Environ. Microbiol.* 46, 549–552.

Augustin, J., and Götz, F. (1990). Transformation of *Staphylococcus* epidermidis and other *Staphylococcal* species with plasmid DNA by electroporation. *FEMS Microbiol. Lett.* 66, 203–207. doi: 10.1111/j.1574-6968.1990.tb03997.x

Aviv, G., Rahav, G., and Gal-mor, O. (2016). Horizontal transfer of the *Salmonella enterica* serovar infantis resistance and virulence plasmid pESI to the gut microbiota of. mBio. *Am. Soc. Microbiol.* 7:e01395-16. doi: 10.1128/mBio.01395-16

Bachmann, H., Pronk, J. T., Kleerebezem, M., and Teusink, B. (2015). Evolutionary engineering to enhance starter culture performance in food fermentations. *Curr. Opin. Biotechnol.* 32, 1–7. doi: 10.1016/j.copbio.2014.09.003

Bates, S., Cashmore, A. M., and Wilkins, B. M. (1998). IncP plasmids are unusually effective in mediating conjugation of *Escherichia coli* and Saccharomyces cerevisiae: involvement of the Tra2 mating system. *J. Bacteriol.* 180, 6538–6543. doi: 10.1128/jb.180.24.6538-6543.1998

Becker, K., Heilmann, C., and Peters, G. (2014). Coagulase-negative staphylococci. *Clin. Microbiol. Rev.* 27, 870–926. doi: 10.1128/CMR.00109-13

Bernardeau, M., Paul, J., Henri-dubernet, S., and Guéguen, M. (2008). Safety assessment of dairy microorganisms: the *Lactobacillus* genus ⋆. *Int. J. Food Microbiol.* 126, 278–285. doi: 10.1016/j.ijfoodmicro.2007.08.015

Börner, R. A., Kandasamy, V., Axelsen, A. M., Nielsen, A. T., and Bosma, E. F. (2019). Genome editing of lactic acid bacteria: opportunities for food, feed, pharma and biotech. *FEMS Microbiol. Lett* 366:fny291. doi: 10.1093/femsle/fny291

Bosma, E. F., Forster, J., and Nielsen, A. T. (2017). *Lactobacilli* and pediococci as versatile cell factories – evaluation of strain properties and genetic tools. *Biotechnol. Adv.* 35, 419–442. doi: 10.1016/j.biotechadv.2017.04.002

Bron, P. A., Marcelli, B., Mulder, J., Van Der Els, S., Morawska, L. P., Kuipers, O. P., et al. (2019). Renaissance of traditional DNA transfer strategies for improvement of industrial lactic acid bacteria. *Curr. Opin. Biotechnol.* 56, 61–68. doi: 10.1016/j.copbio.2018.09.004

Cabezón, E., Ripoll-Rozada, J., Peña, A., de la Cruz, F., and Arechaga, I. (2015). Towards an integrated model of bacterial conjugation. *FEMS Microbiol. Rev.* 39, 81–95. doi: 10.1111/1574-6976.12085

Cano-Garrido, O. C., Seras-Franzoso, J. S., and Garcia-Fruitós, E. G. (2015). Lactic acid bacteria: reviewing the potential of a promising delivery live vector for biomedical purposes. *Microb. Cell Fact.* 14:137. doi: 10.1186/s12934-015-0313-6

Chassy, B. M., and Flickinger, J. L. (1987). Transformation of *Lactobacillus* casei by electroporation. *FEMS Microbiol. Lett.* 44, 173–177.

Costa, S. K., Donegan, N. P., Corvaglia, A. R., François, P., and Cheung, A. L. (2017). Bypassing the restriction system to improve transformation of *Staphylococcus* epidermidis. *J. Bacteriol.* 199:e00271-17. doi: 10.1128/JB.00271-17

de la Cruz, F., and Davies, J. (2000). Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol.* 8, 128–133.

de Vos, W. M. (2011). Systems solutions by lactic acid bacteria: from paradigms to practice. *Microb. Cell Fact.* 10:S2.

Del Rio, B., Alvarez-Sieiro, P., Redruello, B., Martin, M. C., Fernandez, M., Ladero, V., et al. (2018). *Lactobacillus* rossiae strain isolated from sourdough produces putrescine from arginine. *Sci. Rep.* 8:3989. doi: 10.1038/s41598-018-22309-6

Demarre, G., Guérout, A.-M., Matsumoto-Mashimo, C., Rowe-Magnus, D. A., Marlière, P., and Mazel, D. (2004). A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPα) conjugative machineries and their cognate *Escherichia coli* host strains. *Res. Microbiol.* 156, 245–255. doi: 10.1016/j.resmic.2004.09.007

Derkx, P. M. F., Janzen, T., Sørensen, K. I., Christensen, J. E., Stuer-lauridsen, B., and Johansen, E. (2014). The art of strain improvement of industrial lactic acid bacteria without the use of recombinant DNA technology. *Microb. Cell Fact.* 13, 1–13.

Diaz, M., del Rio, B., Ladero, V., Redruello, B., Fernández, M., Martin, M. C., et al. (2015a). Isolation and typification of histamine-producing *Lactobacillus vaginalis* strains from cheese. *Int. J. Food Microbiol.* 215, 117–123. doi: 10.1016/j.ijfoodmicro.2015.08.026

Diaz, M., Ladero, V., Redruello, B., Sanchez-llana, E., Rio, B., Fernandez, M., et al. (2015b). A PCR-DGGE method for the identification of histamine-producing bacteria in cheese. *Food Control* 63, 216–223. doi: 10.1016/j.foodcont.2015.11.035

Diaz, M., Rio, B., Sanchez-llana, E., Ladero, V., Martin, M. C., Alvarez, M. A., et al. (2016). Histamine-producing *Lactobacillus parabuchneri* strains isolated from grated cheese can form bio films on stainless steel. *Food Microbiol.* 59, 85–91. doi: 10.1016/j.fm.2016.05.012

Dominguez, W., and Sullivan, D. J. O. (2013). Developing an efficient and reproducible conjugation-based gene transfer system for bifidobacteria. *Microbiology* 159, 328–338. doi: 10.1099/mic.0.061408-0

Fröhlich-Wyder, M. T., Guggisberg, D., Badertscher, R., Wechsler, D., Wittwer, A., and Irmler, S. (2013). The effect of *Lactobacillus buchneri* and *Lactobacillus parabuchneri* on the eye formation of semi-hard cheese. *Int. Dairy J.* 33, 120–128. doi: 10.1016/j.idairyj.2013.03.004

Gevers, D., Huys, G., and Swings, J. (2003). In vitro conjugal transfer of tetracycline resistance from *Lactobacillus* isolates to other gram-positive bacteria. *FEMS Microbiol. Lett.* 225, 125–130. doi: 10.1016/S0378-1097(03)00505-6

Götz, F., and Schumacher, B. (1987). Improvements of protoplast transformation in *Staphylococcus carnosus*. *FEMS Microbiol. Lett.* 40, 285–288. doi: 10.1111/j.1574-6968.1987.tb02040.x

Grandoso, G., Avila, P., Cayón, A., Hernando, M. A., Llosa, M., and 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. N., Jessee, J., Bloom, F. R., and 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.

Gros, M. F., Te Riele, H., and Ehrlich, S. D. (1987). Rolling circle replication of single-stranded DNA plasmid pC194. *EMBO J.* 6, 3863–3869. doi: 10.1002/j.1460-2075.1987.tb02724.x

Guzmán-Herrador, D. L., and Llosa, M. (2019). The secret life of conjugative relaxases. *Plasmid* 104:102415. doi: 10.1016/j.plasmid.2019.102415

Hansen, P. A., and Lessel, E. F. (1971). *Lactobacillus casei* (Orla-Jensen) comb. nov. *Int. J. Syst. Bacteriol.* 21, 69–71.

Hatti-Kaul, R., Chen, L., Dishisha, T., and El, H. (2018). Lactic acid bacteria: from starter cultures to producers of chemicals. *FEMS Microbiol. Lett.* 365:fny213. doi: 10.1093/femsle/fny213

Herrero-Fresno, A., Martínez, N., Sánchez-Llana, E., Díaz, M., Fernández, M., Martin, M. C., et al. (2012). *Lactobacillus casei* strains isolated from cheese reduce biogenic amine accumulation in an experimental model. *Int. J. Food Microbiol.* 157, 297–304. doi: 10.1016/j.ijfoodmicro.2012.06.002

Horinouchi, S., and Weisblum, B. (1982). Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.* 150, 815–825. doi: 10.1128/jb.150.2.815-825.1982

Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., et al. (2011). Precise manipulation of chromosomes in vivo enables genome- wide codon replacement farren. *Science* 333, 348–353. doi: 10.1016/j.physbeh.2017.03.040

Isolauri, E., and Ouwehand, A. C. (2004). Probiotics. *Best Pract. Res. Clin. Gastroenterol.* 18, 299–313. doi: 10.1053/ybega.2004.443

Jain, A., and Srivastava, P. (2013). Broad host range plasmids. *FEMS Microbiol. Lett.* 348, 87–96. doi: 10.1111/1574-6968.12241

Jany, J., and Barbier, G. (2008). Culture-independent methods for identifying microbial communities in cheese. *Food Microbiol.* 25, 839–848. doi: 10.1016/j.fm.2008.06.003

Johansen, E. (2018). Use of natural selection and evolution to develop new starter cultures for fermented foods. *Annu. Rev. Food Sci. Technol.* 9, 411–428.

Kullen, M. J., and Klaenhammer, T. R. (2000). Genetic modification of intestinal Lactobacilli and Bifidobacteria. *Curr. Issues Mol. Biol.* 2, 41–50.

Landete, J. M., Arqués, J. L., Peirotén, Á, Langa, S., and Medina, M. (2014). An improved method for the electrotransformation of lactic acid bacteria: a

comparative survey. *J. Microbiol. Methods* 105, 130–133. doi: 10.1016/j.mimet.2014.07.022

Lane, D. J. (1991). "16S/23S rRNA sequencing," in *Nucleic Acid Techniques in Bacterial Systematic,* eds E. Stackebrandt, and M. Goodfellow (New York, NY: John Wiley and Sons), 115–175.

Laredo, V., Alvarez-Sieiro, P., Redruello, B., Del Rio, B., Linares, D. M., Martín, M. C., et al. (2013). Draft genome sequence of *Lactobacillus plantarum* strain IPLA 88. *Genome Announc.* 1:e00524-13. doi: 10.1038/nbt.2491

Leer, R. J., van Luijk, N., Posno, M., and Pouwels, P. H. (1992). Structural and functional analysis of two cryptic plasmids from *Lactobacillus pentosus* MD353 and *Lactobacillus plantarum* ATCC 8014. *Mol. Gen. Genet.* 234, 265–274. doi: 10.1007/BF00283847

Löfblom, J., Kronqvist, N., Uhlén, M., Ståhl, S., and Wernérus, H. (2007). Optimization of electroporation-mediated transformation: *Staphylococcus carnosus* as model organism. *J. Appl. Microbiol.* 102, 736–747. doi: 10.1111/j.1365-2672.2006.03127.x

Luzhetskyy, A., Fedoryshyn, M., Gromyko, O., Ostash, B., Rebets, Y., Bechthold, A., et al. (2006). IncP plasmids are most effective in mediating conjugation between *Escherichia coli* and Streptomycetes. *Russ. J. Genet.* 42, 476–481. doi: 10.1134/S1022795406050036

Martin, M. C., Alonso, J. C., Sua, J. E., Alvarez, M. A., and La, P. (2000). Generation of food-grade recombinant lactic acid bacterium strains by site-specific recombination. *Appl. Environ. Microbiol.* 66, 2599–2604.

Martín, M. C., Fernández, M., Martín-Alonso, J. M., Parra, F., Boga, J. A., and Alvarez, M. A. (2004). Nisin-controlled expression of Norwalk virus VP60 protein in *Lactobacillus casei* à Ferna. *FEMS Microbiol. Lett.* 237, 385–391. doi: 10.1016/j.femsle.2004.07.002

Mays, Z. J. S., and Nair, N. U. (2018). Synthetic biology in probiotic lactic acid bacteria: at the frontier of living therapeutics. *Curr. Opin. Biotechnol.* 53, 224–231. doi: 10.1016/j.copbio.2018.01.028

Monk, I. R., and Foster, T. J. (2012). Genetic manipulation of *Staphylococci*-breaking through the barrier. *Front. Cell. Infect. Microbiol.* 2:49. doi: 10.3389/fcimb.2012.00049

Monk, I. R., Shah, I. M., and Xu, M. (2012). Transforming the untransformable: application of direct transformation to manipulate genetically *Staphylococcus aureus* and *Staphylococcus epidermidis*. *MBio* 3, e277–e211. doi: 10.1128/mBio.00277-11

Otto, M. (2009). *Staphylococcus epidermidis* – the "accidental" pathogen. *Nat. Rev. Microbiol.* 7, 555–567. doi: 10.1038/nrmicro2182

Pedersen, M. B., Iversen, S. L., Sørensen, K. I., and Johansen, E. (2005). The long and winding road from the research laboratory to industrial applications of lactic acid bacteria. *FEMS Microbiol. Rev.* 29, 611–624. doi: 10.1016/j.femsre.2005.04.001

Reid, G. (2017). The development of probiotics for women ' s health. *Can. J. Microbiol.* 63, 269–277.

Rio, B., Redruello, B., Fernandez, M., Martin, M. C., Ladero, V., and Alvarez, M. A. (2019). Lactic acid bacteria as a live delivery system for the *in situ* production of nanobodies in the human gastrointestinal tract. *Front. Microbiol.* 9:3179. doi: 10.3389/fmicb.2018.03179

Sadler, J. R., Tecklenburg, M., and Betz, J. L. (1980). Plasmids containing many tandem copies of a synthetic lactose operator. *Gene* 8, 279–300.

Simon, R., Priefer, U., and Pühler, A. (1983). A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. *Nat. Biotechnol.* 1, 784–791.

Smit, G., and Smit, B. A. (2005). Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. *FEMS Microbiol. Rev.* 29, 591–610. doi: 10.1016/j.femsre.2005.04.002

Spath, K., Heinl, S., Egger, E., and Grabherr, R. (2012). *Lactobacillus plantarum* and *Lactobacillus buchneri* as expression systems: evaluation of different origins of replication for the design of suitable shuttle vectors. *Mol. Biotechnol.* 52, 40–48. doi: 10.1007/s12033-011-9471-x

Stephenson, D., Moore, R. J., and Allison, G. E. (2011). Transformation of, and heterologous protein expression in, *Lactobacillus agilis* and *Lactobacillus vaginalis* isolates from the chicken gastrointestinal tract. *Appl. Environ. Microbiol.* 77, 220–228. doi: 10.1128/AEM.02006-10

Syngai, G. G., Gopi, R., Bharali, R., Dey, S., Lakshmanan, G. M. A., and Ahmed, G. (2016). Probiotics – the versatile functional food ingredients. *J. Food Sci. Technol.* 53, 921–933. doi: 10.1007/s13197-015-2011-0

Trieu-Cuot, P., Carlier, C., Martin, P., and Courvalin, P. (1987). Plasmid transfer by conjugation from *Escherichia coli* to gram-positive bacteria. *FEMS Microbiol. Lett.* 48, 289–294.

Vida, T., and Berlec, A. (2020). Safety aspects of genetically modified lactic acid bacteria. *Microorganisms* 8:297.

Waldron, D. E., and Lindsay, J. A. (2006). Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages. *J. Bacteriol.* 188, 5578–5585. doi: 10.1128/JB.00418-06

Wang, C., Cui, Y., and Qu, X. (2020). Optimization of electrotransformation (ETF) conditions in lactic acid bacteria (LAB). *J. Microbiol. Methods* 174:105944. doi: 10.1016/j.mimet.2020.105944

Wang, M., Gao, Z., Zhang, Y., and Pan, L. (2016). Lactic acid bacteria as mucosal delivery vehicles: a realistic therapeutic option. *Appl. Microbiol. Biotechnol.* 100, 5691–5701. doi: 10.1007/s00253-016-7557-x

Wells, J. M., and Mercenier, A. (2008). Mucosal delivery of therapeutic and prophylactic molecules using lactic acid bacteria. *Nat. Rev. Microbiol.* 6, 349–362. doi: 10.1038/nrmicro1840

Winstel, V., Kühner, P., Krismer, B., Peschel, A., and Rohde, H. (2015). Transfer of plasmid DNA to clinical coagulase-negative Staphylococcal pathogens by using a unique bacteriophage. *Appl. Environ. Microbiol.* 81, 2481–2488. doi: 10.1128/AEM.04190-14

Xu, S. Y., Corvaglia, A. R., Chan, S. H., Zheng, Y., and Linder, P. (2011). A type IV modification-dependent restriction enzyme SauUSI from *Staphylococcus aureus* subsp. aureus USA300. *Nucleic Acids Res.* 39, 5597–5610. doi: 10.1093/nar/gkr098

Zhang, Y., Cheng, K., Gao, Q., and Gao, N. F. (2012). Elaboration of an electroporation protocol for *Lactobacillus brevis*. *Adv. Mater. Res.* 393–395, 729–732. doi: 10.4028/www.scientific.net/AMR.393-395.729

Zheng, J., Wittouck, S., Salvetti, E., Franz, C. M. A. P., Harris, H. M. B., Mattarelli, P., et al. (2020). A taxonomic note on the genus *Lactobacillus*: description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. *Int. J. Syst. Evol. Microbiol.* 70, 2782–2858. doi: 10.1099/ijsem.0.004107

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Samperio, Guzmán-Herrador, May-Cuz, Martín, Álvarez and Llosa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

MINISTERIO
DE INDUSTRIA, ENERGÍA
Y TURISMO

Oficina Española
de Patentes y Marcas

Justificante de presentación electrónica de solicitud de patente

Este documento es un justificante de que se ha recibido una solicitud española de patente por vía electrónica utilizando la conexión segura de la O.E.P.M. De acuerdo con lo dispuesto en el art.14.3 del Reglamento para la ejecución de la Ley 11/1986, de 20 de marzo, de Patentes, se han asignado a su solicitud un número de expediente y una fecha de recepción de forma automática. La fecha de presentación de la solicitud a la que se refiere el art. 22 de la Ley le será comunicada posteriormente.

Número de solicitud:	P202030890	
Fecha de recepción:	28 agosto 2020, 13:12 (CEST)	
Oficina receptora:	OEPM Madrid	
Su referencia:	20200828	
Solicitante:	Universidad de Cantabria	
Número de solicitantes:	3	
País:	ES	
Título:	PROTEÍNA QUIMÉRICA RELAXASA-Cas	
Documentos enviados:	Descripcion.pdf (23 p.) Reivindicaciones-1.pdf (2 p.) Resumen-1.pdf (1 p.) Dibujos.pdf (2 p.) OLF-ARCHIVE.zip OTRO-1.pdf (3 p.) OTRO-2.pdf (1 p.) SEQLPDF.pdf (21 p.) SEQLTXT.txt	package-data.xml es-request.xml application-body.xml es-fee-sheet.xml feesheet.pdf request.pdf
Enviados por:	CN=LEON SERRANO JAVIER - DNI 51058553W,givenName=JAVIER,SN=LEON SERRANO,serialNumber=IDCES-51058553W,OU=A05005105855302, OU=VICERRECTORADO DE INVESTIGACION Y TRANSFERENCIA DEL CONOCIMIENTO,OU=CERTIFICADO ELECTRONICO DE EMPLEADO PUBLICO,O=UNIVERSIDAD DE CANTABRIA,C=ES	
Fecha y hora de recepción:	28 agosto 2020, 13:12 (CEST)	

Codificación del envío:	92:AE:41:0C:74:F1:AB:AB:15:BC:BE:97:7A:A3:A7:97:09:A5:79:CF

AVISO IMPORTANTE

Las tasas pagaderas al solicitar y durante la tramitación de una patente o un modelo de utilidad son las que se recogen en el Apartado "Tasas y precios públicos" de la página web de la OEPM (http://www.oepm.es/es/propiedad_industrial/tasas/). Consecuentemente, si recibe una comunicación informándole de la necesidad de hacer un pago por la inscripción de su patente o su modelo de utilidad en un "registro central" o en un "registro de internet" posiblemente se trate de un fraude.

La anotación en este tipo de autodenominados "registros" no despliega ningún tipo de eficacia jurídica ni tiene carácter oficial.

En estos casos le aconsejamos que se ponga en contacto con la Oficina Española de Patentes y Marcas en el correo electrónico informacion@oepm.es.

ADVERTENCIA: POR DISPOSICIÓN LEGAL LOS DATOS CONTENIDOS EN ESTA SOLICITUD PODRÁN SER PUBLICADOS EN EL BOLETÍN OFICIAL DE LA PROPIEDAD INDUSTRIAL E INSCRITOS EN EL REGISTRO DE PATENTES DE LA OEPM, SIENDO AMBAS BASES DE DATOS DE CARÁCTER PÚBLICO Y ACCESIBLES VÍA REDES MUNDIALES DE INFORMÁTICA.

Para cualquier aclaración puede contactar con la O.E.P.M.

/Madrid, Oficina Receptora/

RESUMEN

La presente invención se refiere a la proteína de fusión Relaxasa-Cas12a, al sistema CRISPR/Cas que comprende dicha proteína de fusión y al uso de la proteína de fusión Relaxasa-Cas12a y/o del sistema CRISPR/Cas para la translocación de endonucleasas y/o endonucleasas unidas a moléculas de ADN, a células diana a través del sistema de secreción bacteriano tipo IV, y para la modificación genética de las células diana.