The Conjugative Relaxase TrwC Promotes Integration of Foreign DNA in the Human Genome

Coral González-Prieto,[a] Richard Gabriel,[b] Christoph Dehio,[c] Manfred Schmidt,[b] and Matxalen Llosa,[a#]

Instituto de Biomedicina y Biotecnología de Cantabria (UC-CSIC-SODERCAN), Santander, Spain[a]; Department of Translational Oncology, National Center for Tumor Diseases and German Cancer Research Center, Heidelberg, Germany[b], Focal Area Infection Biology, Biozentrum, Universität Basel, Basel, Switzerland

Running head: TrwC promotes DNA integration in the human genome

Address correspondence to Matxalen LLosa, llosam@unican.es

ABSTRACT

Bacterial conjugation is a mechanism of horizontal DNA transfer. The relaxase TrwC of the conjugative plasmid R388 cleaves one strand of the transferred DNA at the *oriT*, covalently attaches to it and leads the ssDNA into the recipient cell. In addition, TrwC catalyzes site-specific integration of the transferred DNA into its target sequence present in the genome of the recipient bacterium. Here, we report the analysis of the efficiency and specificity of the integrase activity of TrwC in human cells, using the Type IV Secretion System of the human pathogen *Bartonella henselae* to introduce relaxase-DNA complexes. When compared to Mob relaxase from plasmid pBGR1, we found that TrwC mediated a 10-fold increase in the rate of plasmid DNA transfer to human cells, and a 100-fold increase in the rate of chromosomal integration of the transferred DNA. We used linear amplification-mediated PCR and plasmid rescue to characterize the integration pattern in the human genome. DNA sequence analysis revealed mostly reconstituted *oriT* sequences, indicating that TrwC is active and recircularizes transferred DNA in human cells. One TrwC-mediated site-specific integration event was detected, proving that TrwC is capable of mediating site-specific integration in the human genome, albeit with very low efficiency compared to the rate of random integration. Our results suggest that TrwC may stabilize the plasmid DNA molecules in the nucleus of the human cell, probably by recircularization of the transferred DNA strand. This stabilization would increase the opportunities for integration of the DNA by the host machinery.

IMPORTANCE

Different biotechnological applications, including gene therapy strategies, require permanent modification of target cells. Long-term expression is achieved either by extrachromosomal persistence or by integration of the introduced DNA. Here we study the utility of conjugative relaxase TrwC, a bacterial protein with site-specific integrase activity in bacteria, as integrase in human cells. Although not efficient as site-specific integrase, we found that TrwC is active in human cells and promotes random integration of the transferred DNA in the human genome, probably acting as a DNA chaperone until it is integrated by host mechanisms. TrwC-DNA complexes can be delivered to human cells through a Type IV Secretion System involved in pathogenesis. Thus, TrwC could be used in vivo to transfer the DNA of interest into the appropriate cell and promote its integration. If used in combination with a site-specific nuclease, it could lead to site-specific integration of the incoming DNA by homologous recombination.

INTRODUCTION

Bacterial conjugation is an efficient mechanism of horizontal DNA transfer which confers bacteria an elevated level of genomic plasticity (1). DNA is transferred by conjugation from a donor to a recipient bacterium through a protein complex known as conjugative apparatus (2). In gram-negative bacteria, the conjugative machinery is composed of three functional modules (3): i) the relaxosome, a complex formed by the DNA to be transferred - in particular the site known as origin of transfer (*oriT*) - and the proteins responsible for DNA processing, which include a relaxase and one or more accessory proteins; ii) the Type IV Secretion System (T4SS), a multiprotein complex organized in a transmembranal conduit that spans both inner and outer membranes; and iii) the coupling protein (T4CP), a DNA-dependent ATPase which brings together the two previous components and is believed to play a crucial role in substrate selection. The translocated substrate is the relaxase covalently linked to the transferred DNA strand.

R388 is a conjugative plasmid of broad host range that belongs to the IncW incompatibility group (4). The 15 kb transfer region can be separated into an Mpf (for Mating pair formation) region, which encodes the T4SS apparatus, and a Dtr (for DNA transfer and replication) region encoding the T4CP and the relaxosome (5). The latter is composed of an *oriT* of 330 bp length, the relaxase TrwC, and two accessory proteins, the plasmid-encoded TrwA and the host-encoded integration host-factor (IHF) (6). During conjugation, TrwC binds to the *oriT*, cleaves the DNA strand to be transferred at the *nic* site, and makes a covalent bond with its 5´end (7). Then the relaxase-DNA

82 complex is recruited by the T4CP to the T4SS and transported to the recipient cell,
83 where TrwC catalyzes the recircularization of the transferred DNA strand (8, 9).
84 Apart from its role in conjugation, TrwC is able to catalyze site-specific
85 recombination between two *oriT* copies repeated in tandem (10). The reaction takes
86 place in the absence of conjugation, and thus in the absence of single stranded
87 intermediates, and is favored by the accessory protein TrwA. In contrast, IHF was
88 found to exert a negative regulatory role in TrwC-mediated recombination (11). It was
89 proposed that recombination takes place thanks to the single-stranded endonuclease
90 activity of TrwC coupled to the replication machinery of the host cell (10).
91 Once transferred to the recipient cell during conjugation, TrwC can also
92 catalyze site-specific integration of the transferred DNA strand into an *oriT*-containing
93 plasmid in the recipient cells (8). In this case, both TrwA and IHF act as enhancers of
94 the reaction. Integration also occurs when the acceptor *oriT* was located in the
95 chromosomal DNA of the recipient cell (12). A minimal *oriT* core sequence of 17 bp is
96 enough for TrwC to achieve integration. Two human sequences with one single
97 mismatch to that minimal *oriT* were tested as acceptors for TrwC-mediated integration
98 and found to be functional with an efficiency only 2-3 times lower than that obtained
99 with the wild-type minimal *oriT*, indicating that TrwC can act on DNA sequences
100 present in the human genome (12).
101 In addition to the T4SS involved in conjugative DNA transfer, there is another
102 family of T4SS implicated in the secretion of effector proteins during the infection
103 process of several mammalian and plant pathogens (13). Substrate recruitment by
104 T4SS relies on secretion signals present in the protein substrate, and there are several
105 examples of heterologous protein translocation by T4SS upon addition of secretion

106 signals. In particular, conjugative relaxases can be translocated into eukaryotic cells
107 through T4SS of bacterial pathogens, either unmodified due to some similarity in their
108 C-termini with the secretion signal of the specific T4SS - as reported for translocation
109 of MobA of plasmid RSF1010 by the VirB T4SS of *Agrobacterium tumefaciens* (14) -, or
110 upon addition of the corresponding secretion signal, as done with TraA of plasmid
111 pATC58 and the VirB/D4 T4SS of *Bartonella henselae* (15). Moreover, two different
112 reports have demonstrated that relaxase-DNA complexes from two conjugative
113 systems can be translocated into human cells through the VirB/D4 T4SS of *B. henselae*.
114 Those studies reported relaxase-mediated transfer of bacterial plasmids containing the
115 *oriT* and conjugative genes from *Bartonella* cryptic plasmid pBGR1 (16) or from
116 conjugative plasmid R388 (17). For both relaxases, the addition of a BID domain, the
117 translocation signal for the *Bartonella* VirB/D4 T4SS (15), increases DNA transfer (16,
118 18). These reports suggest that trans-kingdom DNA transfer may naturally occur during
119 bacterial infection of human cells.
120 T4SS-mediated DNA transfer to human cells may have biotechnological
121 applications as a tool for in vivo DNA delivery into specific human cells (19). A main
122 concern in genetic modification protocols is the fate of the introduced foreign DNA in
123 the cells. Schröder and co-workers found that the relaxase-driven DNA integrated into
124 the human genome at low frequency, and characterized several integration sites
125 demonstrating that pBGR1 Mob relaxase can protect the 5´end of the mobilizable
126 plasmid, but no preference for specific integration sites could be identified, suggesting
127 random integration of the incoming DNA (16).
128 In contrast to Mob, TrwC has site-specific integrase activity in bacteria,
129 conferring added potential as a tool for genomic engineering (20). In this work, we

analyze TrwC integrase activity into human genomic DNA after the mobilization of TrwC-DNA complexes from *B. henselae*. We show evidence that TrwC is active in the human cell, although the efficiency of site-specific integration is negligible compared to random integration. Interestingly, we find that TrwC promotes a 100-fold increase in the efficiency of integration of the incoming DNA, suggesting it may be protecting DNA from degradation; this feature could be combined with the action of a site-specific nuclease for genomic engineering purposes.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *Escherichia coli* strain D1210 (21) was used for DNA manipulations, while strain β2163 (22) was used as donor for conjugative matings to *B. henselae*. *E. coli* strains were grown at 37ºC in Luria-Bertani broth, supplemented with agar for growth on plates. *B. henselae* strain RSE247 (23) was used for the infection of human cells. *B. henselae* was grown on Columbia blood agar (CBA) plates at 37°C under 5% CO_2 atmosphere. For selection, antibiotics were added at the indicated concentrations: ampicillin (Ap), 100 μg/ml; kanamycin monosulphate (Km), 50 μg/ml; streptomycin (Sm), 300 μg/ml (*E. coli*) or 100 μg/ml (*B. henselae*); gentamicin sulphate (Gm), 10 μg/ml. When needed, media were supplemented with diaminopimelic acid (DAP) at 0.3 mM.

Plasmids. Bacterial plasmids are listed in **Table 1**. Plasmids were constructed using standard methodological techniques (24). Primers used in plasmid constructions are listed in **Table 2**. Plasmids pCOR31, 33, and 35 were constructed by cloning a neomycin resistance cassette amplified from pRS56 with primers adding ClaI restriction sites (Table 2) into the same site of pHP159, pLA24, and pHP181, respectively. Plasmids pMTX708 and 709 were constructed by cloning a Ptac-*oriT* cassette into the NotI site of pTRE2hyg vector, selecting both orientations; Ptac-*oriT* was amplified from plasmid pOD1, which carries an EcoRI-HindIII fragment from pSU1186 (25) containing R388 *oriT* into the same sites of expression vector pKK223-3 (Pharmacia). Restriction enzymes, shrimp alkaline phosphatase, and T4 DNA ligase were purchased from Thermo-Fisher Scientific. Kapa HiFi DNA polymerase was purchased from Kapa Biosystems. Plasmid

DNA was extracted using GenElute Plasmid Miniprep Kit (Sigma Aldrich). DNA sequence of all cloned PCR fragments was determined.

Mating assays. Plasmids were routinely introduced in *B. henselae* by conjugation. *E. coli* donor strain was grown in LB to stationary phase. 200 µl were collected for each mating and resuspended in 1 ml of PBS. Recipient *B. henselae* was grown in CBA plates for 3-4 days. After that time, bacteria from half of the plate were collected with a cotton swab and resuspended in 1 ml of PBS. Both donor and recipient aliquots were centrifuged, pellets were resuspended in 20 µl of PBS, mixed, and the mixture was placed on a cellulose acetate filter on a CBA plate supplemented with DAP. The mating plate was incubated at 37ºC in a 5 % CO_2 atmosphere during 6 h. Transconjugants were selected by recovering the mating mixture and streaking it on a CBA plate with appropriate antibiotics. The plate was incubated for 6-9 days at 37ºC in a 5 % CO_2 atmosphere.

Cell lines and growth conditions. Human cell lines used in this work were immortalized hybridoma EA.hy926 (ATCC CRL-2922), a fusion cell line of human umbilical vein endothelial cells (HUVEC) and adenocarcinomic human alveolar basal epithelial cells (A549), and HeLa (ATCC CCL-2), epithelial cells of cervix adenocarcinoma. HeLa cells containing an integrated copy of the R388 *oriT* were created by transfection of plasmids pMTX708/9 (Table 1) and selection of stable transfectants as explained in the next section.

Cell lines were routinely grown in DMEM medium (Lonza) supplemented with FBS 10 % (Lonza) at 37ºC under 5 % CO_2. When indicated, antibiotics were added to the medium at the following concentrations: G418 disulfate salt (Sigma Aldrich), 500 µg/ml; hygromycin B (Invitrogen), 80 µg/ml; penicillin-streptomycin 1 % (Lonza).

186 **Transfections.** HeLa cells were transfected with the cationic JetPei transfection
187 reagent (Polyplus Transfection). The amounts of DNA and JetPei reagent were adjusted
188 depending on the cell culture format used, following the manufacturer´s instructions.
189 DNA was quantified using a Nano-Drop Spectrophotometer ND-1000 (Thermo
190 Scientific). To generate stably transfected cell lines, HeLa cells were allowed to grow
191 and to express the drug resistance gene under non-selective conditions for 24-48
192 hours after transfection. Then, cells were cultivated in standard medium
193 supplemented with the appropriate drug during 4-5 weeks, until outgrowth of
194 resistant cells. Medium was changed every 2-3 days to avoid loss of selection pressure.
195 To obtain the integration rate of transfected plasmid DNA, transfections were carried
196 out in 6-well plates. To transfect linearized DNA, plasmid DNA was digested with
197 Alw44I (Thermo Scientific) and purified with GeneJet Gel Extraction kit (Thermo
198 Scientific) prior to transfection.
199 **Cell infections.** *B. henselae* containing the appropriate plasmids were grown on
200 CBA plates for 3-4 days. Human cells were seeded the day before the infection. For
201 routine infections, cells were seeded in 6-well plates (80,000 cells per well) in 3 ml of
202 medium. When the purpose of the infection was to select human cells that had stably
203 acquired the plasmid transferred from *B. henselae*, infections were performed in 150
204 mm tissue culture dishes seeded with $1.2x10^6$ cells in 20 ml of medium.
205 The day of infection, DMEM was replaced by M199 medium (Gibco)
206 supplemented with FBS 10 % and appropriate antibiotics to select for the *B. henselae*
207 strains to be added. The bacteria were recovered from the CBA plate and resuspended
208 in 1 ml of PBS. The number of bacteria was calculated considering that an OD_{600}=1
209 corresponds to 10^9 bacteria/ml (26). Bacteria were added to the human cells to get a

multiplicity of infection (MOI) of 400. The mixture of human cells and bacteria was incubated for 72 hours at 37ºC under 5 % CO_2.

Detection of GFP-positive cells. At 72 hours post infection (hpi), infected cells were washed with PBS, trypsinized, and analyzed by flow cytometry using a Cytomics FC500 flow cytometer (Beckman Coulter). Uninfected cells were always used in parallel to set the baseline for detection of GFP-positive cells

Selection of stable integration events. At 72 hpi, G418 disulfate salt (Sigma Aldrich) was added to infected cells, and selection was maintained for 4-5 weeks. Resistant colonies were counted on the plates. G418-resistant cell pools were collected for further analysis of GFP expression and PCR analysis. Genomic high molecular weight DNA was extracted using High Pure PCR Template Preparation Kit (Roche).

Linear amplification mediated PCR. Amplification of genomic integration sites by linear amplification mediated PCR (LAM-PCR) was performed as described in (27). Briefly, it consists of an initial linear amplification of genome-plasmid junctions with a plasmid-specific primer. After synthesis of dsDNA, the PCR product is cut with a restriction enzyme (BfaI or Tsp509I) and a linker cassette of known sequence is ligated. Exponential PCR amplifications are then performed with plasmid- and linker-specific primers. PCR-obtained bands are then analyzed by gel electrophoresis and high-throughput sequencing. Human genomic DNA from human blood (buffy coat; Roche) was analyzed in parallel as negative control.

LAM-PCR template was genomic DNA from the G418-resistant pools. PCR reactions were carried out using Taq DNA polymerase (Genaxxon Bioscience). HPLC-purified primers (Table 2) were designed using Primer3Plus software and ordered from

233 Eurofins Genomics. Details on the primers can be found in **supplemental Materials**
234 **and Methods and Fig. S1**.

235 **High-throughput sequencing of LAM-PCR products.** DNA sequence of purified
236 LAM-PCR products was determined using MiSeq Benchtop next generation sequencing
237 technology (Illumina). The appropriate volumes of different purified samples were
238 mixed together following the manufacturer´s recommendations. Primers used in the
239 second exponential amplification contained the adaptor sequences needed for the
240 sequencing reaction (PE-PCR 1.0 and 2.0, see Table 2). LAM-PCR products were
241 sequenced in both directions. From PE-PCR 1.0 (adaptor present in the primer
242 annealing to the plasmid sequence) 400 nt were sequenced, while only 50 nt were
243 sequenced from PE-PCR 2.0 (adaptor present in the primer annealing to the linker).
244 Information from PE-PCR 1.0 was used for sorting the sequences to the different
245 samples and integration site detection, while information obtained from PE-PCR 2.0
246 was used only for sorting.

247 Bioinformatic analysis to obtain the integration sites was performed by high-
248 throughput insertion site analysis pipeline (HISAP) (28). Briefly, sequences were
249 trimmed by identification and removal of plasmid- and linker-specific sequences.
250 Genomic sequences were aligned to the human genome using stand-alone BLAT
251 (UCSC), using assembly GRCh37/hg19 as reference. Sequences with identities lower
252 than 95% were discarded. For each remaining sequence, the chromosome, the
253 integration site, and the nearest RefSeq protein-coding gene were recorded.

254 **Detection of *oriT*-specific integration events by PCR.** PCR reactions were
255 carried out using Kapa Taq Polymerase (Kapa Biosystems) following manufacturer´s
256 recommendations. 25 ng of plasmid DNA or 250 ng of genomic samples were used as

template. To detect the expected cointegrate molecule, primers NotI_Ptac and Int_pCOR (Table 2) were used for initial amplification. A 1:50 dilution of the initial PCR products served as template for the secondary PCR, carried out with primers NotI_Ptac_2 and Int_pCOR_2 (Table 2), annealing approximately 80 bp closer to the expected integration junction. Primers NotI_Ptac and NotI_oriT1 (Table 2) were used to amplify the chromosomal Ptac-*oriT* cassette.

Recovery of Integrated Plasmids. 5 µg of genomic DNA from G418-resistant cell pools were digested with XmaJI (Thermo-Fisher Scientific), which does not cleave within the integrated plasmid. Digested DNA was treated with T4 DNA ligase at a DNA concentration of 10 µg/ml, to favour self-ligation. The reaction was electroporated into ElectroMAX DH10B *E. coli* cells (Thermo-Fisher Scientific). Plasmid DNA was extracted from gentamicin-resistant *E. coli* transformants, and analyzed by PCR to narrow down the region of the plasmid where the insert of human origin was located. Primers used for PCR mapping reactions are shown in Table 2. The insert in plasmid pCOR52 was sequenced with primers pCOR33_1641F and pCOR33_12445R (Table 2).

Statistical analysis. Unpaired student´s t-test was used to determine statistically significant differences between the mean of at least 3 independent results for each experiment when the data followed a normal distribution. Otherwise, a Wilcoxon-rank-sum analysis was performed for each pair of compared data.

RESULTS

Construction of mobilizable plasmids and target cell lines. TrwC-DNA complexes can be introduced in human cell lines through the T4SS of *B. henselae*. In order to analyze the integration pattern of the transferred DNA upon *Bartonella* infection of human cells, new mobilizable plasmids and cell lines were constructed. The mobilizable plasmids previously used to test DNA transfer from *B. henselae* to human cells (17) contained elements of the R388 Dtr region (*oriT*+*trwABC*), but not the genes of the T4SS, and a eukaryotic *gfp* expression cassette. We added a neomycin phosphotransferase eukaryotic expression cassette in order to be able to select for stable chromosomal integration events. Plasmids were constructed coding for either TrwC or TrwC:BID (TrwC with the secretion signal for *Bartonella* VirB/D4 T4SS fused to its C terminus), and a negative control lacking *trwC*. Plasmid pRS130 (16) encoding Mob:BID relaxase and its cognate *oriT* was always tested in parallel.

The cell lines used for *Bartonella* infections were EA.hy926 and HeLa. The former is derived from fusion of A549 lung carcinoma cells with human vascular endothelial cells, the latter representing the natural target for *Bartonella*, and is efficiently infected by this bacterium (29). HeLa cells represent a cervix-derived epithelial cells line that can be easily manipulated by cell biological and genetic methods and infection by *B. henselae* was reported to occur with 50% efficiency (30). We previously showed DNA transfer to EA.hy cells, but HeLa cells were not tested. EA.hy926 and HeLa cells were tested in parallel in infections with *B. henselae* carrying either pHP161 (*oriT*+*trwABC*) or pHP181 (*oriT*+*trwAB*). DNA transfer efficiency was

lower when using HeLa than when using EA.hy926 cells, but it can be detected robustly in both cell lines (**Fig. 1**).

In order to compare frequencies of TrwC-mediated integration into natural sequences of the human genome with integration when the TrwC target is present in the recipient cell genome, a cell line containing a full length wild-type *oriT* was constructed. We transfected both EA.hy926 and HeLa cells with plasmids pMTX708 and pMTX709 (Table 1), carrying a hygromycin-resistance gene and the R388 *oriT* in both orientations. Both plasmids were used to avoid any bias due to eukaryotic promoters present in the vector, since transcription through the *oriT* has been shown to affect TrwC-mediated recombination (11), and so it could affect integration. Around 100 hygromycin-resistant colonies were obtained in transfections of HeLa cells, while no transfectants appeared for EA.hy926 in spite of several attempts with up to 5 µg of plasmid DNA. Consistent with this finding EA.hy926 cell line has been previously reported to be difficult to transfect (30, 31).

The hygromycin-resistant HeLa colonies obtained were pooled together to establish a polyclonal HeLa::*oriT* cell line, in which the *oriT* is expected to be located in different chromosomal locations and in the two possible orientations with respect to the vector promoter. In this way, we avoid selecting a single clone in which the *oriT* copy may lie in a chromosomal region that could affect integration of the mobilizable plasmid. The presence of the *oriT* was tested by PCR on genomic DNA samples from HeLa and HeLa::*oriT* cells using primers oriT1 and oriT330 (Table 2). Only one band corresponding to the *oriT* was present in the sample obtained from HeLa::*oriT* cells, while no amplification was detected in the sample obtained from unmodified cells (**Fig. S2**).

Transient and permanent expression of transferred DNA in human cells. In order to measure transfer and integration rates of DNA molecules led by different relaxases into human cells, *gfp* and neo-resistance gene expression were measured respectively, as outlined in **Fig. 2a.** EA.hy926, HeLa and HeLa::*oriT* cell lines were infected with *B. henselae* carrying pCOR31 (*trwC*), pCOR33 (*trwC:BID*), pCOR35 (*ΔtrwC*) or pRS130 (*mob:BID*) mobilizable plasmids. Results are shown in Fig. 2 and **Table S2**. Three days post infection, *gfp* expression was measured by flow cytometry (Fig. 2b). DNA transfer occurred to the three different cell lines when a relaxase (TrwC, TrwC:BID, or Mob:BID) was coded in the plasmid, while no DNA transfer was detected when there was no relaxase. The transfer efficiency is higher when using EA.hy926 as host cell, as previously shown (Fig. 1). No significant differences were found in DNA transfer between HeLa and HeLa::*oriT* cell lines, as expected. In all cell lines, DNA transfer rate was significantly lower when using Mob:BID relaxase compared to TrwC or TrwC:BID.

Integration events of the transferred plasmids into the human genome were selected by antibiotic treatment with G418. The drug was added at 72 hpi and selection was maintained for 4-5 weeks. The resistant colonies obtained for each experimental condition were counted and then pooled together. High molecular weight genomic DNA preparations were analyzed by PCR for the presence of *trwA* and *trwC* to confirm the presence of the integrated plasmid (**Fig. S3**). The resistant cell pools were also analyzed by flow cytometry to detect GFP expression as another evidence of integration of the mobilizable plasmids (**Fig. S4**).

Fig. 2c shows the number of resistant colonies obtained after the antibiotic treatment, normalized to the number of cells at the beginning of the experiment. It

was lower when using EA.hy926 than when using HeLa cells, despite the fact that DNA transfer was up to 10-fold higher with EA.hy926. For all cell lines, no resistant colonies were found when using *⊿trwC* plasmid, in concordance with the flow cytometry results, which showed no DNA transfer in the absence of relaxase. When plasmids coded for a relaxase, resistant colonies appeared, but at drastically different rates. Thousands of resistant colonies were obtained in each experiment after mobilization of *trwC*- and *trwC:BID*-carrying plasmids, while only up to 100 resistant colonies were found in infections with *B. henselae* carrying the plasmid coding for Mob:BID (see Table S2).

Fig. 2d shows the ratio between Neo^R colonies and GFP^+ cells, which gives the integration rate, i.e. the proportion of cells receiving the DNA in the nucleus which integrate this DNA into the chromosome. For each cell line, no significant differences were found in the integration rate of TrwC or TrwC:BID plasmids, as expected. No significant differences were found either in the integration rate of each plasmid in HeLa and HeLa::*oriT* cells. The integration rate was higher than 1 in 20 when the transferred DNA was led by TrwC or TrwC:BID, while it was around 1 in 250 in the case of Mob:BID-driven DNA (Table S2).

With the purpose of having a parallel control of random integration, HeLa and HeLa::*oriT* cells were transfected with plasmid pCOR35 (*⊿trwC*). Transient vs. stable expression was determined as outlined before. After transfection of plasmid DNA, we obtained an integration rate of around 1 in 800 when transfecting supercoiled DNA and of close to 1 in 300 when transfecting linearized DNA (Table S2), which is in the range of aour data obtained for Mob-BID. Antibiotic resistant colonies were pooled

372 together and analyzed in parallel with those obtained after relaxase-mediated
373 mobilization of plasmid DNA, to compare both plasmid integration patterns.

374 **Characterization of genomic integration sites**. Relaxases transfer the DNA
375 strand covalently linked to a site known as the *nic* site. In the case of Mob-led DNA, it
376 has been suggested that the relaxase protects the 5´end of this DNA (16). In addition,
377 we know that TrwC acts as a site-specific integrase of the transferred DNA into the
378 genome of recipient bacteria (12), and we observed enhanced integration rate of
379 TrwC-led DNA. Taking together these evidences, we decided to search for integration
380 events occurring at the *nic* site of the R388 *oriT*. For this purpose, we used linear
381 amplification-mediated PCR (LAM-PCR) (32, 33) using a primer annealing close to the
382 *nic* site, as explained in Fig. S1. This strategy would not detect insertions into the full-
383 length *oriT* copy of HeLa::*oriT*, as integration would result in a reconstituted *oriT*, but it
384 would allow the identification of integration events in other chromosomal locations,
385 and comparison with the integration pattern obtained when the *oriT* is not present in
386 the genome to be modified.

387 LAM-PCR was performed as explained in Materials and Methods and
388 supplemental Material and Methods sections. Genomic DNA was extracted from pools
389 of several thousands of resistant colonies obtained after mobilization of *trwC*- or
390 *trwC:BID*-coding plasmids, and this DNA was used as template for the PCR reactions.
391 Genomic DNA was also extracted from resistant colonies obtained by transfection of
392 plasmids pCOR31 (*trwC*) and pCOR33 (*trwC:BID*), which are expected to have a random
393 integration pattern. After LAM-PCR amplification of the integration junctions, two
394 different restriction enzymes were used to avoid any bias due to restriction fragment
395 size. PCR products were checked by electrophoresis in agarose gels (**Fig. 3**).

Fig. 3a-c shows the scheme of the expected band sizes observed in these gels. We expected to see as many bands as different integration sites occurring at the *nic* site (Fig.3a), depending on the location of the nearest restriction site in the genomic junction. If integration did not occur at the *nic* site, the size of the band would be determined by the nearest recognition site in the integrated plasmid (220 bp when using BfaI and 580 bp when using Tsp509I, since this does not cut in the *oriT* but in trwA; Fig 3b). In the case of the HeLa::*oriT* cell line, which has an *oriT* copy integrated in the genome, we expected to obtain in all cases a major band of 220 bp (BfaI) or 345 bp (Tsp509I) corresponding to the sequence of the *oriT*-carrying integrated plasmid (Fig 3c).

As it can be observed in Fig. 3d, a single band was obtained from all samples, obtained either after plasmid transfection or after translocation of TrwC(:BID)-DNA molecules through *B. henselae* T4SS. The size of the band was in all cases the expected for the full length *oriT* present in the mobilizable plasmids or in the genomic *oriT* copy, as explained above and in Fig 3a-c. For the G418-resistant pools obtained either after transfection or infection of HeLa::*oriT* cells, both 345 bp and 580 bp would be visible when using Tsp509I, as observed in the sample obtained after infection and DNA transfer mediated by TrwC (Fig. 3d bottom gel, line 6). A reason for this not being the case for the other samples could be that the smaller amplicon could be preferentially amplified.

These results strongly suggest that the transferred DNA had not become integrated by the *nic* site at the *oriT*. Rather, they presumably reflect illegitimate integration events. Since low-frequency site-specific integration events could be masked by this main band, LAM-PCR products were thus analyzed by high-throughput

sequencing, as explained in Materials and Methods. After identification of linker- and plasmid-specific sequences, the flanked sequences were characterized. As expected from the results in Fig. 3d, most of the 2,000,000 reads obtained were found to be plasmid DNA. This confirms that most of the DNA entering the human cell covalently linked to TrwC is not integrated at the *nic* site, implying that this DNA is recircularized prior to integration.

There were 11,317 reads which could be mapped to the human genome. To discard false positives, identity to the human genome threshold was raised to 98% and integration events obtained less than 15 times were not considered. The resulting 9 integration events (IE) are shown in **Table 3**. IE1 and IE2 were found to occur at the same site of the human genome (the differences in the sequencing reads were assumed to be sequencing errors) so they were considered together as one integration event and named IE2. Most of the integration events showed more than 12 missing base pairs of a total of 41 bp amplified from primer oriTI binding site to the *nic* site, so they were considered as random integration events.

There were only two IE which were not missing any *oriT* sequence 3' to the *nic* site. When aligned with the human DNA sequence, it was found that integration in IE2 had occurred at the position *nic*+1, since this base from the *oriT* sequence was present at the junction and is not present in the UCSC genomic sequence used as reference. We confirmed the genomic sequence of this position in the genome of the HeLa cells used in the experiment, by amplification of the chromosomal region around the integration site (IS) IS2 with primers IS2_Hu11 and Xba_IS2_Hu11 (Table 2) and sequencing the PCR product with the former primer, and this base pair was not present there either. Considering the high specificity of conjugative relaxases for

444 nicking exactly at their *nic* site, this result suggests that this event was yet another
445 illegitimate integration event. Finally, IE7 occurred exactly at the *nic* site, and
446 moreover, the eight nucleotides of the human genome 5' to the integration site are
447 identical to the eight nucleotides 5' to the *nic* site in R388 *oriT* (**Fig. 4**). This integration
448 event took place 1,352,133 bp downstream of the SLITRK1 gene (NM_052910) in
449 human chromosome 13.

450 As LAM-PCR did not allow the detection of the integration events occurring at
451 the *oriT* copy present in the chromosome of HeLa::*oriT* cells, we tried to detect them
452 by PCR amplification of the expected cointegrate molecule; we used a primer
453 annealing in the Ptac promoter located adjacent to the chromosomal *oriT* copy, and
454 another one annealing in the mobilizable plasmid (Table 2). As a control, the
455 chromosomal Ptac-*oriT* cassette was amplified in the same samples analyzed; as
456 expected, the cassette was detected in HeLa::*oriT* and the G418-resistant pools
457 obtained with this cell line, while no amplification was obtained in HeLa and HeLa-
458 derived cell pools. The PCR to amplify the *oriT*-*oriT* cointegrate was negative (data not
459 shown), even after a second round of PCR amplification. Although we cannot discard
460 *oriT*-specific integration occurring at such low frequency that it is not detectable by
461 PCR, this results indicates that it is not occurring efficiently.

462 LAM-PCR can only be used to map those integration events that occurred by a
463 known sequence of the transferred DNA (the *nic* site in our case), but most of the
464 plasmid molecules became integrated in a *nic*-independent manner. Out of the 8
465 human genomic junctions obtained by LAM-PCR (Table 3), 7 did not occur by the *nic*
466 site, and so they represent random integration events. However, they provide
467 information on only one of the integration junctions of the plasmid. We attempted to

468 characterize other random integration sites by recovery of the integrated plasmids
469 together with the flanking genomic sequences, as outlined in Materials and Methods.
470 With this strategy, we were able to characterize one integration event and its
471 corresponding plasmid-genomic DNA junctions (**Fig. 5**). We determined that only a
472 fragment of the plasmid was integrated, which does not include the neomycin
473 resistance gene, so most probably there is another integration event somewhere else
474 in that same cell coding for the *neo*R gene. We also observed that both genomic-
475 plasmid DNA junctions did not occur at the same position of the human genome.
476 Moreover, near one of the junctions (IJ-B in Fig. 5b), a genomic rearrangement was
477 found, when compared to the reference genome (see coordinates in Fig. 5b). The
478 reason could be that the genomic region of chromosome 15 where integration
479 occurred corresponds to a copy of L1MC2, a long interspersed element (LINE), often
480 associated with genomic rearrangements and deletions (34). We tried to sequence
481 that region from genomic DNA of the HeLa cell line used in the experiment using
482 primer Chr15_88728 (Table 2), to determine if the rearrangement was already present
483 or it was a consequence of the illegitimate integration event, but mixed sequences
484 were always obtained.
485 The genomic integration sites of all random integration events characterized
486 were aligned with the R388 *nic* region at the *oriT* (**Table 4**). No homology with the *oriT*
487 was detected, supporting the idea of illegitimate integration.
488

DISCUSSION

The ability to deliver DNA into specific human cell types and to promote its integration in the human genome has high potential as biotechnological tool. In particular, gene therapy strategies ideally should grant in vivo access to specific human tissues and permanent expression of the introduced DNA. In this work, we explore the potential of a bacterial system for genomic modification of the human genome. Our previous work showed that the substrate of a conjugative plasmid, the TrwC-DNA complex, was delivered efficiently to human cells through the T4SS of *B. henselae* (17); the many advantages of such a DNA delivery system in this context have been already discussed (19). We previously showed in a bacterial system that TrwC could catalyze integration of the transferred DNA into DNA sequences of human origin (12), so we aimed to evaluate its potential role as a site-specific integrase in human cells, which would complement the DNA delivery tool. We have analyzed the fate of the DNA in the human cells after translocation as a TrwC-DNA complex through the VirB/D4 T4SS of *B. henselae*.

We measured the efficiency of DNA delivery and integration by the two relaxases previously described to deliver DNA through the T4SS VirB/D4 of *B. henselae*, TrwC and Mob:BID. All mobilizable plasmids carried a eukaryotic expression GFP cassette, allowing us to estimate the efficiency of DNA transfer by measuring the percentage of GFP positive cells. This assay probably underestimates the percentage of cells receiving DNA, since this DNA has to get into the nucleus and be converted into double-stranded form so that it can express the *GFP* gene. Thus, nuclear localization of the relaxase could affect DNA transfer rates. TrwC has been reported to have

513 cytoplasmic localization (35) while a passive entry of the Mob-guided DNA has been
514 suggested (16), so none of the relaxases is expected to have an active role in nuclear
515 import.

516 Our results show that DNA transfer is higher when using TrwC compared to
517 Mob:BID. The differences in DNA transfer rates are probably due to differences in T4SS
518 recruitment efficiency for each relaxase. TrwC could be naturally a better substrate for
519 the *B. henselae* VirB/D4 T4SS than Mob:BID. There is another important factor to take
520 into account: the mobilizable plasmids enconding *trwC* also code for R388 proteins
521 TrwA and TrwB, which could play a role in substrate recruitment. Deletion of *trwB* was
522 shown to affect the transfer of TrwC-DNA complexes significantly (17). Thus, it is likely
523 that TrwB enhances recruitment of TrwC by the VirB/D4 T4SS independently of BID, as
524 previously suggested (16, 18).

525 The plasmids mobilized to human cells carried a eukaryotic resistance marker
526 to select for stable integration events by antibiotic treatment. Selection was carried
527 out for four weeks, discarding the possibility of episomal persistence of the transferred
528 plasmid DNA. Each resistant colony was counted as one integration event. Again, this
529 measure is an underestimation of the integration rate. One single colony could harbor
530 more than one integration event; in fact, the only integration event mapped in its
531 extension (the rescued integrated plasmid) did not include the *neo*R region (Fig. 5a),
532 implying this gene must be integrated somewhere else in the genome. In addition, not
533 all cells integrating the plasmid will thrive to render a colony. This phenomenon was
534 particularly evident with Ea.hy926 cells, which have low viability. Consequently, we
535 obtained less resistant colonies when using EA.hy926 than when using sturdy HeLa
536 cells, despite the fact that DNA transfer was up to 10-fold higher with EA.hy926 (Fig.

537 2). Of course, we cannot rule out that the different integration rates observed in both
538 cell lines are due to intrinsic differences affecting host-mediated integration of foreign
539 DNA.

540 The number of integration events obtained for each experiment was 25-158
541 times higher when either TrwC or TrwC:BID were present, compared to Mob:BID
542 (Table S2). When we measured integration rates, as number of resistant colonies
543 normalized to the number of cells expressing the transferred DNA, we observed that
544 the integration rate for TrwC was on average 1 in 20, while for Mob:BID it went down
545 to about 1 in 250, similar to the integration rate obtained for transfected cells (Fig. 2d
546 and Table S2). Thus, we conclude that TrwC facilitates the integration of the
547 mobilizable plasmids, while Mob does not.

548 A plausible explanation for this difference could be the site-specific integration
549 activity of TrwC, which is presumably absent in the Mob relaxase. To test this
550 hypothesis, we analyzed the integration pattern in the human genome, searching for
551 TrwC-mediated site-specific integration events. Their signature would be the precise
552 integration of the R388 *nic* site into human DNA sequences resembling the natural
553 TrwC target. We analyzed genomic DNA of the resistant cell pools by LAM-PCR,
554 priming from the plasmid DNA into the *nic* site, and subsequent DNA sequencing. The
555 results showed the presence of intact *oriT* sequences in the vast majority of the
556 sequencing reads. One possible explanation could be the integration of plasmid
557 concatemers, as it happens in the integration of T-DNA mediated by *A. tumefaciens*
558 (36), but then *oriT*-host genome junctions would be detected at the end of the
559 concatemer. The DNA extraction kit isolates high molecular weight DNA (30-50 kb)
560 from mammalian cells, while bacterial cells are not lysed, discarding the possibility that

561 plasmid DNA of bacterial origin could be co-isolated. In addition, the absence of
562 transformants when using the same DNA preparations for plasmid rescue (a single
563 transformant was obtained, originated from a rescued integrated plasmid copy) rules
564 out the presence of episomal plasmid molecules. Thus, the most likely possibility is
565 that these reads represent random integration events of the plasmid, which would
566 have been recircularized previously, since it enters the human cell cut at the *nic* site
567 (where TrwC is covalently bound). Recircularization implies that TrwC is active in the
568 human recipient cell, mimicking its activity in the bacterial recipient cell during
569 conjugation (37).

570 Out of the thousands of different integration events present in the analyzed cell
571 pools, we detected one putative site-specific integration event (IE7). It occurred
572 precisely at the *nic* site and in a region of the genome showing 8 base pairs identity
573 with the *oriT* at the 5´ end of the *nic* site (Fig. 4). Since the probability of integration at
574 any position of the human genome is approximately $1/3x10^9$, and the probability of
575 integration occurring by the *nic* site is less than $1/1x10^4$ (the size of the integrated
576 plasmids is around 13 kb), the probability that this event occurred randomly is
577 negligible. From our results we infer that TrwC can act as a site-specific integrase in
578 human cells, but host-mediated random integration is at least 3-4 logs more efficient.
579 Thus, after TrwC-mediated recircularization of the DNA (as inferred from the presence
580 of full-length *oriTs*), most molecules would undergo non-homologous integration
581 events, as observed in the characterized integration sites (Table 4).

582 DNA can also be delivered into human cells by the relaxases Mob and *A.*
583 *tumefaciens* VirD2, and it integrates randomly in the genome (16, 38). It was proposed
584 that these relaxases do not play a role in the integration process, which is likely

585 mediated by the host machinery, but do protect the 5´end of the transferred DNA,
586 based on preservation of the 5´end region of the transferred DNA molecules (16, 38).
587 In our case, we found by LAM-PCR seven integration events occurring within 20 bp
588 from the *nic* site (Table 3). By chance, we would expect around 30 integration events
589 lying in a 20 nt region, from about 20,000 total integration events analyzed, so TrwC
590 does not seem to protect the 5' end of the transferred DNA, but rather to catalyse its
591 conversion to a circular form. Recircularized plasmid DNA will be a more resistant
592 molecular species, showing long-term presence in the nucleus, which could favour its
593 subsequent random integration by the host machinery.
594 From a biotechnological point of view, our results indicate that TrwC is not
595 useful as a site-specific integrase in human cells. However, with the introduction of
596 precision genome editing using RNA-guided endonucleases, such as Cas9 (39), we have
597 entered a new era of genetic engineering and gene therapy which is leaving obsolete
598 the traditional site-specific recombinases and nucleases used for gene targeting in
599 human cells (40). In this new scenario, an improvement in CRISPR-Cas technology
600 would have an immediate impact in the human gene editing field. An RNA-guided
601 nuclease could be translocated simultaneously with TrwC-DNA through the T4SS of
602 bacteria that infect specific human cell types. Delivery of the nuclease protein instead
603 of transfecting the gene could avoid toxicity and off-target activity. The effect of TrwC
604 as DNA chaperone in combination with a site-specific nuclease would promote
605 integration of the incoming DNA molecule by homologous recombination. In support
606 of this approach, it has been reported that concomitant translocation of I-SceI homing
607 site-specific endonuclease together with VirD2 relaxase-T-DNA complexes through *A.*

608 *tumefaciens* T4SS enhanced T-DNA site-specific integration into the yeast chromosome

609 when the I-SceI target site was present (41).

610

AKNOWLEDGMENTS

We are grateful to Anabel Alperi for her assessment on HeLa infection by *B. henselae*. CGP wants to thank members of the Schmidt lab for their help with LAM-PCR.

FUNDING INFORMATION

This work was supported by grant BIO2013-46414-P from the Spanish Ministry of Economy and Competitiveness to ML. CGP was a recipient of a predoctoral fellowship from the University of Cantabria, Spain. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. **de la Cruz F, Davies J.** 2000. Horizontal gene transfer and the origin of species: lessons from bacteria. Trends Microbiol **8:**128-133.
2. **Zechner EL, de la Cruz F, Eisenbrant R, Grahn AM, Koraimann G, Lanka E, Muth G, Pansegrau W, Thomas CM, Wilkins BM, Zatyka M.** 2000. The horizontal gene pool: Bacterial plasmids and gene spread. Harwood Academic Publishers, London.
3. **Llosa M, de la Cruz F.** 2005. Bacterial conjugation: a potential tool for genomic engineering. Res Microbiol **156:**1-6.
4. **Garcillán-Barcia MP, Francia MV, de la Cruz F.** 2009. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev **33:**657-687.
5. **Fernandez-Lopez R, Garcillan-Barcia MP, Revilla C, Lazaro M, Vielva L, de la Cruz F.** 2006. Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. FEMS Microbiol Rev **30:**942-966.
6. **Moncalián G, Valle M, Valpuesta JM, de la Cruz F.** 1999b. IHF protein inhibits cleavage but not assembly of plasmid R388 relaxosomes. Mol Microbiol **31:**1643-1652.
7. **Llosa M, Grandoso G, de la Cruz F.** 1995. Nicking activity of TrwC directed against the origin of transfer of the IncW plasmid R388. J Mol Biol **246:**54-62.
8. **Draper O, Cesar CE, Machon C, de la Cruz F, Llosa M.** 2005. Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. Proc Natl Acad Sci U S A **102:**16385-16390.

9. **Gonzalez-Perez B, Lucas M, Cooke LA, Vyle JS, de la Cruz F, Moncalian G.** 2007. Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. EMBO J **26:**3847-3857.

10. **César CE, Machón C, de la Cruz F, Llosa M.** 2006. A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences. Mol Microbiol **62:**984-996.

11. **César CE, Llosa M.** 2007. TrwC-mediated site-specific recombination is controlled by host factors altering local DNA topology. J Bacteriol **189:**9037-9043.

12. **Agundez L, Gonzalez-Prieto C, Machon C, Llosa M.** 2012. Site-specific integration of foreign DNA into minimal bacterial and human target sequences mediated by a conjugative relaxase. PLoS One **7:**e31047.

13. **Llosa M, Roy C, Dehio C.** 2009. Bacterial type IV secretion systems in human disease. Mol Microbiol **73:**141-151.

14. **Vergunst AC, van Lier MC, den Dulk-Ras A, Stuve TA, Ouwehand A, Hooykaas PJ.** 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.

15. **Schulein R, Guye P, Rhomberg TA, Schmid MC, Schroder G, Vergunst AC, Carena I, Dehio C.** 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.

16. **Schroder G, Schuelein R, Quebatte M, Dehio C.** 2011. Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen Bartonella henselae. Proc Natl Acad Sci U S A **108:**14643-14648.

17. **Fernandez-Gonzalez E, de Paz HD, Alperi A, Agundez L, Faustmann M, Sangari FJ, Dehio C, Llosa M.** 2011. Transfer of R388 derivatives by a pathogenesis-associated type IV secretion system into both bacteria and human cells. J Bacteriol **193:**6257-6265.

18. **Alperi A, Larrea D, Fernandez-Gonzalez E, Dehio C, Zechner EL, Llosa M.** 2013. A translocation motif in relaxase TrwC specifically affects recruitment by its conjugative type IV secretion system. J Bacteriol **195:**4999-5006.

19. **Llosa M, Schroder G, Dehio C.** 2012. New perspectives into bacterial DNA transfer to human cells. Trends Microbiol **20:**355-359.

20. **Gonzalez-Prieto C, Agundez L, Linden RM, Llosa M.** 2013. HUH site-specific recombinases for targeted modification of the human genome. Trends Biotechnol **31:**305-312.

21. **Sadler JR, Tecklenburg M, Betz JL.** 1980. Plasmids containing many tandem copies of a synthetic lactose operator. Gene **8:**279-300.

22. **Demarre G, Guerout AM, Matsumoto-Mashimo C, Rowe-Magnus DA, Marliere P, Mazel D.** 2005. A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalpha) conjugative machineries and their cognate Escherichia coli host strains. Res Microbiol **156:**245-255.

23. **Schmid MC, Schulein R, Dehio M, Denecker G, Carena I, Dehio C.** 2004. The VirB type IV secretion system of Bartonella henselae mediates invasion,

693 proinflammatory activation and antiapoptotic protection of endothelial cells.
694 Mol Microbiol **52:**81-92.
695 24. **Sambrook J, Russell DW.** 2001. Molecular cloning : a laboratory manual, 3rd
696 ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
697 25. **Llosa M, Bolland S, de la Cruz F.** 1991. Structural and functional analysis of the
698 origin of conjugal transfer of the broad-host-range IncW plasmid R388 and
699 comparison with the related IncN plasmid R46. Mol Gen Genet **226:**473-483.
700 26. **Kirby JE, Nekorchuk DM.** 2002. Bartonella-associated endothelial proliferation
701 depends on inhibition of apoptosis. Proc Natl Acad Sci U S A **99:**4656-4661.
702 27. **Gabriel R, Kutschera I, Bartholomae CC, von Kalle C, Schmidt M.** 2014. Linear
703 amplification mediated PCR--localization of genetic elements and
704 characterization of unknown flanking DNA. J Vis Exp
705 doi:10.3791/51543**:**e51543.
706 28. **Arens A, Appelt JU, Bartholomae CC, Gabriel R, Paruzynski A, Gustafson D,**
707 **Cartier N, Aubourg P, Deichmann A, Glimm H, von Kalle C, Schmidt M.** 2012.
708 Bioinformatic clonality analysis of next-generation sequencing-derived viral
709 vector integration sites. Hum Gene Ther Methods **23:**111-118.
710 29. **Rhomberg TA, Truttmann MC, Guye P, Ellner Y, Dehio C.** 2009. A translocated
711 protein of Bartonella henselae interferes with endocytic uptake of individual
712 bacteria and triggers uptake of large bacterial aggregates via the invasome. Cell
713 Microbiol **11:**927-945.
714 30. **Truttmann MC, Rhomberg TA, Dehio C.** 2011. Combined action of the type IV
715 secretion effector proteins BepC and BepF promotes invasome formation of

716 Bartonella henselae on endothelial and epithelial cells. Cell Microbiol **13:**284-
717 299.
718 31. **Hunt MA, Currie MJ, Robinson BA, Dachs GU.** 2010. Optimizing transfection of
719 primary human umbilical vein endothelial cells using commercially available
720 chemical transfection reagents. J Biomol Tech **21:**66-72.
721 32. **Gabriel R, Eckenberg R, Paruzynski A, Bartholomae CC, Nowrouzi A, Arens A,**
722 **Howe SJ, Recchia A, Cattoglio C, Wang W, Faber K, Schwarzwaelder K, Kirsten**
723 **R, Deichmann A, Ball CR, Balaggan KS, Yanez-Munoz RJ, Ali RR, Gaspar HB,**
724 **Biasco L, Aiuti A, Cesana D, Montini E, Naldini L, Cohen-Haguenauer O,**
725 **Mavilio F, Thrasher AJ, Glimm H, von Kalle C, Saurin W, Schmidt M.** 2009.
726 Comprehensive genomic access to vector integration in clinical gene therapy.
727 Nat Med **15:**1431-1436.
728 33. **Schmidt M, Schwarzwaelder K, Bartholomae C, Zaoui K, Ball C, Pilz I, Braun S,**
729 **Glimm H, von Kalle C.** 2007. High-resolution insertion-site analysis by linear
730 amplification-mediated PCR (LAM-PCR). Nat Methods **4:**1051-1057.
731 34. **Han K, Lee J, Meyer TJ, Remedios P, Goodwin L, Batzer MA.** 2008. L1
732 recombination-associated deletions generate human genomic variation. Proc
733 Natl Acad Sci U S A **105:**19366-19371.
734 35. **Agundez L, Machon C, Cesar CE, Rosa-Garrido M, Delgado MD, Llosa M.** 2011.
735 Nuclear targeting of a bacterial integrase that mediates site-specific
736 recombination between bacterial and human target sequences. Appl Environ
737 Microbiol **77:**201-210.

36. **Sullivan TD, Rooney PJ, Klein BS.** 2002. Agrobacterium tumefaciens integrates transfer DNA into single chromosomal sites of dimorphic fungi and yields homokaryotic progeny from multinucleate yeast. Eukaryot Cell **1:**895-905.

37. **Garcillan-Barcia MP, Jurado P, Gonzalez-Perez B, Moncalian G, Fernandez LA, de la Cruz F.** 2007. Conjugative transfer can be inhibited by blocking relaxase activity within recipient cells with intrabodies. Mol Microbiol **63:**404-416.

38. **Kunik T, Tzfira T, Kapulnik Y, Gafni Y, Dingwall C, Citovsky V.** 2001. Genetic transformation of HeLa cells by Agrobacterium. Proc Natl Acad Sci U S A **98:**1871-1876.

39. **Hsu PD, Lander ES, Zhang F.** 2014. Development and applications of CRISPR-Cas9 for genome engineering. Cell **157:**1262-1278.

40. **Maggio I, Goncalves MA.** 2015. Genome editing at the crossroads of delivery, specificity, and fidelity. Trends Biotechnol **33:**280-291.

41. **Rolloos M, Hooykaas PJ, van der Zaal BJ.** 2015. Enhanced targeted integration mediated by translocated I-SceI during the Agrobacterium mediated transformation of yeast. Sci Rep **5:**8345.

FIGURE LEGENDS

Fig. 1. DNA transfer to EA.hy926 and HeLa cell lines. The graph shows the percentage of GFP positive cells detected after 3 days of infection. The cell lines indicated in the x axis were infected with *B. henselae* carrying the mobilizable plasmids pHP181 (containing R388 *oriT+trwAB)* or pHP161 (coding for R388 *oriT+trwABC*)*;* this is indicated as AB or ABC, respectively. Data are the mean of at least 5 independent experiments. **, p<0.01.

Fig. 2. Transient and permanent expression of the transferred DNA. a) Overview of the experimental design to detect transient expression or stable integration of the transferred DNA. After infection of human cell lines with *B. henselae*, the DNA transferred through the T4SS will get to the nucleus where genes will be expressed. At 3 days post infection, transient expression of *gfp* can be detected by flow cytometry. Antibiotic treatment was applied for long-term selection of neomycin-resistant colonies, to detect stable integration events. **b)**, **c)** and **d)** Graphical representation of the percentage of GFP-positive cells obtained 3 days post infection (b) and the number of G418-resistant colonies normalized for the number of cells at the beginning of the selection (c), as well as the Neo^R/GFP^+ ratio (d). The different bars represented for each cell line correspond to the different relaxases under study, following the color code indicated in the squares at the top right. Data represent the mean of at least 3 independent experiments. *, p<0,05.

Fig. 3. Analysis of LAM-PCR products. a-c), Scheme of the expected integration events and the subsequent LAM-PCR products. **a)** If integration takes place by the 5´

779 end of the *nic* site, the size of the LAM-PCR would be determined by the distance to
780 the nearest restriction site in the human genome (in pink). Each integration event
781 occurring in a different locus would generate a band of a different size. The nicked *oriT*
782 is indicated by a slash. **b)** If the plasmid becomes integrated at any other region than
783 the *nic* site, the size of the LAM-PCR product would be always the same and would be
784 determined by the distance to the restriction site in the plasmid sequence. **c)** In
785 HeLa::*oriT* cell line, in addition to the bands generated from the integration events, the
786 *oriT* copy present in pMTX708/9 plasmid generates a single band of a size determined
787 by the distance to the restriction site in the plasmid sequence. *trw* has been omitted
788 from *trwA*, *trwB*, and *trwC* for clarity. Bfa, BfaI. Tsp, Tsp509I. **d)** Gel electrophoresis of
789 LAM-PCR products obtained when using BfaI (top gel) or Tsp509I (bottom gel)
790 restriction enzymes. The cell line is indicated in the top row (EA, EA.hy926; He, HeLa;
791 He:*oriT*, HeLa::*oriT*). Inf, samples obtained after *Bartonella* infection. Tr, samples
792 obtained by transfection of plasmid DNA. LD, 100 bp ladder. T, *trwC*-coding plasmid
793 (pCOR31). T:B, *trwC:BID*-coding plasmid (pCOR33). EA, He, and He::*oriT*, samples from
794 uninfected cell lines. g, Human genomic DNA (Roche), used as negative control. -
795 (1,2,3), negative controls (no DNA) of linear, first, and second exponential PCRs,
796 respectively. The arrows indicate the bands of the expected size according to Fig. 3b
797 (black arrows) and 3c (blue arrow).

798 **Fig. 4. Characterization of integration event IE7.** The genome-plasmid
799 integration junction (IJ) is aligned with the DNA sequence around the *nic* site (*oriT*, on
800 top) and the chromosomal integration site (IS, on bottom). DNA of plasmid origin is
801 shown in blue, and genomic DNA is shown in black. The *nic* site and insertion sites are

802 indicated by a dash. Regions of homology between the plasmid and the genomic
803 sequences are boxed.

804 **Fig. 5. Genomic integration event characterized by recovery of the integrated**
805 **plasmid. a)** Scheme of the mobilizable plasmid coding for *trwC:BID*, and the structure
806 of the integrant in the genomic DNA of HeLa cells. Plasmid DNA is represented as a
807 blue horizontal line, and HeLa genomic DNA, as a black dashed line. *trw* is ommited
808 from *trwA*, *trwB* and *trwC:BID* for clarity. The two dashes in the integrant refer to the
809 genomic reorganization shown in b). Both plasmid-genomic DNA integration junctions
810 resulting from the integration event are named IJ-A and -B. The DNA sequence at the
811 junctions is shown below with their respective coordinates, in black (human genome)
812 and blue (plasmid DNA). The junctions are highlighted in a square. Coordinates of
813 human chromosome 15 are indicated as C-number, where the number corresponds to
814 the coordinates of the human genome in UCSC database (assembly GRCh37/hg19).
815 Coordinates of plasmid DNA are indicated as P-number, where the number
816 corresponds to the nucleotide of the open reading frame of *gfp* (in IJ-A) or *trwA* (in IJ-
817 B). **b)** Genomic rearrangement found near IJ-A. The red dash indicates the genomic
818 junction between non-adjacent human DNA sequences.
819
820

821 **TABLES**

822 **Table 1. Plasmids used in this work**

823

Plasmid	Description	Reference
pCOR31	pBBR6::*oriT trwABC*+*gfp*+*neo*	This work
pCOR33	pBBR6::*oriT trwABC:BID*+*gfp*+*neo*	This work
pCOR35	pBBR6::*oriT trwAB*+*gfp*+*neo*	This work
pCOR52	Rescued integrated plasmid	This work
pHP159	pBBR6::*oriT trwABC*+*gfp*[(a)]	(17)
pHP161	pBBR6::*oriT trwABC*+*gfp*[(a)]	(17)
pHP181	pBBR6::*oriT trwAB*+*gfp*	(17)
pKK223-3	Expression vector	Pharmacia
pLA24	pBBR6::*oriT trwABC:BID*+*gfp*	(17)
pMTX708	pTRE2hyg::Ptac-*oriT*[(b)]	This work
pMTX709	pTRE2hyg::Ptac-*oriT*[(b)]	This work
pOD1	pKK223-3::*oriT*	This work
pRS56	Cre-lox+*neo*	(15)
pRS130	pBGR::*mob:BID*+*gfp*+*neo*	(16)
pSU1186	pUC8::*oriT*	(25)
pTRE2hyg	Mammalian shuttle vector	Clontech

[(a)] pHP159 and pHP161 differ only in the orientation of the *gfp* cassette, which is in the same orientation as the Plac promoter in pHP161 and in the opposite in pHP159.

[(b)] pMTX708/9 differ only in the orientation of the Ptac-*oriT* cassette. In pMTX708, the *oriT* is closer to the hygromycin resistance gene.

824 **Table 2. Oligonucleotides used in this work**

825

Purpose / name	**Sequence (5´to 3´)**
Construction of pCOR31, 33, and 35	
mCla_SnaBI_CMV_NeoF	CCAAATCGATCTACGTATTAGTCATCGCTATT
Cla_EcoRV_NeoR	CCAAATCGATGATATCCGGATATAGTTCC
Construction of pMTX708/9	
NotI_Ptac	CCAGCGGCCGCTTATCGACTGCACGG
NotI_oriT1	CCAGCGGCCGCTCATTTTCTGCATCATTGT
Detection of *oriT*-specific integration events	
Int_pCOR	TCAGGGCGTCCGTTTC
Int_pCOR_2	CTGCATCACATTTGCATC
NotI_Ptac	CCAGCGGCCGCTTATCGACTGCACGG
NotI_Ptac_2	CACTGCATAATTCGTGTC
NotI_oriT1	CCAGCGGCCGCTCATTTTCTGCATCATTGT
PCR mapping of inserts in recovered integrated plasmids	
pCOR33_121F	TGGACAACCCTGCTGGAC
pCOR33_644R	TTTCGCCCTATATCTAGTTC
pCOR33_1641F	CTCGACCTGAATGGAAGCC
pCOR33_2158R	AGCTGGCGTAATAGCGAAG
pCOR33_3157F	CGCAACCCCTTGTAAATGC
pCOR33_3664R	TCTGAACGGCGGTAATCC
pCOR33_10431F	CCTGGCTGACCGCCCAA
pCOR33_10940R	GCTTCTAGAGATCTGACGG
pCOR33_11927F	TCAGGTTCAGGGGGAGGT
pCOR33_12445R	AATACGCAAACCGCCTCTC
Detection of *oriT* in HeLa::*oriT*	
oriT1	CTCATTTTCTGCATCATCA
oriT330	CCTCTCCCGTAGTGTTA
Analysis of G418-resistant cell pools	
670_TrwC	TGTGTGCTAGGTCGAA
BamHI_TrwA_R	AACAGGATCCTCAATCCTCCTTCCCCTCCC

Hind3_TrwA_F	AACAAAGCTTATGGCACTAGGCGACCCC
Hind3_TrwC_F	AACAAAGCTTATGCTCAGTCACATGGTATT
LAM-PCR and high-throughput sequencing	
LC1	GACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG(N)$_{12}$CTA(RO)[a]
LC2	(RO)TAG(N)$_{12}$CCTAACTGCTGTGCCACTGAATTCAGATC[a]
LCI	GACCCGGGAGATCTGAATTC
Mis-LC	(PE-PCR 2.0)AGTGGCACAGCAGTTAGG[b]
Mis-TrwC	(PE-PCR 1.0)(N)$_{10}$CGTCCTTAAAAGCCGGGTTG[c]
oriTI	CGATAACCCAATGCGCATAG
oriTII	TCTTTAGGGTCACGCTGGC
PE-PCR 1.0	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
PE-PCR 2.0	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
Sequencing of human genomic DNA	
Chr15_88728	ATATGAATGTTTGCATTCCTT
IS2_Hu11	AAGAAAGTCAACCTTCATCTT
Xba_IS2_Hu11	CAACTCTAGAGGAAAAGTCAGAAAGACACCAAC

[a] (N)$_{12}$, barcode sequence of linker cassette. (RO), restriction enzyme overhang.

[b] (PE-PCR 2.0), adaptor sequences for high-throughput sequencing.

[c] (PE-PCR 1.0), adaptor sequences for high-throughput sequencing. (N)$_{10}$, barcode sequence introduced in second exponential PCR.

826
827

828 **Table 3. Integration events characterized by LAM-PCR and DNA sequencing**

829

IE	Cell line	Relaxase	Number of sequences	Identity	Chr	Integration locus	Missing bp
1	EA.hy926	TrwC	19	99.75	11	35225119	0
2	EA.hy926	TrwC	402	100	11	35225119	0
3	HeLa	TrwC:BID	15	98,25	2	37383046	15
4	HeLa	TrwC:BID	114	100	2	111118923	16
5	HeLa	TrwC:BID	15	100	6	9173423	18
6	HeLa	TrwC:BID	84	99.48	12	28128063	16
7	HeLa	TrwC:BID	21	100	13	83099211	0
8	HeLa	TrwC:BID	95	99.72	16	68832248	16
9	HeLa	TrwC:BID	15	100	19	18303607	14

The information collected for each integration event is shown. Number of sequences indicates the number of times the sequence read was found. Missing base pairs indicates the number of bp that are missed in the read with respect to the plasmid sequence until the *nic* site. IE, integration event. Chr, chromosome.

830

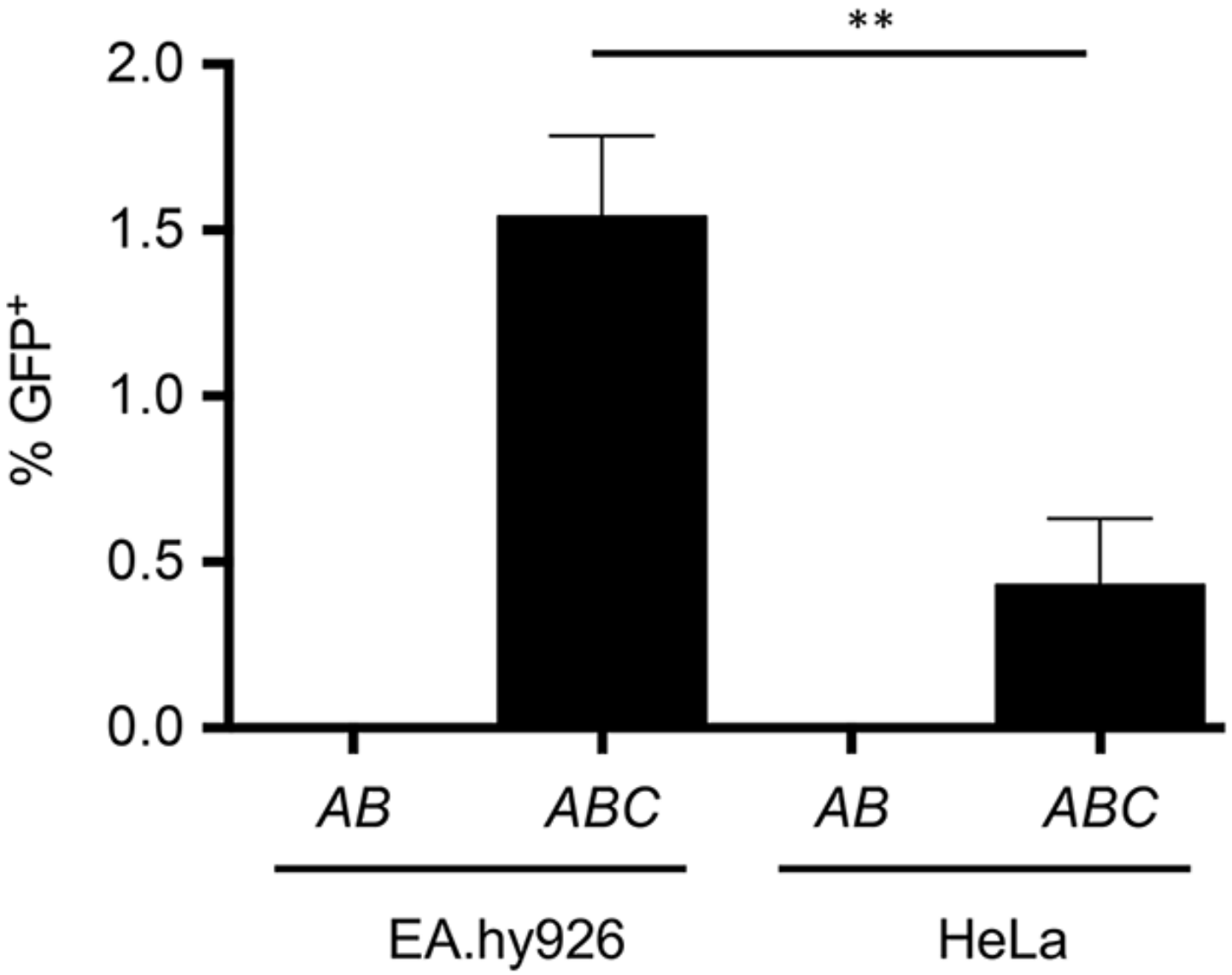
831 **Table 4. Mapped illegitimate integration events**

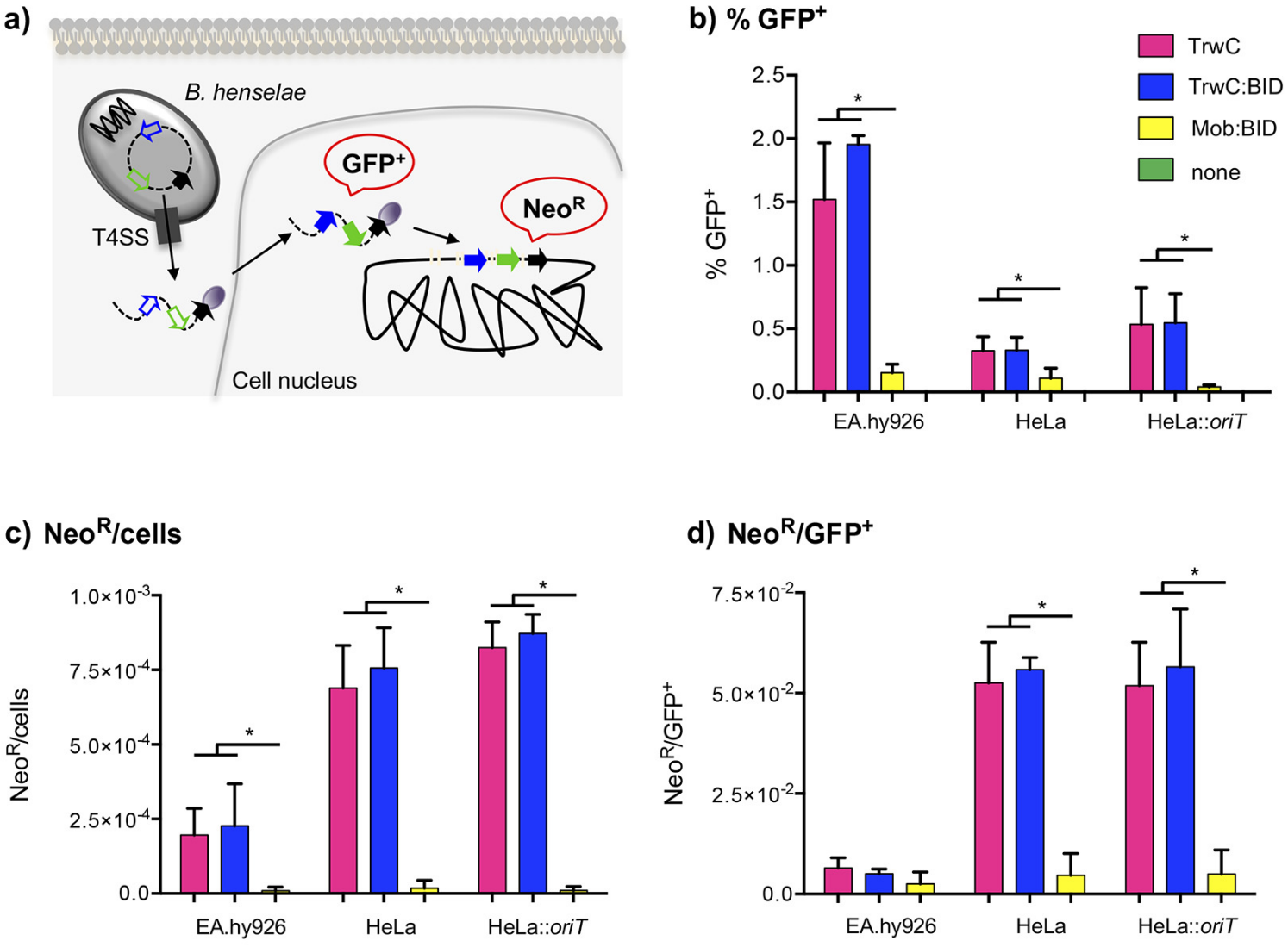
832

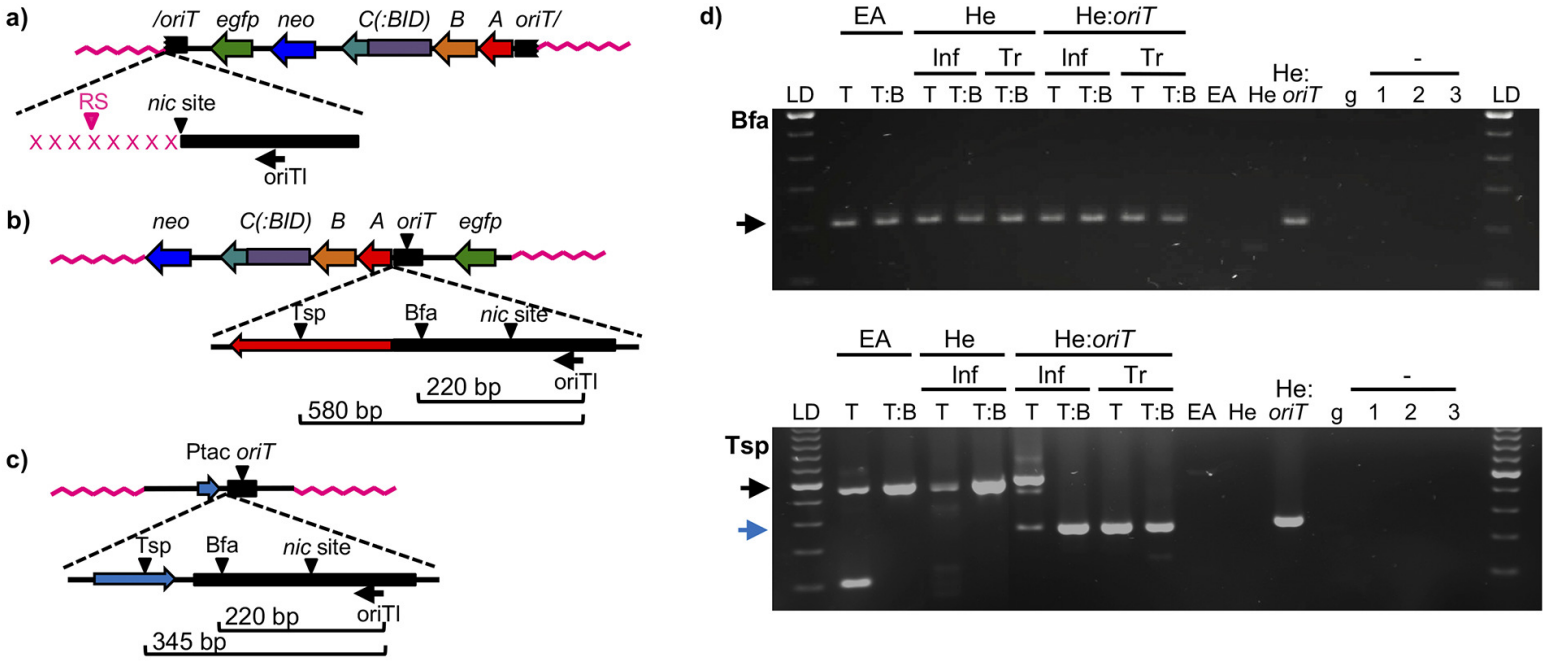
IS	Sequence (5'-3')	Integration junction		
		Chr	Genomic locus	Plasmid
2	AAAATGAGGACAGTT/ATATTTTTTAAATGT	11	35225119	*nic*+1
3	CCAGATCGTGCCACT/GCATTCCAGCCTGGC	2	37383046	*nic*-15
4	TGGGAAACAAATGAA/GAAACAACCCTGCTG	2	111118923	*nic*-16
5	GTTTCCATGGACATT/TGCCACCCCGGCTTC	6	9173423	*nic*-18
6	CGGGTTAGAAACCAA/GCACCCAAGCCGGCG	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*

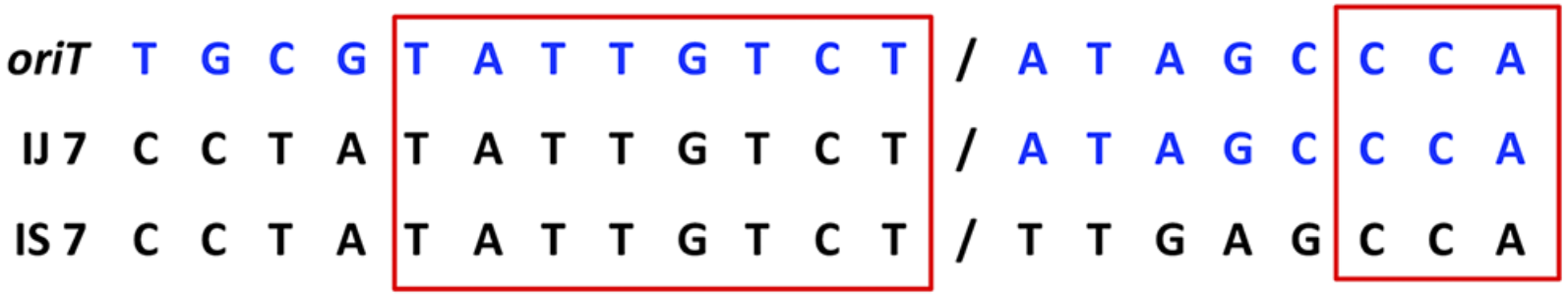
The genomic integration sites are shown aligned with the wild-type target for TrwC, the *oriT*. The dash indicates the integration site (the *nic* site in the *oriT* sequence). The location of the integration site (Chr, chromosome number), as well as the nucleotide of the plasmid by which integration took place, are also displayed. IS, integration site. IS 2-9 were characterized by LAM-PCR. IS 10 and 11 are both integration junctions of the event characterized by recovery of the integrated plasmid. Coordinates of genomic loci correspond to human genome GRCh37/hg19 available in UCSC Genome Browser. Plasmid coordinates refer to the distance from the *nic* site (+ and - indicating 5'or 3' from the *nic* site, respectively) for IS 2-9, or the nucleotide position in the *gfp* and *trwA* ORFs for IS 10-11.

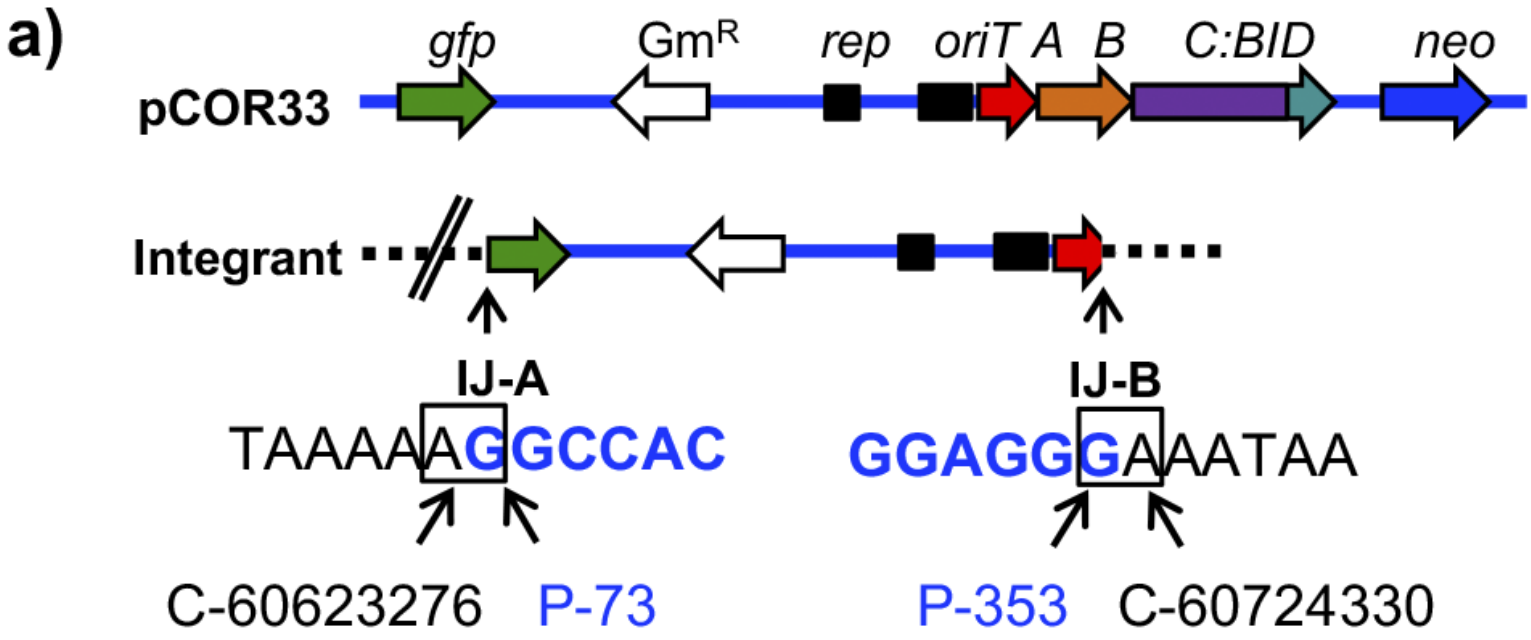
833

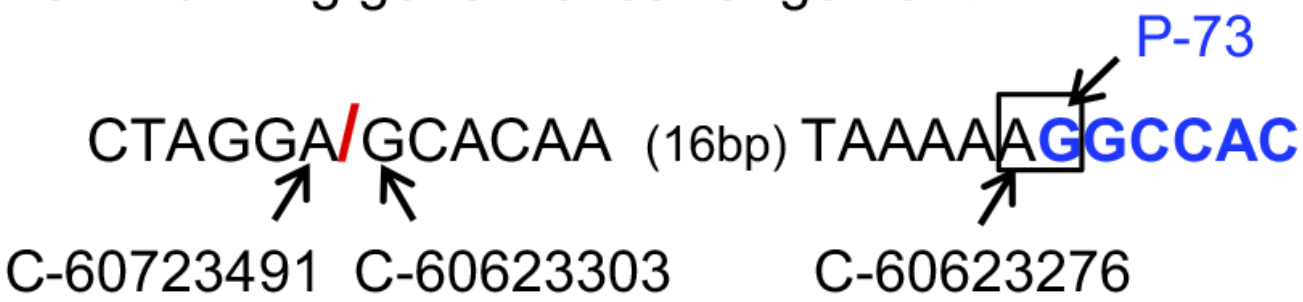

**
2.0
1.5
1.0
0.5
0.0
% GFP+
AB
ABC
AB
ABC
EA.hy926
HeLa


a)
B. henselae
T4SS
GFP+
NeoR
Cell nucleus
b) % GFP+
TrwC
TrwC:BID
Mob:BID
none
% GFP+
2.5
2.0
1.5
1.0
0.5
0.0
*
*
*
EA.hy926
HeLa
HeLa::oriT
c) NeoR/cells
NeoR/cells
1.0×10-3
7.5×10-4
5.0×10-4
2.5×10-4
0.0
*
*
*
EA.hy926
HeLa
HeLa::oriT
d) NeoR/GFP+
NeoR/GFP+
7.5×10-2
5.0×10-2
2.5×10-2
0.0
*
*
*
EA.hy926
HeLa
HeLa::oriT


a)
/oriT egfp neo C(:BID) B A oriT/
RS
nic site
XXXXXXXX
oriTI
b)
neo C(:BID) B A oriT egfp
Tsp
Bfa
nic site
220 bp
oriTI
580 bp
c)
Ptac oriT
Tsp
Bfa
nic site
220 bp
oriTI
345 bp
d)
EA
He
He:oriT
Inf
Tr
Inf
Tr
LD T T:B T T:B T:B T T:B T T:B EA He: He oriT g - 1 2 3 LD
Bfa
EA
He
He:oriT
Inf
Inf
Tr
LD T T:B T T:B T T:B T T:B EA He: He oriT g - 1 2 3
Tsp


oriT T G C G T A T T G T C T / A T A G C C C A
IJ 7 C C T A T A T T G T C T / A T A G C C C A
IS 7 C C T A T A T T G T C T / T T G A G C C A


a)
gfp
Gm^R
rep
oriT A
B
C:BID
neo
pCOR33
Integrant
IJ-A
IJ-B
TAAAAAGGCCAC
GGAGGGAAATAA
C-60623276
P-73
P-353
C-60724330


b)
IJ-A flanking genomic rearrangement
P-73
CTAGGA/GCACAA
(16bp)
TAAAAAGGCCAC
C-60723491
C-60623303
C-60623276