

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Supplementary Information

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

Coral González-Prieto, Richard Gabriel, Christoph Dehio, Manfred Schmidt, and Matxalen Llosa

Contents:

- **Supplementary Materials and Methods:** LAM-PCR primers
- **Supplementary Table S1:** Transient and permanent expression of the transferred DNA.
- **Supplementary Figure S1:** Annealing of LAM-PCR primers
- **Supplementary Figure S2:** Analysis of HeLa::*oriT* cell line
- **Supplementary Figure S3:** PCR amplification of *trwA* and *trwC*
- **Supplementary Figure S4:** GFP expression in G418-resistant cell pools

17 **SUPPLEMENTARY MATERIALS AND METHODS**

18 **LAM-PCR primers.** Primers were designed as described in (1). Table 2 and Fig. S1
19 show the details of the primers used and their annealing sites. The primers used to
20 obtain the linker cassette (LC1 and LC2) contain the proper restriction enzymes
21 overhang (RO) and a barcode sequence (N). The linker cassette used for each sample
22 has a different barcode sequence, to label it and allow the parallel sequencing of
23 different samples. Plasmid-specific primer for linear PCR (oriTI) was designed to align
24 around 120 nt 5' to the *nic* site and is 5' biotinylated. Two additional plasmid-specific
25 primers for first (oriTII; 5' biotinylated) and second (Mis-TrwC; unmodified)
26 exponential PCRs were designed to align between the oriTI primer and the plasmid
27 end. The plasmid-specific primer for second exponential PCR should hybridize at least
28 20 bp away from the plasmid end, leaving enough nucleotides to recognize true
29 integration events. For first and second exponential PCRs, two linker-specific primers
30 were also used, LCI and Mis-LC, respectively. Primers for second exponential PCR
31 contain adaptor sequences (PE-PCR 2/1.0) for the specific sequencing technology used
32 and plasmid-specific primer (Mis-TrwC) also contains an additional barcode sequence
33 (N). A different Mis-TrwC primer (with a different barcode sequence) was used for
34 amplification of each sample.

35

36

37

38

SUPPLEMENTARY TABLE

39

40

Table S1. Transient and permanent expression of the transferred DNA

41

42

Cell line	Infection (relaxase)				Transfection	
	None	TrwC	TrwC:BiD	Mob:BiD	scDNA	dsDNA
EA.hy926						
%GFP ⁺	0	1.52	1.95	0.15	nd	nd
GFP ⁺	0	3.0x10 ⁵	3.9x10 ⁵	3.1x10 ⁴	nd	nd
Neo ^R	0	1,972	1,870	76	nd	nd
Neo ^R /cells	<5x10 ⁻⁸	1.9x10 ⁻⁴	2.3x10 ⁻⁴	9.7x10 ⁻⁶	nd	nd
Neo ^R /GFP ⁺	-	6.6x10 ⁻³	4.8x10 ⁻³	2.5x10 ⁻³	nd	nd
HeLa						
%GFP ⁺	0	0.33	0.33	0.11	6.81	1.28
GFP ⁺	0	6.5x10 ⁴	6.6x10 ⁴	2.2x10 ⁴	5.7x10 ⁴	1.3x10 ⁴
Neo ^R	0	3,428	3,682	102	66	54
Neo ^R /cells	<5x10 ⁻⁸	6.9x10 ⁻⁴	7.6x10 ⁻⁴	1.8x10 ⁻⁵	7.9x10 ⁻⁵	5.4x10 ⁻⁵
Neo ^R /GFP ⁺	-	5.3x10 ⁻²	5.6x10 ⁻²	4.6x10 ⁻³	1.2x10 ⁻³	4.2x10 ⁻³
HeLa::oriT						
%GFP ⁺	0	0.53	0.55	0.04	6.27	1.31
GFP ⁺	0	1.1x10 ⁵	1.1x10 ⁵	8.0x10 ³	6.1x10 ⁴	1.1x10 ⁴
Neo ^R	0	5,542	6,158	39	73	34
Neo ^R /cells	<5x10 ⁻⁸	8.2x10 ⁻⁴	8.7x10 ⁻⁴	1.1x10 ⁻⁵	7.6x10 ⁻⁵	6.8x10 ⁻⁵
Neo ^R /GFP ⁺	-	5.0x10 ⁻²	5.6x10 ⁻²	4.9x10 ⁻³	1.2x10 ⁻³	3.1x10 ⁻³

DNA was introduced in the human cell lines indicated in each sub-heading, either by *Bartonella* T4SS-mediated transfer (Infection) or by plasmid transfection with either supercoiled (sc) or linearized (ds) DNA. Transfections were done with plasmid pCOR35 ($\Delta trwC$). The relaxase encoded by the mobilizable plasmid used in each infection is indicated on top of the Table. For each cell line, the percentage of GFP-positive cells is shown (%GFP⁺) as well as the total number of GFP positive cells (GFP⁺), calculated

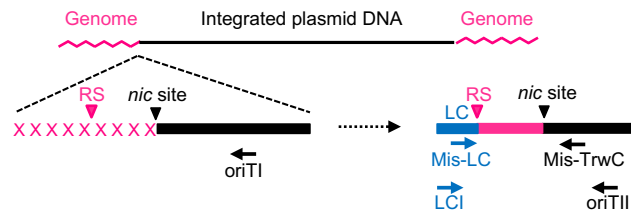
based on the theoretical number of cells on a confluent plate. Neo^R , number of G418-resistant colonies obtained. $\text{Neo}^R/\text{cells}$, the same number divided by total number of cells in the plate. $\text{Neo}^R/\text{GFP}^+$, rate of integration in GFP-positive cells. nd, not determined. Data represent the mean of at least 3 independent experiments.

43

44

45 SUPPLEMENTARY FIGURES

46



47

48

49

50 **Figure S1. Annealing of LAM-PCR primers.** A theoretical integration event is illustrated

51 on the top of the Figure, with the integrated plasmid DNA represented in black and

52 genomic DNA in pink. Below the two amplification steps are represented, a first linear

53 amplification PCR (left) and a second exponential PCR (right). Oligonucleotide names

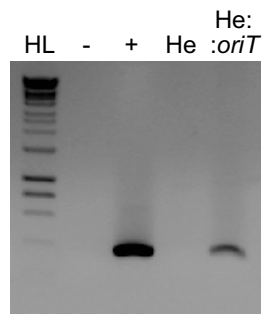
54 and sequences are defined in the text and in Table S1. The linker cassette (LC) added in

55 LAM-PCR is shown in blue. RS, restriction site. The annealing regions for primers used

56 are indicated by arrows.

57

58



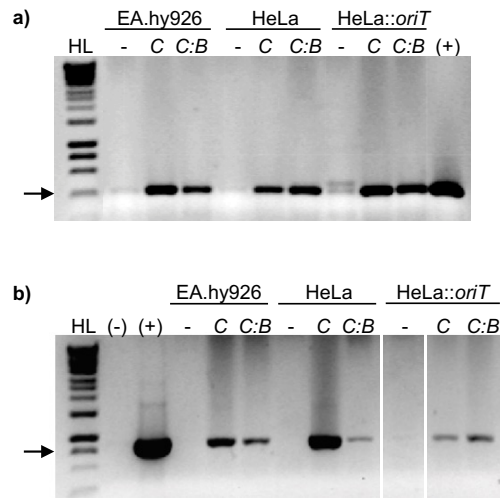
59

60

61

62 **Figure S2. Analysis of HeLa::*oriT* cell line.** PCR amplification of R388 *oriT* with
63 oligonucleotides *oriT1* and *oriT330* (Table S1). Genomic DNA was extracted from
64 unmodified HeLa (He) and from Hyg^R HeLa::*oriT* (He::*oriT*) transfected cell pool. HL,
65 Hyperladder I. -, negative control (no DNA). +, positive control (pMTX708 plasmid
66 DNA).

67



69

70

71

72 **Figure S3. PCR amplification of *trwA* (a) and *trwC* (b).** gDNA was extracted from the

73 G418-resistant cell pools; the cell line is indicated in the top row. Primers used for

74 amplification of *trwA* were Hind3_TrwA_F and BamHI_TrwA_R (Table S1), while *trwC*

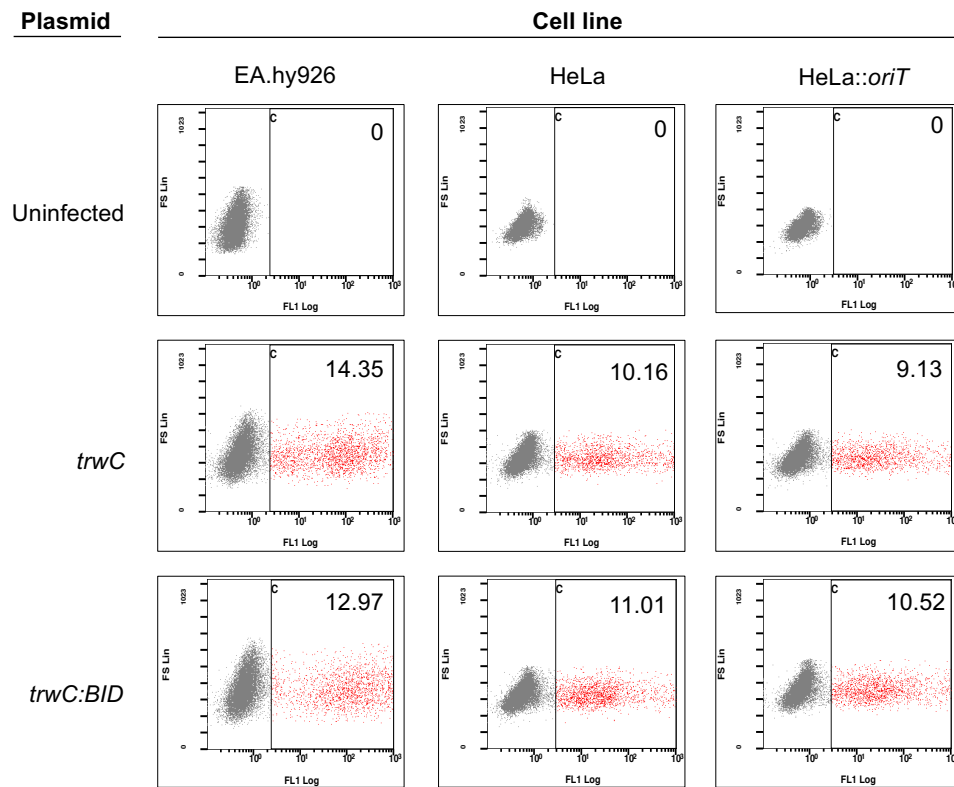
75 was amplified using primer Hind3_TrwC_F and 670_TrwC (Table S1). The arrows

76 indicate the expected bands of 400 bp in a) and 800 bp in b). C, cell pools obtained

77 after mobilization of *trwC*-encoding plasmid. C:B, pools obtained after mobilization of78 *trwC:BD* plasmid. -, uninfected cells. HL, Hyperladder I. (+), positive control (pHP161

79 plasmid DNA). (-), negative control (no DNA).

80



82

83

84

85

Figure S4. GFP expression in G418-resistant cell pools. Pools obtained after

86

integration of both *trwC*- and *trwC:BiD*-coding plasmids were analyzed by flow

87

cytometry using uninfected cells as control. Plots represent cell granularity (FS Lin, in

88

ordinates) versus GFP fluorescence intensity (FL1 Log, in abscissas). The square marks

89

the population considered as positive, whose events are shown in red. The percentage

90

of GFP positive cells is indicated in each plot. Although it is clearly detected, the

91

percentage of GFP positive cells (10-15 %) was lower than expected. This phenomenon

92

has been previously described and associated with the antibiotic treatment used to

93

select the cells (2).

94

95

REFERENCES

96 1.

Paruzynski A, Arens A, Gabriel R, Bartholomae CC, Scholz S, Wang W, Wolf S,

97

Glimm H, Schmidt M, von Kalle C. 2010. Genome-wide high-throughput

98

integrome analyses by nrLAM-PCR and next-generation sequencing. *Nat Protoc*

99

5:1379-1395.

100 2.

Kaufman WL, Kocman I, Agrawal V, Rahn HP, Besser D, Gossen M. 2008.

101

Homogeneity and persistence of transgene expression by omitting antibiotic

102

selection in cell line isolation. *Nucleic Acids Res* **36:e111.**

103

104