Supporting Information

Efficient markerless integration of genes in the chromosome of probiotic *E. coli* Nissle 1917 by bacterial conjugation

Elena M. Seco¹ and Luis Ángel Fernández¹*

¹Department of Microbial Biotechnology, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CNB-CSIC), Darwin 3, Campus UAM Cantoblanco, 28049 Madrid, Spain.

(*) For correspondence: lafdez@cnb.csic.es

This file includes:

- Tables S1 and S2.
- Supporting Experimental Procedures.
- Supporting References.
- Figures S1 to S8.
Table S1. *E. coli* strains and plasmids used in this study

| Name | Genotype and relevant properties | Reference |
|------|----------------------------------|-----------|
| MFDpir | MG1655 RP4-2-Tc::[Mu1::aac(3)Iv-ΔaphA-Δnic35-ΔMu2::deo] ΔdapA:: (erm-pir) ΔrecA | (Ferrières, et al., 2010) |
| BW25141 | (F- Δ(araD-araB)567, ΔlacZ4787::ermB-3, Δ(phoB phoR)580, galU95, ΔuidA3::pir, recA1, endA9(del-ins):FRT, rph-1, Δ(rhaD-rhaB)568, hsdR51) | (Datsenko and Wanner, 2000) |
| *E. coli* Nissle 1917 (EcN) | Serotype O6:K5:H1 | DSM6601/Mutaflor |
| EcN_pACBSR | EcN transformed with pACBSR (CmR) | This work |
| EcN_pACBS | EcN transformed with pACBS (CmR) | This work |
| EcN_pACBSG | EcN transformed with pACBSG (CmR) | This work |
| EcNΔflu_gfp | EcN Δflu::Ptac-gfp | This work |
| EcNΔfim_gfp | EcN ΔfimIALCFGH::Ptac-gfp | This work |
| EcNΔfim_mKATE2 | EcN ΔfimIALCFGH::Ptac-mKATE | This work |
| EcNΔmat_lux | EcN ΔecpA::P2-luxCDABE | This work |
| EcN_Ptet-flhDC | EcN tetR-Ptet-flhDC | This work |
| EcNΔflhDC | EcN ΔflhDC | This work |
| EcNΔfim_mKATE2 Δmat_lux | EcN ΔfimIALCFGH::Ptac-mKATE2 ΔecpA::P2-luxCDABE | This work |
| EcNΔfim_mKATE2 Δmat_lux Ptet-flhDC | EcN TetR_Ptet-flhDC Δfim::Ptac-mKATE2 Δmat::P2-luxCDABE | This work |
| pSEVA237R | KmR, pBBR1-orI, oriT, mCherry | (Martinez-Garcia, et al., 2015) |
| pSEVA241 | KmR, pRO1600/CollE1, oriT | (Martinez-Garcia, et al., 2015) |
| pGE | KmR, R6K-orI, polylinker flanked by two I-Sce restriction sites | (Piñero-Lambea, et al., 2015) |
| pGEC | KmR, oriT, R6K-orI, polylinker flanked by two I-Sce restriction sites | This work |
| pACBSR | p15A-orI, CmR, Pbad promoter, I-Sce endonuclease and λ Red genes | (Herring, et al., 2003) |
| pACBS | p15A-orI, CmR, Pbad promoter, I-Sce endonuclease | This work |
| pACBSG | p15A-orI, CmR, Pbad promoter, I-Sce endonuclease and λ Red gam | This work |
| pGECfluEcN_gfp | pGEC derivative; Ptac-gfp<sup>TCOD</sup> with homology regions of the EcN flu gene | This work |
| pGECfimEcN_gfp | pGEC derivative; Ptac-gfp<sup>TCOD</sup> with homology regions flanking the EcN fimIALCFGH | This work |
| pGECfimEcN_mKATE2 | pGEC derivative; Ptac-mKATE2 with homology regions flanking the fimIALCFGH | This work |
| pGECmatEcNlux | pGEC derivative; P2-luxCDABE with homology regions flanking EcN ecpa | This work |
| pGEC_tetR-Ptet-flhDC' | pGEC derivative; TetR_Ptet-flhDC with homology regions for replacing natural EcN flhDC promoter (2104116-2103813 bases from CP007799.1) by TetR_Ptet. | This work |
| pGECΔflhDC | pGEC derivative with homology regions flanking EcN flhDC (210416-2102879 bases from CP007799.1) | This work |
Table S2. Oligonucleotides used as primers in this work

| Number | Nucleotide sequence (5´-3´) |
|--------|-----------------------------|
| 1      | GCTAATAATAACTCTGTATAGCAAC   |
| 2      | GTCAAGTTTGCGTACGTGCGTC      |
| 3      | CCGCCAGCGGTCCGTTTGTCTGCC    |
| 4      | CCGGCACTTTCATGAGAATCAGAC    |
| 5      | TCATCCGCACATGTTCAGCCATTC    |
| 6      | GTGATGAAACGCTTTTGAAGTGCG    |
| 7      | AGACCCTACATAAGCTGCTGGTC     |
| 8      | AAGGATCCATTAAAGAGGAGAATACTAGATGATGTCAGAATTATAAAGAAATAT |
| 9      | CTGAAATGTAATCTGTAATAATTACAG |
| 10     | GCTTAGACATTTCCTTCTGAATATC   |
| 11     | GATGAAAATGCAGCTAAATCTGTTT   |
| 12     | CCGTCTTCTTGTATTACCGCGTGTC   |
| 13     | GCTATTGATTGTACATTTC        |
| 14     | AAAAAAATCAATTAATGGAAGATGG  |
| 15     | GTTGTAACCTTCCGATTCGACC     |
| 16     | CTGTTCAAATGTTAAACCGTAG     |
| 17     | CCCCCATTACAGCCGCAACATAC    |
| 18     | GCGGCCCTCAGACGCCGCTTCTTCCCTGCCCTCCGTGA |
| 19     | TACACGCGGAGGGAGGAAAGCTCGGCCGAAGCTCCTTTCTGAAATAGCTGTGTTG |
| 20     | GGTCTCGAGAATAGCAATCTGAATAC |
| 21     | TCAACAGCTCATTTCCAGAGAGCTTCGCCAGTCTTTTCCCCCCCCTGTA |
| 22     | ATCTCAATAGTGGACTAATCTAATTAGCTACCTTCTACTGAGAT |
| 23     | GATACAATGTCTGTGATTTCTAAGACTGTTGCTCTGCTTGAAGTACCTGAGAT |
| 24     | GAGCTCAGACCTTCTGTTAGAATGCTGGAAGTCTGGAGTCTGCTGAGT |
| 25     | CGGCAGCATGCCAGAGCTGCTAATCAGG |
| 26     | TCAACAGCTCATTTCCAGAGAGCTTCCGCCAGTCTTTTCCCCCCCCTGTA |
| 27     | ATCTCAATAGTGGACTAATCTAATTAGCTACCTTCTACTGAGAT |
| 28     | GATACAATGTCTGTGATTTCTAAGACTGTTGCTCTGCTTGAAGTACCTGAGAT |
| 29     | GAGCTCAGACCTTCTGTTAGAATGCTGGAAGTCTGGAGTCTGCTGAGT |
| 30     | CGGCAGCATGCCAGAGCTGCTAATCAGG |
| 31     | TCAACAGCTCATTTCCAGAGAGCTTCCGCCAGTCTTTTCCCCCCCCTGTA |
| 32     | ATCTCAATAGTGGACTAATCTAATTAGCTACCTTCTACTGAGAT |
| 33     | GATACAATGTCTGTGATTTCTAAGACTGTTGCTCTGCTTGAAGTACCTGAGAT |
| 34     | GAGCTCAGACCTTCTGTTAGAATGCTGGAAGTCTGGAGTCTGCTGAGT |
| 35     | CGGCAGCATGCCAGAGCTGCTAATCAGG |
| 36     | TCAACAGCTCATTTCCAGAGAGCTTCCGCCAGTCTTTTCCCCCCCCTGTA |
| 37     | ATCTCAATAGTGGACTAATCTAATTAGCTACCTTCTACTGAGAT |
| 38     | GATACAATGTCTGTGATTTCTAAGACTGTTGCTCTGCTTGAAGTACCTGAGAT |
| 39     | GAGCTCAGACCTTCTGTTAGAATGCTGGAAGTCTGGAGTCTGCTGAGT |
| 40     | CGGCAGCATGCCAGAGCTGCTAATCAGG |
| 41     | TCAACAGCTCATTTCCAGAGAGCTTCCGCCAGTCTTTTCCCCCCCCTGTA |
| 42     | ATCTCAATAGTGGACTAATCTAATTAGCTACCTTCTACTGAGAT |
| 43     | GATACAATGTCTGTGATTTCTAAGACTGTTGCTCTGCTTGAAGTACCTGAGAT |
| 44     | GAGCTCAGACCTTCTGTTAGAATGCTGGAAGTCTGGAGTCTGCTGAGT |

* Restriction site is underlined above the oligonucleotide sequence
Supporting Experimental Procedures

Plasmids constructions

**pACBS**: Plasmid for cointegrants resolution constructed by deleting the \textit{Sphl} fragment containing the \(\lambda\). Red genes from plasmid pACBSR.

**pACBSG**: Plasmid for cointegrants resolution constructed by cloning a DNA fragment obtained by PCR from pACBSR with oligonucleotides 43 and 44, digested with \textit{BamHI-Sphl} and cloned into the same sites of pACBS.

**pGEC**: GenBank MZ361915. This suicide plasmid is a pGE-derivative (Piñero-Lambea, et al., 2015) (GenBank MZ361922) containing a 680-bp DNA fragment corresponding to the \textit{oriT} and R6K, amplified by PCR from plasmid pSEVA412S (Martinez-Garcia, et al., 2015) with the oligonucleotides 18 and 19, digested with \textit{XhoI-AscI} and cloned into the same sites of pGE.

**pGECfluEcN\textsubscript{gfp}**: GenBank MZ361916. It is a pGEC-derivative containing a DNA fragment as a fusion PCR product of three individual PCRs: (i) homologous region HR1-flu, amplified from EcN total DNA with oligonucleotides 20 and 21; (ii) Ptac-gfp, containing \textit{gfp\textsuperscript{TCD}} allele (Corcoran, et al., 2010), was amplified from pGEfluPtac-gfp (Ruano-Gallego, et al., 2015) with oligonucleotides 22 and 23; (iii) homologous region HR2-flu, amplified from EcN total DNA with oligonucleotides 24 and 25. Fusion PCR was digested with \textit{XhoI-Sphl} and cloned into the same sites of pGEC.

**pGECfimEcN\textsubscript{gfp}**: GenBank MZ361917. It is a pGEC-derivative containing a DNA fragment obtained as a fusion PCR product of three individual PCRs: (i) the homologous region HR1-fim, amplified from EcN total DNA with oligonucleotides 26 and 27; (ii) Ptac-gfp, amplified from pGEfluPtac-gfp (Ruano-Gallego, et al., 2015) with oligonucleotides 28 and 29; (iii) the homologous region HR2-fim, amplified from EcN total DNA with oligonucleotides 30 and 31. The fusion PCR was digested with \textit{XhoI-Sphl} and cloned into the same sites of pGEC.

**pGECfimEcN\textsubscript{mKATE2}**: GeneBank MZ361918. It is a pGEC-derivative containing a DNA fragment encoding the fluorescent protein mKATE2 under the control of Ptac promoter. This DNA fragment was amplified by PCR from a plasmid encoding mKATE2 (Shcherbo, et al., 2009) with oligonucleotides 8 and 32, digested with \textit{BamHI-SpeI} and cloned into the same sites of pGEC\textsubscript{fimEcN\textsubscript{gfp}}.

**pGECmatEcN\textsubscript{gfp}**: This pGEC-derivative was constructed by simultaneous cloning of these three fragments: (i) the homologous region HR1-mat, amplified by PCR from EcN total DNA with the oligonucleotides 33 and 34 and digested with \textit{XhoI-Sacl}; (ii) Ptac-gfp,
obtained as a *Sac*-*Spe* digested fragment from plasmid pGE*CfluEcN_gfp*; and, (iii) the homologous region HR2-*mat*, amplified by PCR from EcN total DNA with the oligonucleotides 35 and 36 and digested with *Spe*-*SphI*. The three DNA fragments were simultaneously ligated into the *Xhoi/SphI* sites of pGE*C.

**pGE*CmatEcN_lux**: GeneBank MZ361919. It is pGE-C-derivative containing a 5,978 bp *SacI/SpeI* DNA fragment containing the *luxCDABE* operon of *Photobacterium luminescens* under the control of the P2 constitutive promoter isolated from pGE*mat-lux* (Piñero-Lambea, et al., 2015) and cloned in the same restriction sites of pGE*CmatEcN_gfp*, replacing Ptac-*gfp* by P2-*luxCDABE*.

**pGE*C_tetR-Ptet-flhDC**: GeneBank MZ361920. This pGE-C-derivative was constructed in two steps. The first step was the simultaneous cloning of two DNA fragments into pGE*C digested with *XhoI-XbaI*: (i) the homologous region HR1-*flhD*, amplified by PCR from EcN total DNA with oligonucleotides 37 and 38 digested with *XhoI-SacI*; (ii) TetR_Ptet DNA synthesis fragment of 853 bp (GeneArt, Thermo Fisher Scientific) with flanking *SacI-XbaI* sites and digested with these enzymes. In a second step, the homologous region HR2-*flhD*, amplified by PCR from EcN total DNA with oligonucleotides 39 and 40, was cloned into the *XbaI-SphI* restriction sites of the resultant plasmid.

**pGE*CΔflhDC**: GeneBank MZ361921. This pGE-C-derivative was obtained by cloning the homologous region HR3_*flhDC*, amplified by PCR from EcN total DNA with oligonucleotides 41 and 42 and digested with *SacI-SphI* into the same sites of pGE*C_tetR-Ptet-flhDC*.

**E. coli Nissle1917** strain constructions

**EcNΔflu_gfp**: Strain obtained by conjugation of the recipient EcN carrying pACBSR (CmR) with the donor MFDpir carrying the plasmid pGE*CfluEcN_gfp* (KmR). Cointegrants, selected with Cm and Km, were resolved by I-*SceI* endonuclease expression and individual colonies were checked by PCR with oligonucleotides 1 and 2 for *flu* upstream integration and oligonucleotides 3 and 4 for *flu* downstream integration.

**EcNΔfim_gfp**: Strain obtained by conjugation of recipient EcN carrying pACBSR (CmR) with the donor MFDpir carrying the plasmid pGE*CfimEcN_gfp* (KmR). Cointegrants, selected with Cm and Km, were resolved by I-*SceI* endonuclease expression and individual colonies were checked by PCR with oligonucleotides 5 and 2 for *fim* upstream integration and oligonucleotides 3 and 6 for *fim* downstream integration.
EcNΔfim_mKATE2: Strain obtained by conjugation of the recipient EcN carrying pACBSR (Cm<sup>R</sup>) with the donor MFDpir carrying the plasmid pGECfimEcN_mKATE2 (Km<sup>R</sup>). Cointegrants, selected with Cm and Km, were resolved by I-SceI endonuclease expression and individual colonies were checked by PCR with oligonucleotides 5 and 7 for fimAICDFGH upstream integration and oligonucleotides 8 and 6 for fimAICDFGH downstream integration.

EcNΔmat_lux: Strain obtained by conjugation of the recipient EcN carrying pACBSR (Cm<sup>R</sup>) with the donor MFDpir carrying the plasmid pGECmatEcN_lux (Km<sup>R</sup>). Cointegrants, selected with Cm and Km, were resolved by I-SceI endonuclease expression and individual colonies were checked by PCR with oligonucleotides 9 and 10 for ecpA upstream integration and oligonucleotides 11 and 12 for ecpA downstream integration. Positive clones were tested for light emission.

EcN_P<sub>tet</sub>-flhDC: Strain obtained by conjugation of the recipient EcN carrying pACBSR (Cm<sup>R</sup>) with the donor MFDpir carrying the plasmid pGEC<sub>tetR</sub>-Ptet-flhDC (Km<sup>R</sup>). Cointegrants, selected with Cm and Km, were resolved by I-SceI endonuclease expression and individual colonies were checked by PCR with oligonucleotides 13 and 14 for upstream integration and oligonucleotides 15 and 16 for downstream integration.

EcNΔflhDC: Strain obtained by conjugation of the recipient EcN carrying pACBSR (Cm<sup>R</sup>) with the donor MFDpir carrying the plasmid pGEC<sub>ΔflhDC</sub> (Km<sup>R</sup>). Cointegrants, selected with Cm and Km, were resolved by I-SceI endonuclease expression and individual colonies were checked by PCR with oligonucleotides 13 and 17 to select the deletion of flhDC operon.

EcNΔfim_mKATE2 Δmat_lux: Strain obtained by conjugation of the recipient EcNΔfim_mKATE2 carrying pACBSR (Cm<sup>R</sup>) with the donor MFDpir carrying the plasmid pGECmatEcN_lux (Km<sup>R</sup>). Cointegrants, selected with Cm and Km, were resolved by I-SceI endonuclease expression and individual colonies were checked by PCR with oligonucleotides 9 and 10 for ecpA upstream integration and oligonucleotides 11 and 12 for ecpA downstream integration. Positive clones were tested for light emission.

EcNΔfim_mKATE2 Δmat_lux P<sub>tet</sub>-flhDC: Strain obtained by conjugation of the recipient EcNΔfim_mKATE2 Δmat_lux carrying pACBSR (Cm<sup>R</sup>) with the donor MFDpir carrying the plasmid pGEC<sub>tetR</sub>-Ptet-flhDC (Km<sup>R</sup>). Cointegrants, selected with Cm and Km, were resolved by I-SceI endonuclease expression and individual colonies were checked by PCR with oligonucleotides 13 and 14 for ecpA upstream integration and oligonucleotides 15 and 16 for ecpA downstream integration.
Bioinformatic analysis of whole genome sequencing data

The quality of the FASTQ files was analyzed using FastQC v0.11.5 (Andrews, 2010). FastQ Screen v0.14.1 (Wingett and Andrews, 2018) did not detect unexpected contamination with foreign genomic material. Fastp software (Chen, et al., 2018) was applied to remove final poly-Gs of length 12 nt or more, and to filter out reads of less than 50 nt long. Filtered sequences of both samples were aligned with BWA sampe (version 0.7.17) (Li and Durbin, 2010) against the assembly ASM71459v1 (GenBank sequence CP007799.1) for Escherichia coli strain Nissle 1917 with default parameters. Most of the reads were mapped to the reference genome and the mean coverage was around 600 and distributed evenly, with more than 98% of the genome covered by 50 or more reads. Samtools version 1.11 (Li, et al., 2009) was used to compress the alignment files (SAM to BAM format), sort by coordinates and index the files. Optical duplicated reads were removed using Picard MarkDuplicates version 2.18.26-SNAPSHOT (http://broadinstitute.github.io/picard/). Statistics and quality of the alignments were assessed for all samples with Qualimap (Okonechnikov, et al., 2016) software (version v.2.2.1), samtools options flagstats and idxstats and bcftools (version 2.4.0) option stats. Mapped reads (without optical duplicates) of strain_118 and strain_185 were the input of the freebayes (Garrison and Marth, 2012) variant caller to extract SNP, INDEL, multiple nucleotide polymorphism (MNP) and other more complex events, with arguments - -pooled-continuous, that also considers alternative models different from haploid, and -C=5, that filters out detected callings having less than 5 reads supporting the variant allele. In addition, vcffilter (Garrison, et al., 2021) with -Q>20 was used to keep only variants having a probability that the site has a real variant (QUAL) of 20 or more. The resulting multisample variant calling file (VCF) was annotated using SnpEff (Cingolani, et al., 2012) version 4.3t, with the option -ud 500 and using a built database for Escherichia coli strain Nissle 1917. CNVseq (Xie and Tammi, 2009) was used to extract Copy Number Variation (CNV) between samples. The result was a set of three differential CNVs that match with the regions where the three insertions were placed although the borders of the deletions were not precisely defined. Deleted regions in the modified strain were independently evaluated by bedtools genomecov (Quinlan and Hall, 2010) with parameters -bga -split plus grep filtering used on the BAM alignment files to obtain the regions having no coverage of reads in each of the two samples. Inspection of these genomic regions with the genomic browser IGV (Thorvaldsdóttir, et al., 2013) showed only three reliable deletions in the modified strain 185 and none in parental strain 118. The three detected deletions were almost identical (within 1 to 3 bp difference) to those expected by the gene replacements in the modified strain (Supporting Data 1).
These small border differences are due to limitations of the bioinformatic assembly tools but do not represent actual variants in the genome of the modified strain.

To check the existence of possible variants in the inserted regions of sample 185, respect to the expected sequence of the inserts, an additional variant calling analysis was performed (with the same parameters as described above) aligning the filtered FASTQ sequences of sample 185 against an artificial reference file containing the FASTA sequences of the three regions inserted (tetR-Ptet-flhDC, P2-lux, and Ptac-mKATE), plus 100 nt upstream and downstream of the inserts. Results of this variant calling analysis are shown in Supporting Data 1.

**Supporting References**

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. Online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34: i884-i890.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3, Fly (Austin) 6: 80-92.

Corcoran, C.P., Cameron, A.D., and Dorman, C.J. (2010) H-NS silences gfp, the green fluorescent protein gene: gfpTCD is a genetically remastered gfp gene with reduced susceptibility to H-NS-mediated transcription silencing and with enhanced translation, J Bacteriol 192: 4790-4793.

Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products, Proc Natl Acad Sci U S A 97: 6640-6645.

Ferrières, L., Hémery, G., Nham, T., Guerout, A.M., Mazel, D., Beloin, C., and Ghigo, J.M. (2010) Silent mischief: bacteriophage Mu insertions contaminate products of Escherichia coli random mutagenesis performed using suicidal
transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery, *J Bacteriol* **192**: 6418-6427.

Garrison, E., Kronenberg, Z.N., Dawson, E.T., Pedersen, B.S., and Prins, P. (2021) Vcflib and tools for processing the VCF variant call format, *bioRxiv*: 2021.2005.2021.445151.

Garrison, E., and Marth, G. (2012) Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907v1202.

Herring, C.D., Glasner, J.D., and Blattner, F.R. (2003) Gene replacement without selection: regulated suppression of amber mutations in Escherichia coli, *Gene* **311**: 153-163.

Li, H., and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform, *Bioinformatics* **26**: 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009) The Sequence Alignment/Map format and SAMtools, *Bioinformatics* **25**: 2078-2079.

Martinez-Garcia, E., Aparicio, T., Goni-Moreno, A., Fraile, S., and de Lorenzo, V. (2015) SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities, *Nucleic Acids Res* **43**: D1183-1189.

Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data, *Bioinformatics (Oxford, England)* **32**: 292-294.

Piñero-Lambea, C., Bodelón, G., Fernández-Periáñez, R., Cuesta, A.M., Álvarez-Vallina, L., and Fernández, L.A. (2015) Programming Controlled Adhesion of E. coli to Target Surfaces, Cells, and Tumors with Synthetic Adhesins, *ACS synthetic biology* **4**: 463-473.

Quinlan, A.R., and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics* **26**: 841-842.
Ruano-Gallego, D., Álvarez, B., and Fernández, L.Á. (2015) Engineering the Controlled Assembly of Filamentous Injectisomes in *E. coli* K-12 for Protein Translocation into Mammalian Cells, *ACS synthetic biology* **4**: 1030-1041.

Shcherbo, D., Murphy, C.S., Ermakova, G.V., Solovieva, E.A., Chepurnykh, T.V., Shcheglov, A.S., et al. (2009) Far-red fluorescent tags for protein imaging in living tissues, *Biochem J* **418**: 567-574.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, *Briefings in bioinformatics* **14**: 178-192.

Wingett, S.W., and Andrews, S. (2018) FastQ Screen: A tool for multi-genome mapping and quality control, *F1000Res* **7**: 1338-1338.

Xie, C., and Tammi, M.T. (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing, *BMC Bioinformatics* **10**: 80.
Fig. S1. Site-specific markerless integration in EcN chromosome. Scheme showing the integration of a gene of interest (GOI) under the control of a promoter (P) in the EcN chromosome using a suicide conjugative plasmid pGEC derivative containing the origin of transfer (oriT), the replication origin R6K, the KmR gene, and the GOI flanked by two homologous regions (HR1 and HR2) corresponding to regions located upstream and downstream of the target gene. The cassette formed by the GOI and the HRs is flanked by two I-SceI restriction sites. The first recombination event (Recombination I) leads to the cointegrants obtaining which are further resolved by the expression of the I-SceI endonuclease and the \( \lambda \) Red protein from the helper plasmid pACBSR. Double strand breaks generated by the I-SceI endonuclease are repaired a second homologous recombination event (Recombination II) assisted by the \( \lambda \) Red protein that may produce either the wild type allele (i) or the insertion mutant (ii).
Fig. S2. PCR analysis of flu_Ptac-gfp cointegrants.
A. Scheme showing that cointegrants can be generated after homologous recombination using HR1-flu (i) or HR2-flu (ii).
B-C. Colony PCR of five cointegrants using primer oligonucleotides 1 and 2 to check homologous recombination with HR1-flu (B) or with oligonucleotides 3 and 4 to check homologous recombination with HR2-flu (C). PCR of EcN wt colony was used as negative control in both panels B and C.
Fig. S3. PCR analysis of fim_Ptac-gfp cointegrants.
A. Scheme showing that cointegrants can be generated after homologous recombination using HR1-fim (i) or HR2-fim (ii).
B-C. Colony PCR of five cointegrants using primer oligonucleotides 5 and 2 to check homologous recombination with HR1-fim (B) or with oligonucleotides 3 and 6 to check homologous recombination with HR2-fim (C). PCR of EcN wt colony was used as negative control in both panels B and C.
Fig. S4. PCR analysis of the resolution of flu_Ptac-gfp cointegrants.

A. Replica plating of 40 colonies randomly picked after resolution of EcN flu_Ptac-gfp cointegrants using LB Cm plates, with and without Km, to check the loss of Km resistance.

B. Scheme of the cointegrants resolved to the Ptac-gfp insertion.

C-D. Colony PCR of 10 colonies sensitive to Km with oligonucleotides 1 and 2 to check the insertion Ptac-gfp using the upstream region of flu locus (C) or with oligonucleotides 3 and 4 to check the insertion using the downstream region of flu locus (D). PCR of EcN wt colony was used as negative control in both panels C and D.
Fig. S5. PCR analysis of the resolution of fim_Ptac-gfp cointegrants. 
A. Replica plating of 40 colonies randomly picked after resolution of EcN fim_Ptac-gfp cointegrants using LB Cm plates, with and without Km, to check the loss of Km resistance. 
B. Scheme of the cointegrants resolved to the Ptac-gfp insertion. 
C-D. Colony PCR of 10 colonies sensitive to Km with oligonucleotides 5 and 2 to check the insertion Ptac-gfp using the upstream region of fim locus (C) or with oligonucleotides 3 and 6 to check the insertion using the downstream region of fim locus (D). PCR of EcN wt colony was used as negative control in both panels C and D.
**Fig. S6. Expression of GFP in EcNΔflu_Ptac-gfp and EcNΔfim_Ptac-gfp cointegrants.** Flow cytometry analysis of EcN wild type bacteria (EcN) and derivative strains EcNΔflu_gfp and EcNΔfim_gfp carrying insertion of Ptac-gfp replacing flu or fim locus. Bacteria were harvested from cultures of the corresponding strain in LB with IPTG 0.1 mM.
Fig. S7. Integration frequency of pGECfimEcN_gfp in EcN carrying pACBSR or pACBS. Frequencies are calculated as the ratio of cointegrants (C) vs. recipients (R). Horizontal lines indicate means of four independent assays (n=4). Vertical bars indicated standard deviation. Plasmid pACBS lacks the λRed genes found in pACBSR.
Fig. S8. DNA Sanger sequencing of the plasmid preparation of pGEC\textsubscript{mat}EcN-lux used for integration of lux operon in EcN. The mutated nucleotide (G to A) in the start codon of luxC gene is labelled with an arrow.