Supporting Information

Massive integration of large gene libraries in the chromosome of *Escherichia coli*

Lidia Cerdán[1], Beatriz Álvarez[1] and Luis Ángel Fernández[1]*

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

*For correspondence

E-mail: lafdez@cnb.csic.es; Phone:+34 91 585 48 54; Fax:+34 91 585 45 06;

This file includes:

- Supporting Tables S1, S2, and S3.
- Supporting Figures S1, S2, and S3.
- Supporting Experimental procedures.
- Supporting References.

Supporting Table S1. *E. coli* strains used in this work.

Name	Description	References
DH10B-T1^{R}	F- *mcrA* Δ*mrr-hsdRMS-mcrBC* φ*80lacZDM15* Δ*lacX74 recA1 endA1 araD139* Δ*(ara,leu) 7697 galU galK rpsL (Str^{R}) nupG tonA* λ^{-}	(Durfee, et al., 2008)
CC118-λpir	Δ(*ara-leu*) *araD* Δ*lacX74 galE galK phoA20 thi- rpsE rpoB argE(Am) recA1, λpir*	(Herrero, et al., 1990)
T-SACK	W3110 *araD<>tetA-sacB-amp fliC<>cat argG*::Tn5	(Li, et al., 2013)
MG1655	K-12 (F^{-} λ^{-})	(Blattner, et al., 1997)
EcM1	MG1655 Δ*fimA-H rpsL (Str^{R}) rpoB* (Rif^{R})	(Blomfield, et al., 1991, Salema, et al., 2013)
EcM1-SAgfp	EcM1Δ*flu*::P$_{N25}$-Neae [Intimin$_{EHEC}$(1-654)-E-tag]-Vgfp-myc	(Piñero-Lambea, et al., 2015)
EcM1ΔlacI	EcM1 Δ*lacI*	This work
EcM1-Ptac-NVgfp	EcM1 Δ*lacI* Δ*flu*:: *lacI^{q}* P*tac*-RBS$_{T7}$-Neae[Intimin$_{EHEC}$(1-654)-E-tag]-Vgfp-myc; [Apra^{R}]	This work
EcM1-NL	EcM1 Δ*lacI* Δ*flu*:: *lacI^{q}* P*tac*-RBS$_{T7}$-Neae[Intimin$_{EHEC}$(1-654)-E-tag]-*tetA-sacB*; [Tet^{R}]	This work
EcM1-NL-Nb(1 - 7) EGFR	EcM1-NL derivates displaying Nbs 1 to 7 binding EGFR; [Apra^{R}]	This work

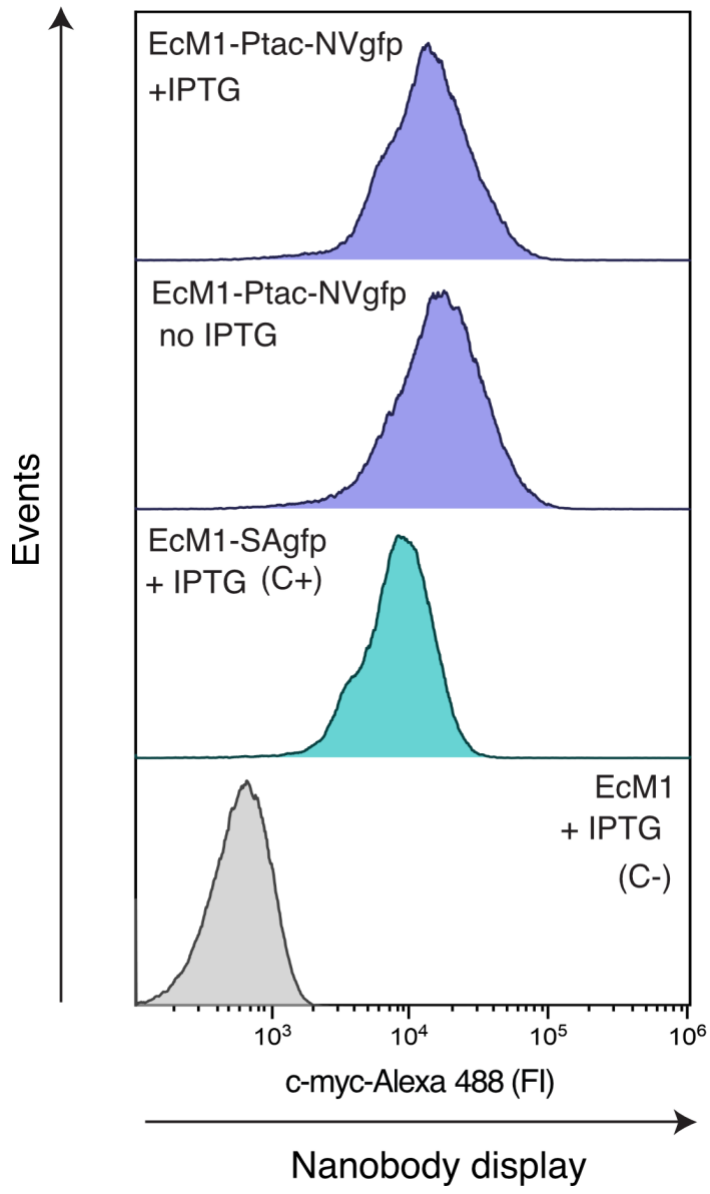
Supporting Table S2. Plasmids used in this work.

Name	Antibiotic resistance	Characteristics	Reference / Genebank
pACBSR	Cm^R	p15A ori, P_{BAD} promoter, *I-SceI* and λRed genes	(Herring, et al., 2003)
pNeae2	Cm^R	pBR322 ori, *lacI*^{q}-Plac promoter, *Neae* [Intimin$_{EHEC}$(1-654)-E-tag]-6xHis-myc tag	(Salema, et al., 2013)
pNVgfp	Cm^R	pNeae2, *lacI*^{q}-Plac promoter, Neae [Intimin$_{EHEC}$(1-654)-E-tag]-Vgfp-myc tag.	(Salema, et al., 2013)
pNeae2-V_{HH} EGFR Library	Cm^R	pNeae2, *lacI*^{q}-Plac promoter, Neae [Intimin$_{EHEC}$(1-654)-E-tag]-VHH-myc tag	(Salema, et al., 2016)
pGE	Km^R	R6K ori, Km^R, muticloning site with I-*Sce*I sites	(Piñero-Lambea, et al., 2015)
pGEΔ*lacI*	Km^R	pGE, Km^R, HRs for deletion of *lacI-lacZ* in *E. coli* K-12	(Asensio-Calavia, et al., 2023)
pGETS	Km^R	pSC101-ts ori, Km^R, multicloning site with I-*Sce*I sites	(Ruano-Gallego, et al., 2015)
pGETSfluNVgfp	Km^R $Apra^R$	pGETS, *flu* HR1, lacI^q Ptac-*Neae* [Intimin$_{EHEC}$(1-654)-E-tag]-*Vgfp*-myc, reversed T7, T0, *aac(3)IV*[$Apra^R$], *flu* HR2	This work / OR359883
pRecomb-TS-tetAsacB	Km^R Tet^R	pGETS, '*Neae* [Intimin$_{EHEC}$(493-654)-E-tag]-*tetA-sacB*, *flu* HR2	This work / OR359884
pRecomb-TS	Km^R $Apra^R$	pGETS, '*Neae* [Intimin$_{EHEC}$(493-654)-E-tag]-*xylE*-myc, reversed T7, T0, *aac(3)IV*[$Apra^R$], *flu* HR2	This work / OR359885
pRecomb-TS-Vgfp	Km^R $Apra^R$	pRecomb-TS, '*Neae* [Intimin$_{EHEC}$(493-654)-E-tag]-Vgfp-myc, reversed T7, T0, *aac(3)IV*[$Apra^R$], *flu* HR2	This work / OR359886
pRecomb-TS-V_{HH} EGFR Library	Km^R $Apra^R$	pRecomb-TS, '*Neae* [Intimin$_{EHEC}$(493-654)-E-tag]-VHH-myc, reversed T7, T0, *aac(3)IV*[$Apra^R$], *flu* HR2	This work

Supporting Table S3. Oligonucleotides used in this work.

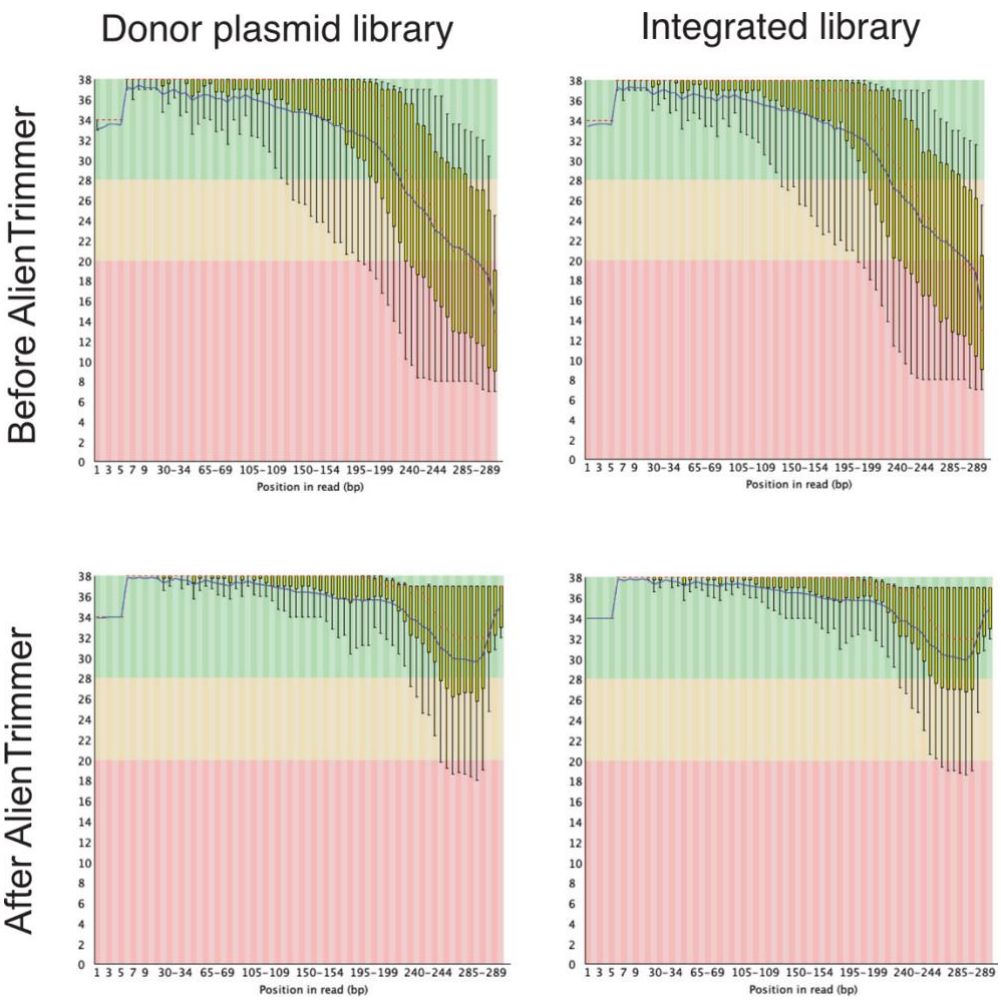
Name/ Forward (F) or Reverse(R)	Nucleotide sequence (5' -3')
5' SacI_XmaI_LacI_Plac (F)	CTGAGAGCTCACTGCCCCGGGCACCATCGAATGGCGCAAAACC
3'_SpeI_HindIII_BmtI_VHH (R)	GGAAACTAGTGTACGAAGCTTCACGAGCTAGCTTATGCAGCTGCATCCTCTTCTGAG
5'_HindIII_FRT_Apra (F)	CCAGAAGCTTGAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCTG CAGCTCACGGTAACTGATGCCG
3'_SpeI_FRT_Apra (R)	GGAAACTAGTGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAGGAACTTCGG AATAGGAACTTATGAGCTCAGCC
F_int_tetAsacB (F)	ATGGTGCGCCGGTGCCGTATCCGGATCCGCTGGAACCGGCCCAGCCGGCCTCCTAATT TTTGTTGACACTCTATC
R_int_tetAsacB (R)	TATCAACAGGAGTCCAAGgtggacTAATACGACTCACTATAGGGgctagcATCAAAGGGAAAA CTGTCCATATGC
eae2 (F)	GCTATAACGTCTTCATTGATCAGGAT
eae5 (F)	GACTTCAGCACTTAATGCCAGTGCGG
HindIII ter T0 (R)	TACGAAGCTTCTGGATTCTCACCAATAAAAAACGCC
VHH-Sfi2 (F)	GTCCTCGCAACTGCGGCCCAGCCGGCCATGGCTCAGGTGCAGCTGGTGGA
VHH-Not2 (R)	GGACTAGTGCGGCCGCTGAGGAGACGGTGACCTGGGT
CS1-E-tag (F)	ACACTGACGACATGGTTCTACAGATCCGCTGGAACCGGC
CS2-c-myc-tag (R)	TACGGTAGCAGAGACTTGGTCTAGCTGCATCCTCTTCTGAGATGAG
VEGFR1-CDR3 (R)	CCAGTAGTCATAGTCAACGCTCCG

Supporting Figure S1



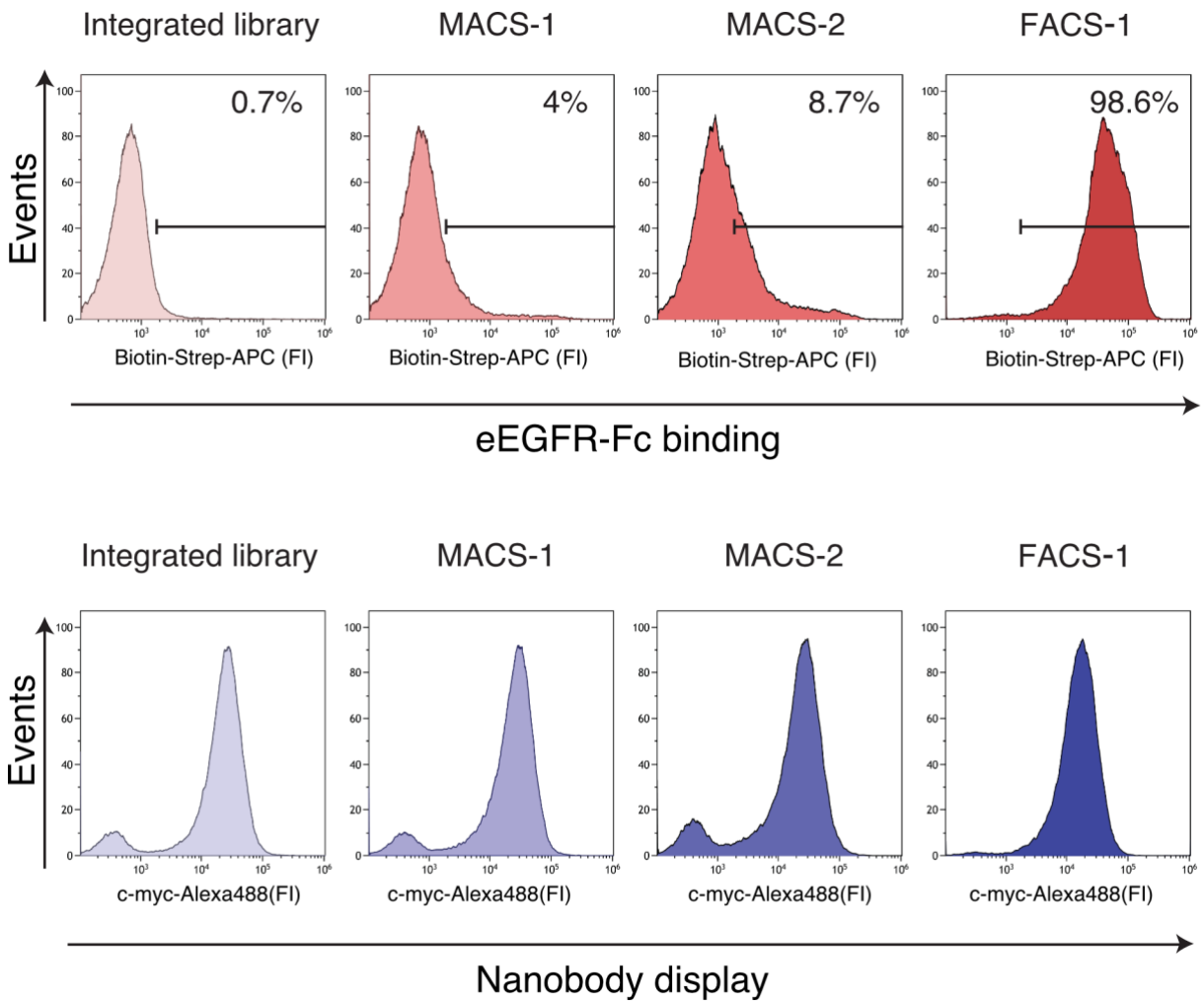
Supporting Figure S1. Surface expression levels of Nbs in EcM1-Ptac-NVgfp with and without IPTG induction. EcM1 bacteria with integrated Neae-Vgfp gene fusion, EcM1 Ptac-NVgfp, were analyzed for Nb display under two different conditions, with and without IPTG induction. Nb expression was detected with an anti-c-myc monoclonal antibody and a secondary goat anti-mouse IgG-Alexa 488 conjugate antibody. Arbitrary fluorescence intensity (FI) units are indicated. EcM1-NL and EcM1-SAgfp strains were used as negative (C-) and positive (C+) controls respectively.

Supporting Figure S2



Supporting Figure S2. Quality analysis of massive DNA sequencing reads from the plasmid and integrated V_{HH} library before and after data cleaning. Quality scores obtained with FASTAQC (Andrews, 2010) across all bases of the reads, before and after data cleaning with AlienTrimmer (Criscuolo and Brisse, 2013). Data of the two pools of V_{HH} sequences from plasmid and integrated library are shown. Quality thresholds are indicated by different colors: High quality ($Q \geq 28$) in green, medium quality ($20 < Q < 28$) in yellow and low quality ($Q \leq 20$) in red.

Supporting Figure S3



Supporting Figure S3. Evolution of the EcM1-NL-V_{HH} population along the selection process with eEGFR-Fc. Flow cytometry analysis of eEGFR-Fc binding (red, top) and Nb surface display (blue, bottom) of bacteria from the integrated V_{HH} library before (left) and after selections MACS-1, MACS-2, and FACS-1, as indicated. The percentage of positive bacteria for eEGFR-Fc binding is indicated for each population on the top right of the antigen-binding histograms**.** Stains for Nb display and antigen binding levels as in Figure 5. The concentration of biotinylated eEGFR-Fc used was 50 nM.

Supporting Experimental procedures

Preparation of biotinylated eEGFR-Fc and Fc proteins.

Purified eEGFR-Fc was from R&D Biosystems. Purified Fc was obtained after cleavage of Nb-Fc fusions (Casasnovas, et al., 2022) with thrombin (Merck) at 22°C O/N and affinity purification using a HiTrap™ IgGselect 1-ml column (GE Healthcare). Purified proteins in PBS (1×) were incubated for 2 h at RT with DMSO-dissolved biotin N-hydroxysuccinimide (Biotin-NHS ester, Sigma, Ref: B2643) at a molar ratio of 1:20 (antigen:label). After incubation, the conjugation reaction was blocked for 1 h at 4°C by adding Tris-HCl pH 7.5 at 50 mM final concentration. Finally, excess labeling molecules were removed using centrifugal filters with a 10-kDa cutoff (Amicon, Merck).

Construction of plasmids.

The plasmid backbones and their derivatives built in this study are listed in Supporting Table S2. The plasmids were constructed using standard techniques of PCR amplification, digestion with restriction enzymes and ligation (Ausubel, et al., 2002). The proofreading DNA polymerase Herculase II Fusion (Agilent Technologies) was used to amplify DNA fragments for cloning. The sequences of the primers used for the amplification of the main genetic elements of the plasmids are listed in Supporting Table S3.

pGETSfluNVgfp is derived from pGETSfluPtac-eLEE5 (Ruano-Gallego, et al., 2015). The expression cassette of Ptac-*eLEE5* was replaced by the fusion *lacI*^{q}-Plac-*Neae-Vgfp-myc*, encoding Intimin$_{EHEC}$ residues 1 to 654, followed by E-tag and Vgfp-myc, under the control of *lacI*^{q}-Plac promoter, which was amplified with the primer pair 5' SacI_XmaI_LacI_Plac / 3' SpeI_HindIII_BmtI_VHH using the plasmid pNVgfp as a template. The promoter Plac was replaced by the stronger promoter Ptac. Lastly, an apramycin resistance cassette was amplified with the primers 5'_HindIII_FRT_Apra / 3'_SpeI_FRT_Apra and cloned downstream of Neae-Vgfp.

pRecomb-TS-Vgfp is derived from pGETSfluNVgfp. The 5' *flu* HR1 and the expression cassette of lacI^{q}-Ptac-Neae-Vgfp-myc was replaced by the HR1 'Neae corresponding to Intimin$_{EHEC}$ residues 493 to 654, E-tag and Vgfp-myc, amplified with the primer pair 5'_XhoI_HR_int / 3' SpeI_HindIII_BmtI_VHH using the plasmid pGETSfluNVgfp as a template.

pRecomb-TS-tetAsacB is derived from pRecomb-TS-Vgfp**.** The Vgfp-Apra^{R} segment was replaced by the *tetAsacB* cassette amplified with the primer pair F_int_tetAsacB / R_int_tetAsacB using genomic DNA of the strain T-SACK as a template.

pRecomb-TS is derived from pRecomb-TS-Vgfp. pRecomb-TS-Vgfp was digested with *Sfi*I and *Not*I restriction enzymes to replace the Vgfp by a ~1 kb stuffer DNA from the *xylE* gene of *Pseudomona putida* mt-2 (de Lorenzo, et al., 1990) cloned between *Sfi*I and *Not*I restriction sites.

Construction of *E. coli* strains.

EcM1Δ*lac.* The genes *lacZ and lacI* from the *lac* operon were deleted using a strategy based on the generation of double-strand breaks *in vivo* with I-*Sce*I endonuclease. The EcM1 strain with pACBSR was transformed with the pGEΔlacI suicide vector (Km^{R}) (Asensio-Calavia, et al., 2023) carrying the HRs within the lac operon for its integration into the *E. coli* chromosome and the I-*Sce*I restriction sites. The resulting Cm^{R} / Km^{R} colonies were grown O/N at 37ºC with shaking (250 rpm) in LB supplemented with Km and Cm. The next day, the cultures were diluted 1:100 in LB supplemented with Cm and incubated at 37ºC and 250 rpm. When the cultures reached the exponential phase of growth (OD_{600} 0.4-0.6), 0.4% (w/v) L-arabinose was added and the cultures were incubated for a further 5 h to induce expression of I-*Sce*I and the cleavage of the chromosome at the integration site. This cleavage promoted a second step of homologous recombination that led to the deletion of the vector sequences and the corresponding gene. Individual colonies of the induced cultures were isolated on LB agar plates containing Cm. The colonies were streaked onto LB agar plates with and without Km to confirm that they are

susceptible to Km due to the loss of vector sequences. The gene deletion was assessed by PCR screening.

EcM1-Ptac-NVgfp. The EcM1ΔlacI strain with pACBSR was transformed with the plasmid pGETSfluNVgfp (Km^R), which allows integration of the cassette *lacI*^{Q}-Ptac-*Neae-Vgfp-myc* into the *flu* locus. The $Apra^R$ marker was incorporated to facilitate the selection of the integrants. Individual colonies were grown O/N in liquid LB medium with Km and Cm at 30°C and 250 rpm. The next day, the cultures were diluted 1:100 in LB with Apra and Cm and incubated under the same conditions until reaching an OD_{600} 0.4-0.6. Then, 0.4% (w/v) L-arabinose was added and the temperature was shifted to 37°C to prevent plasmid replication. The cultures were further grown for 5 h and then plated onto LB agar plates with Apra and Cm. Individual colonies were tested for sensitivity to Km as above and integration of the corresponding cassette was confirmed by PCR.

EcM1-NL. Bacteria EcM1-Ptac-NVgfp carrying pACBSR were transformed with plasmid pRecomb-TS-tetAsacB to replace the Vgfp coding sequence and $Apra^R$ marker with the *tetA-sacB* counter-selection cassette. We followed the same strategy as above except that tetracycline was used instead of apramycin to select integrants. The plasmid pACBSR was cured from the final strains by passaging the cultures in LB without antibiotic.

Supporting References

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

Asensio-Calavia, A., Ceballos-Munuera, Á., Méndez-Pérez, A., Álvarez, B., and Fernández, L. (2023) A tuneable genetic switch for tight control of tac promoters in *Escherichia coli* boosts expression of synthetic injectisomes, *Microb Biotechnol*. 00, 1–14. Available from: https://doi.org/10.1111/1751- 7915.14328

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (2002) *Short Protocols in Molecular Biology*. New York: John Wiley & Sons, Inc.

Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., et al. (1997) The complete genome sequence of *Escherichia coli* K-12, *Science* **277**: 1453-1462.

Blomfield, I.C., McClain, M.S., and Eisenstein, B.I. (1991) Type 1 fimbriae mutants of *Escherichia coli* K12: characterization of recognized afimbriate strains and construction of new fim deletion mutants, *Mol Microbiol* **5**: 1439-1445.
Casasnovas, J.M., Margolles, Y., Noriega, M.A., Guzmán, M., Arranz, R., Melero, R., et al. (2022) Nanobodies Protecting From Lethal SARS-CoV-2 Infection Target Receptor Binding Epitopes Preserved in Virus Variants Other Than Omicron, *Frontiers in Immunology* **13**: 863831.
Criscuolo, A., and Brisse, S. (2013) AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads, *Genomics* **102**: 500-506.
de Lorenzo, V., Herrero, M., Jakubzik, U., and Timmis, K.N. (1990) Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria, *J Bacteriol* **172**: 6568-6572.
Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., 3rd, Burland, V., Mau, B., et al. (2008) The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse, *J Bacteriol* **190**: 2597-2606.
Herrero, M., de Lorenzo, V., and Timmis, K.N. (1990) Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria, *J Bacteriol* **172**: 6557-6567.
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, X.T., Thomason, L.C., Sawitzke, J.A., Costantino, N., and Court, D.L. (2013) Positive and negative selection using the *tetA-sacB* cassette: recombineering and P1 transduction in Escherichia coli, *Nucleic Acids Res* **41**: e204.
Piñero-Lambea, C., Bodelón, G., Fernández-Periañez, R., Cuesta, A.M., Álvarez-Vallina, L., and Fernández, L.Á. (2015) Programming controlled adhesion of *E. coli* to target surfaces, cells, and tumors with synthetic adhesins, *ACS synthetic biology* **4**: 463-473.
Ruano-Gallego, D., Álvarez, B., and Fernández, L.A. (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.
Salema, V., Mañas, C., Cerdán, L., Piñero-Lambea, C., Marín, E., Roovers, R.C., et al. (2016) High affinity nanobodies against human epidermal growth factor receptor selected on cells by *E. coli* display, *MAbs* **8**: 1286–1301.
Salema, V., Marín, E., Martínez-Arteaga, R., Ruano-Gallego, D., Fraile, S., Margolles, Y., et al. (2013) Selection of single domain antibodies from immune libraries displayed on the surface of *E. coli* cells with two β-domains of opposite topologies, *PLoS ONE* **8**: e75126.