Received: 31 July 2023 | Accepted: 22 October 2023

DOI: 10.1111/1751-7915.14367

MICROBIAL BIOTECHNOLOGY Applied Microbiology International

RESEARCH ARTICLE

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

Lidia Cerdán | **Beatriz Álvarez** | **Luis Ángel Fernández**

Department of Microbial Biotechnology, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain

Correspondence
Luis Ángel Fernández, Department of Microbial Biotechnology, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CSIC), Campus UAM Cantoblanco, 28049 Madrid, Spain.
Email: lafdez@cnb.csic.es

Funding information
MCIN/AEI and NextGeneration EU/PRTR PLEC2021-007739; European Commission, Grant/Award Number: H2020 FET Open 965018-BIOCELLPHE; Ministerio de Ciencia e Innovación, Grant/Award Number: MCIN/AEI BIO2017-89081-R

Abstract

Large gene libraries are frequently created in *Escherichia coli* plasmids, which can induce cell toxicity and expression instability due to the high gene dosage. To address these limitations, gene libraries can be integrated in a single copy into the bacterial chromosome. Here, we describe an efficient system for the massive integration (MAIN) of large gene libraries in the *E. coli* chromosome that generates in-frame gene fusions that are expressed stably. MAIN uses a thermosensitive integrative plasmid that is linearized in vivo to promote extensive integration of the gene library via homologous recombination. Positive and negative selections efficiently remove bacteria lacking gene integration in the target site. We tested MAIN with a library of 10^7 V_{HH} genes that encode nanobodies (Nbs). The integration of V_{HH} genes into a custom target locus of the *E. coli* chromosome enabled stable expression and surface display of the Nbs. Next-generation DNA sequencing confirmed that MAIN preserved the diversity of the gene library after integration. Finally, we screened the integrated library to select Nbs that bind a specific antigen using magnetic and fluorescence-activated cell sorting. This allowed us to identify Nbs binding the epidermal growth factor receptor that were not previously isolated in a similar screening of a multicopy plasmid library. Our results demonstrate that MAIN enables large gene library integration into the *E. coli* chromosome, creating stably expressed in-frame fusions for functional screening.

INTRODUCTION

Cloning of large gene libraries from natural or synthetic repertoires is an essential step in combinatorial biology and directed evolution strategies for the selection of improved enzymes and antibodies (Bornscheuer et al., 2012; Simon et al., 2019; Tiller et al., 2017). *Escherichia coli* is the preferred host for the cloning of gene libraries because of its high efficiency of transformation and because of the variety of available expression vectors, which are mostly based on multicopy plasmids (Rosano & Ceccarelli, 2014). Nonetheless, the use multicopy plasmids confers a significant physiological burden for bacteria because of the elevated expression levels of the recombinant proteins (Dumon-Seignovert et al., 2004). Plasmid-based vectors are also inherently unstable and require selective pressure (e.g. antibiotics) to ensure maintenance. This can impose limitations on the selection of low-frequency clones from large gene libraries, for which loss of functional expression and overgrowth of non-expressing or plasmid-free clones should be minimized.

The insertion of genes into the bacterial chromosome provides a simple way to overcome many of the drawbacks of plasmid-based systems, by stabilizing the exogenous DNA and reducing gene dosage to a single copy (Kuhlman & Cox, 2010; St-Pierre et al., 2013; Zucca et al., 2013). Various techniques have been devised for integrating genes into the bacterial chromosome randomly or site specifically (Li et al., 2019; Ou

et al., 2018). Inserting genes at permissive sites is generally a good choice to ensure comparable bacterial growth and equitable comparison of gene expression levels across different clones. Site-specific transposases (e.g. Tn7) and integrases from bacteriophages (e.g. λInt) can be used to insert genes at specific attachment (*att*) sites in the chromosome (e.g. *attTn7*, *attB*) (Landy, 2015; Peters, 2019). Engineering these sites at various positions of the chromosome also enables integration at multiple loci (Egger et al., 2020). Another option is the use of transposon-encoded CRISPR-Cas systems, which utilize designed crRNA molecules to guide the insertion of the exogenous DNA into a custom target site in the genome (Vo et al., 2020). However, all these systems require the cargo DNA to be flanked with short DNA sequences (~20–40 bp) recognized by the transposase or integrase, which are inserted together with the cargo DNA. The presence of these 'scar' sequences complicates the generation of in-frame fusions of the exogenous DNA with a target gene in the chromosome.

More versatile scarless integration systems rely on homologous recombination between the chromosome and exogenous DNA molecules carrying homology regions (HRs) of the target site in the chromosome. Cellular and bacteriophage recombinases such as RecA and λRed can recombine the chromosome with the HRs of a circular or linear DNA that is introduced into the bacterium (Datsenko & Wanner, 2000; Kuhlman & Cox, 2010). As homologous recombination events are rare, double crossovers of the HRs flanking a cargo DNA must be positively selected, often with an antibiotic resistance (Ab^R) gene marker. These Ab^R markers may subsequently be removed by flanking them with short sequences (e.g. *FRT*) that are recognized by site-specific recombinases (e.g. FLP) (Datsenko & Wanner, 2000).

Strategies for markerless integrations have also been reported (Feher et al., 2008; Posfai et al., 1999). These integrations rely on the creation of double-strand breaks (DSBs) in the chromosome, which are lethal lesions for bacteria unless repaired by a homologous recombination with the exogenous DNA. DSBs can be induced in vivo through the expression of a restriction endonuclease such as I-SceI, which specifically recognizes a long target sequence that is absent in bacterial chromosomes but can be incorporated into a suicide plasmid vector carrying the HRs (Herring et al., 2003; Posfai et al., 1999). Following the initial homologous recombination event, which integrates the circular suicide plasmid vector, the chromosome becomes vulnerable to cleavage by I-SceI. The DSBs generated by the endonuclease in the chromosome are lethal for bacteria unless they are repaired by a subsequent homologous recombination event, which eliminates all vector sequences and leaves a markerless integration of the cargo DNA. Similar to I-SceI, DSBs can be generated by the expression of CRISPR-Cas9 nuclease and crRNAs directed at a target sequence of the chromosome (Wang et al., 2016; Wang et al., 2022).

Despite the diversity of available integration systems, there are no reports of the integration of large gene libraries into the chromosome of bacteria (Li et al., 2019). To the best of our knowledge, the largest integrated gene library in bacteria utilized random transposition (e.g. mini-Tn5) and contained ~1.4×10^5 independent clones (Scholz et al., 2019). Other integrated libraries are limited to a maximum of 3×10^4 clones (Biggs et al., 2020; Cowie et al., 2006; Elmore et al., 2017; Parisutham et al., 2022; Saleski et al., 2021). Developing efficient systems that can integrate large gene libraries would facilitate more effective screening of combinatorial libraries and the implementation of continuous selection processes in synthetic evolution (Simon et al., 2019).

In this study, we developed an efficient system for the massive integration (MAIN) of large gene libraries into the chromosome of *E. coli*. MAIN uses homologous recombination to generate scarless in-frame fusions with the target gene in the chromosome. These fusions produce stable protein expression after integration, allowing for continuous selection of clones with functional activity. MAIN is based on a thermosensitive suicide plasmid with I-*Sce*I sites, which generates a linear double-stranded DNA fragment for homologous recombination in vivo. The plasmid also has strong selection and counter-selection systems for removing non-integrants effectively. The linear DNA formed in vivo contains HRs flanking the cargo gene (segment) of interest (GOI), which lacks a transcriptional promoter to ensure expression only after its integration into the chromosome. As a proof of concept, we integrated into the chromosome of *E. coli* a library of V_{HH} genes raised against the human epidermal growth factor receptor (EGFR), containing ~1×10^7 independent clones (Salema et al., 2016).

V_{HH} domains are the variable domains of the heavy chain-only antibodies naturally found in camelids (Hamers-Casterman et al., 1993; Muyldermans et al., 2001). These V_{HH} gene segments encode functional single-domain antibody fragments called nanobodies (Nbs), which are attractive molecules for therapeutic and diagnostic applications (De Meyer et al., 2014; Jovčevska & Muyldermans, 2019; Muyldermans, 2013; Yang & Shah, 2020). We used MAIN to integrate a large V_{HH} library in the chromosome of an 'acceptor' *E. coli* strain. The strain was customized to contain a gene segment in its chromosome coding for the outer membrane (OM)-anchoring domain of intimin, also known as Neae (Bodelón et al., 2009; Salema et al., 2013). Gene integration led to in-frame fusions with Neae, which ultimately led to the display of Nbs on the bacterial surface. Massive DNA sequencing of integrated clones demonstrated that MAIN preserved

the clonal diversity of original gene library. Finally, we show that EGFR-binding Nbs can be selected by screening the integrated library for antigen-binding clones using magnetic and fluorescence-activated cell sorting (MACS and FACS) (Salema et al., 2016; Salema & Fernández, 2017). Taken together, our results show that MAIN facilitates the extensive integration of large gene libraries into the chromosome of *E. coli*, generating in-frame gene fusions that remain stable, and enabling the identification of clones of interest through functional screenings of the integrated library.

EXPERIMENTAL PROCEDURES

Bacterial strains, media and growth conditions

The *E. coli* strains used in this work are described in Supporting Table S1. The strain DH10BT1R (Durfee et al., 2008) was used for plasmid propagation and cloning. Bacterial strains were grown in liquid or solid agar Lysogeny broth (LB) media (Miller, 1992) at 30°C or 37°C, as indicated. Bacteria for electrocompetent cell preparation were grown at 37°C with shaking (250 rpm). Bacteria carrying cloned V_{HH} genes in the plasmids pNeae2 or pRecomb-TS (Supporting Table S2) were grown at 30°C with shaking (170 rpm), unless indicated otherwise. For preparation of LB solid medium, 1.5% (w/v) agar (Gibco, Thermo Fisher Scientific) was added. Bacteria grown on solid agar media were spread to obtain individual colonies in conventional Petri dishes (90 mm). In the case of libraries, bacteria were grown as lawns on large plates (150 mm) and serial dilutions were also plated on conventional Petri dishes for CFU counting. Starter liquid cultures of 10 mL were inoculated with individual colonies when working with single clones or, in the case of libraries, from a mixture of bacteria freshly harvested from plates (initial OD_{600} of the culture ~0.05). Starter cultures were grown overnight (O/N) at 30°C under static conditions, unless otherwise indicated. For depletion of non-integrants, bacteria were grown for 48 h at 37°C on Tet/SacB counter-selection agar medium, prepared as described previously (Li et al., 2013). This solid medium contains per litre: 15 g of agar, 4 g of tryptone, 4 g of yeast extract, 8 g of NaCl, 8 g of NaH_2PO_4 H_2O, 0.11 g $ZnCl_2$, 24 mg fusaric acid (Fus) and 60 g sucrose (Suc), referred in this work as Fus + Suc medium. When required, antibiotics and inducers were added at the following concentrations: ampicillin (Amp) at 150 μg/mL, chloramphenicol (Cm) at 30 μg/mL, kanamycin (Km) at 50 μg/mL, apramycin (Apra) at 50 μg/mL, tetracycline (Tet) at 15 μg/mL, isopropylthio-β-D-galactoside (IPTG) at 0.1 mM and L-Ara at 0.2% (w/v), unless otherwise indicated. Antibiotics were obtained from Duchefa Biochemie. Chemical reagents were obtained from Merck-Sigma.

Plates and starter liquid cultures of bacteria carrying derivatives of pNeae2 (Supporting Table S2) contained 2% (w/v) glucose to repress the *lac* promoter before induction. For IPTG induction, bacteria corresponding to an OD_{600} of 0.5 were harvested by centrifugation (4000*g*, 5 min) from a starter culture, washed twice with 1 volume of liquid LB and then re-suspended in 10 mL of LB with 0.1 mM IPTG. Induced bacteria were incubated at 30°C for 3 h with shaking (170 rpm).

Plasmids, DNA cloning and oligonucleotides

Plasmids used in this work are detailed in Supporting Table S2. Cloning procedures were performed using standard DNA techniques (Ausubel et al., 2002). Details of DNA constructs are described in Supporting Experimental procedures. All DNA constructs were sequenced using the chain-termination method (Macrogen). The thermosensitive plasmid pRecomb-TS (Km^R) is a derivative of pGETS (Km^R, pSC101-ts origin of replication) (Ruano-Gallego et al., 2015) used for cloning the V_{HH} gene library and chromosomal integration. The insert of pRecomb-TS comprises: ~500 bp HR1 of the 3′ end of intimin *N*-terminal domain ('Neae) corresponding to residues 493–654 from enterohemorrhagic *E. coli* (EHEC) intimin (Salema et al., 2013), ~1 kb stuffer DNA (*xylE*) between the unique *Sfi*I and *Not*I sites, a c-myc-epitope tag, a stop codon, a reverse T7 promoter (taatacgactcactataggg), a transcriptional terminator (T0) (https://parts.igem.org/Part:BBa_B0010), the Apramycin resistance ($Apra^R$) marker (Magalhaes & Blanchard, 2005) flanked by FRT sites and ~500 bp HR2 of the 3' end of *E. coli* K-12 *flu* gene (van der Woude & Henderson, 2008). Oligonucleotides used for DNA amplification and sequencing were synthesized by Sigma and are listed in Supporting Table S3.

Genome modifications of *E. coli* strains EcM1-Ptac-Vgfp and EcM1-NL

The *E. coli* strain EcM1-Ptac-NVgfp (Supporting Table S1) was generated from an EcM1Δ*lacI* strain (Supporting Experimental procedures) by integrating a gene cassette containing the $lacI^q$-*Ptac* region into the chromosomal *flu* locus (Amann et al., 1988), controlling expression of an intimin Neae-Nb fusion binding GFP (NVgfp) (Salema et al., 2013) followed by the T0 terminator and the $Apra^R$ marker. This integration was performed using the thermosensitive plasmid pGETS*flu*NVgfp-$Apra^R$ (Supporting Table S2). The acceptor strain EcM1-NL originated from the EcM1-Ptac-NVgfp strain by replacement of the DNA region comprising Vgfp and $Apra^R$ with the *tetA-sacB* counter-selection cassette (Li et al., 2013)

using pRecomb-TS-tetAsacB (Supporting Table S2). Details of construction of EcM1-Ptac-NVgfp and EcM1-NL strains are described in the Supporting Experimental procedures.

Cloning of the V_{HH} gene library

The DNA sequences of V_{HH}s from the immune library against human EGFR were excised from the plasmid pool of the pNeae2-V_{HH} EGFR library (Salema et al., 2016) by *Sfi*I and *Not*I digestion and cloned into the same sites of pRecomb-TS (Km^R), replacing the stuffer DNA *xylE*. Digested V_{HH} DNA fragments were run on an agarose gel (1% w/v), stained with SYBR safe and visualized under a blue light transilluminator (Thermo Fisher Scientific). The band of ~400 bp corresponding to the V_{HH} fragments was excised and purified using the PureLink™ Quick Gel Extraction Kit (Thermo Fisher Scientific). Five hundred nanograms of purified V_{HH} fragments (in ddH_2O) were ligated with the backbone of pRecomb-TS vector (Supporting Table S2) previously digested with *Sfi*I (Thermo Fisher Scientific) and *Not*I (New England Biolabs) and gel purified as earlier. The ligation reaction was prepared in a final volume of 0.2 mL at a 3:1 insert:vector molar ratio using a final vector concentration of ~2 ng/μL and 5 U of T4 DNA ligase (Merck–Roche). After O/N incubation at 16°C, the ligation products were ethanol precipitated and re-suspended in ddH_2O to a final DNA concentration of ~50 ng/μL. Ligation products were electroporated into *E. coli* DH10B-$T1^R$ cells (~500 ng of DNA per aliquot of 100 μL of competent cells). Six 100 μL aliquots of electrocompetent cells were used to reach the library size of $\sim 1.3 \times 10^7$ clones. Plasmids from the library were purified (Midi-prep kit, Qiagen) from the pool of transformed *E. coli* DH10B-$T1^R$ (grown as a lawn on LB agar) and transferred by electroporation into *E. coli* EcM1-NL electrocompetent cells containing the pACBSR (Cm^R) (Herring et al., 2003) to obtain $>1.3 \times 10^7$ clones. Library size was determined by plating serial dilutions on LB-Km or LB-Km-Cm plates followed by CFU counting.

Integration of the V_{HH} gene library in the EcM1-NL chromosome

EcM1-NL/pACBSR transformants carrying the pRecomb-TS-V_{HH} EGFR library (Supporting Table S2) were harvested from 150-mm LB-Km-Cm plates and used to prepare a single 5 mL starter LB liquid culture supplemented with Apra and Cm at an initial OD_{600} of 0.05. This culture was incubated at 30°C with shaking (250 rpm) until reaching an OD_{600} of 0.5. The culture was then induced for integration by adding L-Ara at 0.2% (w/v) and the temperature was increased to 37°C. After 3.5 h of induction, 2 mL of bacterial culture was plated as a lawn (0.5 mL/plate) on LB-Apra-Cm 150-mm plates (First selection step). In parallel, serial dilutions were plated on conventional Petri dishes with the same medium for CFU counting. Once a library size of $>1.3 \times 10^7$ clones was obtained, bacteria were harvested from the plates using 4 mL of liquid LB medium per plate and a cell spreader (Drigalski spatula). Finally, 1.5 mL of harvested bacteria were re-plated as a lawn (0.5 mL/plate) on Fus + Suc medium 150-mm plates for depletion of non-integrant bacteria by counter-selection (Second selection step). In parallel, serial dilutions were plated on conventional Petri dishes with the same selective medium. After 48 h of incubation at 37°C, the library size was assessed by CFU counting ($>1.3 \times 10^7$) and the bacterial lawn was harvested.

High-throughput DNA sequencing of V_{HH} sequences for diversity analysis

High-throughput DNA sequencing of the V_{HH} sequences was performed to analyse the diversity of the V_{HH} library. To do this, the V_{HH} sequences (~400 bp) were amplified by PCR using primers CS1-E-tag and CS2-c-myc-tag (Supporting Table S3). The DNA template was the plasmid pRecomb-TS-V_{HH} EGFR library (before integration) or purified chromosomal DNA from the integrated EcM1-NL-V_{HH} EGFR library (Supporting Table S1). Plasmid Midi-prep kit (Qiagen) and GENOME® DNA isolation kit (MP Biomedicals) were used for the purification of plasmid and genomic DNA respectively. Each PCR reaction contained the following: 1.5 μL of CS1-E-tag oligo at a concentration of 10 μM, 1.5 μL of reverse CS2-c-myc-tag oligo at a concentration of 10 μM, 5 μL of GoTaq® G2 Flexi Reaction Buffer (10×) (Promega), 4.5 μL of $MgCl_2$ (Promega), 1 μL of dNTP mix (dA;dC;dG;dT) at a concentration of 2.5 mM, 0.5 μL of GoTaq® G2 Flexi DNA Polymerase (Promega) and ddH_2O to a final volume of 50 μL. For these reactions, 18.9 pg of the V_{HH} library in pRecomb-TS plasmid or 31.2 ng of genomic DNA from the V_{HH} library integrated in EcM1-NL were used as templates in the PCR. These amounts correspond to $\sim 3 \times 10^5$ and 6×10^5 molecules of pRecomb-TS-NL (5500 bp) and *E. coli* chromosomal DNA (4.6×10^6 bp) respectively. The PCR programme was: 1 cycle of 2 min at 94°C, 30 cycles of 1 min at 94°C and 2 min at 72°C (amplification) and a final cycle of 10 min at 72°C. The amplicons of ~500 bp were run into agarose gels (0.8% w/v) and the corresponding band was excised and purified with PureLink™ Quick Gel Extraction Kit (Thermo Fisher Scientific) and sequenced using the Illumina Miseq platform with paired end (length $>2 \times 300$ bp) to acquire ~500,000 reads per sample. This was performed in the Genomic Unit service of the Technologic Park of Madrid.

For this study, ~500,000 reverse reads containing the CDR3 regions of the V_{HH} sequences were analysed per pool (pool 1 = Plasmid library, pool 2 = Integrated library). The reads were filtered and adaptor trimmed to improve their quality using AlienTrimmer 0.4.0 software (Criscuolo & Brisse, 2013) with the following parameters: minimum quality (q) = 28, conservativity (k) = 10, maximum mismatch (m) = 10 and minimum length (l) = 175. The quality of the reads was determined by FastQC software (Leggett et al., 2013; Patel & Jain, 2012) before and after trimming. Improved sequences were then clustered based on their sequence identity (ID) using the CD-HIT-EST online server (Huang et al., 2010) and setting an ID threshold of 98%. Obtained data were subsequently used to estimate the diversity of the library using two different parameters: (i) the diversity value (DV); [DV = (total no. of clusters/total analysed reads) × 100] and (ii) the percentage (%) of low membership clusters. Arbitrary clusters with 20 or less members were considered as low membership clusters.

Escherichia coli magnetic cell sorting and fluorescence-activated cell sorting

We used magnetic cell sorting (MACS) and fluorescence-activated cell sorting (FACS) to select bacteria binding eEGFR-Fc (the ectodomain of human EGFR fused to the Fc domain of human IgG1) (R&D Systems). First, MACS was performed to enrich binders from the library. Bacteria of the integrated EcM1-NL-V_{HH} EGFR library were harvested from lawns grown on Fus + Suc 150 mm plates and used as inoculum of a 10-mL LB-Apra liquid culture at an initial OD_{600} of 0.5 (>100 times greater CFU than the clonal size of the library). After 3 h of growth with shaking (170 rpm) at 30°C (final OD_{600} ~1.5), ~5×10^8 bacteria were harvested by centrifugation (4000*g*, 3 min, RT) and washed twice with 2 mL of PBS (1×). Then, bacteria were re-suspended in 100 μL of PBS-BSA [PBS (1×) supplemented with 0.5% (w/v) BSA] and mixed with 100 μL of biotinylated eEGFR-Fc in the same buffer at a final concentration of 100 nM. Biotinylation of eEGFR-Fc protein was performed as described previously (Salema et al., 2016). After 1 h of incubation at RT, bacteria were harvested by centrifugation, washed twice with 2 mL of PBS-BSA and re-suspended in 100 μL of the same buffer with 20 μL of anti-biotin paramagnetic beads (Miltenyi Biotec). After 20 min of incubation at 4°C, bacteria were harvested by centrifugation, washed, re-suspended in 500 μL of PBS-BSA and loaded onto a MACS MS column (Miltenyi Biotec) that had been previously equilibrated with 500 μL of the PBS-BSA and then placed on an OctoMACS Separator (Miltenyi Biotec). Unbound bacteria were collected and the MACS column was washed 3 times with 500 μL of PBS-BSA. Unbound bacteria and washed volumes were combined as 'unbound fraction'. Next, the MACS column was removed from the separator and placed onto a collector tube for elution of the column-bound bacteria with 2 mL of LB. This 'bound fraction' was plated as a lawn on LB-Apra 150 mm plates (0.5 mL/plate) for later bacterial harvesting. In parallel, serial dilutions of bound and unbound fractions were plated on the same medium for CFU counting. For the subsequent MACS round, the bacterial lawn grown from the previous MACS was harvested and used as inoculum for a new starter culture, which was grown and processed as earlier.

After MACS, the final selection of EGFR-binding bacteria was performed by FACS. Bacteria captured by MACS were used to inoculate a starter LB liquid culture (OD_{600} ~0.05), which was incubated O/N at 30°C under static conditions followed by 2 h at the same temperature and 170 rpm (final OD_{600} ~1.5–2). Bacteria were harvested by centrifugation, washed and re-suspended as in the MACS analysis, followed by a double staining for surface expression and antigen binding as detailed below for flow cytometry analysis. Finally, samples were re-suspended in 1 mL of PBS and sorted on a FACS vintage SE sorter cytometer (Becton Dickinson). The bacterial population positive for both fluorophores was collected in a sterile tube containing 2 mL of liquid LB-Apra medium and grown as a lawn (0.5 mL/plate) on LB-Apra 150 mm plates. A total of ~1×10^7 bacteria were processed by FACS.

Flow cytometry analysis of bacteria for Nb surface display and antigen binding

The Nb display levels and antigen binding capacity of selected bacteria (individual clones or libraries) were analysed by flow cytometry. For these assays, ~10^9 *E. coli* cells from LB liquid cultures inoculated with a single colony (individual clones) or with lawn-harvested bacteria at an OD_{600} of ~0.5 (libraries) and grown for 3 h at 30°C with shaking (170 rpm) were harvested by centrifugation (4000*g*, 3 min, RT), washed 3 times with 500 μL of PBS (1×) and re-suspended in 400 μL of the same buffer. Then, 90 μL of cell suspension (~2×10^8 bacteria) was removed and incubated for 1 h at RT with 10 μL of primary antibody (for Nb display analysis) and/or labelled proteins diluted in PBS (1×) (for binding analysis). Mouse anti-c-myc monoclonal antibody (1:500; 9B11 clone; Cell Signalling Technology, Ref: 2276) was used for staining of bacteria displaying Nbs. Biotinylated eEGFR-Fc and human Fc (prepared as described in Supporting Experimental procedures) were used at 50 or 100 nM for antigen-binding analysis. Then, bacteria were washed 3 times with 500 μL of PBS (1×) and incubated for 1 h at 4°C in the dark with 100 μL of PBS (1×) containing the corresponding secondary reagent. Goat anti-mouse IgG-Alexa 488-conjugated polyclonal

antibody (1:500; Thermo Fisher Scientific, Ref: A11029) was used for the detection anti-myc mAb (Nb surface display) on bacteria, whereas Streptavidin-APC (1:100; Beckman Coulter, Ref: 733001) was used for the detection of biotinylated antigen on bacteria (antigen binding). Finally, samples were washed 3 times with PBS (1×) and re-suspended in 500 μL of the same buffer for analysis in a Gallios cytometer (Beckman Coulter). Around 5×10^4 bacteria were analysed per sample.

Identification and DNA sequencing of specific V_{HH} clones

The V_{HH} sequences of the specific anti-EGFR clones integrated in the EcM1-NL chromosome were amplified by colony PCR with oligonucleotides eae5 and HindIII-Ter T0 (Supporting Table S3) using the GoTaq® G2 Flexi DNA Polymerase (Promega). Obtained PCR products of ~600 bp were run into an agarose gel (0.8% w/v) and purified from agarose bands. The PCR amplicons were sequenced by the chain termination method (Macrogen) using the primer eae5. To identify low-frequency clones different from the most frequent clone of the library (Nb1-EGFR), individual colonies were screened by PCR to detect colonies carrying this Nb with the reverse oligo VEGFR1-CDR3 and the universal forward oligo VHH-Sfi2 (Supporting Table S3). Colony PCR reactions were performed using the NZYTaq II 2× Green Master Mix (NZYtech, Ref: MB358). Colonies that did not have amplification bands (~400 bp) were analysed by flow cytometry for antigen binding capacity and the cloned V_{HH} sequenced as described earlier.

RESULTS

Design of the MAIN system

We designed MAIN by combining the approach for scarless gene integration by homologous recombination with selection and counter-selection markers to efficiently recover integrants and eliminate non-integrants. The 'donor' vector component of MAIN is a thermosensitive plasmid with two I-SceI sites and a multiple cloning site (MCS) into which the GOI library is cloned flanked by the HR sequences for chromosomal insertion. Replication of the donor plasmid at low temperatures (30°C) enables the propagation and cloning of the gene library in *E. coli*. By shifting the incubation temperature to 37°C and inducing I-SceI endonuclease and λRed proteins (Exo, Bet, Gam) in vivo (Herring et al., 2003), plasmid replication is prevented and a linear double-stranded DNA fragment is generated. This fragment can undergo homologous recombination with the target site in the chromosome of the acceptor strain. To positively select integrants, we incorporated the apramycin resistance ($Apra^R$) gene *aac(3)IV* (Magalhaes & Blanchard, 2005) into the donor vector, downstream of the GOI (Figure 1). This marker is not commonly found in *E. coli* plasmids, enabling the transformation of $Apra^R$ strains with multiple vectors without further manipulation. Nonetheless, we added FRT sites flanking the *aac(3)IV* gene to facilitate its (optional) removal via FLP recombinase (Datsenko & Wanner, 2000). To effectively eliminate non-integrants, we utilized the counter-selection cassette *tetA-sacB* (Li

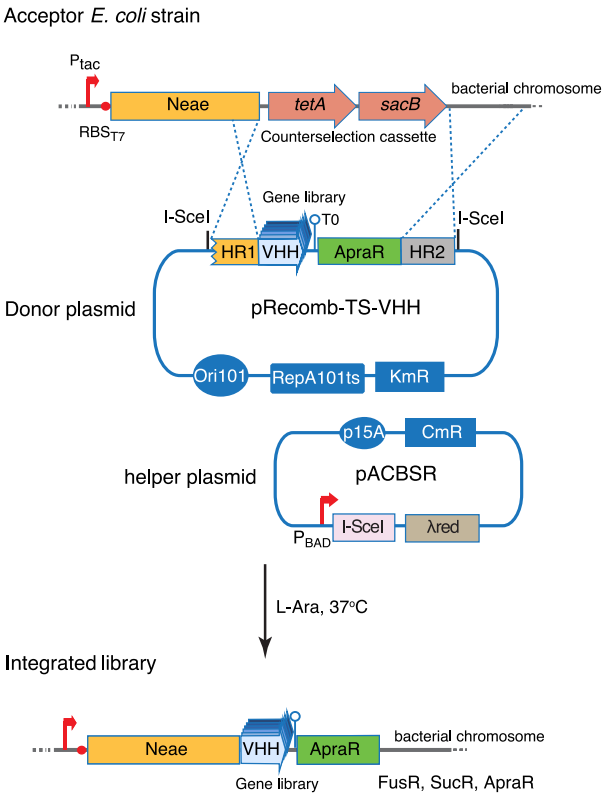

FIGURE 1 Schematic representation of the genetic elements comprising the MAIN system for integration of V_{HH} gene libraries. The acceptor *Escherichia coli* strain (EcM1-NL) contains the Neae gene construct and the counter-selection cassette *tetA-sacB* inserted into the chromosomal *flu* gene. The thermosensitive donor vector (pRecomb-TS) carries the cloned V_{HH} gene library and a downstream apramycin resistance marker ($Apra^R$) flanked by homology regions (HR1 and HR2) and I-SceI restriction sites. The inducible expression of I-SceI and the λ-Red products with L-Ara from a helper plasmid (pACBSR) mediates the in vivo digestion of the donor plasmid, releasing a linear DNA fragment (HR1-V_{HH}-$Apra^R$-HR2) that can be integrated into the chromosome of the acceptor strain by a double homologous recombination event. A temperature shift from 30°C to 37°C hinders further replication of undigested donor plasmids. The resulting bacteria carry the Neae-V_{HH} in-frame fusion integrated in single copy at the *flu* site of *E. coli* chromosome. The following regulatory elements are indicated: *tac* promoter (P_{tac}) and T7 ribosome binding site (RBS_{T7}) in the acceptor strain, the transcriptional terminator T0 (T0) and the thermosensitive replication (ori101, *repA101ts*) in the donor vector and the arabinose promoter (P_{BAD}) in the helper plasmid. The resistance (R) of the bacterial strains and plasmids to fusaric acid (Fus), sucrose (Suc), kanamycin (Km), apramycin (Apra) and chloramphenicol (Cm) are indicated.

et al., 2013). The *tetA-sacB* cassette contains TetA, a cytoplasmic membrane protein that provides resistance to tetracycline (Tet^R) and susceptibility to fusaric acid (Fus), as well as SacB, a levansucrase that transforms sucrose (Suc) to levan, which is toxic for the bacteria when accumulated in the periplasm. Accordingly, *E. coli* containing *tetA-sacB* in the chromosome is eliminated in a rich medium containing Fus and Suc. Because both proteins exert their toxic activities independently, the occurrence of spontaneous mutants resistant to the selective Fus + Suc medium ($<10^{-6}$) is orders of magnitude lower than with either of these counter-selection genes used alone (Li et al., 2013). This characteristic was employed for the effective elimination of bacteria lacking the expected homologous recombination; that is, when *tetA-sacB* was not replaced by the GOI.

Implementing donor vector and acceptor strain for the integration of a V_{HH} gene library

As a proof of concept of the MAIN system, we integrated a V_{HH} gene library of ~1×10^7 independent clones (Salema et al., 2016) into the chromosome of an acceptor *E. coli* strain (EcM1-NL; Supporting Table S1) that was modified with a genetic construct encoding the intimin Neae fragment (Salema et al., 2013; Salema & Fernández, 2017). This construct was integrated into the *flu* gene in the *E. coli* K-12 chromosome (van der Woude & Henderson, 2008) under the control of the Ptac promoter (de Boer et al., 1983) and a strong ribosome binding sequence from T7 bacteriophage (RBS_{T7}) (Figure 1). This allowed constitutive leaky expression of Neae-V_{HH} fusions in the absence of the inducer IPTG (Supporting Figure S1). We previously showed that the constitutive expression of Neae-V_{HH} gene fusions integrated into the *flu* site are stable and express the fusion proteins in the OM, displaying functional Nbs on the bacterial surface (Al-ramahi et al., 2021; Piñero-Lambea et al., 2015). Finally, the acceptor EcM1-NL strain was engineered with the *tetA-sacB* cassette downstream of the Neae gene segment (Figure 1) for counter-selection of bacteria for which a V_{HH} gene was not integrated replacing the *tetA-sacB* cassette.

The donor vector pRecomb-TS (Figure 1, Supporting Table S2) was based on the backbone of the thermosensitive plasmid pGETS (Ruano-Gallego et al., 2015), which contains a kanamycin resistance marker (Km^R), Ori101, RepA101(ts) and a MCS flanked by two I-SceI sites. Ori101 confers a low-copy phenotype (~5 copies per bacterium at 30°C) (Hashimoto-Gotoh & Sekiguchi, 1977) and exhibits class A theta replication (Lilly & Camps, 2015). The MCS of pRecomb-TS was modified with a DNA insert comprising the following: (1) the first homology region (HR1) of ~500 bp corresponding to the 3′ end of the Neae gene segment; (2) the GOI (i.e. a 1-kb stuffer DNA between unique *Sfi*I and *Not*I sites to be replaced by a 0.4-kb V_{HH} gene in frame with the upstream Neae); (3) a short DNA segment encoding an in-frame C-terminal myc-tag, a translation stop codon and a downstream transcriptional terminator (T0); (4) the $Apra^R$ gene marker with flanking FRT sequences and (5) the second homology region (HR2) of ~500 bp corresponding to the 3′ end of the *flu* locus of the *E. coli* K-12 chromosome. The whole DNA insert can be released in vivo by I-SceI cleavage, generating a linear DNA fragment carrying the HRs, the V_{HH} and $Apra^R$ gene, which can be integrated through a double recombination event in the target site of the chromosome of EcM1-NL (i.e. *Neae-tetA-sacB*). This results in the assembly of full-length in-frame Neae-V_{HH} fusions that are expressed from the chromosomal Ptac promoter (Figure 1). The donor plasmid lacks a promoter driving transcription of the cloned V_{HH}, which prevents potential toxicity due to Nb expression from this plasmid. Further, in the case of spurious transcription from the plasmid, the polypeptide encoding the Nb does not contain an RBS for translation and cannot be displayed on the bacterial surface as it is fused only to a truncated extracellular region of Neae (residues 493 to 654 of intimin), which lacks both the N-terminal signal peptide and the β-barrel domain required for OM localization (Bodelón et al., 2009; Fairman et al., 2012).

To minimize the size of the donor vector, the genes encoding the I-*Sce*I enzyme and λ-Red products were not included in pRecomb-TS and were instead co-expressed from the chloramphenicol-resistant (Cm^R) helper plasmid pACBSR (Figure 1, Supporting Table S2) upon induction with L-arabinose (L-Ara) (Herring et al., 2003). After L-Ara induction at the restrictive temperature (37°C), bacteria with the correct double recombination event were expected to have in-frame Neae-Nb fusions and the $Apra^R$ marker (Figure 1). These recombinants would have lost the *tetA-sacB* cassette, and therefore any remaining non-integrant bacteria in the $Apra^R$ population could be cleared in the Fus + Suc medium.

Validation of the MAIN system using an immune V_{HH} library against human EGFR

To validate the system, we chose an immune V_{HH} library against human EGFR that contains ~1.3×10^7 clones (Salema et al., 2016). The V_{HH} sequences were extracted from the original replicative plasmid library in pNeae2 (Salema et al., 2016) and ligated between the *Sfi*I and *Not*I sites of pRecomb-TS. A library of ~1.5×10^7 clones was produced in the highly competent *E. coli* DH10BT1R strain (Supporting Table S1) at 30°C, a permissive temperature for pRecomb-TS replication. The DNA plasmid library, named pRecomb-TS-V_{HH} EGFR (Supporting Table S2), was purified from

the bacterial pool and electroporated into the EcM1-NL acceptor strain carrying the pACBSR helper plasmid (Figure 2). We obtained ~8×10^7 independent transformants at 30°C in LB agar plates containing Cm and Km. To integrate the V_{HH} library, the bacteria were harvested and grown to exponential phase at 30°C in LB liquid culture supplemented with Cm and Apra. At this stage, Apra is employed instead of Km for donor plasmid selection since only the $Apra^R$ marker is maintained after integration. Subsequently, the expression of I-*Sce*I and λ-Red from pACBSR was induced with L-Ara and the temperature was raised to 37°C (Figure 2) to activate the in vivo generation of the linear DNA containing HR1-V_{HH}-$Apra^R$-HR2 for integration into the bacterial chromosome.

We performed two successive selection steps to select the integrant clones. In step one, bacteria were plated on large (150-mm) agar plates containing LB + Apra + Cm and were grown at 37°C for positive selection of bacteria with the $Apra^R$ marker integrated into the chromosome (Figure 2, step 5). Based on the parallel bacteria counting of serial dilutions, a total of ~2×10^7 Cm^R and $Apra^R$ colony-forming units (CFUs) grew on plates, a number greater than the size of the original V_{HH} library. After incubation, bacteria were collected from the plates and directly plated onto Fus + Suc agar medium for the second selection step, which depleted non-integrant bacteria containing the *tetA-sacB* cassette (Figure 2, step 6). Bacterial counting revealed a total of ~9×10^8 CFUs growing on plates with Fus + Suc, which again exceeded the size of the original library.

As the surface display of Nbs depends on the correct in-frame fusion of the V_{HH} to the chromosomal Neae gene segment, we analysed the percentage of bacteria displaying Nbs in the different bacterial populations to estimate the percentage of correctly integrated bacteria after each selection step. The Nb display was analysed by flow cytometry of the myc-tag located at the C-terminus of the Nb (Figure 3A). Results revealed that ~33% of bacteria obtained after the first selection step (Apra + Cm) displayed Nbs on their surface (Figure 3B), and this increased to ~88% after the counter-selection step (Fus + Suc medium) (Figure 3B). In this final population we found ~12% of bacteria negative for Nb display, which is similar to the number of non-expressing clones (~10%) found in the original library in pNeae2 (Figure 3B). Notably, similar fluorescence signals of Nb display were observed from the replicative high-copy plasmid pNeae2 (Salema et al., 2016) and from the

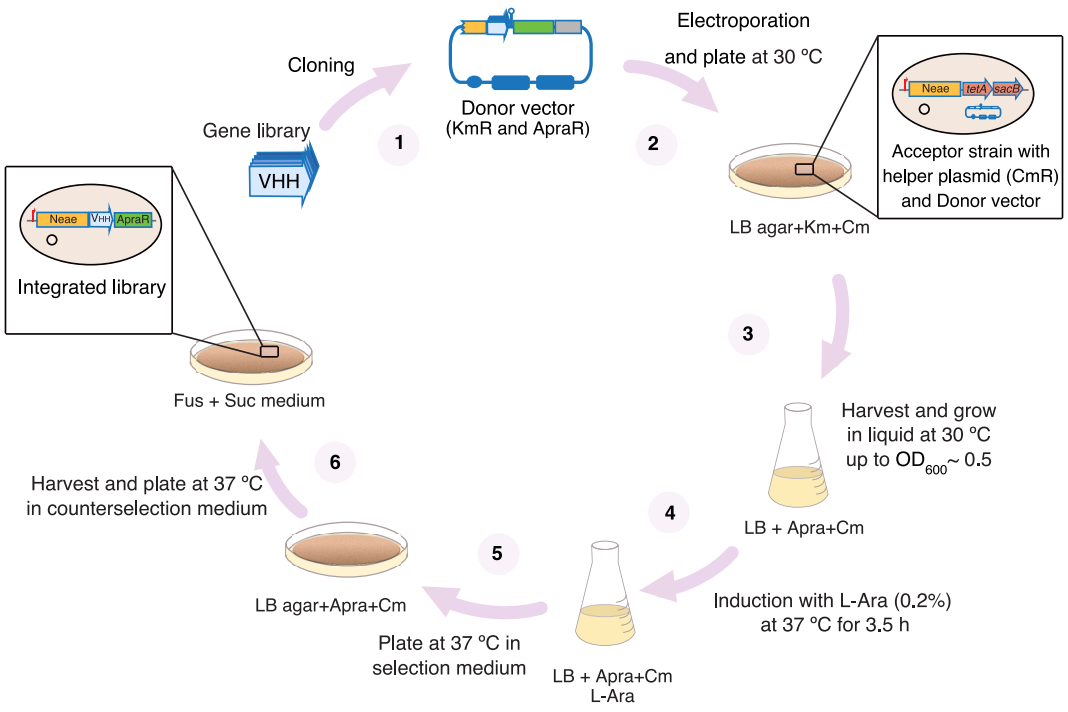

FIGURE 2 Scheme of the integration process for the V_{HH} gene library using MAIN system. A V_{HH} gene library was built in pRecomb-TS using *Escherichia coli* cloning strain DH10BT1R at 30°C (step 1). The pRecomb-TS-V_{HH} plasmid library was electroporated into the acceptor strain EcM1-NL carrying the pACBSR helper plasmid (step 2). Transformants with pRecomb-TS-V_{HH} library and pACBSR plasmid were harvested from plates and grown in liquid LB medium supplemented with Apra and Cm at 30°C up to exponential phase (step 3). The I-SceI meganuclease and λRed recombination system were induced from pACBSR with L-Ara 0.2% (w/v) and the temperature was increased to 37°C to hinder pRecomb-TS replication (step 4). After 3.5 h of induction bacteria were plated as a lawn in LB agar supplemented with Apra and Cm for the first selection step of bacteria containing the GOI-$Apra^R$ cassette integrated into the chromosome (step 5). Bacteria from the first selection step were harvested and directly plated onto Fus + Suc counter-selection medium to remove non-integrant bacteria carrying *the tetA-sacB* cassette (step 6).

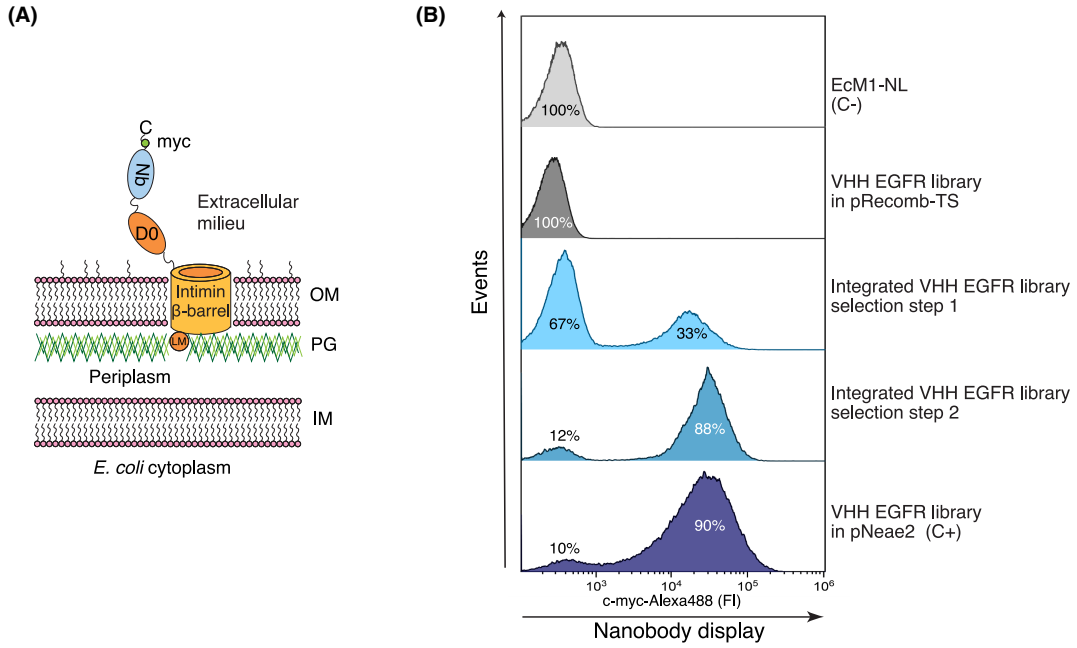

FIGURE 3 Nb display levels of the integrated library. (A) Scheme representing the display of Nbs on the outer membrane (OM), indicating the periplasmic LysM domain (LM) binding the peptidoglycan (PG), the β-barrel domain anchored in the OM and the extracellular Ig-like (D0) and Nb domains with C-terminal myc-tag (myc). (B) Flow cytometry analysis of Nb display levels in the indicated bacterial populations stained with an anti-myc monoclonal antibody and secondary anti-mouse IgG-Alexa488 conjugate. Bacteria analysed from top to bottom: EcM1-NL strain as a negative control (C−); EcM1-NL with pRecomb-TS-V_{HH} EGFR library grown at 30°C; integrated EcM1-NL-V_{HH} EGFR library grown at 37°C after the first selection step in Apra + Cm LB agar and after the second selection step in Fus + Suc medium; EcM1 with pNeae2-V_{HH} EGFR library grown at 30°C and induced with IPTG, as a positive control (Salema et al., 2016).

TABLE 1 Diversity of the plasmid and integrated libraries.

	Total reads	HQ reads	No. of clusters[a]	Diversity value[b] (%)
pRecomb-TS-V_{HH} library	697,458	518,361	50,813	9.80%
Integrated library	495,202	377,690	26,989	7.15%

[a]A cluster is defined as a group of sequences having ≥98% of identity.
[b]Diversity value (%) = (No. of clusters/High-quality reads) × 100.

single-copy integration in the bacterial chromosome (Figure 3B), which indicated a similar surface display level of Nbs in both expression systems. These results suggest that the integration process of the V_{HH} genes had occurred as anticipated, generating in-frame fusions that are displayed on the bacterial surface.

Diversity analysis of the integrated library

To assess whether the integration process influenced the library diversity, we performed high-throughput DNA sequencing to compare the sequence diversity of the V_{HH} library before and after integration. To achieve this, we obtained two ~400 bp amplicons by PCR that included the V_{HH}s from the library before (EcM1-NL with pRecomb-TS-V_{HH} EGFR library) and after integration (EcM1-NL-V_{HH} EGFR library after the second selection step). The DNA amplicons were sequenced on an Illumina Miseq platform utilizing paired ends (length ~2 × 300 bp) to acquire ~500,000 reads per sample (~5% of the estimated library size of ~1×10^7 clones). As the highest variability in the sequence of Nbs concentrates in the CDR3 region, which is located close to the 3′ end of the V_{HH}, only the reverse sequence reads were analysed. High-quality reads (~75% of the raw reads; Supporting Figure S2) were organized into clusters based on sequences with a ≥98% of sequence identity (ID). A diversity value (DV) was calculated by dividing the number of identified clusters by the total number of analysed sequences, expressed as a percentage. Thus, a DV of 100% would indicate that all the sequences in the collection are unique, whereas a decrease in the DV would indicate the presence of repeated sequences. With this approach two sequences with an ID ≥98% are considered the same. Data analysis revealed that the non-integrated plasmid library had a DV of 9.80%, whereas the integrated library had an estimated DV of 7.15% (Table 1). Therefore, only a small reduction in the DV (%) was found after integration,

indicating that the overall diversity of the V_{HH} library is maintained after its chromosomal integration.

Another variability parameter analysed was the size of the clusters. Clusters with a large number of members would indicate the presence of repeated V_{HH} sequences in the library, signifying reduced variability. We compared the pattern of size distribution of the clusters before and after chromosomal integration, finding that cluster distribution was similar in the plasmid and in the integrated library (Figure 4). We also observed that most clusters from the plasmid (~92%) and the integrated (~86%) libraries had less than 20 members, indicating a high-sequence diversity in both cases. In accordance with the DV values, the gene cluster distribution also indicated that the diversity of the V_{HH} library was well maintained after the MAIN process.

Selection of Nbs binding to EGFR antigen using the integrated library

We next evaluated whether Nbs binding EGFR could be selected from the integrated V_{HH} library. To do this, bacteria from the integrated EcM1-NL-V_{HH} EGFR library were incubated with the ectodomain of human EGFR fused to human Fc of IgG1 (eEGFR-Fc) labelled with biotin. Antigen-binding bacteria were enriched using MACS and FACS (Experimental procedures). Approximately 2×10^8 bacteria underwent two consecutive rounds of MACS using biotinylated eEGFR-Fc as bait. Enriched bacteria after the MACS cycles were subjected to a single round of FACS by incubation with biotinylated eEGFR-Fc and a c-myc monoclonal antibody, which enabled the selection of the double positive population in antigen binding and Nb display (Figure 5A).

Nb display levels and binding to biotinylated eEGFR-Fc of the bacterial populations were analysed by flow cytometry after the MACS and FACS cycles (Supporting Figure S3). The results demonstrated a steady gradual enrichment in positive eEGFR-Fc binders: from a baseline of ~0.7% of the integrated library to 8.7% after the second round of MACS. Notably, about 98.6% of bacteria showed positivity for antigen binding after FACS (Supporting Figure S3). Nb display levels remained constant in the library across MACS selections with a significant reduction in the number of non-expressing bacteria after the FACS (Supporting Figure S3).

For the isolation of specific EGFR binders, we analysed 195 individual colonies from the FACS selection by flow cytometry for eEGFR-Fc binding and V_{HH} sequence determination (see Experimental procedures). Eighty percent of the positive binders corresponded to a highly frequent clone also found in the original pNeae2 library, known as VEGFR1 (Salema et al., 2016), which we named Nb1-EGFR for this study. Six additional Nbs were identified and named sequentially from Nb2- to Nb7-EGFR based on their isolation frequency. The frequency of these Nbs and the amino acid sequence of their CDR3 region are summarized in Table 2. Nb3-EGFR was the only binder previously isolated from the original plasmid library (Salema et al., 2016). As determined by flow cytometry (Figure 5B), all of the Nbs identified in the screening of the integrated library showed specific binding to eEGFR-Fc and not to human Fc, as the non-specific control antigen. Notably, Nbs with both intermediate and high antigen-binding signals were identified (Figure 5B), suggesting that Nbs of different affinities were selected in this screening. Overall, these findings demonstrate that MAIN facilitates the expression of the gene fusions generated after the massive V_{HH} gene integration, enabling the screening of the integrated *E. coli* library to select Nbs with specific antigen-binding capabilities.

DISCUSSION

The MAIN system is designed to facilitate the efficient integration of large gene libraries into a target gene of the bacterial chromosome, resulting in large collections

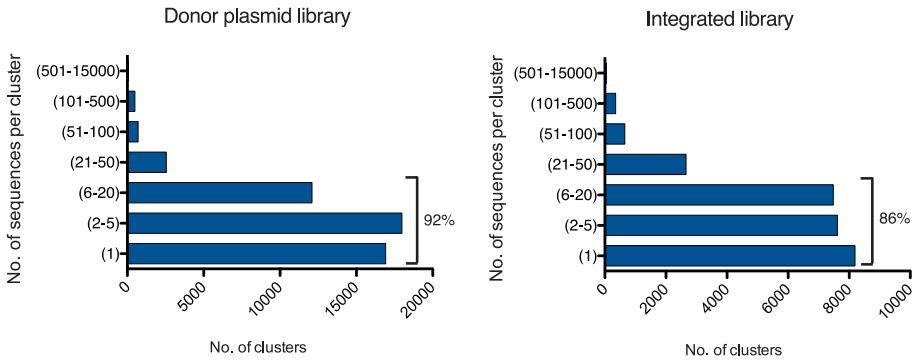

FIGURE 4 Cluster distribution of V_{HH} sequences in the library before and after integration. Graphs represent the cluster size distribution. Data of the V_{HH} library in plasmid (left) and integrated (right) are shown. The percentage of clusters with ≤20 members is indicated in each graph.

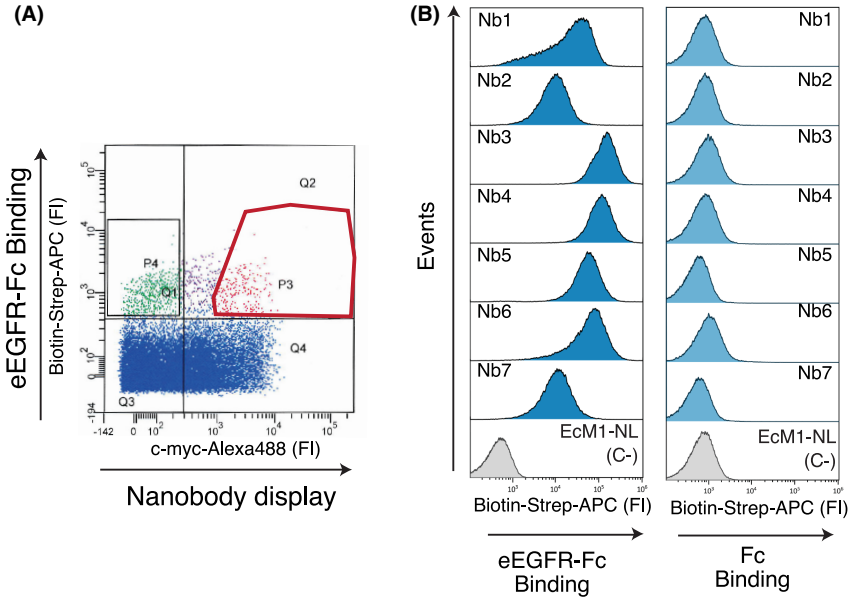

FIGURE 5 Selection of Nbs binding EGFR from the integrated V_{HH} library. (A) Fluorescence-activated cell sorting (FACS) for selection of bacteria displaying Nbs binding the eEGFR-Fc antigen. Histogram shows fluorescence intensity (FI) signals of bacteria from the integrated V_{HH} EGFR library enriched by MACS and incubated with biotinylated eEGFR-Fc (100 nM) and an anti-c-myc monoclonal antibody, followed by Streptavidin-APC and anti-mouse IgG-Alexa 488 as secondary reagents. The double-labelled sorted population is indicated as P3. (B) Flow cytometry analysis of *Escherichia coli* clones displaying anti-EGFR Nbs (Nb1–Nb7) identified from the integrated library. Binding analyses of bacterial clones against the target antigen (eEGFR-Fc) and a negative control antigen (human Fc) were performed to assess the specificity. The parental strain EcM1-NL was used as a negative control (C−). Nb display and antigen binding were stained as in A, but the concentration of biotinylated antigen (either eEGFR-Fc or Fc) was reduced to 50 nM.

TABLE 2 Nanobodies binding EGFR from the integrated *E. coli* library.

Nb	Frequency	CDR3
Nb1-EGFR	167/195	DKWSSSRRSVDYD
Nb2-EGFR	10/195	STTWGRPSYVYR
Nb3-EGFR	6/195	STYSRDTIFTKWANYN
Nb4-EGFR	3/195	DKWASSTRSIDYD
Nb5-EGFR	2/195	SRIIYSYVNYVNPGEYD
Nb6-EGFR	1/195	STYSRDTIFTNRANYN
Nb7-EGFR	1/195	DRRSTTDLKTLRAD

of in-frame fusions expressed in single copy. The MAIN system requires two customized elements: (1) a thermosensitive donor plasmid with I-SceI sites and the HRs of the target gene in the chromosome for cloning the library of the GOI and (2) an acceptor *E. coli* strain carrying the *tetA-sacB* cassette integrated into the target gene of the chromosome. Using these elements, we demonstrate that the MAIN system enables the MAIN of large gene libraries of $>10^7$ clones in a custom target gene of the *E. coli* chromosome. The acceptor *E. coli* strain used in our study is RecA+, a cellular recombinase that promotes homologous recombination (del Val et al., 2019). Nonetheless, since the Bet recombinase from λRed is co-expressed with I-SceI from the helper plasmid, homologous recombination between HRs of plasmid and chromosome could also occur in *E. coli* hosts having a *recA* mutation (Murphy Kenan, 2016).

MAIN exploits the in vivo expression of I-*Sce*I endonuclease, which linearizes the thermosensitive plasmid at I-SceI sites but not the bacterial chromosome, which does not contain I-SceI sites (Herring et al., 2003). Additionally, the co-expression of the λRed recombination system (*exo gam* and *bet* genes) from the helper plasmid protects the linear double-stranded DNA from host exonucleases and facilitate homologous recombination (Pines et al., 2015). We found no defect in the growth of bacteria carrying the helper plasmid pACBSR in the absence of the L-Ara inducer, and so curing this helper plasmid is not needed after the integration process. In vivo linearization of plasmids combines the advantages of using plasmids for propagation of the GOI and linear DNA for integration. This strategy was originally reported for the site-specific integration of a GOI into the bacterial chromosome (Herring et al., 2003). As the double recombination event occurs between two different homology regions (HR1 and HR2) flanking the GOI, the integration process is stable, mono-directional and irreversible.

The MAIN system includes positive selection of integrant bacteria using the $Apra^R$ marker (Magalhaes & Blanchard, 2005), located in the donor plasmid downstream of the GOI and the counter-selection cassette *tetA-sacB* (Li et al., 2013) in the target locus of the acceptor strain for removal of non-integrants. This

combination of positive selection using the antibiotic Apra and counter-selection of non-integrant bacteria on Fus + Suc medium leads to a highly efficient selection process, resulting in a bacterial population composed entirely of integrant clones. Our findings indicate that a single selection step with antibiotic (Apra) is not sufficient to obtain a large gene library composed entirely of integrant bacteria. We found that only 33% of the bacteria harvested from Apra plates were integrants correctly displaying the Nb. The high percentage of non-integrant bacteria after growth in Apra plates could be attributed to an insufficient selection pressure of the antibiotic on the plates due to the high density of bacteria plated when working with large libraries. However, we cannot rule out the presence of undigested plasmids in some of these bacteria or off-target integrations that could confer resistance to the antibiotic. We found that counter-selection in the Fus + Suc medium was crucial to completely eliminate the non-integrant bacteria still carrying *tetA-sacB* in the chromosome. The effectiveness of the *tetA-sacB* cassette in Fus + Suc medium is likely due to the low frequency of resistant bacteria ($\leq 6 \times 10^{-7}$) using this double counter-selection system (Li et al., 2013).

We have demonstrated the potential of the MAIN system by successfully integrating an immune V_{HH} library of $\sim 1 \times 10^7$ clones (Salema et al., 2016), representing the largest integrated library in a bacterial chromosome. Gene libraries of $\geq 10^7$ clones are essential in combinatorial biology and synthetic evolution strategies for the successful selection of novel peptides, enzymes and antibodies (Dufner et al., 2006; Löfblom, 2011; Rees, 2020; Simon et al., 2019). An important aspect of MAIN is its scalability by increasing the volume of *E. coli* cultures and the plating surface for $Apra^R$ selection and $Fus^R + Suc^R$ counter-selection. Here we used agar plates of 150 mm diameter for selection and counter-selections, but for the integration of larger libraries (e.g. $\sim 10^9$ clones) the plating surface should be increased to ensure that integrant clones are correctly selected. In theory, the maximum size of libraries that can be integrated using MAIN is only limited by the cloning efficiency of *E. coli* ($\sim 10^{10}$ clones for the largest reported libraries), provided that sufficient plating area is used for selections.

In our proof-of-concept study, the V_{HH} genes of the immune library were integrated in frame with the gene segment Neae to display the Nbs on the bacterial surface (Salema & Fernández, 2017). The Neae fragment corresponds to the N-terminal OM-anchoring domain (residues 1–654) of EHEC intimin (Bodelón et al., 2009; Fairman et al., 2012; Salema et al., 2013). Flow cytometry analysis revealed that ~12% of the integrants failed to correctly display the Nbs with the C-terminal myc tag. This was likely due to the presence of truncated V_{HH} genes (i.e. having premature stop codons and out-of-frame nucleotide insertions) in the original pNeae2 library (Salema et al., 2016) rather than generated during the integration process itself. Indeed, a similar number of non-expressing clones was detected by flow cytometry in the original pNeae2 library. As the display of Nbs on the bacterial surface depends on the correct in-frame fusion of the V_{HH} with the chromosomal Neae gene segment, any duplication or gene rearrangement during integration would result in a substantial reduction in Nb display levels in the integrated library. The comparable display levels between both libraries – the replicative pNeae2 and the integrated library – suggest that duplications or gene rearrangements after integration are either absent or extremely rare. This demonstrates the robustness of the MAIN system for creating in-frame fusions in the chromosome by homologous recombination, which is very useful for integrating gene libraries encoding specific protein domains while leaving the rest of the target gene unaffected.

Importantly, we have also demonstrated using high-throughput DNA sequencing that the MAIN system does not compromise the library diversity. In the case of V_{HH} and antibody libraries, high diversity is essential for the identification of high-affinity binders during the selection process, as a correlation exists between library size and the probability of finding high-affinity clones in the population (Bradbury & Marks, 2004). In this work, the large size of the integrated V_{HH} library allowed the identification of seven different Nbs that specifically bind EGFR. Remarkably, only two of these Nbs had previously been isolated from this immune V_{HH} library in the replicative plasmid pNeae2 (Salema et al., 2016). The identification of novel Nbs binders in the integrated library suggests that low-frequency V_{HH} sequences could be more effectively selected in the integrated library.

The expression of gene libraries from the bacterial chromosome results in more stable gene expression than with those constructed in high-copy plasmids. In our work, flow cytometry peaks associated with the Nb surface expression levels were more heterogeneous in bacteria carrying the pNeae2 plasmid than in the integrated library, where more uniform and narrower peaks were reproducibly observed. In fact, plasmid expression from the Plac promoter required IPTG for induction, whereas for the integrated library leaky expression from the Ptac promoter (Wilson et al., 2007) was sufficient to achieve good levels of expression in the absence of IPTG. Chromosomal integration of V_{HH} libraries displayed on *E. coli* could also facilitate downstream affinity maturation processes of selected Nbs by ssDNA recombineering (Al-ramahi et al., 2021) or the use of base deaminase-T7 RNA polymerase fusions (Álvarez et al., 2020), as pRecomb-TS contains a reverse T7 promoter downstream of the V_{HH} (Supporting Table S2).

Another important aspect of the MAIN system is its ability to be customized for different applications beyond

V_{HH} libraries. To do this, the HRs of pRecomb-TS donor plasmid need simply to be changed to those of the new target integration site in the chromosome. The acceptor *E. coli* strain should also be modified to contain the *tetA-sacB* cassette in the new integration site, as shown here for the synthetic *flu::Neae* site. MAIN could be applied to the chromosomal insertion of large gene libraries encoding enzymes of industrial interest (Intasian et al., 2021), to optimize screens from metagenomic studies (Ngara & Zhang, 2018) and for directed evolution approaches after in vitro or in vivo mutagenesis (Zeymer & Hilvert, 2018). Metabolic enzymes are often found in chromosomal operons and their optimization for metabolic engineering requires their correct expression within the operon (Fisher et al., 2014). MAIN allows gene libraries of enzymes to be inserted into their natural chromosomal context, ensuring balanced expression with other enzymes in the metabolic pathway. Other potential applications of MAIN include the integration of synthetic DNA sequences for information storage in bacterial populations (Bencurova et al., 2023; Hao et al., 2020) or the tagging of synthetic bacteria with DNA barcodes for tracking individual strains (Tellechea-Luzardo et al., 2022).

CONCLUSIONS

MAIN represents a powerful, scalable and customizable strategy for MAIN of large gene libraries into the *E. coli* chromosome, enabling the generation of precise gene fusions for the screening and selection of protein variants of interest. Adaptation of MAIN to other bacterial species beyond *E. coli* (and closely related enterobacteria) will require appropriate selection and counter-selection markers for the different bacteria, but the basic principles of I-SceI cleavage to linearize a conditional replicative plasmid for homologous recombination can be applied in many bacterial hosts amenable to genetic manipulation, such as *Pseudomonas* (Martínez-García & de Lorenzo, 2011), *Streptomyces* (Fernández-Martínez & Bibb, 2014), *Bacillus* (Wang et al., 2018), *Clostridium* (Zhang et al., 2015), *Lactobacillus* (Van Zyl et al., 2019) and *Mycoplasma* (Piñero-Lambea et al., 2022).

AUTHOR CONTRIBUTIONS

Lidia Cerdán: Conceptualization (equal); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); validation (supporting); visualization (lead); writing – original draft (equal); writing – review and editing (supporting). **Beatriz Alvarez:** Conceptualization (equal); formal analysis (supporting); investigation (supporting); methodology (supporting); supervision (supporting); validation (supporting); visualization (supporting); writing – original draft (equal); writing – review and editing (supporting). **Luis Ángel Fernández Herrero:** Conceptualization (equal); formal analysis (supporting); funding acquisition (lead); methodology (supporting); supervision (lead); validation (lead); writing – review and editing (lead).

ACKNOWLEDGEMENTS

We thank the excellent technical support of CNB-CSIC core scientific facility 'Flow Cytometry' and of the Flow Cytometry service at SidI-UAM. We also thank to the Genomic Unit of 'Parque Científico de Madrid' for their technical work in massive DNA sequencing. We thank Kenneth McCreath for the professional edition of English in this manuscript. This work was supported by research grants to L.A.F.: MCIN/AEI BIO2017-89081-R and NextGeneration EU/PRTR (PLEC2021-007739), and the European Union's Horizon 2020 Future and Emerging Technologies research and innovation programme (FET Open 965018-BIOCELLPHE).

FUNDING INFORMATION

This work was supported by research grants to L.A.F.: MCIN/AEI BIO2017-89081-R and NextGeneration EU/PRTR (PLEC2021-007739), and the European Union's Horizon 2020 Future and Emerging Technologies research and innovation programme (FET Open 965018-BIOCELLPHE).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

Sequencing data from high-throughput DNA sequencing experiments are deposited in Sequencing Read Archive (SRA) in the Bioproject ID PRJNA1000930 and are freely available from the following URL https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1000930.

AlienTrimmer 0.4.0 software (Criscuolo & Brisse, 2013) was used to trim adapter sequences and low-quality bases from raw Illumina reads. The software is freely available for download from the following URL https://bioweb.pasteur.fr/packages/pack@AlienTrimmer@0.4.0. The source code is also available upon request.

FastQC software (Andrews, 2010) was used to evaluate the quality of the trimmed reads. The software is freely available for download from the following URL: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. The source code is also available upon request.

CD-HIT-EST online server (Li & Godzik, 2006) was used to cluster the assembled transcripts. The server is freely available for use at the following URL https://sites.google.com/view/cd-hit/web-server. The online version is not available, but the program can be run as a command line tool or a local CD-HIT server can also be downloaded from this URL: https://github.com/weizhongli/cdhit-web-server.

GenBank accession numbers for plasmids: pGETS-fluNVgfp (OR359883), pRecomb-TS-tetAsacB (OR359884), pRecomb-TS (OR359885) and pRecomb-TS-Vgfp (OR359886).

ORCID

Lidia Cerdán https://orcid.org/0000-0003-4938-5848
Beatriz Álvarez https://orcid.org/0000-0002-9613-5473
Luis Ángel Fernández https://orcid.org/0000-0001-5920-0638

REFERENCES

Al-ramahi, Y., Nyerges, A., Margolles, Y., Cerdán, L., Ferenc, G., Pál, C. et al. (2021) ssDNA recombineering boosts *in vivo* evolution of nanobodies displayed on bacterial surfaces. *Communications Biology*, 4, 1169.

Álvarez, B., Mencía, M., de Lorenzo, V. & Fernández, L.Á. (2020) *In vivo* diversification of target genomic sites using processive base deaminase fusions blocked by dCas9. *Nature Communications*, 11, 6436.

Amann, E., Ochs, B. & Abel, K.J. (1988) Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene*, 69, 301–315.

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

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. et al. (2002) *Short protocols in molecular biology*. New York: John Wiley & Sons, Inc.

Bencurova, E., Akash, A., Dobson, R.C.J. & Dandekar, T. (2023) DNA storage – from natural biology to synthetic biology. *Computational and Structural Biotechnology Journal*, 21, 1227–1235.

Biggs, B.W., Bedore, S.R., Arvay, E., Huang, S., Subramanian, H., McIntyre, E.A. et al. (2020) Development of a genetic toolset for the highly engineerable and metabolically versatile *Acinetobacter baylyi* ADP1. *Nucleic Acids Research*, 48, 5169–5182.

Bodelón, G., Marín, E. & Fernández, L.Á. (2009) Role of periplasmic chaperones and BamA (YaeT/Omp85) in folding and secretion of intimin from enteropathogenic *Escherichia coli* strains. *Journal of Bacteriology*, 191, 5169–5179.

Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C. & Robins, K. (2012) Engineering the third wave of biocatalysis. *Nature*, 485, 185–194.

Bradbury, A.R. & Marks, J.D. (2004) Antibodies from phage antibody libraries. *Journal of Immunological Methods*, 290, 29–49.

Cowie, A., Cheng, J., Sibley, C.D., Fong, Y., Zaheer, R., Patten, C.L. et al. (2006) An integrated approach to functional genomics: construction of a novel reporter gene fusion library for *Sinorhizobium meliloti*. *Applied and Environmental Microbiology*, 72, 7156–7167.

Criscuolo, A. & 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.

Datsenko, K.A. & Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 6640–6645.

de Boer, H.A., Comstock, L.J. & Vasser, M. (1983) The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 80, 21–25.

De Meyer, T., Muyldermans, S. & Depicker, A. (2014) Nanobody-based products as research and diagnostic tools. *Trends in Biotechnology*, 32, 263–270.

del Val, E., Nasser, W., Abaibou, H. & Reverchon, S. (2019) RecA and DNA recombination: a review of molecular mechanisms. *Biochemical Society Transactions*, 47, 1511–1531.

Dufner, P., Jermutus, L. & Minter, R.R. (2006) Harnessing phage and ribosome display for antibody optimisation. *Trends in Biotechnology*, 24, 523–529.

Dumon-Seignovert, L., Cariot, G. & Vuillard, L. (2004) The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purification*, 37, 203–206.

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. *Journal of Bacteriology*, 190, 2597–2606.

Egger, E., Tauer, C., Cserjan-Puschmann, M., Grabherr, R. & Striedner, G. (2020) Fast and antibiotic free genome integration into *Escherichia coli* chromosome. *Scientific Reports*, 10, 16510.

Elmore, J.R., Furches, A., Wolff, G.N., Gorday, K. & Guss, A.M. (2017) Development of a high efficiency integration system and promoter library for rapid modification of *Pseudomonas putida* KT2440. *Metabolic Engineering Communications*, 5, 1–8.

Fairman, J.W., Dautin, N., Wojtowicz, D., Liu, W., Noinaj, N., Barnard, T.J. et al. (2012) Crystal structures of the outer membrane domain of intimin and invasin from enterohemorrhagic *E. coli* and enteropathogenic *Y. pseudotuberculosis*. *Structure*, 20, 1233–1243.

Feher, T., Karcagi, I., Gyorfy, Z., Umenhoffer, K., Csorgo, B. & Posfai, G. (2008) Scarless engineering of the *Escherichia coli* genome. *Methods in Molecular Biology*, 416, 251–259.

Fernández-Martínez, L.T. & Bibb, M.J. (2014) Use of the Meganuclease I-SceI of *Saccharomyces cerevisiae* to select for gene deletions in actinomycetes. *Scientific Reports*, 4, 7100.

Fisher, A.K., Freedman, B.G., Bevan, D.R. & Senger, R.S. (2014) A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories. *Computational and Structural Biotechnology Journal*, 11, 91–99.

Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E.B. et al. (1993) Naturally occurring antibodies devoid of light chains. *Nature*, 363, 446–448.

Hao, M., Qiao, H., Gao, Y., Wang, Z., Qiao, X., Chen, X. et al. (2020) A mixed culture of bacterial cells enables an economic DNA storage on a large scale. *Communications Biology*, 3, 416.

Hashimoto-Gotoh, T. & Sekiguchi, M. (1977) Mutations of temperature sensitivity in R plasmid pSC101. *Journal of Bacteriology*, 131, 405–412.

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

Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. (2010) CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, 26, 680–682.

Intasian, P., Prakinee, K., Phintha, A., Trisrivirat, D., Weeranoppanant, N., Wongnate, T. et al. (2021) Enzymes, *in vivo* biocatalysis, and metabolic engineering for enabling a circular economy and sustainability. *Chemical Reviews*, 121, 10367–10451.

Jovčevska, I. & Muyldermans, S. (2019) The therapeutic potential of Nanobodies. *BioDrugs*, 34, 11–26.

Kuhlman, T.E. & Cox, E.C. (2010) Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Research*, 38, e92.

Landy, A. (2015) The λ integrase site-specific recombination pathway. *Microbiology Spectrum*, 3, Mdna3-0051-2014.

Leggett, R.M., Ramirez-Gonzalez, R.H., Clavijo, B.J., Waite, D. & Davey, R.P. (2013) Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Frontiers in Genetics*, 4, 288.

Li, L., Liu, X., Wei, K., Lu, Y. & Jiang, W. (2019) Synthetic biology approaches for chromosomal integration of genes and pathways in industrial microbial systems. *Biotechnology Advances*, 37, 730–745.

Li, W. & Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22, 1658–1659.

Li, X.T., Thomason, L.C., Sawitzke, J.A., Costantino, N. & Court, D.L. (2013) Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in *Escherichia coli*. *Nucleic Acids Research*, 41, e204.

Lilly, J. & Camps, M. (2015) Mechanisms of theta plasmid replication. *Microbiology Spectrum*, 3, Plas-0029-2014.

Löfblom, J. (2011) Bacterial display in combinatorial protein engineering. *Biotechnology Journal*, 6, 1115–1129.

Magalhaes, M.L. & Blanchard, J.S. (2005) The kinetic mechanism of AAC3-IV aminoglycoside acetyltransferase from *Escherichia coli*. *Biochemistry*, 44, 16275–16283.

Martínez-García, E. & de Lorenzo, V. (2011) Engineering multiple genomic deletions in gram-negative bacteria: analysis of the multi-resistant antibiotic profile of pseudomonas putida KT2440. *Environmental Microbiology*, 13, 2702–2716.

Miller, J.H. (1992) *A short course in bacterial genetics: a laboratory manual and handbook for Escherichia coli and related bacteria*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Murphy Kenan, C. (2016) λ recombination and Recombineering. *EcoSal Plus*, 7. Available from: https://doi.org/10.1128/ecosalplus.ESP-0011-2015

Muyldermans, S. (2013) Nanobodies: natural single-domain antibodies. *Annual Review of Biochemistry*, 82, 775–797.

Muyldermans, S., Cambillau, C. & Wyns, L. (2001) Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains. *Trends in Biochemical Sciences*, 26, 230–235.

Ngara, T.R. & Zhang, H. (2018) Recent advances in function-based metagenomic screening. *Genomics, Proteomics & Bioinformatics*, 16, 405–415.

Ou, B., Garcia, C., Wang, Y., Zhang, W. & Zhu, G. (2018) Techniques for chromosomal integration and expression optimization in *Escherichia coli*. *Biotechnology and Bioengineering*, 115, 2467–2478.

Parisutham, V., Chhabra, S., Ali, M.Z. & Brewster, R.C. (2022) Tunable transcription factor library for robust quantification of regulatory properties in *Escherichia coli*. *Molecular Systems Biology*, 18, e10843.

Patel, R.K. & Jain, M. (2012) NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*, 7, e30619.

Peters, J.E. (2019) Targeted transposition with Tn7 elements: safe sites, mobile plasmids. *CRISPR/Cas and beyond, Mol Microbiol*, 112, 1635–1644.

Piñero-Lambea, C., Bodelón, G., Fernández-Periañez, R., Cuesta, A.M., Álvarez-Vallina, L. & 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.

Piñero-Lambea, C., Garcia-Ramallo, E., Miravet-Verde, S., Burgos, R., Scarpa, M., Serrano, L. et al. (2022) SURE editing: combining oligo-recombineering and programmable insertion/deletion of selection markers to efficiently edit the *Mycoplasma pneumoniae* genome. *Nucleic Acids Research*, 50, e127.

Pines, G., Freed, E.F., Winkler, J.D. & Gill, R.T. (2015) Bacterial Recombineering: genome engineering via phage-based homologous recombination. *ACS Synthetic Biology*, 4, 1176–1185.

Posfai, G., Kolisnychenko, V., Bereczki, Z. & Blattner, F.R. (1999) Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome. *Nucleic Acids Research*, 27, 4409–4415.

Rees, A.R. (2020) Understanding the human antibody repertoire. *MAbs*, 12, 1729683.

Rosano, G.L. & Ceccarelli, E.A. (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5, 172.

Ruano-Gallego, D., Álvarez, B. & 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. & Fernández, L.Á. (2017) *Escherichia coli* surface display for the selection of nanobodies. *Microbial Biotechnology*, 10, 1468–1484.

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.

Saleski, T.E., Chung, M.T., Carruthers, D.N., Khasbaatar, A., Kurabayashi, K. & Lin, X.N. (2021) Optimized gene expression from bacterial chromosome by high-throughput integration and screening. *Science Advances*, 7, eabe1767. Available from: https://doi.org/10.1126/sciadv.abe1767

Scholz, S.A., Diao, R., Wolfe, M.B., Fivenson, E.M., Lin, X.N. & Freddolino, P.L. (2019) High-resolution mapping of the *Escherichia coli* chromosome reveals positions of high and low transcription. *Cell Systems*, 8(212–225), 212–225.e9.

Simon, A.J., d'Oelsnitz, S. & Ellington, A.D. (2019) Synthetic evolution. *Nature Biotechnology*, 37, 730–743.

St-Pierre, F., Cui, L., Priest, D.G., Endy, D., Dodd, I.B. & Shearwin, K.E. (2013) One-step cloning and chromosomal integration of DNA. *ACS Synthetic Biology*, 2, 537–541.

Tellechea-Luzardo, J., Hobbs, L., Velázquez, E., Pelechova, L., Woods, S., de Lorenzo, V. et al. (2022) Versioning biological cells for trustworthy cell engineering. *Nature Communications*, 13, 765.

Tiller, K.E., Chowdhury, R., Li, T., Ludwig, S.D., Sen, S., Maranas, C.D. et al. (2017) Facile affinity maturation of antibody variable domains using natural diversity mutagenesis. *Frontiers in Immunology*, 8, 986.

van der Woude, M.W. & Henderson, I.R. (2008) Regulation and function of Ag43 (flu). *Annual Review of Microbiology*, 62, 153–169.

Van Zyl, W.F., Dicks, L.M.T. & Deane, S.M. (2019) Development of a novel selection/counter-selection system for chromosomal gene integrations and deletions in lactic acid bacteria. *BMC Molecular Biology*, 20, 10.

Vo, P.L.H., Ronda, C., Klompe, S.E., Chen, E.E., Acree, C., Wang, H.H. et al. (2020) CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nature Biotechnology*, 39, 480–489.

Wang, H., La Russa, M. & Qi, L.S. (2016) CRISPR/Cas9 in genome editing and beyond. *Annual Review of Biochemistry*, 85, 227–264.

Wang, J.Y., Pausch, P. & Doudna, J.A. (2022) Structural biology of CRISPR–Cas immunity and genome editing enzymes. *Nature Reviews Microbiology*, 20, 641–656.

Wang, T., Wang, D., Lyu, Y., Feng, E., Zhu, L., Liu, C. et al. (2018) Construction of a high-efficiency cloning system using the Golden Gate method and I-SceI endonuclease for

targeted gene replacement in *Bacillus anthracis*. *Journal of Biotechnology*, 271, 8–16.
Wilson, C.J., Zhan, H., Swint-Kruse, L. & Matthews, K.S. (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cellular and Molecular Life Sciences*, 64, 3–16.
Yang, E.Y. & Shah, K. (2020) Nanobodies: next generation of cancer diagnostics and therapeutics. *Frontiers in Oncology*, 10, 1182.
Zeymer, C. & Hilvert, D. (2018) Directed evolution of protein catalysts. *Annual Review of Biochemistry*, 87, 131–157.
Zhang, N., Shao, L., Jiang, Y., Gu, Y., Li, Q., Liu, J. et al. (2015) I-SceI-mediated scarless gene modification via allelic exchange in clostridium. *Journal of Microbiological Methods*, 108, 49–60.
Zucca, S., Pasotti, L., Politi, N., Cusella De Angelis, M.G. & Magni, P. (2013) A standard vector for the chromosomal integration and characterization of BioBrick™ parts in *Escherichia coli*. *Journal of Biological Engineering*, 7, 12.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Cerdán, L., Álvarez, B. & Fernández, L.Á. (2024) Massive integration of large gene libraries in the chromosome of *Escherichia coli*. *Microbial Biotechnology*, 17, e14367. Available from: https://doi.org/10.1111/1751-7915.14367