BMC Methods

PROTOCOL

Open Access

Gap-repair recombineering for efficient retrieval of large DNA fragments from BAC clones and manipulation of large high-copy number plasmids

Thomas C. Dix[1,2†], Anna Lassota[1,2†], Ulrike Brauer[1†], Irmgard U. Haussmann[3] and Matthias Soller[1,2,4*]

Abstract

Background Bacterial high-copy number plasmids are the preferred DNA source for reporter constructs used in cell transfection experiments and making transgenic invertebrates to study gene expression, develop gene therapies and biotechnological application. They can be quickly validated from small cultures and easily generated in large quantities. However, manipulating plasmids above 10 kb can become very tedious.

Method Here, we devised simple and highly efficient gap-repair recombineering methodology in *E. coli* to manipulate high-copy number plasmids up to 20 kb with up to 100% efficiency. This method utilises rare cutting restriction enzymes to introduce a gap, which is then subsequently repaired through homologous recombination from a provided template. Unlike traditional cloning methods, large concentration differences among fragments are tolerated. Moreover, CRISPR-Cas9-mediated in vitro DNA scission can be sufficiently efficient to overcome limitations from finding rare cutting restriction enzymes.

Discussion Gap-repair recombineering provides a significant advancement in generating recombinant high-copy number DNA plasmids through enhancing efficiency, speed, and robustness. We validated this technology by generating reporter transgenes of the highly repetitive *Drosophila Down Syndrome Cell Adhesion Molecule* (*Dscam*) gene to analyse alternative splicing.

Keywords Recombineering, High-copy number plasmid, Gap-repair, CRISPR-Cas9, *Drosophila*, *C. elegans*

Background

Efficient manipulation of recombinant DNA molecules is essential for the analysis of gene expression using transfection of reporter constructs into mammalian cell culture cells and to make transgenic organisms. Standard DNA ligase or Gibson assembly based cloning procedures are used for making smaller vectors such as plasmids, while manipulation of larger DNA molecules like bacterial artificial chromosomes (BACs) rely on homologous recombination in *E. coli*, termed recombineering (homologous recombination-based genetic engineering) [1, 2]. Recombineering is key to manipulation of BACs for use in mouse transgenesis [3, 4], but whether it can be

†Thomas C. Dix, Anna Lassota and Ulrike Brauer contributed equally to this work.

*Correspondence:
Matthias Soller
matthias.soller@manchester.ac.uk
[1] School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[2] Birmingham Centre for Genome Biology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[3] Department of Life Science, School of Health Sciences, Birmingham City University, Birmingham B5 3TN, UK
[4] Division of Molecular and Cellular Function, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK

applied to high-copy number (HCN) plasmids containing large inserts has not been explored.

Recombineering in *E. coli* is dependent on λ Red bacteriophage proteins Exo, Beta, and Gam. Gam acts as an inhibitor of *E. coli*'s endogenous RecBCD and SbcCD nucleases, which would otherwise counteract Exo and Beta activity. Exo is a 5' to 3' double stranded (ds) DNA exonuclease generating the single-stranded (ss) DNA 3' overhangs required homologous recombination initiation. Beta binds these 3' overhangs to protect them from degradation and facilitates their annealing to complementary DNA sequences [5].

HCN plasmids, such as *pUC19*, can be efficiently replicated in *E. coli* to quickly yield large quantities of high-quality DNA for cell transfection experiments and the generation of transgenic model organisms, such as *Drosophila* and *C. elegans*. Moreover, HCN plasmids can be easily manipulated, and alterations validated by restriction digests, making them the preferred option for recombinant DNA technology. As HCN plasmids are maintained at approximately 500–700 copies per cell, their size is limited to 20 kb (backbone plus insert) to avoid an intolerable metabolic burden, which if exceeded, reduces the plasmids final yield and quality [6–8]. Although 20 kb is sufficient for most applications, this maximal size is rarely utilised because cloning and further manipulation can become very difficult with plasmids above 8–10 kb. This is due to various factors, including difficulty to have equimolar concentrations of fragments, reduced transformation efficiency, lack of unique restriction sites and undesired recombination within the plasmid requiring identification of rare recombinant clones by filter-lifts and hybridisation of a ^{32}P-labelled probe [9–13].

To facilitate manipulation of large HCN plasmids, we developed a recombineering strategy based on the generation of a "gap", termed "gap-repair recombineering". Here, rare cutting restriction enzymes cut out a fragment and the resulting gap is then repaired by homologous recombination from a template provided. This method enables efficient editing of large HCN plasmids (up to 100% in some instances) and takes advantage of their maximal size, using the common DH5α *E. coli* strain [14]. In addition, we have demonstrated that in vitro DNA scission by CRISPR-Cas9 is efficient enough to be used for cloning [14].

Gap-repair recombineering exploits λ Red-mediated homologous recombination to generate the desired plasmid by simply transforming cells with a plasmid that encodes the λ Red proteins. The small, low-copy *pSC101-BAD-gbaA* plasmid can be efficiently transformed and possesses all three *λ Red* genes under an L-arabinose inducible operon [15]. This plasmid also encodes the RecA DNA repair enzyme to temporarily compensate for the *recA* minus genetic background of DH5α *E. coli* [16]. Expression is induced by the addition of L-arabinose to generate the molecular machinery for homologous recombination to retrieve a sequence from a BAC containing a large genomic fragment or gap-repair recombineering between a parent vector and donor plasmid.

As an example, we used a BAC containing the *Drosophila Down Syndrome Cell Adhesion Molecule* (*Dscam*) gene to retrieve the part encompassing exons 3 to 5 to generate a splicing reporter transgene (Fig. 1a and b). To retrieve a sequence from a BAC, two short (300–600 bp) homology arms of the beginning and the end of the sequence are cloned into this vector. These homology arms are designed to introduce a blunt cutting restriction site for vector linearization when fused together (Fig. 1a and b). For example, the 3'-end of the 5' homology arm ended with GAT, whereas the 5'-end of the 3' homology arm started with an ATC generating an *EcoR*V blunt cutting restriction site (Fig. 1a and b).

To make maximal use of plasmids for *Drosophila* transgenesis we generated a minimal *pUC19*-based transformation vector containing an *attP* attachment site for phiC31-mediated genome integration and a photoreceptor-specific *3xP3* promoter-driven GFP visible marker to identify transformant flies, flanked by *LoxP* sites for possible later removal [14]. Once integrated, heterologous expression of the transgene can be initiated by co-expressing the GAL4 transcriptional activator that will bind to the *Upstream Activation Sequence* (*UAS*) promoter [17].

To then introduce mutations into the parent vector we employed gap-repair recombineering between the parent vector and a smaller donor vector, where mutations in variable exons were introduced (Fig. 1c and d). For this manipulation of large HCN plasmids, a gap is generated in the parent vector by rare cutting restriction enzymes. The cut plasmid is transformed into recombineering electrocompetent cells alongside a dsDNA donor substrate (either a small easily manipulatable cut plasmid or PCR product) with short homology arms flanking the gap. The gap is then repaired by *λ Red* proteins, incorporating the desired mutations into the large HCN plasmid (Fig. 1c and d). Gaps can also be introduced by in vitro CRISPR-Cas9 DNA scission; however, sgRNAs need to be optimally designed to achieve complete digestion of the parent vector [18].

For gap-repair recombineering the *λ Red* genes are provided on plasmid *pSC101-BAD-gbaA,* which offers distinct advantages compared to the use of specific *E. coli* strains, which have the *λ Red* genes inserted into their genome (e.g. DY380). The temperature sensitive origin, SC101, in *pSC101-BAD-gbaA* enables complete removal

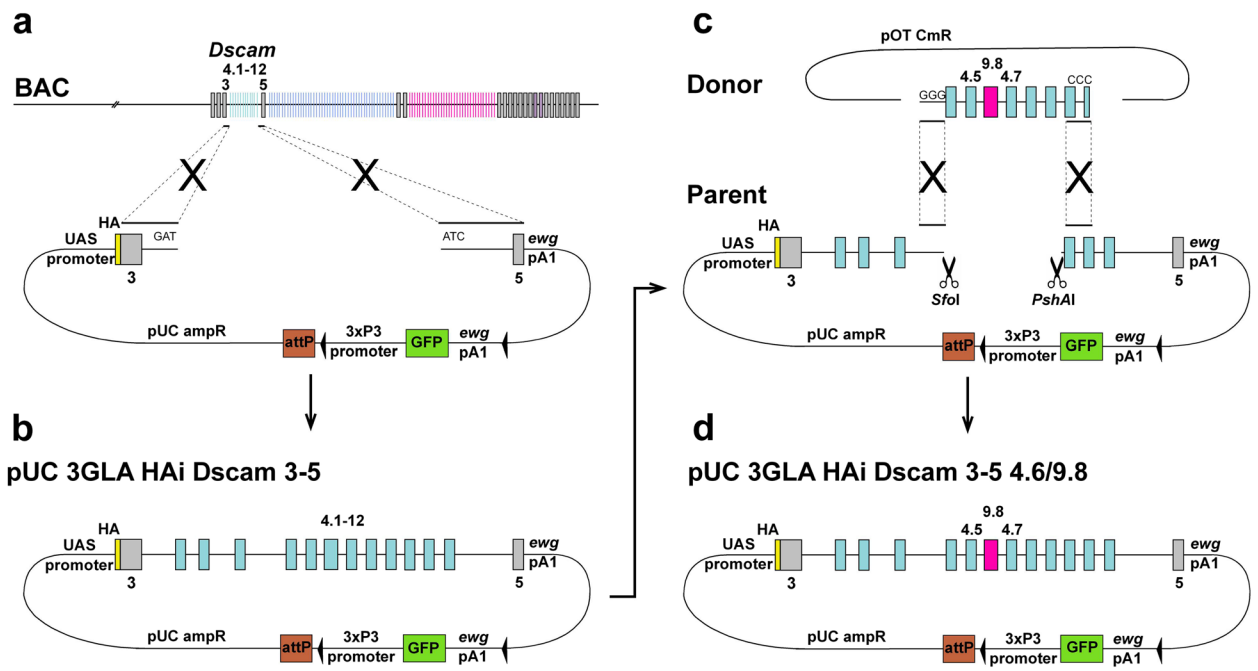

Fig. 1 Schematic for BAC sequence retrieval and gap-repair recombineering of large HCN plasmids for *Drosophila* transgenesis. **a**, **b** Schematic for retrieval of a DNA fragments from a BAC clone. Here, the homology arms from *Dscam* exon 3 and exon 5 were sub-cloned into a *UAS* transgenesis vector (*pUC 3GLA HAi)* and the linearized vector was used to obtain the intervening sequence by recombineering to generate *pUC 3GLA Hai Dscam 3–5*. Transgenes from this plasmid were then generated to analyse heterologous expression in neurons using *elavGAL4* for the analysis of *Dscam* alternative splicing. Note, the retrieval vector possessed all necessary sequence elements for heterologous expression from a *UAS* (*Upstream Activation Sequence*) promoter, an N-terminal haemaglutinin (HA) tag to visualize the protein a short *polyA* tail (from the *erect wing* (*ewg*) transcription factor gene (pA1), and *Drosophila* phiC31-mediated transgenesis with the *attP* attachment site for genome integration, and a photoreceptor-specific *3xP3*-driven Green Fluorescent Protein (GFP) reporter for identification of transformant flies. The *3xP3-GFP* visible marker is flanked by *LoxP* sites for later removal, if necessary. The short 300–600 bp homology arms of the beginning and the end of the sequence for retrieval of the *Dscam* variable exon 4 cluster from the BAC were linearized by *EcoRV* digestion of the GAT^ATC motif. **c**, **d** To introduce alterations (e.g. swapping exon 4.6 with exon 9.8) the parent *pUC 3GLA Hai Dscam 3–5* plasmid is cut by restriction enzymes and the alterations are retrieved by gap-repair recombineering to generate the *pUC 3GLA Hai Dscam 3–5 4.6/9.8* plasmid, where exon 4.6 is exchanged by exon 9.8. Please note that the fragment cut out from the parent vector is not shown, but present in the recombineering reaction

of the *Red* genes by simply increasing the culture temperature above 30° C. Firstly, this removes the possibility of leaky expression of the λ Red proteins, preventing any deleterious consequences of the prolonged presence of the recombineering enzymes. Secondly, as BACs are typically large (e.g. 150–350 kb), transformation even with specialised protocols can be very inefficient [9]. By transforming the small *pSC101-BAD-gbaA* plasmid directly into the BAC-containing cells, the loss of efficiency associated with BAC transformation is eliminated. Thirdly, introduction of a gap and its repair by a linear fragment eliminates the need to introduce a selection cassette and counter-selection for its removal [19]. Therefore, this gap-repair recombineering protocol is both efficient and quick.

This article is split into three distinct basic protocols and three support protocols. Basic Protocol 1 describes the retrieval of large sequences directly from a BAC by gap-repair recombineering into a HCN plasmid retrieval vector (Fig. 1a and b, Fig. 2). Basic Protocol 2 describes manipulation of high-copy number plasmids by gap-repair recombineering (Fig. 1c and d, Fig. 3) and Basic Protocol 3 describes the generation of in vitro transcribed sgRNAs for in vitro DNA scission by CRISPR-Cas9 to overcome a lack of unique restriction sites (Fig. 4).

In addition, Support Protocol 1 describes how to efficiently extract BAC DNA from the host cell (Additional file 1). Support Protocol 2 describes the generation of gap-repair recombineering electrocompetent cells, which possess all the enzymes from the *pSC101-BAD-gbaA* plasmid required for gap-repair recombineering and Support Protocol 3 describes a simple and optimised boiling Miniprep method for the rapid purification of plasmid DNA for restriction digest fingerprint screening (Additional file 1).

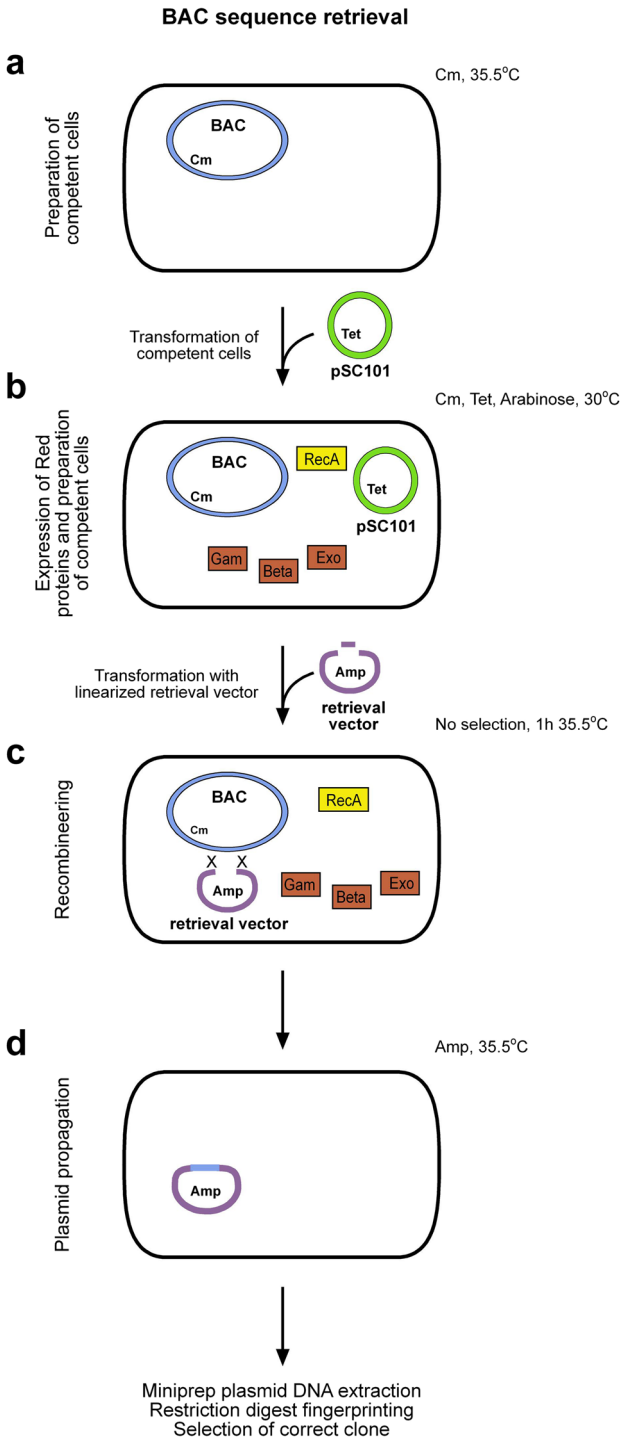

Fig. 2 Schematic for retrieval of a DNA fragment from a BAC clone by gap-repair recombineering. **a**-**d** As described in Basic Protocol 1, electrocompetent cells are first prepared from the BAC containing cells (35.5° C, 12.5 µg/ml chloramphenicol, (**a**) and transformed with the *pSC101-BAD-gbaA* plasmid containing the RecA and λ Red recombineering proteins (30° C, 12.5 µg/ml chloramphenicol, 3 µg/ml tetracycline). Then, RecA and λ Red recombineering proteins are induced and electrocompetent cells are made (30° C, 12.5 µg/ml chloramphenicol, 3 µg/ml tetracycline, 0.3% arabinose, (**b**). Finally, the linearized retrieval vector is electroporated into the recombineering competent cells harbouring the BAC clone and incubated at 35.5° C with no antibiotic selection for 60 min for recombineering to occur (**c**). Then, the final HCN plasmid is propagated (35.5° C, 100 µg/ml ampicillin) to make plasmid DNA (see Support Protocol 3) for diagnostic fingerprint restriction digest to identify positive clones (**d**)

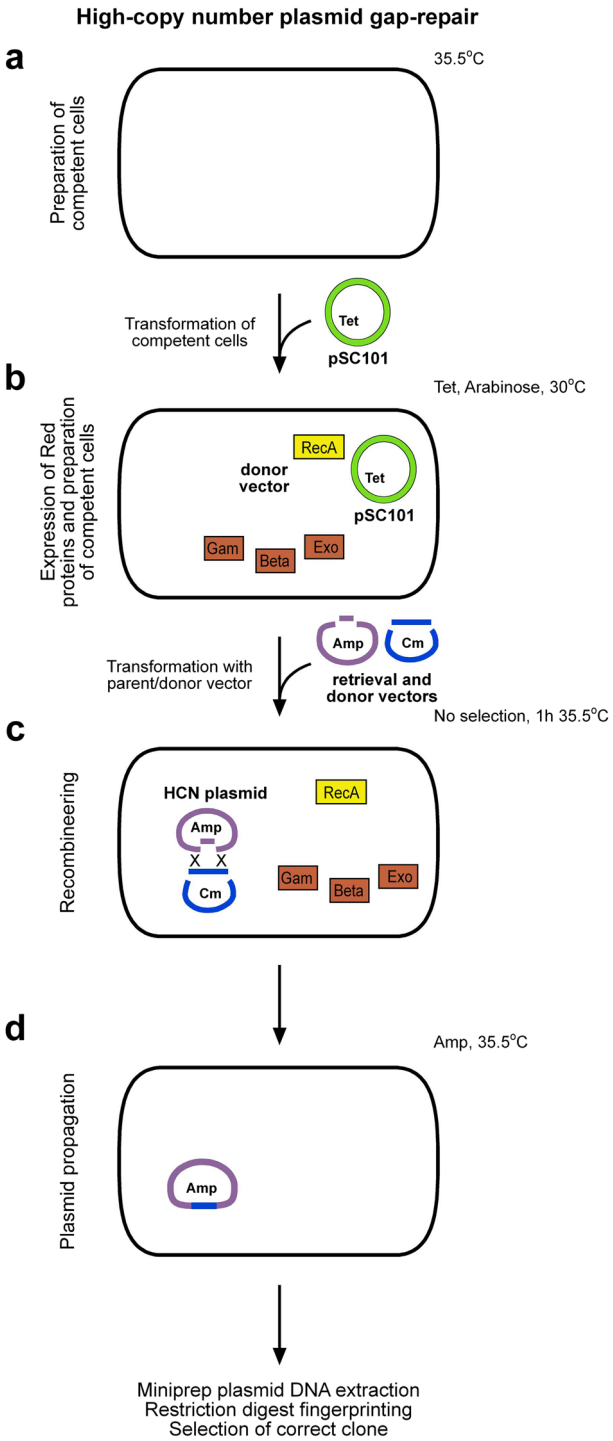

Fig. 3 Schematic for the manipulation of large HCN plasmids by gap-repair recombineering. **a**-**d** As described in Basic Protocol 2, a batch of DH5α cells containing the RecA and λ Red recombineering proteins needs to be prepared by making electrocompetent DH5α cells for transformation with *pSC101-BAD-gbaA* plasmid (30º C, 3 µg/ml tetracycline, (**a**). Then, RecA and λ Red recombineering proteins are induced and again electrocompetent cells are made (30º C, 3 µg/ml tetracycline, 0.3% arabinose, (**b**). Finally, the digested parent and donor vectors are electroporated into the recombineering competent cells and incubated at 35.5º C with no antibiotic selection for 60 min for recombineering to occur (**c**). The manipulated HCN plasmid is propagated (35.5º C, 100 µg/ml ampicillin) to make plasmid DNA (see Support Protocol 3) for diagnostic fingerprint restriction digest to identify positive clones (**d**)

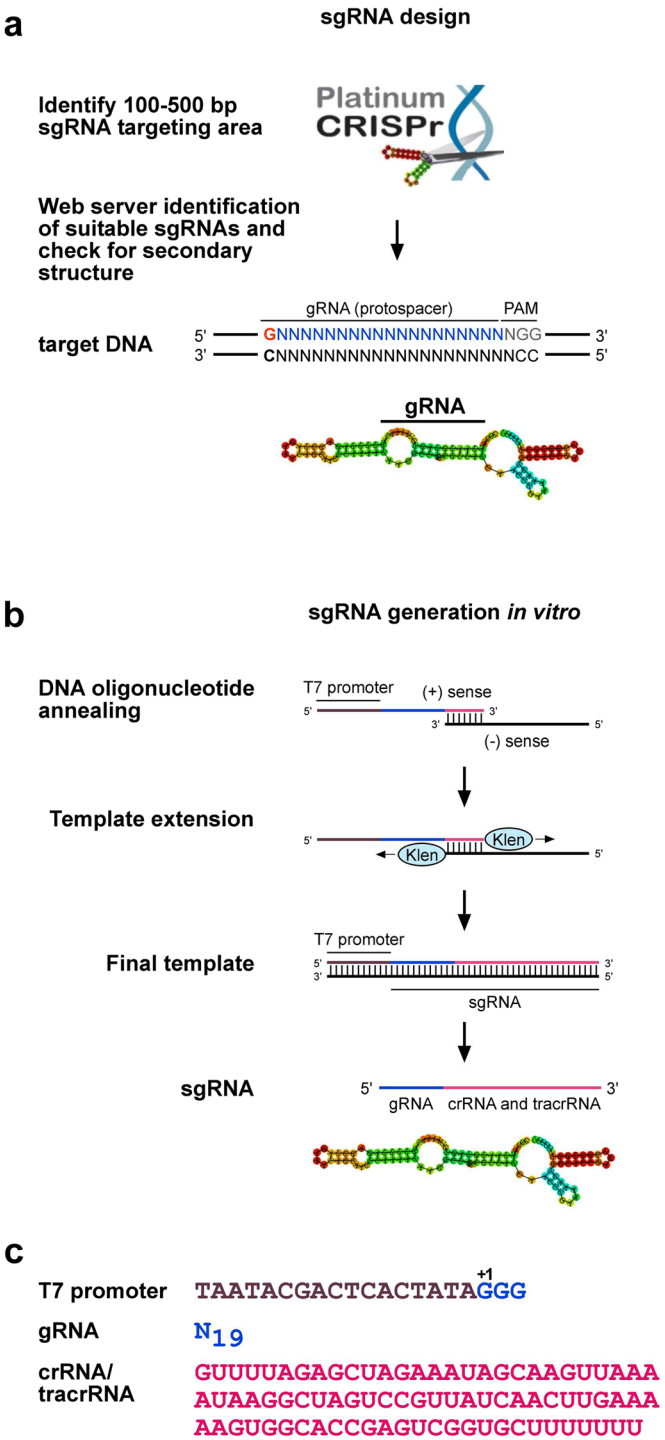

Fig. 4 Schematic for sgRNA selection, generation of dsDNA template and in vitro transcription of sgRNA. As described in Basic Protocol 3, first a gRNA (blue) is identified within the genomic target sequence using PlatinumCRISPR and analysed for efficiency based on adopting the correct RNA secondary structure and presence of detrimental sequences (**a**). A G (red) needs to be present at the start to allow for efficient in vitro transcription by bacteriophage polymerase T7. The PAM site present in the sgRNA target site is shown in grey. Then, two oligonucleotides are designed to incorporate the T7 promoter (brown), the gRNA sequence (blue) and homology to the constant part of the sgRNA (magenta). These oligos are then annealed and extended by the Klenow fragment of bacterial RNA polymerase to form a double stranded template for in vitro transcription of the sgRNA (**b** and **c**)

Materials

Basic protocol 1: BAC clone retrieval

Biological materials

- NEB® 5-alpha Competent *E. coli* (NEB #C2987) (provided in HiFi DNA Assembly cloning kit (NEB E5520S)).
- Suitable BAC clone containing gene of interest, identified from flybase.
- Retrieval vector (e.g. pUC19 or similar).
- *pSC101-BAD-gbaA* plasmid (GeneBridges).

Reagents

- Q5 HiFi DNA polymerase (NEB M0491S).
- BAC clone DNA (See Support Protocol 1).
- Primers for cloning homology arms (e.g. from IDT).
- HiFi DNA Assembly cloning kit (NEB E5520S).
- 10% (v/v) UltraPure glycerol (Invitrogen 15,514,011).
- Boiling Miniprep buffers and reagents (See Support Protocol 2).
- Midiprep kit (QIAfilter Plasmid Midi prep kit 12,145, Qiagen).
- Selection of blunt cutting restriction endonucleases.
- Agarose gels for DNA electrophoresis (Sigma A9539).
- Ice.

Solutions

10% (w/v) L-arabinose solution in ddH_2O (SIGMA A3256).

Set of bacterial growth media (See Recipes):

- LB agar bacterial growth plates supplemented with ampicillin (100 µg/ml), chloramphenicol (12.5 µg/ml), tetracycline (3 µg/ml), or combinations thereof.
- 2YT bacterial growth media supplemented with ampicillin (100 µg/ml), chloramphenicol (12.5 µg/ml), tetracycline (3 µg/ml), or combinations thereof.
- LB bacterial growth media supplemented with ampicillin (100 µg/ml), chloramphenicol (12.5 µg/ml), tetracycline (3 µg/ml), or combinations thereof.
- SOC bacterial growth media (Outgrowth media).

Phenol/$CHCl_3$extraction and precipitation:

- Phenol: $CHCl_3$: isoamyl alcohol (50:49:1) solution (100 ml): 50 ml phenol saturated with 10 mM Tris–HCL pH 8, 1 mM EDTA pH 8 (Sigma P4457), 49 ml $CHCl_3$ (VWR, 83,627), 1 ml isoamyl alcohol (Sigma W205702).
- $CHCl_3$: Isoamyl alcohol (49:1) solution (50 ml).
- 3 M sodium acetate (pH 5.2): 3 M NaAc (Fisher BP333-500) dissolved in ddH_2O. pH adjusted with glacial acetic acid (VWR36289).
- Molecular biology grade glycogen (20 µg/µl Roche R0561).
- Absolute ethanol (96%, Fisher 11,384,064).
- 75% (v/v) ethanol.

Recipes

All media were autoclaved at 121° C for 30 min. Liquefied LB agar was allowed to cool below 60° C prior to the addition of the respective selection antibiotic(s).

- Ampicillin stock (100 mg/ml, Sigma A9518) in ddH_2O. Dilute 1:1000 for final concentration (100 µg/ml).
- Chloramphenicol stock (36 mg/ml, Sigma C0857) in ethanol. Dilute 1:2900 for final concentration (12.5 µg/ml).
- Tetracycline stock (6 mg/ml, Sigma 87,128) in ethanol. Dilute 1:2000 for final concentration (3 µg/ml). Use fresh, colour is orange and will turn yellow with age, which indicates that tetracycline is no longer active (Additional file 2)
- LB bacterial growth media (500 ml): 5 g peptone (Milipore 70,169), 2.5 g yeast extract (Fisher BP9727), 1.25 g sodium chloride, 300 µl 5 M sodium hydroxide.
- 2YT bacterial growth media (250 ml): 4 g peptone, 2.5 g yeast extract, 1.25 g sodium chloride, 375 µl 5 M sodium hydroxide.
- SOC bacterial growth media (250 ml): 5 g peptone, 1.25 g yeast extract, 0.1251 g sodium chloride dissolve and add 2.5 ml of 100 mM KCl, 1.25 ml of 2 M $MgCl_2$ and adjust pH to 7. Add 5 ml of 1 M glucose.
- LB agar plates (500 ml, 20×25 ml plates): 5 g peptone, 2.5 g yeast extract, 5 g sodium chloride, 7.5 g technical agar (Fisher A360500), 900 µl 5 M sodium hydroxide.

Laboratory equipment

- Thermocycler.
- Electroporation apparatus.
- Electroporation cuvettes (1 mm, Geneflow Cellprojects E6-0050).
- Static incubator (set to 35.5° C).
- Static incubator (set to 30° C).
- Orbital incubator (set to 35.5° C, 180—200 rpm).
- Orbital incubator (set to 30° C, 180—200 rpm).
- Agarose gel electrophoresis apparatus.
- Benchtop microfuge.

- Water bath.
- -20° C freezer.
- Ice machine.

Basic protocol 2: HCN plasmid gap-repair recombineering

Biological materials

- Parent vector (e.g. made in Basic Protocol 1).
- Donor vector.
- Electrocompetent cells containing λ Red proteins (See Support Protocol 2).

Reagents

- Boiling Miniprep buffers and reagents (See Support Protocol 2).
- Midiprep kit (QIAfilter Plasmid Midi prep kit, Qiagen recommended).
- Selection of restriction endonucleases (suitable for diagnostic fingerprint digestion screening).
- Agarose gels for DNA electrophoresis.
- Ice.

Solutions

- Set of bacterial growth media (See Basic Protocol Solutions).
- Phenol/$CHCl_3$ extraction and precipitation solutions (See Basic Protocol Solutions).

Laboratory equipment

- Electroporation apparatus.
- Electroporation cuvettes (1 mm).
- Static incubator (set to 35.5° C).
- Static incubator (set to 30° C).
- Orbital incubator (set to 35.5° C, 180—200 rpm).
- Agarose gel electrophoresis apparatus.
- 2 L flask.
- Cold room.
- Ice machine.

Basic protocol 3: Generation of sgRNAs by in vitro transcription

Biological materials

Recombinant *Streptococcus pyogenes* (*Spy*) Cas9 (NEB, M0386) with provided reaction buffer (NEBuffer 3.1: 100 mM NaCl, 50 mM Tris HCL,10 mM $MgCl_2$, 100 µg/ ml BSA, pH 7.9 @ 25° C).

Reagents

- DNA Polymerase I, large (Klenow) fragment (NEB, M0210) and reaction buffer (NEBuffer 2: 50 mM NaCl, 10 mM Tris–HCl, 10 mM $MgCl_2$,1 mM DTT, pH7.9 @ 25° C).
- G-50 AutoSeq Sephadex spin column (GE Life Sciences, 27,534).
- G-50 ProbeQuant Sephadex spin column (GE Life Sciences, 28,903,408).
- Diethyl Pyrocarbonate (DEPC) (Sigma D5758).
- Custom sense oligo: GGCTAATACGACTCACTATA**G**GG N_{19} (gRNA) GUUUUAGAGCUAGAAAUAGC (with T7 promoter (underlined, first G bold), protospacer sequence (N_{19}) and tracrRNA overlap).
- Constant anti-sense (complementary) oligo:

AAAAAAAGCACCGACTCGGTGCCACTTTTT CAAGTTGATAACGGACTAGCCTTATTTTAACTT GCTATTTCTAGCTCTAAAAC (tracrRNA).

- T7 MEGASCRIPT Kit (Ambion, AM1333).
- RNase inhibitor (Promega N2111).
- DNase I (Ambion, AM2222).
- Urea gel solutions (National diagnostics).
- Standard agarose gels for DNA electrophoresis.

For ^{32}P alpha-ATP (radioactive) trace labelling for sgRNA quantification.

^{32}P alpha-ATP (0.1-3 µl, 800 Ci/mmol, 12.5 µM, Perkin Elmer, NEG003X).

For non-radioactive sgRNA quantification.

- 5'EndTag™ Nucleic Acid Labelling System, Vector Laboratories, MB-9001.
- IRDye® 800CW Maleimide, Li-Cor, 929–80,020.
- 20 nt DNA oligonucleotide (e.g. the gRNA sequence) as standard to quantify the RNA.

Solutions

- DEPC-treated (1 ml DEPC per 1 l ddH_2O incubated at room temperature overnight, autoclaved 121° C 15 min) OR nuclease-free water.
- Phenol/$CHCl_3$ extraction and precipitation solutions (See Basic Protocol Solutions).

Laboratory hardware

- Heat block/ water bath 40/ 80/ 85/ 95° C.
- Refrigerated benchtop microfuge.
- 20×20 cm gel apparatus (alternatively, BioRad or similar Minigels).

For ^{32}P alpha-ATP (radioactive) trace labelling for sgRNA quantification.

- Phosphoimager (BioRad).
- Scintillation counter.
- Agarose gel electrophoresis apparatus.

For non-radioactive sgRNA quantification.

Li-Cor 9141-WP Odyssey® CLx Infrared Imaging System.

Procedure

Critical parameters

Linearization of vectors for recombineering

For efficient retrieval of DNA from BAC clones the vector needs to be 100% linearized. This is normally achieved by 20–40-fold over-digestion (1 unit of enzyme cuts 1 µg of DNA per hour), but not all enzymes cut equally well. In our experience, the following enzymes from NEB: EcoRV (GAT^ATC), NruI (TCG^CGA), AfeI (AGC^GCT), StuI (AGG^CCT, methylation sensitive), SmaI (CCC^GGG, at 25° C) work well for this application. If background persisst from incomplete linearization, two restriction sites separated by a few nucleotides can be incorporated into the retrieval vector.

Growth temperature

Although 37° C is given in standard protocols for growing bacteria, the temperature in most incubators fluctuates and goes above 37° C. If the temperature goes above 37° C, bacteria initiate a heat shock response fundamentally altering their physiology resulting in increased recombination [20]. Moreover, under these conditions, DNA becomes indigestible if using the boiling prep method (but not for the alkaline lysis method). Therefore, we set our incubators at 35.5° C.

Cold temperatures can also induce recombination. Therefore, plates containing large constructs should be kept at 12-18° C for a day until positives colonies have been identified. For large constructs, we strongly recommend to immediately analyse the Midiprep DNA by restriction digest fingerprinting to validate that no recombination has happened, which mostly results in loss of large parts of the plasmid. When making *Drosophila* transgenes by phiC31 mediated transformation, only one molecule of the construct is inserted, and we recommend to validate transformants for the correct construct. We have not observed plasmid heterogeneity from DNA retrieval from BACs (see below).

Tetracycline

Tetracycline (orange colour when fresh) loses its activity with age and changes its colour to yellow (Additional file 2). Inactive tetracycline is a common reason for recombineering failure.

Competent cells and gap-repair recombineering efficiency

When growing *E. coli* for making gap-repair recombineering competent cells (Support protocol 3) it is important that they are well aerated, e.g. we use 500 ml in a 2 l Erlenmeyer flask and shake well (200 rpm or more). We have generated both electro- and chemical-competent cells (with rubidium chloride) cells [14]. Generally, if cells were competent to take up large plasmids and tetracycline was active, positive colonies were observed from gap-repair recombineering ($n > 10$). For parameters affecting gap-repair recombineering efficiency, please see Haussmann et al. (2019).

Design of donor vector with mutation(s) for gap-repair recombineering

To be able to identify the recombinant clone with the desired mutation efficiently, it is recommended to include an additional restriction site, or delete an existing site in the donor vector to distinguish the recombinant from the parent. For releasing the insert of the donor vector for *Dscam* gap-repair (Fig. 1c), we use the SmaI blunt cutter, because we could find three guanosines at the beginning of the left and three cytidines at the end of right homology arm to make up the palindrome of the restriction site, respectively.

Preparation of parent and donor vectors for gap repair recombineering

After digestion, phenol/$CHCl_3$ extraction and precipitation, parent and donor vectors are mixed without gel purification. Concentrations of those two fragments required for recombineering are calculated as described in Support Protocol 4 (Additional File 1). Ensure that the newly generated plasmid has a new restriction site included, or a restriction site has been removed such that it can be distinguished from the parent plasmid.

Fingerprinting of plasmids

To evaluate the integrity of large plasmids restriction enzymes are chosen such that distinct fragments are observed (e.g. one larger fragment and smaller ones form 0.5–4 kb). Ideally, this ladder of fragments encompasses the insert. To distinguish parent clones from manipulated

clones, an additional restriction site needs to be inserted or an existing restriction site needs to be altered. If a combination of enzymes is required, they need to have the same buffer conditions. Generally, classic restriction enzymes (e.g. EcoRI, EcoRV, XhoI, BamHI, XbaI and HinDIII) are good choice to cut boiling prep DNA, but we noticed that low salt cutters (e.g. KpnI and SacI) will not cut boiling prep DNA. If a restriction enzyme has not been used for boiling prep DNA, it is best to test it first. Also, we noticed that PvuII cuts slower than other enzymes.

Heterogeneity of gap-repair recombineered bacterial colonies

Occasionally, we observed both parental and gap-repair recombineered plasmids, possibly from cells taking up more than one plasmid. However, plasmid fingerprinting would show a mix of fragments from parental and gap-repair recombineered plasmids. In addition, we have observed concatenated plasmids, likely from high plasmid concentrations used for electroporation. To detect concatenated plasmids, run undigested plasmids on a 0.7% agarose gel.

Single molecule selection by phiC31 mediated transformation

For *Drosophila* transgenesis, phiC31 mediated insertion using *attB-attP* recombination is routinely used. This system has the advantage that at a given *attB* site only one plasmid will be inserted. Hence, it is important to establish transgenic lines from a single chromosome. If plasmid heterogeneity is a problem, this step could be used for selection as transformation efficiency with plasmids is reasonably high (1 transformant in 2–5 G_0 crosses).

Basic protocol 1: Retrieval of large DNA fragments from a BAC by gap-repair recombineering

To subclone a large genomic fragment, a suitable BAC needs first to be identified. BAC clones for *Drosophila* can be found in flybase by selecting the JBrowse viewer under Genomic Libraries and ordered from BACPAC resource centre [21]. The BACPAC resource centre has extensive libraries of clones available for many eukaryotic species with the exception of plants, but does not list available BAC clones in a catalogue-like fashion, so they must be identified on species-specific genome browsers or Ensembl [22]. Other sources of BAC clones are Geneservice [23], the Sanger Institute or commercial suppliers (e.g. Thermo Fisher Scientific). It is important that the HCN retrieval vector contains a different antibiotic resistance gene to the BAC (usually chloramphenicol resistant). Typically, we use *pUC*-based plasmids, which provide ampicillin resistance, and chloramphenicol-resistant *pOT* plasmids available from Addgene (Accession number: KM977569 [14]). For schematic representation of this procedure, see Figs. 1A, B and 2. For phiC31-mediated transgenesis in *Drosophila*, our retrieval vector, *pUC 3GLA UAS HAi* (Accession number: KM253740 [14]), is available from AddGene.

Generation of BAC retrieval vector

To retrieve a large fragment from a BAC, the ends of this large fragment are cloned such that they are separated by a restriction site composed of nucleotides present in the genome. If using *EcoR*V for example, the left homology arm sequence ends with GAT and the right homology arm sequence starts with ATC.

These short homology arms (300–600 bp) are PCR amplified from the BAC using Q5 HiFi DNA polymerase (see Support Protocol 1) and cloned into the Multiple Cloning Site (MCS) of the pre-engineered retrieval vector by Gibson Assembly [24]. We validate the concentration of fragments on agarose gels stained with DNA stain (e.g. ethidium bromide, SYBR Safe and others), using a DNA marker for quantification (see Support Protocol 4).

1. For cloning, mix the two homology arms and the digested retrieval vector (60–100 ng/µl) in a 1:1:1 equimolar ratio in a total volume of 2–4 µl using at least 60 ng of a 3–4 kb vector. Note: run fragments on an agarose gel to quantify, see Support Protocol 4.
2. Add an equal volume of the 2 × HiFi DNA assembly master mix to the homology arm/retrieval vector mix (e.g. 4 µl DNA mixture, 4 µl master mix).
3. Leave on ice for 5 min.
4. Incubate at 50° C for 15 min and put on ice afterwards.
5. Take 2 µl of the Gibson assembly mix and add to 15 µl DH5α cells (included in the kit). Gently flick to mix and incubate on ice for 2–5 min.
6. Transform the cells by heat-shock at 42° C for 1 min in a water bath, then put on ice for 2 min.
7. Add 300 µl SOC outgrowth media (included in the kit) and incubate for 5 min for ampicillin-resistant vectors (Note: for chloramphenicol and tetracycline a 30 min incubation period is required).
8. Plate 50 µl and 250 µl of transformed cells onto LB agar bacterial growth plates (ampicillin 100 µg/ml, or corresponding selection antibiotic) incubate at 35.5° C overnight (16 h) until colonies have a diameter of 1 mm, then keep at ambient temperature to prevent overgrowth.
9. Take 12 single colonies (use a small Eppendorf pipette tip to scoop the entire colony, regrow the

colony to set up fresh cultures for a later Midiprep) to inoculate 3 ml 2YT bacterial growth media (with selective antibiotic) and incubate at 35.5° C overnight (16 h).
10. The following day perform boiling Miniprep plasmid DNA extraction (see Support Protocol 3) and screen for positive clones by a diagnostic restriction enzyme digest.
11. Once positive colonies have been identified, take the re-grown single positive colony to inoculate a 5 ml LB bacterial growth media (with selective antibiotic) starter culture and incubate at 35.5° C, 200 rpm for 4 h. For plasmid Midiprep, use this starter culture to inoculate 45 ml LB bacterial growth media (with selective antibiotic, final volume 50 ml) and incubate at 35.5° C, 200 rpm, overnight (16 h).
12. Harvest cells by centrifuging at 3000 g for 15 min and pour the supernatant off. Either store the pellet at -20° C or immediately proceed to Midiprep (QIAfilter Midiprep kit according to the manufacturer's instructions).
13. Validate the final construct by diagnostic restriction digest and validate the sequence the around homology arm junction by Sanger sequencing.
14. Linearize 3 μg of the retrieval vector with the blunt cutter between the two homology arms in a 50 μl digestion mix (20–40 fold overdigestion, e.g. for 30 U EcoRV incubate for 2–4 h).
15. Increase the volume to 200 μl with ddH_2O and add an equal volume of phenol: $CHCl_3$: isoamyl alcohol (50:49:1). Vigorously vortex.
16. Centrifuge at 16,400 g for 1 min.
17. Transfer supernatant to a fresh Eppendorf tube and add an equal volume of $CHCl_3$:isoamyl alcohol (49:1). Vigorously vortex.
18. Centrifuge at 16,400 g for 1 min.
19. Transfer supernatant to a fresh Eppendorf tube. Add 24 μl 3 M sodium acetate (pH 5.2), 1 μl (20 μg) glycogen (Roche R0561), and 700 μl of absolute ethanol (96%, 3 Volumes). Mix by tube inversion.
20. Incubate at -80° C for 30 min (or at -20° C for 90 min).
21. Centrifuge at 16,400 g for 10 min then pour the supernatant off and tip the inverted tube on a paper tissue to get rid of excess ethanol
22. Wash the pellet with 750 μl 75% (v/v) ethanol.
23. Centrifuge at 16,400 g for 10 min then pour the supernatant off.
24. Air dry the pellet (15 min in 35.5° C incubator, longer at room temperature), ensure no residual ethanol is left behind as this will inhibit subsequent enzymatic reactions.
25. Dissolve in 50 μl water (60 ng/μl, do not use TE or other buffers as salt compromises electroporation, the resistance of distilled water should by around 14–18 Ohm).

Transforming BAC host cells with *pSC101-BAD-gbaA* plasmid

1. Take a single colony of BAC-containing cells to inoculate 3 ml LB bacterial growth media (with corresponding BAC selection antibiotic, typically chloramphenicol) and incubate at 35.5° C, 200 rpm, overnight.
2. Take 30 μl of the overnight starter culture and inoculate 1.4 ml fresh LB bacterial growth media (with selective antibiotic) in a 1.5 ml Eppendorf tube, and incubate at 35.5° C, 200 rpm for 2–3 h (optical density OD_{600} should be 0.6).
3. Pre-cool benchtop microfuge to 0° C.
4. Centrifuge cells at 10,000 g for 30 s. Pour the supernatant off and tap the tube on to a paper towel to remove residual fluid.
5. Resuspend in 1 ml ice cold 10% UltraPure Glycerol (Invitrogen) then centrifuge again at 10,000 g for 30 s. Discard the supernatant as above.
6. Repeat above washing procedure 3 times.
7. After the final wash, discard the supernatant as above, but leave approximately 30 μl 10% UltraPure glycerol (Invitrogen). Gently resuspend the cells (Use 10 μl cells in the next step and put the reminder at -80° for future use).
8. Add 1 μl (1–5 ng) of *pSC101-BAD-gbaA* plasmid (tetracycline resistant) to 10 μl resuspended cells.
9. Transfer the mix of resuspended cells and *pSC101-BAD-gbaA* to a 1 mm electroporation cuvette on ice**.**
10. Transform by electroporation with the following parameters 1.8–2.5 kV, 200 Ω, 25 μF. Time capacitance extender set to 125 μFD. Time constant should be 4.5–5.0 ms. A lower time constant indicates presence of salt. If the samples "pop", arcing occurs due to too much salt is present leading to dead cells).
11. Resuspend transformed cells in 300 μl SOC outgrowth media.
12. Incubate transformed cells at 30° C for 70 min. Note, tetracycline is light sensitive so ensure the incubator is dark.
13. Plate 50 μl and 250 μl of transformed cells on to LB agar bacterial growth plates (with selective antibiotic for the BAC and tetracycline 3 μg/ml). Incubate at 30° C for 48 hours.

Preparation of BAC/*pSC101-BAD-gbaA* host cells for DNA sequence retrieval

1. Take a single colony of BAC/*pSC101-BAD-gbaA*-containing cells to inoculate 3 ml LB bacterial growth media (with chloramphenicol 12.5 µg/ml and tetracycline 3 µg/ml) and incubate at 30° C, 200 rpm, for 32 h in the dark.
2. The following morning, take 30 µl of the starter culture and inoculate 1.4 ml fresh LB bacterial growth media (with selective antibiotic for the BAC and tetracycline 3 µg/ml) in 1.5 ml Eppendorf tubes, and incubate at 30° C, 200 rpm for 2 h until OD_{600} is 0.15–0.2.

From the remining starter, make a glycerol stock (e.g. 800 µl the starter poured to 200 µl glycerol) and freeze at -80° C for future use.

3. Add 30 µl 10% L-arabinose to one tube and return to the incubator (30° C, 200 rpm) for 45–60 min until OD_{600} is 0.35–0.4.
4. Pre-cool bench top microfuge to 0° C.
5. Centrifuge cells at 10,000 g for 30 s and pour the supernatant off and tap the tube on to a paper towel to remove residual fluid.
6. Resuspend in 1 ml ice cold 10% UltraPure Glycerol (Invitrogen) then centrifuge again at 10,000 g for 30 s. Discard the supernatant as above.
7. Repeat above washing procedure 3 times.
8. After the final wash, discard the supernatant as above, but leave approximately 30 µl 10% UltraPure glycerol (Invitrogen). Gently resuspend the cells.
9. Add 50–300 ng linearized retrieval vector to the resuspended cells.
10. Transfer the mix of resuspended cells and linearized retrieval vector to a 1 mm electroporation cuvette on ice**.**
11. Transform by electroporation with the following parameters 1.8–2.5 kV, 200 Ω, 25 µF. Time capacitance extender set to 125 µFD. Time constant should be 4.5–5.0 ms.
12. Resuspend transformed cells in 300 µl SOC outgrowth media and transfer to a fresh Eppendorf tube.
13. Incubate transformed cells at 35.5° C for 1 h. This is the stage where the gap-repair recombineering occurs. The *pSC101-BAD-gbaA* plasmid will be lost due to the *SC101* temperature-sensitive origin of replication, the recombineering enzymes will remain.
14. Plate 5 µl, 50 µl, and 245 µl transformed cells on to LB agar bacterial growth plates (ampicillin 100 µg/ml, or corresponding retrieval vector antibiotic).
15. Incubate at 35.5° C overnight.
16. Screen 12 colonies as per Support Protocol 2. Typically, 60–98% of colonies have the insert. If this is not the case, either the retrieval vector was not 100% linearized or tetracycline was inactive. See Troubleshooting section.

Basic protocol 2: Manipulation of large HCN plasmids by gap-repair recombineering

For manipulation of a large HCN plasmid (the parent vector) by gap-repair recombineering, a small, easily manipulatable vector (the donor vector) is required. For this, we recommend using the *pOT2* (BDGP, chloramphenicol resistant) HCN plasmid which is chloramphenicol resistant enabling selection of the manipulated parent vector with ampicillin. For schematic representation of this procedure, see Fig. 1c and d, and Fig. 3. For this procedure, a dsDNA template (e.g. PCR amplicon, synthesised gene fragment) may also be used; however, we prefer the use of a small vector for validation of sequences by Sanger sequencing and as a template to introduce mutations.

Preparation of parent vector for manipulation

1. Digested the large parent vector by rare cutting enzymes or cleave in vitro with CRISPR-Cas9 DNA (See Basic Protocol 3) at positions flanking the region to be manipulated.
2. Analyse the digested vector by a standard DNA agarose gel electrophoresis to ensure 100% digestion has occurred (otherwise there will be considerable non-manipulated parent vector contamination during screening).
3. Purify the digested parent vector by phenol:$CHCl_3$ extraction and precipitation (see Basic Protocol 1, steps 15–25) and resuspend in water.

Preparation of donor vector

The donor vector should be designed to have short (300–600 bp) homology arms overlapping the digestion sites of the parent vector.

1. Digest the donor vector with unique restriction enzymes that flank the homology arms (see Basic Protocol 1, step 14) to liberate a linear dsDNA template for gap-repair recombineering.

2. Analyse the digested vector by standard DNA agarose gel electrophoresis to ensure 100% digestion has occurred.
3. Purify the digested donor vector by phenol:$CHCl_3$ extraction and precipitation (see Basic Protocol 1, steps 15–25) and resuspend in water.

Transformation of recombineering electro-competent cells with the donor and parent vectors

1. Prepare a mixture of 100 ng digested parent vector and donor vector in a 1:2 molar ratio and incubate on ice for 5 min (if unsuccessful, try 1:2–10, see Troubleshooting section).
2. Add 5 µl parent/ donor vector mixture to 30 µl recombineering electro-competent cells (see Support Protocol 2). Gently flick to mix. Incubate on ice for 1 min.
3. Transfer the mix of cells and parent/ donor vector to a1 mm electroporation cuvette on ice.
4. Transform by electroporation with the following parameters 1.8–2.5 kV, 200 Ω, 25 µF. Time capacitance extender set to 125 µFD. Time constant should be 4.5–5.0 ms.
5. Resuspend transformed cells in 300 µl SOC outgrowth media and transfer to a fresh Eppendorf tube.
6. Incubate transformed cells at 35.5° C for 1 h. This is the stage where the gap-repair recombineering occurs. The *pSC101-BAD-gbaA* plasmid will be lost at this stage due to the *SC101* temperature-sensitive origin of replication.
7. Plate 50 µl and 250 µl of transformed cells on to LB agar bacterial growth plates (ampicillin 100 µg/ml, selection antibiotic for the parent vector).
8. Incubate at 35.5° C overnight.
9. Screen colonies as described in Support Protocol 3.

Basic protocol 3: Generation of in vitro transcribed sgRNAs for in vitro DNA scission by CRISPR-Cas9

To overcome the lack of unique restriction sites in desired positions, DNA can be cut by CRISPR-Cas9 in vitro by providing in vitro transcribed sgRNAs. To design sgRNAs identify a target region and enter the sequence to the PlatinumCRISPr web server [18, 25]. This server will select suitable 20 nt sequences termed protospacer followed by a NGG Protospacer Adjacent Motif (PAM, where N denotes any nucleotide, and G denotes guanosine), which is required for *S. pyogenes* (*Spy*) Cas9 to cut DNA. This 20 nucleotides protospacer is taken as the guide RNA portion of the sgRNA. The first nucleotide of the protospacer needs to be a G for transcription from the T7 promoter, but this G does not need to be in the endogenous sequence and can be added artificially. For efficient in vitro transcription with T7 RNA polymerase, we add three Gs, which do not need to be present in the template. PlatinumCRISPr will also assess the efficacy of the sgRNA based on RNA folding and presence of activity compromising motifs [18]. For evaluation of sgRNA folding by PlatinumCRISPr the entire gRNA sequence needs to be added including the G's added for in vitro transcription. The selection and screening of sgRNAs and the generation of the dsDNA substrate for in vitro transcription of sgRNAs is summarised in Fig. 4. For quantification of in vitro transcribed sgRNA we use ^{32}P alpha-ATP trace labelling, but the included non-radioactive labelling protocol provides an alternative if no ^{32}P alpha-ATP is available. If the concentration is high enough a Nanodrop can also be used.

Protocol steps

Synthesis of dsDNA substrate for in vitro sgRNA transcription

1. Design sgRNAs as detailed in [18] and order a 60 nt oligonucleotide containing the T7 promoter, the guide RNA (gRNA) and 20 nt overlap with the tracrRNA (GGCTAATACGACTCACTATAG-N_{19} sgRNA-GUUUUAGAGCUAGAAAUAGC, the tracrRNA complementary sequence is underlined) and the anti-sense tracrRNA complementary oligo (AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC) for extension with the Klenow fragment of DNA Pol I.
2. Mix the T7/sgRNA oligonucleotides and tracrRNA complementary oligos (2 µM) in a 1:1 ratio and incubate at 40° C for 15 min to facilitate annealing.
3. Make double stranded using the DNA Pol I Klenow fragment according to the manufacturer's instructions.
4. Heat-inactivate Klenow by incubation at 85° C for 10 min and desalt using a G-50 Autoseq Sephadex spin column. Complete removal of salt is required as otherwise in vitro transcription will be inhibited.

In vitro transcription of sgRNAs

1. Transcribe sgRNA from oligo template (1 µl of 2 µM stock in 20 µl) using T7 MEGASCRIPT kit including RNAse inhibitor and trace-label with ^{32}P alpha-

ATP (0.1–3 µl, 800 Ci/mmol, 12.5 µM, Perkin Elmer) according to the manufacturer's instructions.
2. Digest away the oligonucleotides with DNAse I according to the manufacturer's instructions.
3. Remove free nucleotides with a G-50 Probequant Sephadex spin column.
4. Quantify sgRNAs by scintillation counting (dilute 1 µl in 100 µl water and add 10 µl to a scintillation vial containing scintillation fluid).
5. Validate sgRNA by running on an 8% urea denaturing polyacrylamide gel according to the manufacturer's instructions (National Diagnostics).
6. Dry the gel, expose to a phosphoimager screen (Bio-Rad) and scan the screen in a suitable phosphoimager (Cytiva).

Calculating the amount of sgRNA from trace labelling

3 µl of ^{32}PATP (800 Ci/mmole, 10 miCi/ml, 12.5 µM) corresponds to 30 miCi (0.66×10^8dpm) on the reference date (adjust accordingly from table).

Calculate the incorporation rate (X): total counts after nucleotide removal in cpm×200×2×100 (dpm)/input (dpm) = X % (200: dilution factor (0.1 µl measured from 20 µl), 2: dpm conversion, 100: percent).

Amount of RNA synthesized:

> Combine the amount of unlabeled (10 µl of 1 mM corresponds to 10 nmole) and labelled (3 µl of 12.5 µM ATP in 10 µl reaction results in 3.75 µM which corresponds to 37.5 pmole) ATP. Here the total is 10,037.5 pmole. Then adjust the amount to the incorporation rate (amount of ATP (Y) in pmole).

To get the amount of RNA, Y pmole is divided by the number of ATPs in the RNA, which yields the amount of RNA in pmole (this amount divided by volume yields the concentration).

Example: 60117 cpm were measured in 0.1 µl of 20 µl, 4 days before reference date.

60'117 x 200 x 2 100 divided by 660'00'000 x 1.214 = 30 % incorporation.

30 % of 10,037.5 pmole are 3011.25 pmole.

The sgRNA contains 5 ATPs: 3011.25 pmole divided by 5 results in 602.25 pmole (30 pmole/µl, which is a 30 µM solution).

Non-radioactive labelling and quantification of sgRNAs

Although quantification and validation of RNAs is easiest with radioactive ^{32}P labelling, this can be substituted as described below using an infrared dye and visualisation with the Li-Cor Odyssey Imaging system [26] or using a nanodrop.

1. After in vitro transcription of the RNA, remove the DNA by digestion with DNAse I according to the manufacturer's instructions.
2. Increase volume to 20 µl and remove free nucleotides with a G-50 Probequant Sephadex spin column.
3. Dephosphorylate the RNA arctic phosphatase, phenol/$CHCl_3$ extract and precipitate, then kinase with γS-ATP and react with maleimide-IRDye at room temperature according to the manufacturer's instructions (Vector Laboratories). Note that the phosphatase used in the kit is unable to remove the thiophosphate group. For higher labelling efficiency, increase the temperature of the maleimide reaction to 50-65° C in the presence of 5 mM EDTA. As a standard for quantification, use a DNA oligo of known concentration and label in parallel (no dephosphorylation required for synthetic oligos).
4. Purify by phenol/$CHCl_3$ extraction and precipitation (see Basic Protocol 1)
5. Validate and quantify labelled sgRNAs and DNA oligonucleotide by running serial dilutions on an 8% urea denaturing polyacrylamide gel according to the manufacturer's instructions (National Diagnostics). Compare sizes to accordingly labelled marker. Visualize and quantify using the Li-Cor Odyssey infrared imaging system.

Cleaving parent plasmid DNA with the in vitro transcribed sgRNA for gap-repair recombineering

1. Heat sgRNA for 2 min to 95° C and leave at room temperature until ambient temperature to adopt folding.
2. Prepare a mixture of 100 nM recombinant *Spy*Cas9 with 100 nM in vitro transcribed sgRNA in DEPC-treated ddH_2O.
3. Incubate at 25° C for 10 min.
4. Add 10 nM plasmid DNA and continue to incubate at 25° C for 2 h or until digestion is complete (e.g. 24 h). Note, the final ratio of plasmid DNA:sgRNA:*Spy* Cas9 should be 1:10:10 with a final reaction volume of 10 µl.
5. Analyse an aliquot on an agarose gel to check whether 100% digestion has occurred.
6. Phenol/$CHCl_3$extract and precipitate (see Basic Protocol 1)

7. Analyse an aliquot on an agarose gel to determine the concentration and use for gap-repair recombineering as appropriate (Basic protocol 2).

General notes

Large low copy DNA vectors cannot be efficiently manipulated by standard cloning methods using restriction enzymes. To overcome this limitation, homologous recombination based DNA manipulation, termed recombineering was adopted for manipulation of BACs for use in mouse transgenesis [3, 27], or copy control vectors for *Drosophila* or *C. elegans* transgenesis [7, 27–33]. However, for use with HCN plasmids, recombineering despite its high efficiency to clone large fragments remains underused. We used recombineering for making *Drosophila* HCN transformation constructs of over 10 genes a with very high efficiency [14, 18, 34–36].

The protocol for retrieval of large sequences directly from a BAC represents significant advancement due to its superior efficiency compared to first PCR amplify fragments that are then cloned. The limitations here are that PCR for large fragments is difficult and often inefficient to obtain the quantities needed for standard cloning.

For retrieval from a BAC clone, we generally achieve efficiencies above 75% even for fragments of 20 kb which is the limit for HCN plasmids. A further advantage of HCN plasmids is that substantial quantities can be obtained very easily facilitating validation and transgenesis. Moreover, *Drosophila* transformation efficiency by phiC31 integration with HCN plasmids is generally high and we obtain about one transformant per four injected G_0 flies. These advantages prevail over plasmids larger than 20 kb, which require a copy number control to maintain them in the host at one copy [37]. Upon induction plasmids are amplified to obtain larger quantities, but their large size is not compatible with bacterial growth at high copy numbers. Often a sufficient amount can be obtained for validation and transgenesis, but the procedure is not as straight forward to validate cloning out-come by fingerprinting Minipreps because the large size will yield many fragments.

For recombineering, two options are available to provide the Red proteins either integrated in the bacterial host or on a plasmid [15, 38]. The preferred method is to use the *pSC101-BAD-gbaA* plasmid, because it contains its own selection marker (tetracycline) and a temperature control for maintenance only at 30° C [15]. By using the plasmid for recombineering, the presence of Red proteins is limited to a very short time after their induction limiting any unwanted recombination events. In fact, we have been using recombineering extensively to manipulate the highly repetitive *Dscam* gene and have not observed any unwanted recombination events [14].

Gap-repair recombineering offers a distinct advantage in efficiency compared to other conventional cloning strategies such as traditional cloning, Golden Gate cloning and Gibson assembly to clone large DNA fragments due to its direct modification of DNA within the host cell via homologous recombination. Conventional methods often require several PCR products to clone a large DNA fragments. In addition, conventional methods including Gibson assembly rely on equimolar concentrations of insert and host plasmid for successful cloning. In contrast, gap-repair recombineering yields robust success even with tenfold excess of insert over host plasmid [14]. In fact, a tenfold excess of insert increased gap-repair recombineering efficiency significantly while there is no enhancement option for the classic cloning methods [14]. Ultimately, the high success rate of cloning large DNA fragments essentially contributes to the speed of gap-repair recombineering as conventional methods for cloning large fragments generally fail to produce a high number of positive colonies or often fail completely and need to be repeated. The main reason is the difficulty to obtain equimolar concentrations of insert and vector, which is further exaggerated when using multiple fragments.

Although Gateway cloning can be used for efficient transfer of large fragments between vectors, but uses a recombinase and specific sequences, thus lacking adaptability to sub-clone large DNA fragments for BAC clones [39, 40].

The other major advantage of gap-repair recombineering in the modification of a large plasmid as detailed in Fig. 1 lies in the efficient modification of existing large constructs for gene function analysis. Here, a smaller fragment flanked by unique restriction sites is subcloned into a vector with a different antibiotic resistance (e.g. chloramphenicol-resistant pOT) to introduce the desired mutations. Introducing mutations in a small DNA fragment is much more efficient than in a large one. Afterwards, the small fragment can then be reintroduced into the larger construct efficiently by gap-repair recombineering.

The gap-repair recombineering protocol is very efficient in retrieving large DNA fragments form BAC clones or for manipulating large HCN plasmids, there are several limitations to consider. The success of the protocol relies on maintaining precise conditions, including carefully controlled inducer concentrations and the use of high-quality competent cells. Variations in these conditions can impact the overall effectiveness of the recombineering process. Additionally, tetracycline's light sensitivity can complicate the experimental

process and necessitates specific handling and storage conditions to maintain its effectiveness. Potentially, proteins required for homologous recombination could be inefficient as a result from suboptimal expression or activity of the recombination enzymes, but we haven't observed this case. When recombineering failed in our hands, either tetracycline was inactive or the cells were not competent.

Taken together, the gap-repair recombineering protocol described here is a very efficient, fast and reliable method to clone large DNA fragments into HCN plasmids. Moreover, the high efficiency of recombineering allows to manipulate HCN plasmids and avoids the lengthy procedures associated with copy number control vectors. Due to smaller introns, invertebrate genes are much smaller than vertebrate genes [41]. As introns will be spliced reliably, most genes from model organisms like *Drosophila* or *C. elegans* can be directly incorporated into a plasmid vector [14, 41].

Troubleshooting

Low transformation efficiency/ unknown recombinant clones

For manipulation of large HCN plasmids by gap-repair recombineering (Basic Protocol 2), low transformation efficiency can be improved by increasing vector backbone (parent vector) concentration. By increasing the vector backbone from 50 to 100 ng the number of colonies increased ~ fourfold, with 60–100% containing the correct clone [14]. Moreover, the frequency of concatemerized, unknown or otherwise incorrect recombinant clones can be reduced by using a 1:5–10 vector: insert ratio.

If no positive colonies are obtained, repeat the preparation of making recombineering competent cells following exactly the protocol.

Phenol/chloroform/ethanol contamination

Following phenol:$CHCl_3$ extraction and precipitation, the presence of residual phenol, $CHCl_3$ or ethanol will inhibit enzymatic reactions such as Gibson Assembly and restriction digests. Make sure there is no phenol:$CHCl_3$ carry over and ensure the pellet is thoroughly air dried prior to resuspension in TE. Occasionally, over-drying pure DNA results in loss when tubes are kept inverted for drying.

DNA quality and yield

Use 2YT for growing Miniprep cultures as 2YT is richer than LB and gives more DNA. For QIAfilter Midipreps, use LB, as 2YT increase impurities, mostly polysaccharides.

For Midipreps use the QIAfilter version to remove the SDS precipitate, rather than using the kit without the filter where the SDS precipitate is removed by centrifugation.

We use 3 Vol of ethanol for precipitation because the purpose is to precipitate rather than selectively precipitate for cleaning the sample. Also, precipitation is faster. If the DNA is lost after precipitation, extend precipitation times and reduce temperature. Sometimes pellets of pure DNA are lost. Do not overdry as the pellet can fall out of inverted tubes.

We use 75% ethanol (v/v) made with commercial ethanol (96%) by adding 12.5 ml water to a 50 ml falcon tube and filling up with ethanol.

DNA minipreps

We favour the boiling prep method, because more preps can be done in less time than with commercial Miniprep kits, and it is very cost effective. Generally, we calculate 6–8 preps/construct, but for more complicated constructs we do up to 48 preps (which takes not much longer than 6–10 preps). If no positive colony is found in 48 preps, repeat the procedure evaluating possible errors including checking primer sequences and complete digestion of parent vector.

We normally do not use colony PCR to identify correct colonies, because with PCR only a portion of the clone is analysed and often artefacts are amplified. Also, if clones are identified by colony PCR, a Miniprep should be done to fingerprint clones with a restriction digest to test the overall integrity.

In rare occasions, a construct can be toxic to bacteria, e.g. when trying to express recombinant proteins [42]. For this, the cloning strategy needs to be changed. To make toxic recombinant proteins, we cloned them into a vector with a T7 promoter, such as *pET26b*. As the DH5α *E. coli* strain does not naturally possess the T7 polymerase, leaky expression of the toxic product does not occur. Once the correct clone was identified, it can then be transformed in to a specific *E. coil* strain, which contains an inducible T7 polymerase [42]. Alternatively, an FRT flanked stop cassette can be used, that can then be removed once transgenic flies have been generated [10, 34].

Indigestible DNA

When incubators are set at 37° C, boiling prep plasmid DNA becomes indigestible because the heat-shock response is induced if temperatures go above 37° C. The alkaline lysis protocol used in commercial Miniprep/Midiprep kits is very robust against growing cultures at temperatures above 37° C [40, 43].

Isopropanol is a strong precipitant, which precipitates impurities as well as the DNA [40]. The presence of these impurities can inhibit restriction enzyme digestion. This can be avoided by limiting the incubation time with isopropanol by centrifugation for 10 min immediately after mixing well by inversion. We also noted that phenol/$CHCl_3$ extraction does not solve the problem.

No or little DNA

Cell lysis in the boiling prep DNA protocol can be insufficient. This can be recognised by a brittle pellet after lysozyme treatment, boiling and centrifugation (Support Protocol 2, step 9). The pellet of the cellular debris should have a sticky, glue-like consistency. The main reason for this is when incubators are set at 37° C rather than 35.5° C, but insufficient boiling or low lysozyme activity can be the cause as well. When preparing the lysozyme stock, dissolve in 10 mM Tris HCl pH 7.5 and incubate on ice. Let sit and mix occasionally by inversion until completely dissolved. Aliquot and freeze in liquid nitrogen, then store at -20° C. Do not refreeze lysozyme aliquots.

Suboptimal growth of cultures can result in low amounts of DNA. If grown for too short, not enough plasmid DNA is present, while if grown too long, the antibiotic is used up and bacteria quickly get rid of the plasmid. For minipreps we use an entire 1 mm colony to set a 3 ml culture and let it grow for 16 h at 35.5° C. For Midipreps, we grow the 5 ml starter for 4 h and then pour the entire starter to 45 ml LB and grow for 16 h.

Low yield in vitro transcription

Make sure template is desalted as salt inhibits in vitro transcription. The efficiency of transcription is lower from a short oligo. If high efficiency is required, clone into a suitable vector. The MEGASCRIPT kit (Ambion) includes pyrophosphatase to increase yield as pyrophosphate inhibits the RNA polymerase.

sgRNA/Cas9 did not cut

Design new sgRNA according to [18]. All the sgRNAs we designed accordingly have been successful ($n > 15$). If a polymorphism is present in the sgRNA target region, the activity is inhibited. Hence, we recommend sequencing of the target region.

The presence of restriction enzymes inhibits sgRNA/Cas9. If it is necessary to better visualize sgRNA/Cas9 activity by cutting out a DNA fragment rather than just linearizing a plasmid, we do this sequentially and add the restriction enzyme after heat-inactivation of sgRNA/Cas9 (2 min at 95° C).

Understanding results

The results of diagnostic restriction digests of extracted plasmid DNA are generally straight forward to interpret using the predesigned plasmid files generated in Seqbuilder (Lasergene) or similar. Since the donor vector has a different selection marker, which is usually chloramphenicol, it is not present. Rarely, the empty parent vector is detected. However, one has to be aware of concatemerized plasmid, which can be detected by running undigested plasmid on an 0.7% agarose gel. Occasionally, we observe mixed clones, which originate from transformation of multiple plasmids.

Time considerations

The protocols detailed in this article have been optimised to be as quick and as efficient as possible. The time scales for each protocol are as follows:

Basic protocol 1

Subcloning a DNA fragment from a BAC clone into a HCN plasmid takes 8 days (Table 1). From streaking out the BAC clone it takes 5 days until transformation with the linearized retrieval vector containing the homology arms, which can be prepared in parallel. To identify correct clones will take 2 days and 1 day to obtain high quality Midiprep DNA.

For phiC31-mediated transgenesis in *Drosophila*, we developed a small *pUC19* based retrieval vector, *pUC 3GLA UAS HAi* (Accession number: KM253740), which is available from AddGene. For expression from UAS transgenes, genomic DNA can be used as introns are spliced efficiently. If a genomic rescue construct is required the *UAS* promoter can be cut out of *pUC 3GLA UAS Hai* beforehand [14, 34].

Basic protocol 2

To alter a HCN plasmid by gap-repair recombineering takes 4 days (Table 2), if competent cell containing the Red proteins have been prepared (Support protocol 2,

Table 1 Timeline for subcloning a DNA fragment from a BAC clone

Day 1: streak BAC clone
Day 2: Set up BAC clone culture
Day 3: Prepare competent cells and transform with *pSC101-BAD-gbaA* plasmid
Day 4: Set up Bac/*pSC101-BAD-gbaA* clone culture
Day 5: Induce Red proteins, prepare competent cells and transform with linearized retrieval vector
Day 6: Set up Miniprep cultures
Day 7: Minipreps, set up Midiprep culture
Day 8: Prepare Midiprep high quality DNA

Table 2 Timeline for altering a HCN plasmid by gap-repair recombineering

Day 1: streak DH5α
Day 2: Set up culture
Day 3: Prepare competent cells and transform with *pSC101-BAD-gbaA* plasmid
Day 4: Set up culture
Day 5: Induce Red proteins, prepare competent cells and transform with cut donor and retrieval vector
Day 6: Set up Miniprep cultures
Day 7: Minipreps, set up Midiprep culture
Day 8: Prepare Midiprep high quality DNA

5 days), To identify correct clones after transformation will take 3 days and 1 day to obtain high quality Midiprep DNA.

Basic protocol 3

Generation of template DNA, in vitro transcription and clean-up of sgRNAs can be done in 1 day. For testing DNA scission activity of sgRNAs a 2 h incubation period is sufficient, but for cloning it is advisable to extend the digestion time to 24 h. Note that presence of restriction enzymes inhibits sgRNA/Cas9.

Support protocol 1

Generation of BAC DNA. DNA can be obtained in about 90 min.

Support protocol 2

Generation of recombineering competent cells. This procedure takes 5 days (Table 2), with actual preparation time of approximately 4 h on day 4. The rest of the time consists mainly of incubation periods.

Support protocol 3

Boiling prep Minipreps. Plasmid DNA extraction from 48 3 ml overnight cultures and analysis of restriction digests on agarose gels can be achieve in 3 h.

Expected results

Generally, we obtain the correctly manipulated plasmid with a frequency of 60–80% when using recombineering. The newly obtained plasmids are validated by diagnostic restriction digests of extracted plasmid DNA and compared to the predesigned plasmid generated in Seqbuilder (DNAstar) or similar. Since the donor vector has a different selection marker, usually chloramphenicol, it is not present on the plate. The empty parent vector is rare because an excess of insert is used, unless it is insufficiently cut by restriction enzymes of sgRNA/Cas9. However, one has to be aware of concatemerized plasmid, which can be detected by running undigested plasmid on a low percentage agarose gel (e.g. 0.7%). Occasionally, we observe mixed clones, which originate from transformation of multiple plasmids.

Abbreviations

BAC	Bacterial artificial chromosome
C. elegans	*Caenorhabditis elegans*
Dscam	*Drosophila Down Syndrome Cell Adhesion Molecule*
DEPC	Diethyl pyrocarbonate
ddH_2O	Double-distilled H_2O
dsDNA	Double-stranded DNA
E. coli	*Escherichia coli*
EDTA	Ethylenediamine tetraacetic acid
ewg	*erect wing*
GFP	Green fluorescent protein
gRNA	Guide RNA (protospacer, the 20 nt complementary to the target)
HA	Haemaglutinin
HCN	High copy number (plasmid)
MFE	Minimum free energy
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
sgRNA	Single guide RNA (gRNA fused to tracrRNA and crRNA)
Spy	*Streptococcus pyogenes*
ssDNA	Single-stranded DNA
TAE	Tris–acetate-EDTA
tracrRNA	Trans-activating CRISPR RNA
UAS	Upstream activation sequence

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s44330-024-00011-6.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

We thank Heiner Schrewe for methodological support and discussions, and David W. J. McQuarrie, Kinga Winczura and Neale Harrison for comments on the manuscript.

Authors' contributions

M.S. and I.U.H. conceived and directed the project. All authors performed experiments. A.L. compiled and analysed data. T.C.D., A.L., I.U.H. and M.S. contributed to writing the manuscript draft and all authors contributed to the final manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council and the Leverhulme trust.

Availability of data and materials

All data generated and protocols required for this study are included within this manuscript or in the additional files 1 and 2.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 25 June 2024 Accepted: 16 September 2024
Published online: 30 September 2024

References

1. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A. 2000;97(12):6640–5.
2. Zhang Y, Buchholz F, Muyrers JP, Stewart AF. A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet. 1998;20(2):123–8.
3. Copeland NG, Jenkins NA, Court DL. Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet. 2001;2(10):769–79.
4. Muyrers JP, Zhang Y, Stewart AF. Techniques: ecombinogenic engineering–new options for cloning and manipulating DNA. Trends Biochem Sci. 2001;26(5):325–31.
5. Mosberg JA, Lajoie MJ, Church GM. Lambda Red Recombineering in Escherichia coli Occurs Through a Fully Single-Stranded Intermediate. Genetics. 2010;186(3):791–9.
6. Lin-Chao S, Chen W-T, Wong T-T. High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. Mol Microbiol. 1992;6(22):3385–93.
7. Alves-Silva J, Sanchez-Soriano N, Beaven R, Klein M, Parkin J, Millard TH, Bellen HJ, Venken KJ, Ballestrem C, Kammerer RA, et al. Spectraplakins promote microtubule-mediated axonal growth by functioning as structural microtubule-associated proteins and EB1-dependent +TIPs (tip interacting proteins). J Neurosci. 2012;32(27):9143–58.
8. Silva F, Queiroz JA, Domingues FC. Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli. Biotechnol Adv. 2012;30(3):691–708.
9. Hanahan D. Studies on transformation of Escherichia coli with plasmids. J Mol Biol. 1983;166(4):557–80.
10. Haussmann IU, White K, Soller M. Erect wing regulates synaptic growth in Ddrosophilaby integration of multiple signaling pathways. Genome Biol. 2008;9(4):R73.
11. Haussmann IU, Li M, Soller M. ELAV-Mediated 3′-End Processing of ewg Transcripts Is Evolutionarily Conserved Despite Sequence Degeneration of the ELAV-Binding Site. Genetics. 2011;189(1):97–107.
12. Soller M, White K. ELAV inhibits 3′-end processing to promote neural splicing of ewg pre-mRNA. Genes Dev. 2003;17(20):2526–38.
13. Soller M, White K. ELAV multimerizes on conserved AU4-6 motifs important for ewg splicing regulation - PubMed. Mol Cell Biol. 2005;25(17):7580–91.
14. Haussmann IU, Ustaoglu P, Brauer U, Hemani Y, Dix Thomas C, Soller M. Plasmid-based gap-repair recombineered transgenes reveal a central role for introns in mutually exclusive alternative splicing in down syndrome cell adhesion molecule exon 4. Nucleic Acids Res. 2019;47(3):1389–403.
15. Wang J, Sarov M, Rientjes J, Fu J, Hollak H, Kranz H, Xie W, Stewart AF, Zhang Y. An improved recombineering approach by adding RecA to lambda red recombination. Mol Biotechnol. 2006;32(1):43–53.
16. van den Bosch M, Lohman PH, Pastink A. DNA double-strand break repair by homologous recombination. Biol Chem. 2002;383(6):873–92.
17. Duffy JB. GAL4 system in drosophila: a fly geneticist's swiss army knife. Genesis. 2002;34(1–2):1–15.
18. Haussmann IU, Dix TC, Mcquarrie DWJ, Dezi V, Hans AI, Arnold R, Soller M. sgRNA constraints and genetic limitations for efficient Cas9 genome editing to generate knock-outs. 2024; Genome Research, in press.
19. Thomason LC, Sawitzke JA, Li X, Costantino N, Court DL. Recombineering: genetic engineering in bacteria using homologous recombination. Curr Protoc Mol Biol. 2014;106:1 16 11-11 16 39.
20. Lüders S, Fallet C, Franco-Lara E, Lüders S, Fallet C, Franco-Lara E. Proteome analysis of the Escherichia coli heat shock response under steady-state conditions. Proteome Sci. 2009;7(1):36.
21. BACPAC Resources Center (BPRC). https://bacpacresources.org/.
22. Ensembl. https://www.ensembl.org/index.html.
23. Gene Services. https://www.geneservice.co.uk/.
24. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO, Gibson DG, Young L, Chuang R-Y, Venter JC, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6(5):343–5.
25. PlatinumCRISPR. https://platinum-crispr.bham.ac.uk/predict.pl.
26. Zarnegar BJ, Flynn RA, Shen Y, Do BT, Chang HY, Khavari PA, Zarnegar BJ, Flynn RA, Shen Y, Do BT, et al. irCLIP platform for efficient characterization of protein–RNA interactions. Nat Methods. 2016;13(6):489.
27. Kumagai K, Takanashi M, Ohno SI, Kuroda M, Sudo K. An improved Red/ET recombineering system and mouse ES cells culture conditions for the generation of targeted mutant mice. Exp Anim. 2017;66(2):125–36.
28. Venken KJ, He Y, Hoskins RA, Bellen HJ. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science. 2006;314(5806):1747–51.
29. Sarov M, Schneider S, Pozniakovski A, Roguev A, Ernst S, Zhang Y, Hyman AA, Stewart AF. A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. Nat Methods. 2006;3:839–44.
30. McElroy KA, Jung YL, Zee BM, Wang CI, Park PJ, Kuroda MI. upSET, the Drosophila homologue of SET3, is required for viability and the proper balance of active and repressive chromatin marks. G3 (Bethesda). 2017;7(2):625–35.
31. Urban JA, Doherty CA, Jordan WT, Bliss JE, Feng J, Soruco MM, Rieder LE, Tsiarli MA, Larschan EN. The essential *Drosophila* CLAMP differentially regulates non-coding roe RNAs in males and females. Chromosome Res. 2017;25:101–13.
32. Carreira-Rosario A, Scoggin S, Shalaby NA, Williams ND, Hiesinger PR, Buszczak M. Recombineering homologous recombination constructs in Drosophila. J Vis Exp. 2013;77:e50346.
33. Chan CC, Scoggin S, Hiesinger PR, Buszczak M. Combining recombineering and ends-out homologous recombination to systematically characterize Drosophila gene families: Rab GTPases as a case study. Commun Integr Biol. 2012;5(2):179–83.
34. Zaharieva E, Haussmann IU, Brauer U, Soller M. Concentration and localization of coexpressed ELAV/Hu proteins control specificity of mRNA Processing. Mol Cell Biol. 2015;35(18):3104–15.
35. Solomon DA, Stepto A, Au WH, Adachi Y, Diaper DC, Hall R, Rekhi A, Boudi A, Tziortzouda P, Lee YB, et al. A feedback loop between dipeptide-repeat protein, TDP-43 and karyopherin-alpha mediates C9orf72-related neurodegeneration. Brain. 2018;141(10):2908–24.
36. Dix TC. Characterisation of Elav multimerisation in neuron development and function. PhD thesis. University of Birmingham, School of Biosciences; 2021.
37. Wild J, Hradecna Z, Szybalski W. Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. Genome Res. 2002;12(9):1434–44.
38. Datta S, Costantino N, Zhou X, Court DL. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. Proc Natl Acad Sci U S A. 2008;105(5):1626–31.
39. Walhout AJ, Temple GF, Brasch MA, Hartley JL, Lorson MA, van den Heuvel S, Vidal M. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. Methods Enzymol. 2000;328:575–92.
40. Green MR, Sambrook J. Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press; 2012.
41. Soller M. Pre-messenger RNA processing and its regulation: a genomic perspective. Cell Mol Life Sci CMLS. 2006;63(7):796–819.
42. Koushika SP, Soller M, DeSimone SM, Daub DM, White K. Differential and inefficient splicing of a broadly expressed Drosophila erect wing transcript results in tissue-specific enrichment of the vital EWG protein isoform. Mol Cell Biol. 1999;19(6):3998.
43. Birnboim HC, Doly J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 1979;7(6):1513–23.

Publisher's Note