Iterative *in vivo* cut'n'paste of chromosomal loci in *Escherichia coli* K-12 using synthetic DNA

Authors: Sebastian Gude[1,*], Andreas B. Bertelsen[1], and Morten H. H. Nørholm[1,2]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark

[2] current address: Novo Nordisk A/S, Hørsholm, Denmark

***Corresponding author:** Sebastian Gude, The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark, segu@biosustain.dtu.dk

Research area:

Biological sciences, microbiology

Keywords:

easy-to-use, *in vivo*, genome engineering, genome refactoring, CRISPR Cas9, λ Red recombineering, homologous recombination, bacteria, *Escherichia coli*

Preprint:

bioRxiv: https://doi.org/10.1101/2025.06.19.660532

CC-BY-NC-ND 4.0 International license

Abstract

We have developed an easy-to-use *in vivo* method to iteratively relocate functional chromosomal loci onto an episome in *Escherichia coli* by utilizing synthetic DNA fragments. In this *in vivo* cut'n'paste procedure, "cutting" is executed by the RNA-guided DNA endonuclease Cas9 and a set of guides, while "pasting" is facilitated by the phage λ Red recombinase, which are all synthesized on easily curable helper plasmids. To demonstrate the utility of *in vivo* cut'n'paste, we commercially obtained synthetic DNA fragments containing locus-specific homology regions, antibiotic marker cassettes, and standardized Cas9 target sequences, and successfully relocated seven functional chromosomal loci. Scarless relocation mutants of the *trg*, *aer*, *tsr*, *malHM*, *malQT*, *macB-nadA*, and *folA* chromosomal loci were obtained as antibiotic-resistant isolates by combining Cas9 counterselection with the restoration of an antibiotic marker cassette. The additional antibiotic marker cassettes and standardized Cas9 target sequences present in the synthetic DNA fragments are inherently eliminated upon completion of the procedure, enabling iterative processing of chromosomal loci. *In vivo* cut'n'paste should be widely useful, especially in genome (re-)engineering efforts of *E. coli* and other bacteria because the procedure can be performed by non-specialists in unmodified wild-type cells.

Significance

Editing of genomes is crucial to understand the inner workings of bacteria and to alter them for biotechnological applications. *In vivo* cut'n'paste, in analogy to the

49 frequently used keyboard shortcuts Ctrl-X and Ctrl-V which allow to move text, images,
50 or files on computers, makes it easy to relocate large chunks of the bacterial genome
51 without requiring skills to manipulate DNA outside of the bacterial cell. Like Ctrl-X and
52 Ctrl-V, *in vivo* cut'n'paste can be executed several times to iteratively edit multiple
53 genomic locations in the same bacterium. Since *in vivo* cut'n'paste entirely takes place
54 inside the bacterial cell, it is particularly well suited to edit hard-to-tackle genomic
55 features such as essential genes.
56
57 **Introduction**
58
59 Democratising access to methodological advancements, i.e., making them
60 accessible to a broad audience of non-specialists, is a key component of method
61 development. The introduction of the point-and-click graphical user interface by Apple
62 and Microsoft[1], the creation of the "batteries included" programming language
63 Python[2], and the development of the one-step inactivation of chromosomal genes
64 procedure[3] are a few prominent examples of such endeavours. Graphical user
65 interfaces remove the hurdle to know how to interact with a computer via a text
66 console. The extensive Python Standard Library supports nearly every aspect of
67 computer science, thus, freeing programmers from having to reimplement many
68 commonly required features over and over again. Similarly, the one-step inactivation
69 of chromosomal genes procedure enables researchers to generate gene knockouts
70 without the need to assemble integrative plasmids.
71 The ability to manipulate genomes is key for uncovering and leveraging the
72 potential of (micro-)biology. A milestone in our ability to edit bacterial genomes was the
73 development of the one-step inactivation of chromosomal genes procedure[3]. Among

many other scientific and engineering endeavours, it facilitated the creation of a comprehensive single gene knockout library and the charting of essential genes in *E. coli*[4], the study of bacterial motility[5,6], stringent response[7], the interplay of growth and gene expression[8], two-component systems[9], evolution of antibiotic resistsance[10], and chromosomal organisation[11] as well as the development of strains with reduced genomes[12] and highly titratable induction systems[13] while also playing a crucial role in metabolic engineering applications[14,15]. Though the one-step inactivation procedure[3] was preceded by numerous other bacterial genome editing techniques[16–23], its ease of use and broad applicability ensured its still continuing success. The required short homology extensions to create genomic deletions are easily attached to an antibiotic marker cassette via PCR, the recombineering machinery is expressed from an easily curable plasmid, and the procedure does not require any prior modifications to the bacterial genome[3]. Furthermore, the antibiotic marker cassette can subsequently be eliminated via FLP-FRT recombination, resulting in an unmarked, yet not fully scarless, mutant that is compatible with another round of genome editing[3].

With the rise of CRISPR-Cas technologies[24], λ Red recombineering was successfully supplemented with sequence-specific RNA-guided counterselection[25–27]. Combining λ Red recombineering and CRISPR-Cas with intricate multi-layered selection, counterselection, and colour screening enables large-scale genome editing such as genome recoding[28,29], global genome rearrangement[30,31], and interspecies gene transfer[32]. While these sophisticated large-scale CRISPR-Cas genome editing techniques accomplish feats vastly beyond the capabilities of the one-step inactivation of chromosomal genes procedure[3], they unfortunately fall short in one aspect. Namely, they are relatively difficult to implement in non-specialist laboratories due to their

98 technical complexity, thus impeding the democratisation of large-scale genome
99 editing.

100 Building upon established large-scale CRISPR-Cas genome editing methods[28–
101 32], we present here *in vivo* cut'n'paste - an attempt to democratise large-scale genome
102 editing. *In vivo* cut'n'paste is an easy-to-use procedure to relocate chromosomal loci
103 onto an episome in wild-type cells that was designed to be performed by non-
104 specialists without expert skills in molecular biology. *In vivo* cut'n'paste solely relies on
105 commercially synthesized DNA fragments and a set of easily curable helper plasmids.
106 Leveraging CRISPR-Cas cutting and λ Red recombineering with commercially
107 synthesized DNA fragments completely removes the need for *in vitro* DNA
108 manipulations. Furthermore, easily curable helper plasmids and an antibiotic marker
109 recycling scheme inherently enable the iterative creation of multiple large-scale
110 genome edits. We successfully relocated seven functional genomic loci, including
111 metabolic and essential genes, in *E. coli* of up to 140 kb in size.

112

113 **Results**

114

115 *In vivo* cut'n'paste

116

117 *In vivo* cut'n'paste enables to iteratively relocate genomic loci from the native
118 chromosome onto an episome (Fig. 1A, Supp. Movie 1). In each round, a genomic
119 locus, defined by two pairs of tandem homology regions HR^A-HR^B and HR^C-HR^D is
120 excised from the native chromosome and subsequently integrated in between the
121 homology regions HR^E and HR^{TET} on an episome (Fig. 1B). The entire *in vivo*

122 cut'n'paste procedure occurs inside the cell without requiring any *in vitro* DNA
123 manipulations.

124 Each round of *in vivo* cut'n'paste consists of two phases: (1) pre-loading and
125 (2) relocation (Fig. 1B). During pre-loading, two commercially synthesized DNA
126 fragments, PRE^{epi} and PRE^{chr}, are sequentially integrated via homologous
127 recombination into the episome and the native chromosome, respectively, to equip
128 cells with the required modification for the subsequent relocation of the genomic locus
129 (Fig. 1C; see Supplementary Materials and Supp. Fig. 1-3 for a detailed description of
130 pre-loading fragment design choices). Both pre-loading events are accompanied by
131 positive antibiotic selection.

132 PRE^{epi} is flanked by the homology regions HR^{E} and HR^{TET}. Directly downstream
133 of HR^{E}, PRE^{epi} contains the inside homology region HR^{B} of the genomic locus. The
134 two outside homology regions HR^{A} and HR^{D} of the genomic locus are located next to
135 HR^{B} followed by a chloramphenicol marker cassette (Fig. 1C). Integration of PRE^{epi} at
136 HR^{E} and HR^{TET} into the episome truncates the tetracycline maker of the episome and,
137 thus, renders the cells sensitive to tetracycline yet resistant to chloramphenicol (Fig.
138 1B).

139 PRE^{chr} is flanked by the homology region HR^{C} and HR^{D} of the genomic locus,
140 enabling its integration directly adjacent to the genomic locus of interest in the native
141 chromosome (Fig. 1B). In PRE^{chr}, HR^{C} is directly followed by 5'TET^{R}, encoding a short,
142 non-functional fragment of the tetracycline marker cassette, and HR^{TET}. A kanamycin
143 marker cassette completes PRE^{chr} (Fig. 1C). Notably, the pre-loading fragments can
144 be commercially obtained in their entirety without requiring further assembly.

145 Once pre-loading is completed, cells contain all features required for relocation
146 (Fig. 1B,C). Firstly, the genomic locus in the native chromosome now contains HR^{B} on

147 one side and HR^{TET} on its other side. Secondly, the episome also contains both HR^{B}
148 and HR^{TET}. Thirdly, a “repair band-aid”, consisting of HR^{A}-HR^{D}, which can fix the
149 double-stranded break that will be created by the excision of the genomic locus from
150 the native chromosome, is present on the episome (see Supp. Movie 1 for a detailed,
151 step-by-step illustration of the pre-loading process). Optionally, additional features
152 may be introduced during pre-loading. For example, a terminator can be inserted in
153 between HR^{C} and 5’TET^{R} in EPI^{chr}, which will, once relocation is completed, partially
154 insulate the relocated locus from its new local environment on the episome.

155 Relocation is initiated by transforming the helper plasmid pCut (Fig. 1B). pCut
156 constitutively expresses standardized, functionally tested guides targeting constant
157 cut sites (CS^{1}, CS^{2}, and CS^{3}) included in the pre-loading fragments from a CRISPR
158 array, as well as a single guide (CS^{L}) targeting a locus-specific junction between HR^{A}
159 and HR^{B} (Fig. 1B,C). The sequence encoding the locus-specific single guide can
160 effectively be inserted as a third synthetic DNA fragment into the helper plasmid pCut
161 via I-CreI cutting counterselection-aided plasmid recombineering, thus completely
162 eliminating the need for *in vitro* DNA manipulations (see Supp. Fig. 4 for details
163 regarding I-CreI cutting counterselection-aided plasmid recombineering; a technical
164 description of the procedure is provided in the Methods). In total, RNA-guided Cas9
165 endonuclease activity now introduces double-stranded breaks into six targeted
166 locations across the episome and the native chromosome (Fig. 1B,C). These double-
167 stranded breaks release the genomic locus from the native chromosome and the
168 “repair band-aid” from the episome while simultaneously exposing the required
169 homology regions for integration of the genomic locus into the episome and for
170 integration of the “repair band-aid” into the native chromosome. Importantly,
171 integration of the genomic locus into the episome not only fixes its double-stranded

172 break but, additionally, leads to the restoration of the tetracycline marker cassette,
173 enabling positive selection of successful relocation events in the presence of
174 tetracycline (see Supp. Movie 1 for a detailed step-by-step illustration of the relocation
175 process).

176 The chloramphenicol and kanamycin marker cassettes, along with the constant
177 cut sites, which were all introduced during pre-loading as components of PRE^{epi} and
178 PRE^{chr}, are inherently lost during the relocation process. Finally, the helper plasmid
179 pCut encodes rhamnose-inducible, self-targeting guides enabling its straightforward
180 elimination from cells (see Methods for details), thus immediately enabling relocation
181 mutants to undergo additional rounds of *in vivo* cut'n'paste (Fig. 1B). We typically
182 tested four isolates by diagnostic PCR at each editing step (episome pre-loading,
183 native chromosome pre-loading, and relocation) of the *in vivo* cut'n'paste procedure.
184 An individual isolate possessing the desired genomic alternation was chosen for
185 further processing.

186

187 Relocating a genomic locus

188

189 In a first attempt, we employed *in vivo* cut'n'paste to relocate the *trg* locus from
190 its native genomic location in strain CNP^{empty} (*E. coli* K-12 MG1655 WT + EPI^{empty})
191 onto an empty episome (EPI^{empty}). Candidates of successful relocations, CNP^{trg} (*E.*
192 *coli* K-12 MG1655 Δ*trg* + EPI^{trg}), were obtained under tetracycline selection. After
193 curing the helper plasmids, we sent a CNP^{trg} isolate for short-read whole genome re-
194 sequencing (Supp. Table 1). Mapping reads of the CNP^{trg} isolate to the parental
195 CNP^{empty} reference depicted a depth of coverage of about 400x across the native
196 chromosome and the episome (Fig. 2A, middle row). No mutations relative to the

197 parental strain CNP^{empty} were observed. Yet, notably, pairs of adjacent bases at three
198 genomic locations were not spanned by any reads. Namely, both edges of the *trg* locus
199 on the native chromosome (HR^A-HR^B and HR^C-HR^D) and one edge of HR^E on the
200 episome lacked spanning reads (Fig. 2A, middle row). No lack of spanning reads at
201 these locations was observed when reads of the parental CNP^{empty} strain were
202 mapped to the CNP^{empty} reference (Fig. 2A, top row). The lack of spanning reads in
203 the CNP^{trg} isolate suggested that though the *trg* locus was present, it was not located
204 in its native genomic location, and, additionally, that something had likely been inserted
205 into the episome next to HR^E. We updated the reference map to include the desired
206 relocation of the *trg* locus based on our *in silico* design of the *in vivo* cut'n'paste
207 procedure. When the reads of the CNP^{trg} isolate were mapped to this updated
208 reference, the lack of spanning reads disappeared (Fig. 2A, bottom row), suggesting
209 that the *trg* locus had indeed been relocated from the native chromosome onto the
210 episome. Long-read sequencing of an extracted episome further supported the
211 successful relocation of the *trg* locus (Supp. Table 2). In total, we observed eleven
212 reads that spanned from the backbone of the episome across the entire, mutation-free
213 *trg* locus.

214 The *trg* locus (≈1.8 kb) contains a single gene encoding the transmembrane
215 Trg chemoreceptor. Trg mediates directed motile behaviour (i.e., chemotaxis) toward
216 D-ribose[33]. We assayed chemotactic colony expansion dynamics in glycerol minimal
217 medium soft agar plates in the presence and absence of low supplementations of D-
218 ribose as proxy for ribose sensing (see Methods for details). Initially, a small drop of
219 bacterial culture was spotted into the centre of the soft agar plates. Upon growing,
220 bacteria locally deplete nutrients and thus create gradients enabling chemotactic
221 sensing and expansion[34]. Expansion dynamics of *trg*-relocated CNP^{trg} isolates (blue)

222 in glycerol minimal medium soft agar plates in the presence (sold lines) and absence
223 (dashed lines) of D-ribose were comparable to the behaviour of their ancestor
224 CNP^{empty} (black), clearly indicating that *trg* retained its functionality after relocation
225 onto the episome (Fig. 2B,C). When the episome EPI^{trg} was removed from the CNP^{trg}
226 isolates (red), the expansion dynamics in presence of D-ribose were reduced to a
227 slower aerotactic behaviour[35] as also observed in glycerol minimal medium in the
228 absence of D-ribose, indicating that the ability to sense D-ribose and utilizing its
229 concentration gradients for chemotactic expansion was lost along with the episome
230 (Fig. 2B,C). As an additional test, we separately transferred the episomes EPI^{trg} and
231 EPI^{empty} into a strain background in which the *trg* gene had been disrupted by a
232 kanamycin marker cassette (*E. coli* K-12 MG1655 Δtrg::KanR). As expected, faster
233 chemotactic expansion dynamics in presence of D-ribose were restored only when
234 EPI^{trg}, containing the relocated *trg* locus, was present (Fig. 2D).
235
236 <u>Iteratively relocating genomic loci</u>
237
238 In accordance with the design principles of the *in vivo* cut'n'paste procedure,
239 the additional antibiotic marker cassettes and constant cut sites acquired during pre-
240 loading were inherently lost, and the tetracycline marker cassette on the episome was
241 restored, upon completion of the *trg* relocation (Fig. 1B). Thus, CNP^{trg} was capable of
242 undergoing further rounds of *in vivo* cut'n'paste. We iteratively applied *in vivo*
243 cut'n'paste to relocate two additional genomic loci onto the episome using CNP^{trg} as
244 a starting point (Fig. 3A). First, the *aer* locus (≈1.7 kb) encoding the aerotaxis receptor
245 Aer was relocated, resulting in $CNP^{trg,aer}$ (*E. coli* K-12 MG1655 Δ*trg* Δ*aer* + $EPI^{trg,aer}$).
246 Then, the *tsr* locus (≈1.8 kb) encoding the chemoreceptor Tsr, mediating aerotaxis and

247 sensing of L-Serine, was relocated, resulting in $CNP^{trg,aer,tsr}$ (*E. coli* K-12 MG1655 Δ*trg*
248 Δ*aer* Δ*tsr* + $EPI^{trg,aer,tsr}$).
249 The relocated loci were mutation-free (Supp. Table 2). The triple-relocated
250 $CNP^{trg,aer,tsr}$ qualitatively retained the colony expansion behaviour of its ancestor
251 CNP^{empty} in presence of D-ribose, L-Serine, or in the absence of supplementations to
252 glycerol minimal medium (Fig. 3 B,C). Furthermore, separately transferring the
253 episomes EPI^{empty} (Fig. 3D), EPI^{trg} (Fig. 3 E), and $EPI^{trg,aer}$ (Fig. 3F) into a Δ*trg* Δ*aer*
254 Δ*tsr* chromosomal background (i.e., a $CNP^{trg,aer,tsr}$ derivative from which $EPI^{trg,aer,tsr}$ had
255 been eliminated) partially restored the colony expansion behaviour in presence of D-
256 ribose (sensed via Trg), L-Serine (sensed via Tsr), and in absence of any
257 supplementations (energytaxis via Aer and Tsr) as expected[33,35].
258 In an independent attempt, using CNP^{empty} as a starting point, we successfully
259 iteratively relocated two genomic loci (≈8.1 kb and ≈7.9 kb) encoding 10 genes
260 involved in maltose utilisation onto an episome (Fig. 4A). Again, the relocated loci were
261 mutation-free. The resulting relocation mutants CNP^{malHM} (*E. coli* K-12 MG1655
262 Δ(*malH-malM*) + EPI^{malHM}) and $CNP^{malHM,malQT}$ (*E. coli* K-12 MG1655 Δ(*malH-malM*)
263 Δ(*malQ-malT*) + $EPI^{malHM,malQT}$) retained their ability to uptake and ferment maltose as
264 indicated by producing a dark red phenotype when spotted and grown on maltose-
265 MacConkey/agar plates (Fig. 4B).
266
267 <u>Relocating a large genomic locus and an essential gene</u>
268
269 To illustrate the utility of *in vivo* cut'n'paste to process large genomic loci, we
270 attempted to relocate the *macB-nadA* locus (≈140 kb) containing 135 non-essential

271 genes from its native genomic location onto an episome. The relocation was
272 successful and mutation-free (Supp. Table 1).

273 As opposed to traditional sequential deletion and complementation procedures,
274 all genes, including those located inside the genomic locus that is edited, remain
275 present throughout the entire *in vivo* cut'n'paste process. Hence, *in vivo* cut'n'paste
276 should be well suited for manipulating hard-to-edit genomic features such as essential
277 genes. Indeed, we successfully relocated the essential gene *folA* onto an episome,
278 resulting in CNP^{folA} (*E. coli* K-12 MG1655 Δ*folA* + EPI^{folA}).

279 An episome containing an essential gene that is simultaneously absent from
280 the native chromosome is expected to be stably maintained even in the absence of
281 antibiotic selection because cells are unable to survive without the essential gene. As
282 a direct test, we removed the tetracycline antibiotic marker cassette from EPI^{folA} and
283 inserted a highly expressed green fluorescent reporter (GFP) in its place, resulting in
284 $CNP^{folA,TetR::gfp}$ (*E. coli* K-12 MG1655 Δ*folA* + $EPI^{folA,TetR::gfp}$). We then assayed the
285 retention of the antibiotic marker-free episome $EPI^{folA,TetR::gfp}$ by serial culturing of
286 $CNP^{folA,TetR::gfp}$ in LB medium without any antibiotic supplementation (see Methods for
287 details). Plating on non-selective LB agar plates was performed at each transfer. After
288 overnight incubation, plates were imaged in GFP and brightfield channels (Fig. 5,
289 Supp. Fig. 5). At each transfer, all fluorescent colonies visible in the GFP channel (Fig.
290 5A) had a corresponding colony in the associated brightfield image (Fig. 5B) and vice
291 versa. Thus, 100% of the bacterial population stably retained the antibiotic marker-free
292 episome $EPI^{folA,TetR::gfp}$ throughout ten serial passages in the absence of antibiotic
293 selection.

294

295 <u>Unintended side effects</u>

296

297 All relocated genomic loci were mutation-free. Yet, some of the relocation
298 mutants acquired background mutations in their native chromosome while undergoing
299 *in vivo* cut'n'paste (Supp. Table 7). In general, it remains unclear whether these
300 background mutations have any functional relevance or are merely coincidental
301 genetic drift.

302

303

304 **Discussion**

305

306 We developed *in vivo* cut'n'paste - an easy-to-use procedure to relocate
307 chromosomal loci onto an episome in *E. coli* that does not rely on any *in vitro* DNA
308 manipulations. We successfully relocated genomic loci encoding metabolic and
309 essential genes, generated a genomic alteration >100 kb, and performed up to three
310 consecutive edits in a single strain. All relocated loci, including the 140 kb *macB-nadA*
311 relocation, were mutation-free. Due to its inherent robustness, as each editing step is
312 accompanied by positive antibiotic selection, *in vivo* cut'n'paste is well-suited for
313 implementation in non-specialist laboratories or via unsupervised lab automation.

314 *In vivo* cut'n'paste can be leveraged to precisely characterise gene function.
315 One round of *in vivo* cut'n'paste effectively creates a complementation mutant in which
316 the gene of interest is relocated onto a single-copy extra-chromosomal replicon. Thus,
317 *in vivo* cut'n'paste inherently overcomes the limitations of traditional sequential "delete
318 and complement" procedures hampering the investigation of essential genes while
319 also circumventing any convoluting copy number effects present when using
320 traditional multi-copy plasmid backbones.

321 By iterating the *in vivo* cut'n'paste procedure, arbitrary modules of genes, like
322 those obtained here (Fig. 3 & 4), can be created. For example, genes that are co-
323 modulated or functionally connected but natively dispersed across the chromosome
324 can easily be clustered onto a single extra-chromosomal replicon[36]. Utilising the origin
325 of transfer (oriT) encoded on the episome, such modules can be transferred to other
326 strains or species via conjugation without the size-limitations imposed by
327 transformation[37], and, hence, potential future adaptations of *in vivo* cut'n'paste to other
328 bacterial species will only broaden its utility (see Supplementary Text).
329 Its ability to relocate genomic loci makes *in vivo* cut'n'paste particularly well-
330 suited for studying how genome architecture impacts cell functioning, a topic of which
331 we are only beginning to scratch the surface[38,39]. Ultimately, *in vivo* cut'n'paste can
332 facilitate alternative avenues to obtain minimal genomes via bottom-up approaches to
333 complement traditional, ineffectual top-down genome reduction attempts.
334 Moreover, relocations of essential genes, such as demonstrated here in
335 CNP^{folA}, are highly desirable in the context of large-scale fermentation applications in
336 green biotechnology as they allow to stably maintain transferable extra-chromosomal
337 replicons in production strains without incurring the metabolic costs and legal
338 requirements associated with continual antibiotic selection[40].
339 Akin to the graphical user interface in personal computing[1], *in vivo* cut'n'paste
340 does not enable any feature that is not already present in existing techniques. We also
341 do not claim that *in vivo* cut'n'paste is faster or more efficient compared to other large-
342 scale genome editing procedures. Instead, the main contribution of *in vivo* cut'n'paste
343 lies in its ease of use, as it does not require any DNA manipulations outside of the
344 bacterial cell, not even for the engineering of helper plasmids. While specialists in
345 molecular biology may view multi-fragment DNA assemblies and PCRs of large, and

346 potentially complex, DNA pieces as trivial undertakings, such tasks can appear as
347 insurmountable hurdles to researchers utilising bacteria in fields such as biophysics,
348 protein structure characterisation, evolutionary biology, or large-scale fermentation. *In*
349 *vivo* cut'n'paste is an attempt to democratise large-scale genome editing and, thus, to
350 open up the world of large-scale genome editing to non-specialists in molecular
351 biology. In essence, *in vivo* cut'n'paste solely requires (1) to decide which genomic
352 locus to relocate while identifying the homology regions HR^A-HR^D (see Fig. 1B), (2) to
353 complete the annotated pre-loading fragment genbank file templates (supplied in the
354 Supplementary Materials), (3) to commercially obtain standardised synthetic DNA
355 fragments, (4) to transform these synthetic DNA fragments and the easily curable
356 helper plasmids into the strain of interest while culturing in appropriate selective
357 conditions, (5) to cure helper plasmids, and, finally, (6) to send the mutants for whole
358 genome re-sequencing. Thus, *in vivo* cut'n'paste empowers experimenters to refactor
359 genomes, akin to the seven genomic edits demonstrated here, without requiring
360 specialized knowledge in molecular biology methodologies.
361
362 **Materials and Methods**
363
364 See "Methods" document.
365
366 **Data availability**
367
368 All data is publicly available (see Supp. Table 8 for details). All custom-written
369 scripts are supplied in the Supplementary Materials. Sequencing data was deposited
370 at NCBI (PRJNA1271273; see Supp. Table 1, 2, and 3 for assession numbers).

371 Sequence-verified, annotated maps of episomes and helper plasmids are supplied in
372 the Supplementary Materials. All *in vivo* cut'n'paste homology regions are listed in
373 Supp. Table 6. Strains are available for non-commercial use through the Belgian
374 Coordinated Collections of Microorganisms (https://bccm.belspo.be; see Supp. Table
375 1 for assession numbers). Helper plasmids are available for non-commercial use
376 through Addgene (https://addgene.org; see Supp. Table 3 for assession numbers).
377

378 **Authors' contributions**

379

380 SG, ABB, and MHHN developed the initial concept of *in vivo* cut'n'paste. SG
381 implemented, tested, and refined *in vivo* cut'n'paste iteratively. SG performed all
382 experiments, analysed the data, and wrote the manuscript. All authors read the final
383 version of the manuscript.
384

385 **Conflicts of interest**

386

387 The authors do not have any conflicts of interest to disclose.
388

389 **Acknowledgement and funding information**

390

391 The authors would like to acknowledge Valeria Stavila (The Novo Nordisk
392 Foundation Center for Biosustainability, Technical University of Denmark, Kongens
393 Lyngby, Denmark) for technical assistance. S.G. would like to acknowledge Alex T.
394 Nielsen (The Novo Nordisk Foundation Center for Biosustainability, Technical
395 University of Denmark, Kongens Lyngby, Denmark) for hosting the final period of

research activity in his laboratory. *E. coli* K-12 MG1655 WT was a kind gift of Tom Shimizu (AMOLF, Amsterdam, The Netherlands). The title and the abstract of this manuscript pay homage to the article “One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products”. This work was supported by The Novo Nordisk Foundation.

References

1. Abbate, J. Getting small: a short history of the personal computer. *Proc. IEEE* **87**, 1695–1698 (1999).
2. Dubois, P. F. Guest Editor’s Introduction: Python: Batteries Included. *Comput. Sci. Eng.* **9**, 7–9 (2007).
3. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97**, 6640–6645 (2000).
4. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
5. Chapman, M. R. *et al.* Role of *Escherichia coli* Curli Operons in Directing Amyloid Fiber Formation. *Science* **295**, 851–855 (2002).
6. Smith, K. D. *et al.* Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. *Nat. Immunol.* **4**, 1247–1253 (2003).
7. Traxler, M. F. *et al.* The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol. Microbiol.* **68**, 1128–1148 (2008).

8. Klumpp, S., Zhang, Z. & Hwa, T. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell* **139**, 1366–1375 (2009).

9. Zhou, L., Lei, X.-H., Bochner, B. R. & Wanner, B. L. Phenotype MicroArray Analysis of *Escherichia coli* K-12 Mutants with Deletions of All Two-Component Systems. *J. Bacteriol.* **185**, 4956–4972 (2003).

10. Liu, J., Gefen, O., Ronin, I., Bar-Meir, M. & Balaban, N. Q. Effect of tolerance on the evolution of antibiotic resistance under drug combinations. *Science* **367**, 200–204 (2020).

11. Lau, I. F. *et al.* Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Mol. Microbiol.* **49**, 731–743 (2003).

12. Kolisnychenko, V. *et al.* Engineering a Reduced *Escherichia coli* Genome. *Genome Res.* **12**, 640–647 (2002).

13. Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J. & Voigt, C. A. Escherichia coli “Marionette” strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol.* **15**, 196–204 (2019).

14. Lee, K. H., Park, J. H., Kim, T. Y., Kim, H. U. & Lee, S. Y. Systems metabolic engineering of *Escherichia coli* for L -threonine production. *Mol. Syst. Biol.* **3**, 149 (2007).

15. Atsumi, S. *et al.* Metabolic engineering of Escherichia coli for 1-butanol production. *Metab. Eng.* **10**, 305–311 (2008).

16. Winans, S. C., Elledge, S. J., Krueger, J. H. & Walker, G. C. Site-directed insertion and deletion mutagenesis with cloned fragments in Escherichia coli. *J. Bacteriol.* **161**, 1219–1221 (1985).

17. Hamilton, C. M., Aldea, M., Washburn, B. K., Babitzke, P. & Kushner, S. R. New method for generating deletions and gene replacements in Escherichia coli. *J. Bacteriol.* **171**, 4617–4622 (1989).

18. Russell, C. B., Thaler, D. S. & Dahlquist, F. W. Chromosomal Transformation of Escherichia coli recD Strains with Linearized Plasmidst. *J BACTERIOL* **171**, (1989).

19. Metcalf, W. W. *et al.* Conditionally Replicative and Conjugative Plasmids Carrying lacZa for Cloning, Mutagenesis, and Allele Replacement in Bacteria. *Plasmid* **35**, 1–13 (1996).

20. Link, A. J., Phillips, D. & Church, G. M. Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: application to open reading frame characterization. *J. Bacteriol.* **179**, 6228–6237 (1997).

21. Kato, C., Ohmiya, R. & Mizuno, T. A Rapid Method for Disrupting Genes in the Escherichia coli Genome. *Biosci. Biotechnol. Biochem.* **62**, 1826–1829 (1998).

22. Zhang, Y., Buchholz, F., Muyrers, J. P. P. & Stewart, A. F. A new logic for DNA engineering using recombination in Escherichia coli. *Nat. Genet.* **20**, 123–128 (1998).

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

24. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).

25. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).

26. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified

467 Chromosomal Gene Replacement in Escherichia coli. *Appl. Environ. Microbiol.* **81**,
468 5103–5114 (2015).
469 27. Jiang, Y. *et al.* Multigene Editing in the Escherichia coli Genome via the CRISPR-
470 Cas9 System. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).
471 28. Wang, K. *et al.* Defining synonymous codon compression schemes by genome
472 recoding. *Nature* **539**, 59–64 (2016).
473 29. Zürcher, J. F. *et al.* Continuous synthesis of E. coli genome sections and Mb-
474 scale human DNA assembly. *Nature* **619**, 555–562 (2023).
475 30. Wang, K., De La Torre, D., Robertson, W. E. & Chin, J. W. Programmed
476 chromosome fission and fusion enable precise large-scale genome rearrangement
477 and assembly. *Science* **365**, 922–926 (2019).
478 31. Su, J. *et al.* A CRISPR-based chromosomal-separation technique for Escherichia
479 coli. *Microb. Cell Factories* **21**, 235 (2022).
480 32. Teufel, M., Klein, C. A., Mager, M. & Sobetzko, P. A multifunctional system for
481 genome editing and large-scale interspecies gene transfer. *Nat. Commun.* **13**, 3430
482 (2022).
483 33. Kondoh, H., Ball, C. B. & Adler, J. Identification of a methyl-accepting chemotaxis
484 protein for the ribose and galactose chemoreceptors of Escherichia coli. *Proc. Natl.*
485 *Acad. Sci.* **76**, 260–264 (1979).
486 34. Adler, J. Chemotaxis in Bacteria. *Science* **153**, 708–716 (1966).
487 35. Zhulin, I. B., Rowsell, E. H., Johnson, M. S. & Taylor, B. L. Glycerol elicits energy
488 taxis of Escherichia coli and Salmonella typhimurium. *J. Bacteriol.* **179**, 3196–3201
489 (1997).

36. Tan, J. *et al.* Independent component analysis of E. coli's transcriptome reveals the cellular processes that respond to heterologous gene expression. *Metab. Eng.* **61**, 360–368 (2020).

37. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).

38. El Houdaigui, B. *et al.* Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Res.* **47**, 5648–5657 (2019).

39. Dame, R. T., Rashid, F.-Z. M. & Grainger, D. C. Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat. Rev. Genet.* **21**, 227–242 (2020).

40. Sevillano, L., Díaz, M. & Santamaría, R. I. Development of an antibiotic marker-free platform for heterologous protein production in Streptomyces. *Microb. Cell Factories* **16**, 164 (2017).

Figures and figure legends

See "Figures" document.

Supplementary Materials

See "SupplementaryMaterials" file.

Iterative *in vivo* cut'n'paste of chromosomal loci in *Escherichia coli* K-12 using synthetic DNA

Figures and figure legends

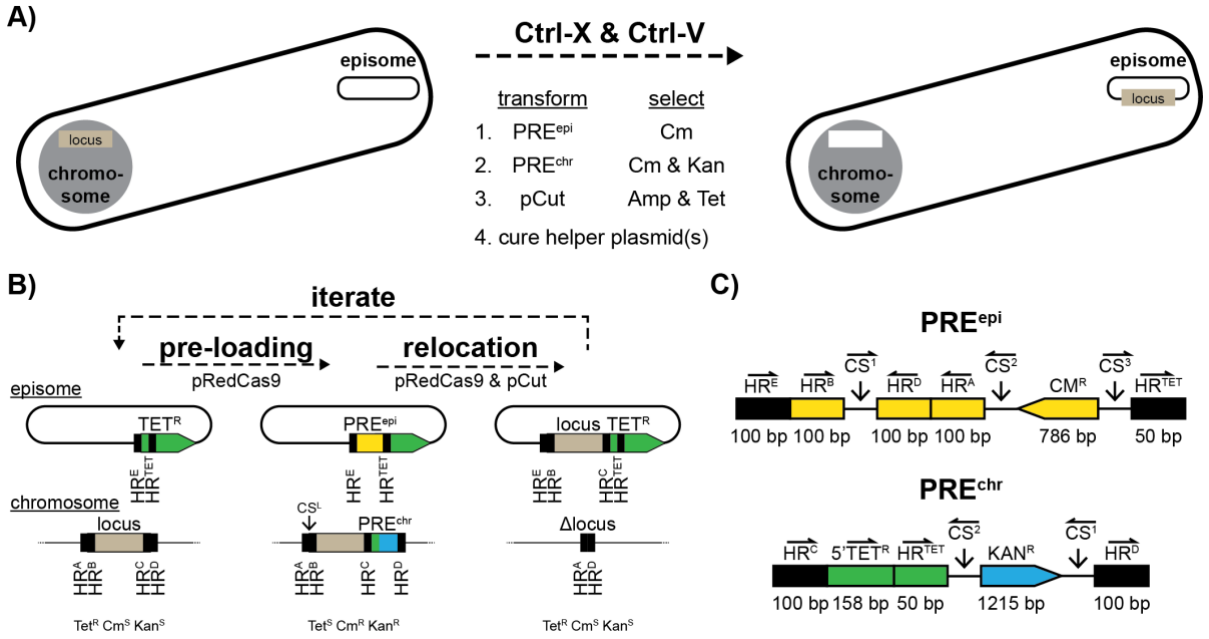


Figure 1: Iterative *in vivo* cut'n'paste

a) Graphical overview of *in vivo* cut'n'paste. **b)** Outline of the individual steps (pre-loading and relocation) of the *in vivo* cut'n'paste procedure. **c)** Features of commercially synthesized DNA fragments PRE^{epi} and PRE^{chr}. A more detailed description of the design choices is provided in the Supplementary Materials. Half-arrows indicate the relative orientations of features. Solid, downwards arrows mark constant cut sites. pRedCas9 and pCut: easily curable helper plasmids, HR: homology region, CS: cut site, TET^R: tetracycline marker cassette, CM^R: chloramphenicol marker cassette, KAN^R: kanamycin marker cassette, $5'TET^R$: non-functional fragment of TET^R, Tet: tetracycline, Cm: chloramphenicol, Kan: kanamycin, R: resistant, S: susceptible, and bp: base pairs. Not to scale.

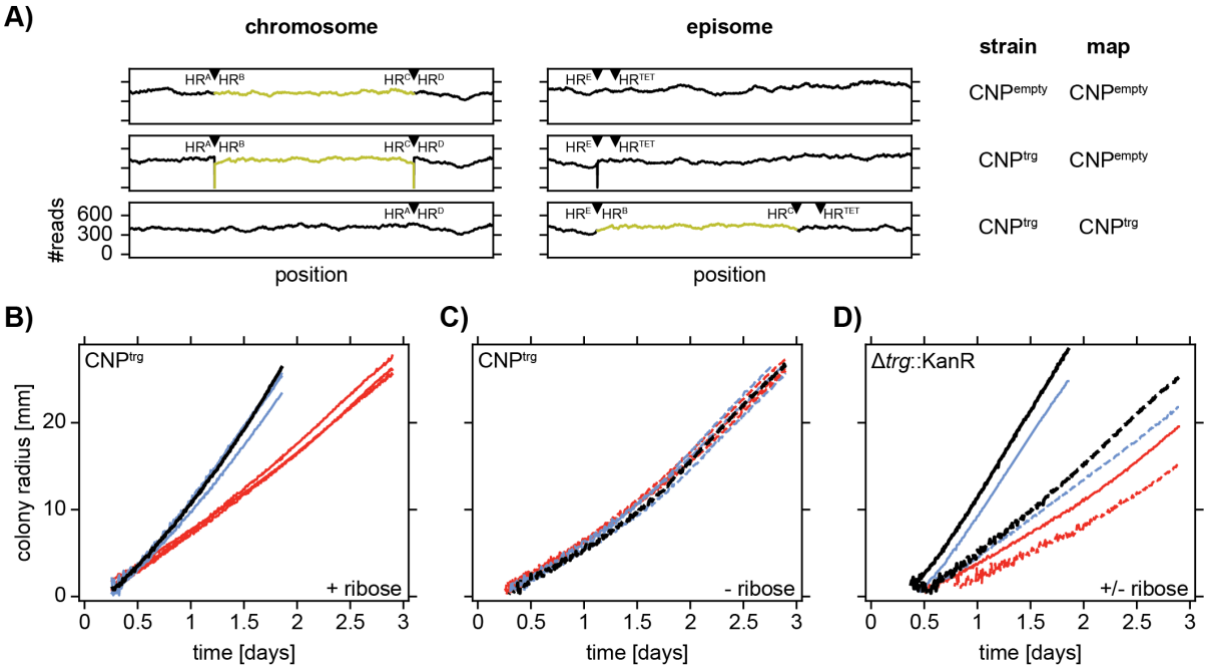


Figure 2: *In vivo* cut'n'paste of the *trg* locus

a) Number of mapped reads in 3.2 kb segments of the native chromosome and the episome. Only reads mapping to a position and to both of its direct neighbors are counted. Strain identities and reference maps as indicated within the panel. The *trg* locus is highlighted in yellow. Triangles mark locations of homology regions. **b-d)** Motility phenotypes in soft agar. Colony expansion as function of time in glycerol minimal medium with (solid) or without (dashed) 100 μM D-ribose. **b-c)** CNP^{trg} (blue), CNP^{trg} without EPI^{trg} (red), and CNP^{empty} (black). Three independent CNP^{trg} isolates and a single replicate of CNP^{empty}. **d)** *E. coli* K-12 MG1655 Δtrg::KanR with EPI^{trg} (blue), *E. coli* K-12 MG1655 Δtrg::KanR with EPI^{empty} (red), and CNP^{empty} (black). A single replicate each.

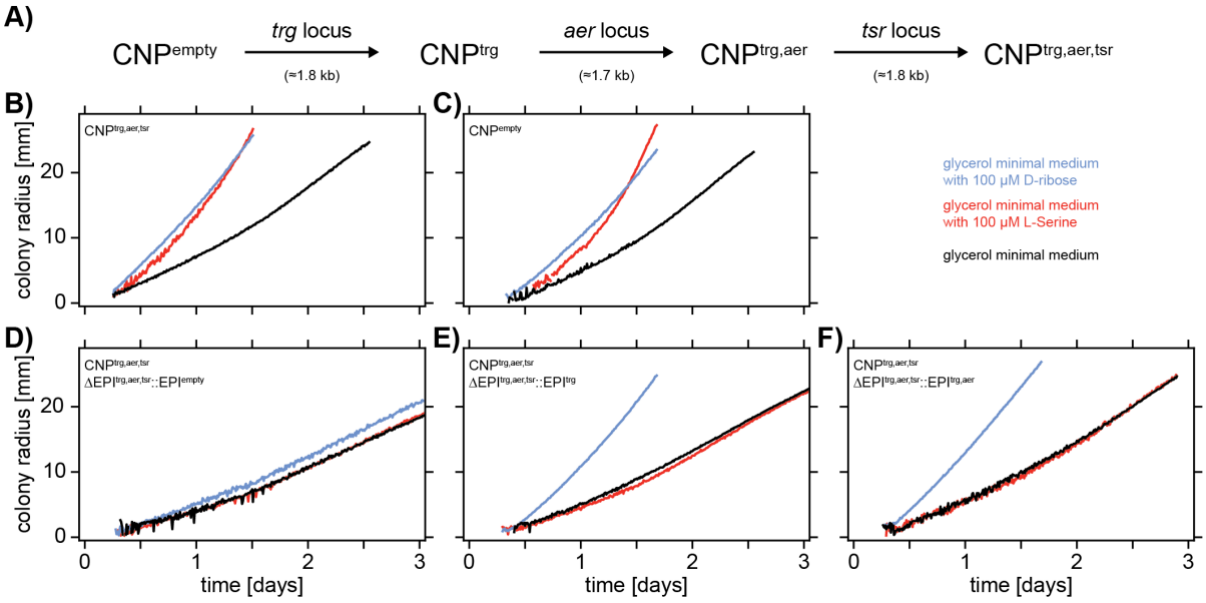


Figure 3: Iterative *in vivo* cut'n'paste of three chemoreceptor loci

a) Illustration of iterative *in vivo* cut'n'paste of the *trg*, *aer*, and *tsr* chemoreceptor loci.

b-f) Motility phenotypes in soft agar. Colony expansion as function of time in glycerol minimal medium with 100 µM D-ribose (blue), with 100 µM L-Serine (red), or without any additions (black). Strain identities as defined within the individual panels.

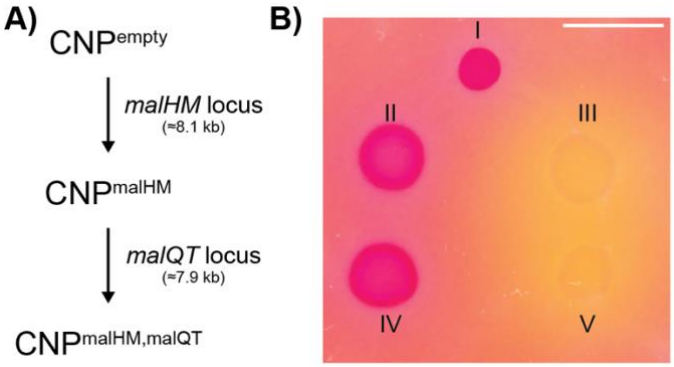


Figure 4: Iterative *in vivo* cut'n'paste of two maltose-utilization loci

a) Illustration of iterative *in vivo* cut'n'paste of the *malHM* (i.e., *malEFGH* and *malKlamBmalM*) and *malQT* (i.e., *malPQ* and *malT*) maltose-utilization loci. **b)** Maltose utilization of CNP strains. Spotting of liquid culture onto maltose-MacConkey+Tet/agar. Imaged after 9 hours. Dark red disks: cells can utilize maltose, pale yellow disks with halo: cells cannot utilize maltose. The orange-reddish background depicts MacConkey/agar. I: CNP^{empty}, II: CNP^{malHM}, III: CNP^{malHM} ΔEPI^{malHM}::EPI^{empty}, IV: $CNP^{malHM,malQT}$, and V: $CNP^{malHM,malQT}$ $\Delta EPI^{malHM,malQT}$::EPI^{empty}. Scale bar: 10 mm.

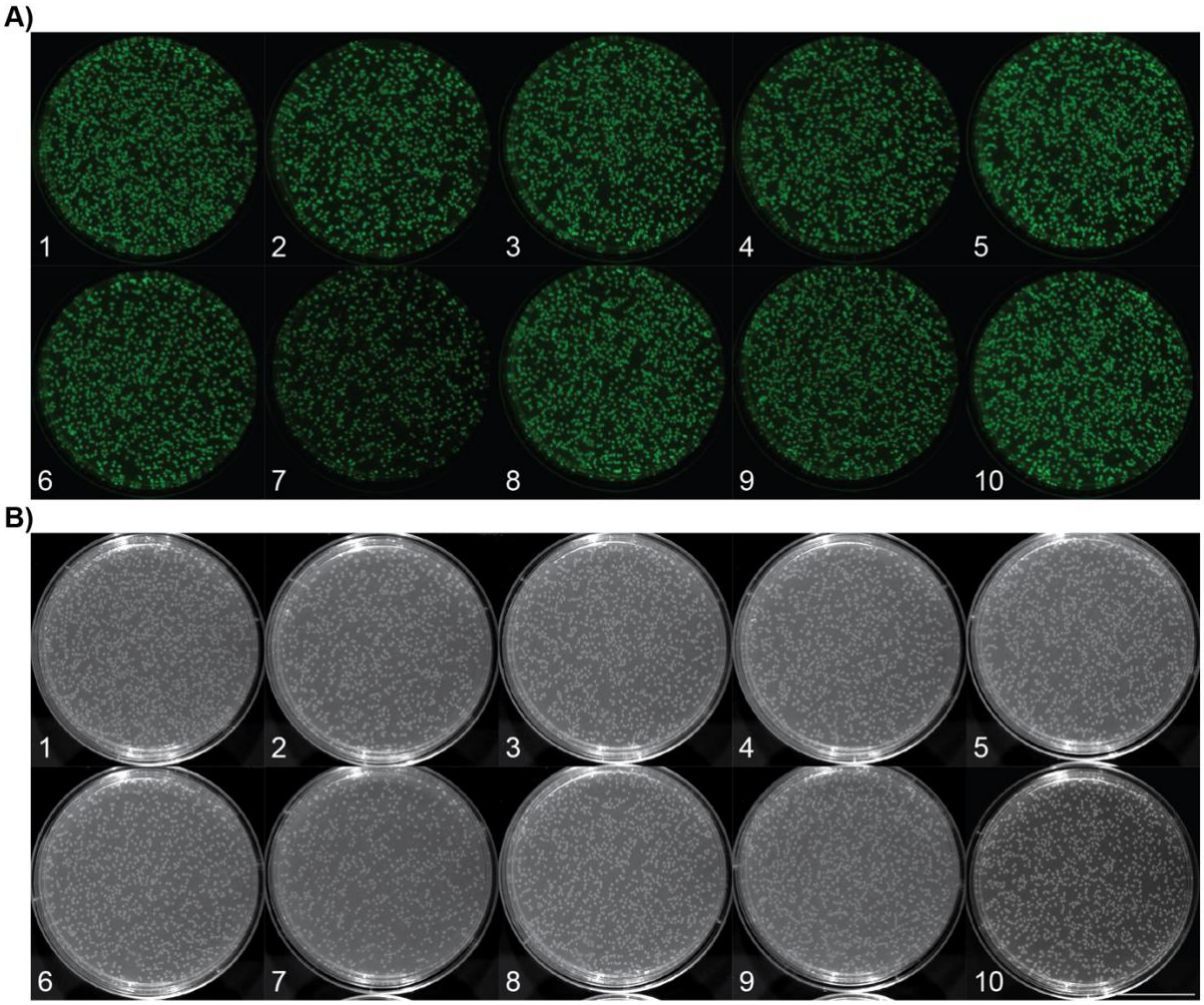
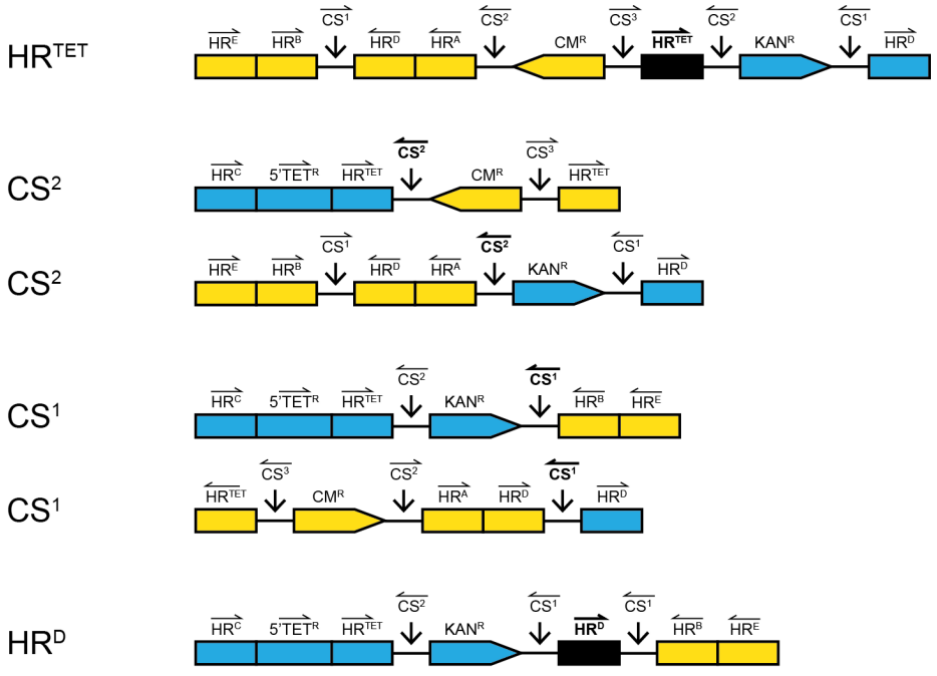


Figure 5: Episome retention in absence of antibiotic selection

a) False-colored fluorescence (GFP; excitation: 455-485 nm, emission: 508-557 nm, exposure time: 15 msec) and **b)** brightfield images of $CNP^{folA,TetR::gfp}$ plated onto LB agar at transfers during serial culturing in LB without any antibiotic supplementation. Numerals within the panels indicate transfer numbers. Cultures reached comparable densities of $OD_{600} \approx 4$ at each transfer. Scale bar: 30 mm.

60



61

62

63 **Supplementary Figure 1:** Potential for “merging” of synthetic DNA fragments during

64 simultaneous pre-loading

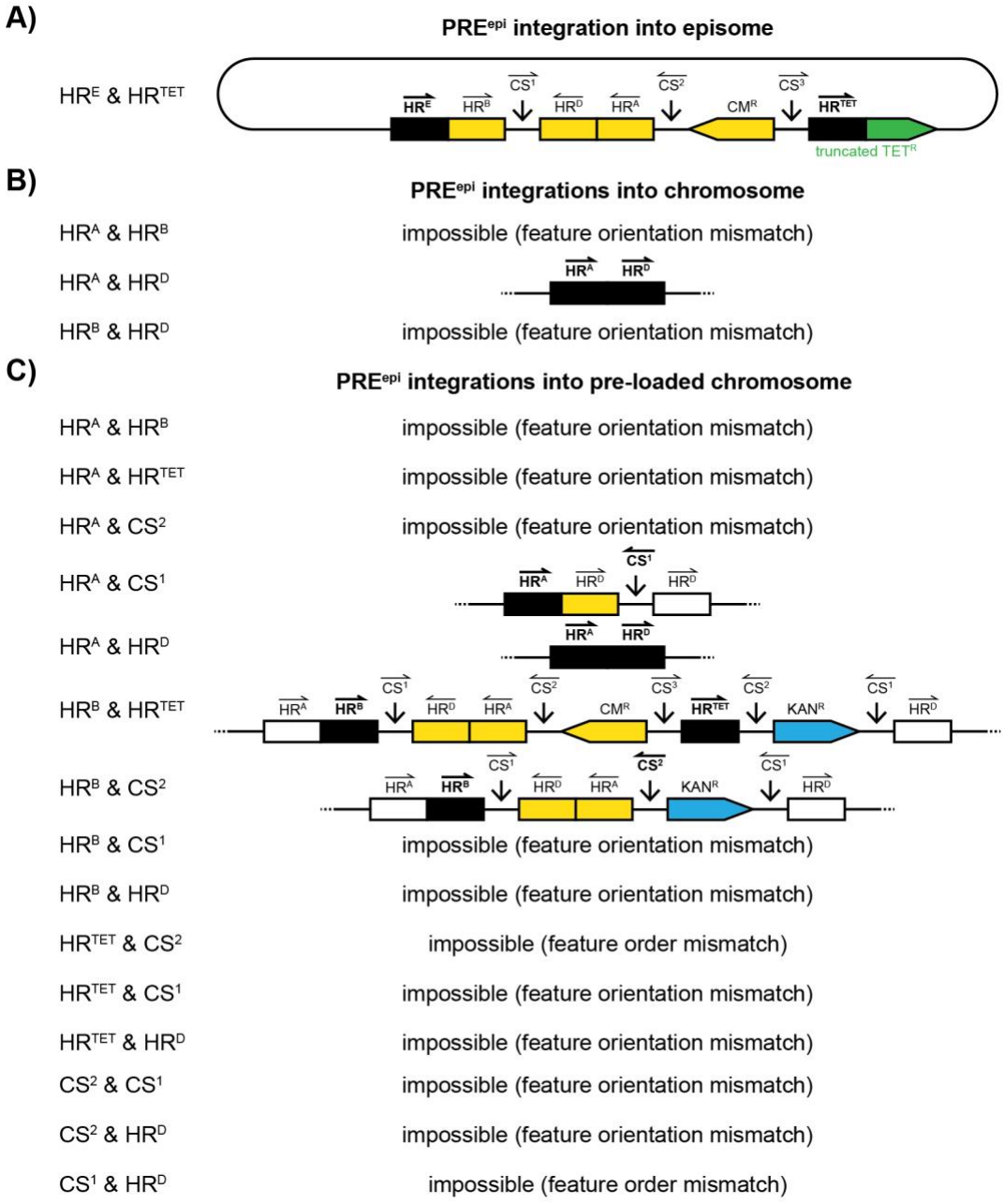
65 Illustration of “merged” dimers of the synthetic DNA fragments PRE^{epi} (yellow) and

66 PRE^{chr} (blue) due to sequence homology at shared features. “Merging” of PRE^{epi} and

67 PRE^{chr} causes feature loss. Locations of “merging” are highlighted in bold and,

68 additionally, stated on the left of the illustrations. Definitions as defined in Fig. 1. Not

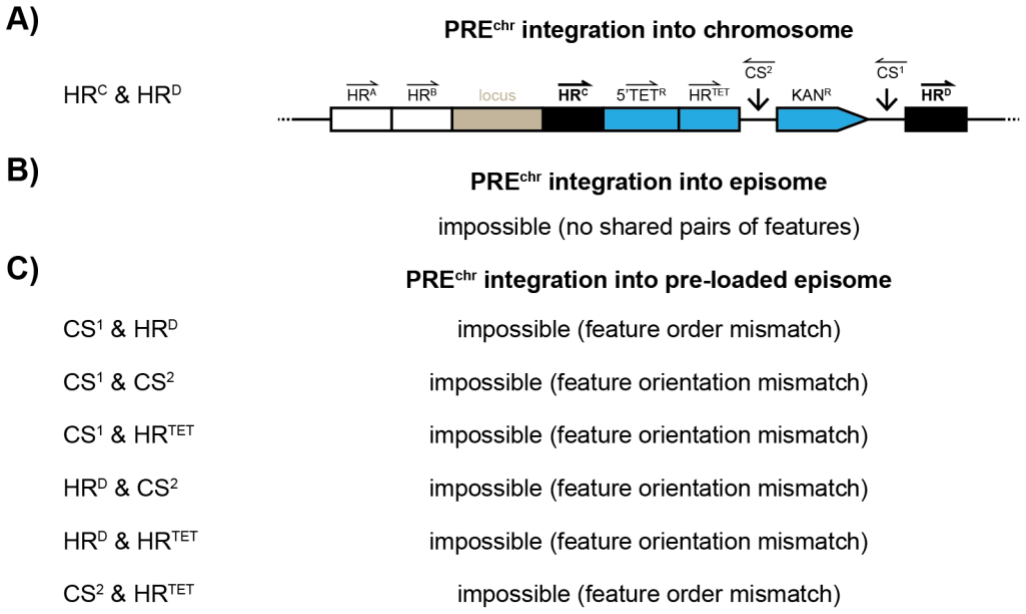
69 to scale.



Supplementary Figure 2: Potential integration events of synthetic DNA fragment PRE^{epi} during sequential pre-loading

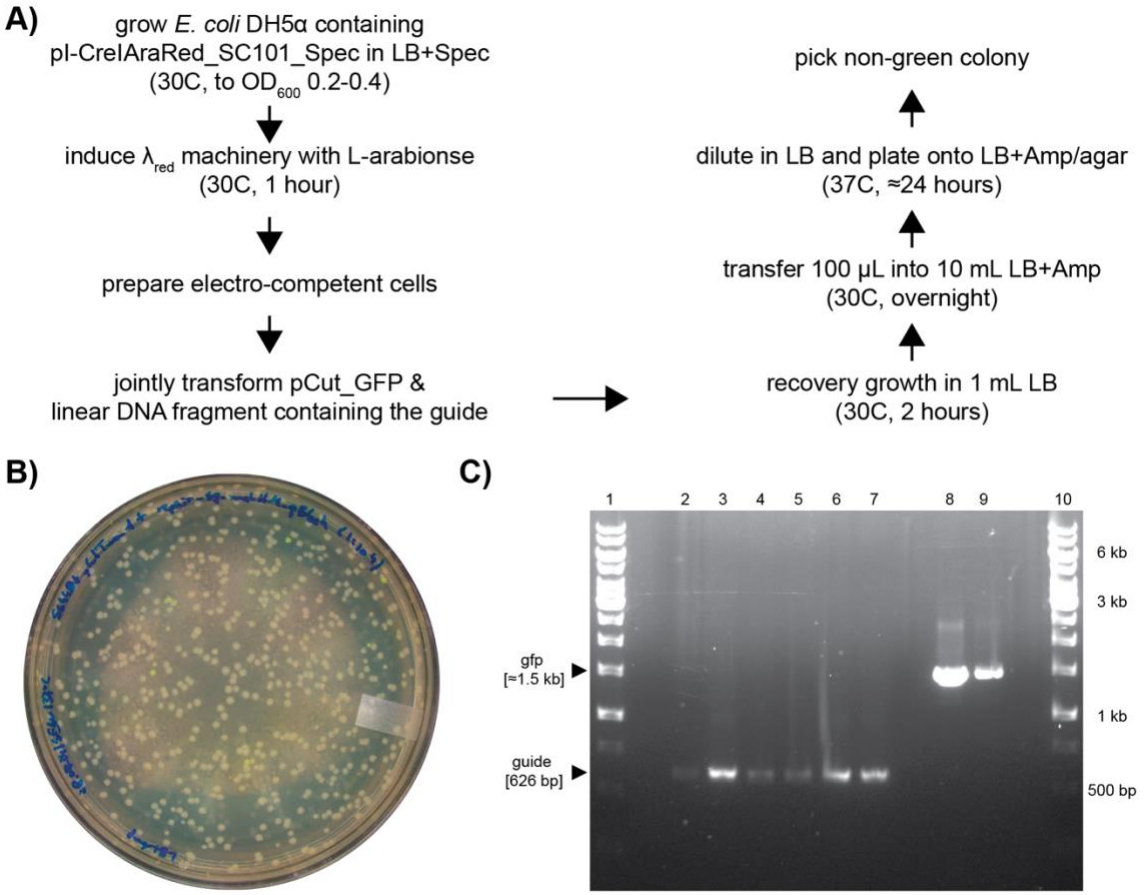
Illustration of integration events of the synthetic DNA fragment PRE^{epi} (yellow) due to homology at pairs of shared features into **a)** the episome, **b)** the native chromosome, or **c)** a native chromosome containing PRE^{chr} (blue). Integration events may cause feature loss. Many pairs of shared features cannot undergo successful integration due

to their relative orientations (feature orientation mismatch) or relative locations (feature order mismatch). Integration homologies are highlighted in bold and, additionally, stated on the left side of the illustrations. Definitions as defined in Fig. 1. Not to scale.



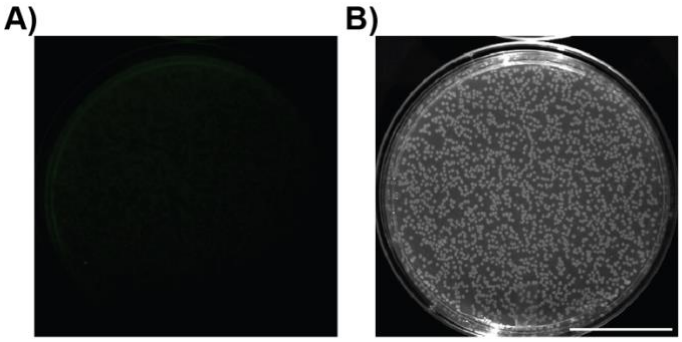
Supplementary Figure 3: Potential integration of synthetic DNA fragment PRE^{chr} during sequential pre-loading

Illustration of integration events of the synthetic DNA fragment PRE^{chr} (blue) due to homology at pairs of shared features into **a)** the native chromosome, **b)** the episome, or **c)** an episome containing PRE^{epi}. Most pairs of shared features cannot undergo successful integration due to their relative orientations (feature orientation mismatch) or relative locations (feature order mismatch). Homologies used for integration are highlighted in bold and, additionally, stated on the left side of the illustrations. Definitions as defined in Fig. 1. Not to scale.



Supplementary Figure 4: Insertion of locus-specific guide sequence into pCut_GFP via I-CreI cutting counterselection-aided plasmid recombineering

a) Overview of I-CreI cutting counterselection-aided plasmid recombineering workflow. **b)** Image of transformants (1:10^4 dilution) plated onto LB-agar containing ampicillin [100 mg/L] (plate diameter: 90 mm). **c)** Agarose gel electrophoresis of diagnostic PCRs across the engineering location in pCut (primers: diag_guide_A, diag_guide_B). Non-green transformants (lanes 2-7), green transformants (lanes 8 and 9), and “gene ruler [1 kb]” DNA ladder (lanes 1 and 10). Expected band sizes are indicated by black triangles.



Supplementary Figure 5: *E. coli* K-12 MG1655 WT plated on LB agar

a) False-colored fluorescence (GFP; excitation: 455-485 nm, emission: 508-557 nm, exposure time: 15 msec) and **b)** brightfield images of *E. coli* K-12 MG1655 WT plated onto LB agar. Identical contrast settings as in Fig. 5. Scale bar: 30 mm.