

Knock-in/Knock-out (KIKO) vectors for rapid integration of large DNA sequences, including whole metabolic pathways, onto the *Escherichia coli* chromosome at well-characterised loci

Sabri *et al.*

TECHNICAL NOTES

Open Access

Knock-in/Knock-out (KIKO) vectors for rapid integration of large DNA sequences, including whole metabolic pathways, onto the *Escherichia coli* chromosome at well-characterised loci

Suriana Sabri[†], Jennifer A Steen[†], Mareike Bongers, Lars K Nielsen and Claudia E Vickers^{*}

Abstract

Background: Metabolic engineering projects often require integration of multiple genes in order to control the desired phenotype. However, this often requires iterative rounds of engineering because many current insertion approaches are limited by the size of the DNA that can be transferred onto the chromosome. Consequently, construction of highly engineered strains is very time-consuming. A lack of well-characterised insertion loci is also problematic.

Results: A series of knock-in/knock-out (KIKO) vectors was constructed for integration of large DNA sequences onto the *E. coli* chromosome at well-defined loci. The KIKO plasmids target three nonessential genes/operons as insertion sites: *arsB* (an arsenite transporter); *lacZ* (β -galactosidase); and *rbsA-rbsR* (a ribose metabolism operon). Two homologous 'arms' target each insertion locus; insertion is mediated by λ Red recombinase through these arms. Between the arms is a multiple cloning site for the introduction of exogenous sequences and an antibiotic resistance marker (either chloramphenicol or kanamycin) for selection of positive recombinants. The resistance marker can subsequently be removed by flippase-mediated recombination. The insertion cassette is flanked by hairpin loops to isolate it from the effects of external transcription at the integration locus. To characterize each target locus, a xylanase reporter gene (*xynA*) was integrated onto the chromosomes of *E. coli* strains W and K-12 using the KIKO vectors. Expression levels varied between loci, with the *arsB* locus consistently showing the highest level of expression. To demonstrate the simultaneous use of all three loci in one strain, *xynA*, green fluorescent protein (*gfp*) and a sucrose catabolic operon (*cscAKB*) were introduced into *lacZ*, *arsB* and *rbsAR* respectively, and shown to be functional.

Conclusions: The KIKO plasmids are a useful tool for efficient integration of large DNA fragments (including multiple genes and pathways) into *E. coli*. Chromosomal insertion provides stable expression without the need for continuous antibiotic selection. Three non-essential loci have been characterised as insertion loci; combinatorial insertion at all three loci can be performed in one strain. The largest insertion at a single site described here was 5.4 kb; we have used this method in other studies to insert a total of 7.3 kb at one locus and 11.3 kb across two loci. These vectors are particularly useful for integration of multigene cassettes for metabolic engineering applications.

Keywords: Chromosomal integration, Homologous recombination, Plasmid, Recombineering, *E. coli*, Xylanase, GFP, *csc* genes

* Correspondence: c.vickers@uq.edu.au

[†]Equal contributors

Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, Brisbane QLD 4072, Australia

Background

Integration of large constructs into the *Escherichia coli* chromosome is desirable both for applied engineering projects and for examining metabolic control and other aspects of biological pathways. In particular, metabolic engineering often requires introduction of many genes to achieve production of industrially relevant biochemicals. Whole new pathways may be introduced, as well as genes to increase flux to metabolic precursors, decrease competition from competing pathways, and balance cofactor and redox requirements [1]. The resulting bioprocesses use renewable, sustainable feedstocks such as sucrose and glucose [2,3]; ultimately, lignocellulosic feedstocks will also become available [4]. Recent advances in engineering approaches are allowing for delivery of these 'green' replacement products at commercially-viable levels [5].

In *E. coli*, foreign genes may be introduced and expressed from extra-chromosomal elements or integrated onto the host chromosome. While relatively large fragments can be introduced on plasmids, plasmid-based expression has a number of disadvantages. These include the metabolic burden placed on the cell for plasmid maintenance and expression of gene products from multiple copies, the potential for internal rearrangements, segregational instability, fluctuations in copy number and resulting effects on interpretation of data, and the burden placed on cell in presence of antibiotic or other selective agents required for plasmid maintenance [6-11]. Furthermore, unpredictable pleiotropic effects can result from use of plasmids [12], and inclusion of antibiotics for plasmid maintenance increases the overall bioprocess cost. In particular, high copy-number plasmids are known to exert a much stronger metabolic drain than low copy-number vectors [13], and for this and associated reasons, low copy-number vectors are often found to be superior for engineering and other applications [11,14,15].

Integration onto the chromosome solves many of the inherent problems associated with plasmid use, and is therefore a preferable approach in many cases. Constructs can be stably maintained in the absence of selective pressure, the metabolic burden from plasmid maintenance is removed, copy number and segregation are stable and rearrangements are much less likely to occur. Despite the decrease in copy number, productivity can in fact be improved by engineering of genes on the chromosome [16,17].

Chromosomal integration in *E. coli* may be achieved through use of transposons (either random or site-specific; [18,19]) or recombination mediated by phages/phage-derived elements [20-25]. Phages can be used to integrate large DNA fragments. However, the integration sites are fixed, large regions of non-target DNA may be integrated in parallel, and dedicated labs are required for

phage manipulation due to the risk of contaminating phage-free strains. Use of phage-derived elements for homologous recombination circumvents several of these problems. For example, specifically-designed integrative plasmids (CRIM plasmids) carrying phage attachment (*attP*) site can be used to insert large DNA fragments at bacterial phage-attachment (*attB*) sites via *in trans* expression of phage-derived integration (*int*) and excision (*xis*) genes [21]. An alternative approach that allows greater flexibility in the insertion locus is the λ Red recombinase system [25-28]. A PCR-based version of this method is widely used for the targeted deletion of chromosomal genes in *E. coli* [23]. This method relies on the replacement of a chromosomal sequence with an antibiotic marker that is amplified by PCR using primers with ~50 bp homology extensions to the flanking regions of the target sequence. The antibiotic resistance gene template is a plasmid; flippase recognition target (FRT) sites are included in order to remove the antibiotic resistance gene after chromosomal integration, allowing recycling of the resistance gene for later engineering steps. Recombination onto the chromosome is mediated by the Red recombinase derived from the λ phage. This recombination system consists of three genes (γ , β , *exo*), which encode the phage recombinase functions as well as an inhibitor of the host RecBCD exonuclease V (which normally mediates degradation of linear DNA in the cell) [29].

The PCR/ λ Red method can also be used for co-integration of relatively short sequences that can be generated through PCR, e.g. introduction of point mutations, promoter replacement, and promoter-less reporter genes for promoter tagging experiments [16,30-34]. However, the efficiency of integration drops sharply for DNA fragments above about ~1500 bp and integration of fragments larger than 2500 bp using 50 bp homologous arms is very difficult [23,35], though integration of fragments up to 3,500 bp has been successful [36-38]. Since the resistance gene cassette is 1000-1500 bp (depending on which resistance gene is used), integration of genes of interest is functionally limited to a few genes at a time. This necessitates many iterative rounds of engineering for introduction of large pathways or multiple genes, even when synthetic operons are constructed to minimise sequence length. It has been demonstrated that increasing the region of homology serves to increase the efficiency of integration for large fragments by an order of magnitude [24]. However, this quickly becomes impractical using PCR approaches as the homology region is conferred by the primer sequences, and synthesis of very long primers is both expensive and more error-prone. Furthermore, the frequency of errors in amplified constructs increases with increasing amplicon length, and the yield of amplicons decreases with increasing length.

More recently, two-step techniques that combine the use of the λ Red recombinase with the yeast mitochondrial homing endonuclease I-SceI have been developed for introduction of large DNA fragments onto the *E. coli* chromosome [35,39]. I-SceI has an 18-bp recognition site which is not found in the *E. coli* genome; since double-stranded DNA breaks stimulate *in vivo* recombination, the efficiency of recombination can be increased by several orders of magnitude by using I-SceI sites at the target locus [40]. Two approaches have been described. In the first method, $\phi 80$ -*attB* sites are introduced onto the chromosome, allowing subsequent integration of a CRIM plasmid bearing *attP* sites and the target DNA fragments (Dual In/Out strategy; [41]). I-SceI sites are used to improve cloning efficiency of large DNA fragments into the CRIM plasmids prior to integration [39]. This approach avoids use of PCR for cloning, and allowed cloning and integration of an 8 kb fragment. The second method involves first using λ Red recombinase to introduce a 'landing pad' onto the genome that includes a tetracycline resistance gene flanked by I-SceI sites and short sequences homologous to the ends of the desired integration cassette [35]. Once this landing pad is integrated, the cassette containing the sequences of interest is introduced by digestion with plasmid-expressed I-SceI in the presence of λ Red recombinase. The utility of this method was demonstrated by introduction of a 7 kb fragment at six different loci, and by combinatorial integration at two different loci in one strain. Using these two-step techniques, it is possible to insert very large DNA fragments onto the chromosome in any desired location.

Where flexibility of location is less important than the ability to quickly and efficiently integrate large DNA fragments at well-characterised loci (for example, many metabolic engineering projects), a one-step method of integration is preferable. A second consideration is the lack of availability of well-characterised integration sites for insertion of multiple genes. Insertion at different sites on the chromosome has a significant effect on expression levels; the effect is mediated through interference from local sequences, gene orientation, and insertion position relative to the chromosomal origin of replication [42]. It is therefore of value to know the relative expression levels when genes are inserted at different loci. In addition, it is necessary to avoid disruption of normal growth.

Here we present an alternative, one-step method for integration of large DNA fragments onto the *E. coli* chromosome. The method relies on a series of knock-in/knock-out (KIKO) vectors which facilitate the integration (knock-in) of large DNA sequences onto the *E. coli* chromosome and the parallel interruption (knock-out) of native genes at defined, well-characterised loci using the λ Red system. To increase efficiency of integration, we used large homologous regions (~ 500 bp). Three

non-essential loci were characterised as integration sites; combinatorial integration at all three sites in one strain is successfully demonstrated.

Results and discussion

Selection of chromosomal sites for recombination

The target loci used for development of the KIKO vectors were selected with laboratory/industrial growth in mind. Two of the loci are involved in PTS uptake of compounds typically not used in laboratory or industrial growth conditions: the arsenite transporter *arsB* [43], and the ribose transporter operon *rbsA-rbsR* [44]. The third locus was the well-characterised lactose catabolic enzyme gene β -galactosidase, *lacZ*. It has been previously demonstrated that interruption of these ORFs is not deleterious for *E. coli* growth or cell division under standard cultivation conditions [43,45]. To ensure broader application to different *E. coli* strains, these loci were also screened and found to be conserved across a large cross-section of *E. coli* strains, including B, C, W, and K-12 as well as many pathogenic strains (data not shown). The *rbs* operon (*rbsABCKR*) is incomplete and non-functional in *E. coli* W, an industrially relevant sucrose utilizing strain [2], however the target sequences for homologous recombination are conserved. The *lacZ* locus is often subjected to modification in laboratory strains, and has been successfully used as a site for integration of sucrose utilization genes in *E. coli* K-12, C and B [46]. Furthermore, when using *lacZ* as a target locus, blue/white screening can be used to facilitate identification of insertion mutants. The positions of all three loci on the *E. coli* MG1655 genome are shown in Figure 1.

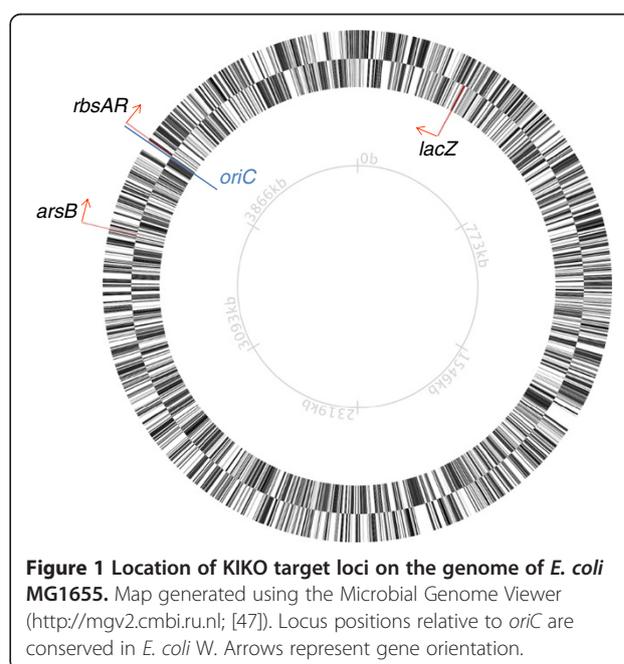


Figure 1 Location of KIKO target loci on the genome of *E. coli* MG1655. Map generated using the Microbial Genome Viewer (<http://mgv2.cmbi.ru.nl>; [47]). Locus positions relative to *oriC* are conserved in *E. coli* W. Arrows represent gene orientation.

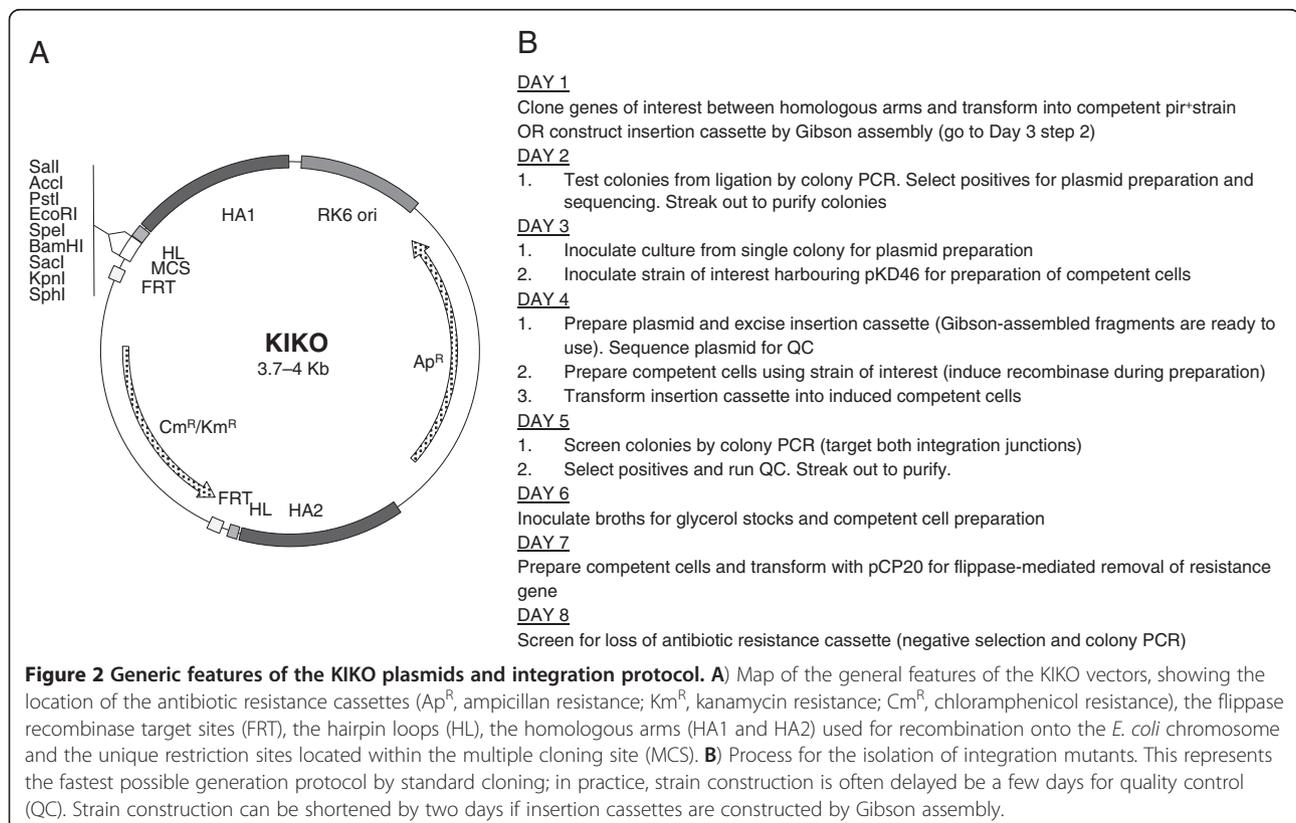
Design and construction of the KIKO vectors

The generic features of the KIKO vectors are shown in Figure 2A. They share a common backbone encoding the R6K *pir*⁺-dependent origin of replication (*ori*) and an ampicillin antibiotic resistance cassette. The R6K *ori* requires the π protein, which is supplied in *trans* by the λ *pir* gene [48,49]. Use of the R6K *ori* ensures that the plasmid cannot be maintained in target strains lacking λ *pir*; therefore, antibiotic resistance should not occur in the absence of integration. Homologous arms of ~500 bp each for each target locus are included in the KIKO vectors; this large amount of homology serves to increase recombination efficiency [24]. Embedded between the homologous arms is a multiple cloning site (to facilitate the insertion of genes of interest) as well as an antibiotic resistance gene cassette (chloramphenicol or kanamycin). The antibiotic resistance cassette is flanked by flippase recombination target (FRT) sites to allow for the removal of the antibiotic resistance cassette after insertion of the construct onto the chromosome so that these markers maybe reused as necessary [50,51]. There are three main influencers on gene expression for inserted DNA: local sequence context, gene location, and orientation [42]. Of these, the effect of local sequence context is the most pronounced when expression is controlled by constitutive promoters and the use of transcriptional terminators is absent [42]. Therefore,

the entire insertion construct was flanked with sequences encoding hairpin loops [52] to act as transcriptional terminators. These repeats effectively isolate the integrated genes from the influence of transcriptional events at adjacent loci [42]. The process for one round of integration using these vectors is shown in Figure 2B.

Integration at the target loci does not affect growth rate

We tested the KIKO vectors using two different *E. coli* strains: K-12 (MG1655) and *W. E. coli* K-12 (MG1655) and its derivatives are widely used in both laboratory and industry [53-55] and chromosomal integration using the λ Red functions is routine in this strain. *E. coli* W is of interest as it is particularly fast-growing on sucrose (1.2 h⁻¹; [56]), an especially attractive renewable carbon source for industrial applications [2]. *E. coli* W also demonstrates low acetate production enabling cultivation to high cell densities and high tolerance to environmental stresses (reviewed in [2]). To examine the effect of integration at the three selected loci, a FRT:xylanase:FRT cassette was inserted at each site in both W and K12 using the λ Red recombination system [23]. Integration did not affect growth rate in either strain for any of the loci (Figure 3), indicating that the integration of exogenous DNA at these loci is not deleterious for growth rate or viability under standard laboratory conditions. As we have previously observed [46], strain W has a relatively



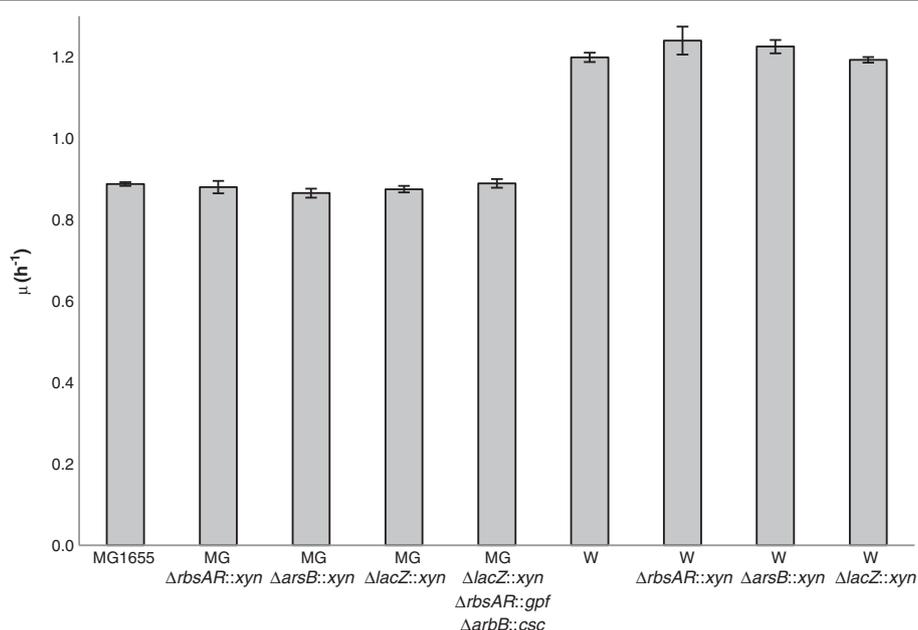


Figure 3 Specific growth rate (μ ; hr^{-1}) of engineered strains. The growth rate of *E. coli* K-12 (MG1655) and W strains harbouring exogenous DNA integrated into various different loci was measured on LB medium in the absence of selection to allow comparison with wild type strains. Bars are means; errors are standard deviations ($n = 3$). No significant differences were identified for either strain (1-way ANOVA; $p > 0.05$).

high growth rate; in this case, on LB medium, W grows at 1.2 hr^{-1} compared to 0.9 hr^{-1} for K-12(MG1655). The average growth rates were slightly lower than values reported in the literature for shake flask and bioreactor experiments. For example, reported growth rates for MG1655 range from 1.2 hr^{-1} to 2.0 hr^{-1} [57-61]. These variations are due to a number of factors, including variations between physiological behaviour of different strain stocks [54], the exact sampling time during the growth curve [58,60], and particular growth conditions (e.g. pH-controlled bioreactor, shake flask, relative aeration, etc.). The lower growth rates observed in the microtitre plates are consistent with our previous observations [46] and with other literature (e.g. [62]), showing a slight growth rate retardation in microtitre plate assay compared with shake-flask analysis; this is due to mild oxygen limitation [62].

Expression strength is variable at different loci

The xylanase insertion cassette was used to assess the difference in expression strength at each locus. Xylanase is used industrially to assist in degradation of plant fibre in the feed, food processing, pulp and paper industries [63]. In addition, xylanase has previously been developed as a reporter gene and its reaction properties have been well characterised [64]. Xylanase activity varied significantly between different integration sites (Figure 4). Xylanase activity also varied between the two strains; however, the relative expression strength between the sites was consistent between strains: xylanase activity was the

highest when *xynA* was integrated at the *arsB* locus, followed by *rbsAR* and then *lacZ* ($arsB > rbsAR > lacZ$). Variation was highest in strain K-12(MG1655), where a 2.3-fold difference was seen between the *arsB* locus and the *lacZ* locus. Expression levels were also higher in K-12(MG1655).

Expression of the xylanase gene should be isolated from the influence of local transcription units due to the presence of transcription terminators [42] in our constructs. However, the proximity of the insertion to the bi-directional origin of replication (*oriC*) also influences gene expression: due to constant genome replication, genes closer to the *oriC* exist at functionally higher copy number [42,65]. Given the locations of our target loci relative to *oriC* (Figure 1), one might therefore expect expression strength to fall in the order $rbsAR > arsB > lacZ$. However, gene orientation also influences expression levels, with genes on the leading strand being more highly expressed [42]. In *E. coli*, *arsB* and *rbsAR* are encoded on the leading strand, whereas *lacZ* is encoded on the lagging strand. However, in the case of the *xynA* insertions in each locus, the *xynA* gene cassette was always cloned in the opposite orientation relative to the natural expression of the target gene encoding the homologous arms. Thus, *arsB* and *rbsAR* loci *xynA* knock-ins are encoded on the lagging strand, whereas the *lacZ* locus *xynA* knock-in is encoded on the leading strand. Clearly, the expression level advantage enjoyed by genes encoded on the leading strand is insufficient to compensate for relatively greater distance from the *oriC* that *lacZ* is located (Figure 1).

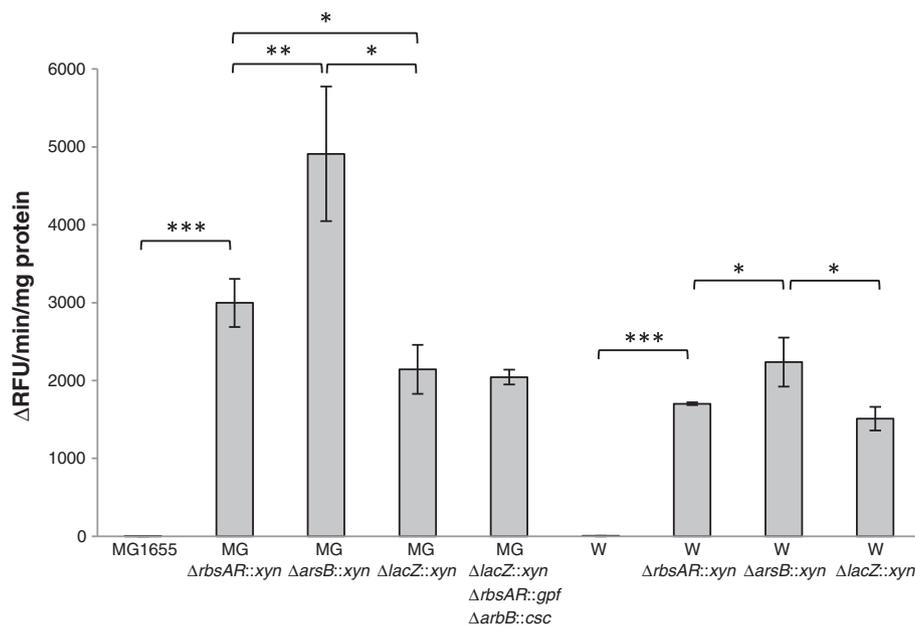


Figure 4 Xylanase assay. Xylanase activity was measured in extracts from *E. coli* K-12 (MG1655) and W strains harbouring *xynA* integrated into various different loci. Xylanase is expressed in relative fluorescence units (RFU) per minute per mg protein. Bars represent means; errors are standard deviations ($n = 3$). One-way ANOVA followed by Tukey's HSD post-hoc analysis was used to resolve differences in means; *, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$.

However, since the *rbsAR* locus is closer to *oriC* than *arsB*, but *arsB* consistently showed higher expression levels for both strains, other unknown factors must also influence gene expression levels from these loci. In such a small sample size it is not possible to determine what these factors are, and investigation of these phenomena requires further research. It is worth noting that expression of inserted transgenes at loci is relatively predictable [42] as are relative gene expression levels from native genes under a wide variety of conditions [66]. This former observation is supported by the conservation of rank order and relative expression levels between the two strains (Figure 4) and suggests that transgenes integrated at these loci will behave in a predictable fashion.

It should be noted that the actual orientation of the gene being expressed for any given construct made from the KIKO vectors depends on the direction of insertion of the gene cassette between the homologous arms, and this may affect relative expression strength between the loci. It should also be noted that leading strand fidelity is higher than lagging strand fidelity during DNA replication [67]; this, in addition to the generalised expectation of higher expression levels from leading strand-encoded genes [42], suggests that integration in leading strand orientation would be preferable when designing an experiment.

Finally, the higher xylanase activity measured in K-12 relative to W may be due to the relatively higher growth rate of W, which can result in dilution of cellular protein concentrations [68].

Integration of different exogenous sequences at all three loci in a single strain

To demonstrate the functional use of all three integration sites simultaneously, a sucrose metabolism operon encoded by the chromosomally-encoded sucrose catabolism genes (*cscAKB*, 3.9 kb) and green fluorescent protein gene (*gfp*, 0.8 kb) were introduced into the MG1655 strain harbouring the xylanase gene at the *lacZ* locus to produce a triple knock-in strain. Strain K-12(MG1655) cannot normally utilize sucrose [2]. Introduction of the *cscAKB* genes on a high copy number plasmid causes severe phenotypic instability, and we have previously demonstrated that successful sucrose utilization occurs after spontaneous integration of the *csc* genes the *lacZ* locus (this occurs at low frequency in the absence of the λ Red recombinase functions) [46]. While *cscA* alone is sufficient to confer a sucrose-positive phenotype [69-71], *cscAKB* together are required for maximal growth rates on sucrose [72]. Following each integration event the chloramphenicol antibiotic resistance gene was removed via FLP-mediated recombination so that it could be recycled in subsequent rounds of integration. The successful integration of *cscAKB* and *gfp* in strain *MGΔlacZ::xyn ΔrbsAR::gfp ΔarsB::csc* was demonstrated by a positive xylanase assay (Figure 4), the ability to ferment sucrose (Figure 5A) and expression of GFP (Figure 5B). The growth rate and xylanase activity in this triple knock-in strain were indistinguishable from the single knock-in strain carrying only the xylanase (Figure 3, Figure 4),

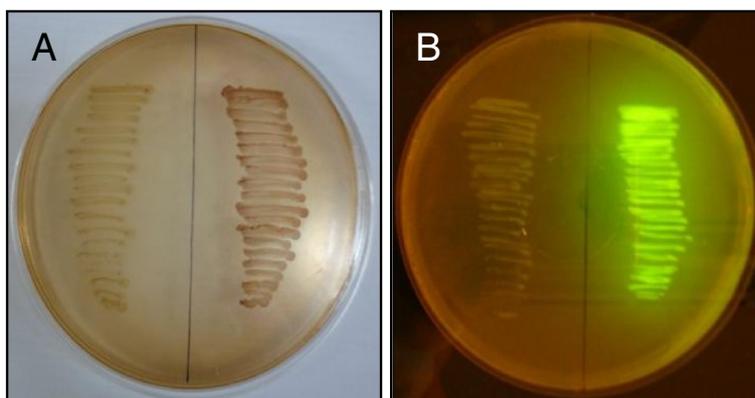


Figure 5 Sucrose utilisation and GFP activity in strain *MGΔlacZ::xyn ΔrbsAR::gfp ΔarsB::csc*. **A**) MacConkey agar supplemented with 1% sucrose was used to examine sucrose fermentation in wild type MG1655 (left) and engineered strain *MGΔlacZ::xyn ΔrbsAR::gfp ΔarsB::csc*. The ability to utilise sucrose is indicated by a pH change resulting from fermentation products; this is evidenced by red colonies on the MacConkey sucrose plate. **B**) Wild type MG1655 exhibits no fluorescence in the presence of UV light (left), whereas the engineered strain *MGΔlacZ::xyn ΔrbsAR::gfp ΔarsB::csc* fluoresces bright green (right).

indicating that knock-in at all three loci has no negative effect on normal cell growth under these conditions.

It should be noted that iterative rounds of knock-in can result in unintended effects (including deletions and rearrangements) resulting from FRT scars remaining in the genome from previous integration events. In other experiments, we have observed unintended deletions when subsequent secondary DNA constructs are introduced at the same locus [72]. In the case of the KIKO vectors described here, the target loci are well separated in the genome (see Figure 1) and recombination between them would result in large deletions which would presumably be fatal. Strains containing rearrangements and deletions can be excluded from analysis by screening both integration junctions (as described in the Materials and Methods). Use of asymmetric sites for removal of flanking markers, such as λ attL/attR [41,73] or the mutant FRT [74] sites available in *E. coli* or mutant loxP sites [75,76] currently used in eukaryotic systems, would preclude such unintended effects when integrating further constructs at the same locus.

Comparison of linearised plasmids and PCR products as integration templates

While performing the integration experiments described above using linearised plasmids, we observed a relatively high frequency of single cross-over events, whereby the entire plasmid (instead of just the target region between the homologous arms) was integrated at the target locus. This could be identified by screening at each insertion junction (only a single junction could be amplified by PCR; see Materials and Methods) and was further evidenced by persistent resistance to ampicillin (encoded on the plasmid backbone). The resistance gene cassette could not be cured from these strains because of an

inability to select for the pCP20 plasmid [51], which expresses the flippase required to remove sequences between the FRT sites. Moreover, we also observed a high rate of false positives which were resistant to ampicillin but had a wild type locus. This represented substantial background. In order to quantify this background and examine if an alternative approach could be used to decrease background, we performed a direct comparison to investigate the differing efficiencies of integration between linearised plasmids and PCR-amplified fragments (encompassing the two homologous arms and the regions between the arms; see Figure 2A).

The number of colonies obtained for each transformation event was often higher for MG1655 than for W (Table 1), an observation at odds with the competency of the cells: W was an order of magnitude higher in competency than MG1655 (10^9 CFU/ μ g plasmid DNA vs. 10^8 CFU/ μ g, respectively). We observed this consistently through a variety of other experiments (data not shown), and it means that experiments are generally more facile in MG1655. A generalised dilution of cellular protein (including the recombinase functions) and of transformed DNA fragments as a result of the faster growth rate of W (as described above) may also provide an explanation for the lower transformation efficiency observed. Differences may also be due to variations in the activity of the recombinase functions between the two strains. Regardless, we anticipate that similar variations in transformation and integration efficiency will be seen in other strains.

Transformations using linearised plasmid typically produced more colonies than those using PCR product (Table 1). The percentage of false positives and single crossover events was highly variable in the samples examined; however in most cases it was higher in plasmid

Table 1 Direct comparison of integration using linearised plasmid and pcr products

Locus	Strain	Template	Colonies	WT locus	Single X	Double X	Background
<i>arsB</i>	W	Plasmid	38	33	2	3	92%
		PCR	2	0	2	0	100%
	MG1655	Plasmid	38	4	15	19	50%
		PCR	14	0	2	12	14%
<i>lacZ</i>	W	Plasmid	8	5	1	2	75%
		PCR	6	0	1	5	17%
	MG1655	Plasmid	32	25	6	1	97%
		PCR	39	5	12	22	44%
<i>rbsAR</i>	W	Plasmid	22	0	16	6	73%
		PCR	7	0	5	2	71%
	MG1655	Plasmid	8	0	8	0	100%
		PCR	15	0	14	1	93%

Templates for each transformation (linearised plasmid or PCR) are noted, as well as the number of colonies obtained from each transformation, the number of colonies having a wild type locus (no integration), the number of colonies having a single cross-over (Single X; integration of entire plasmid) and the number of colonies having a double cross-over (Double X; integration of only the region contained between the homologous arms). Background % is calculated based on the number of WT locus + Single X colonies.

transformations than in PCR transformations. While the lower number of colonies meant that PCR transformation occasionally did not yield any double crossover events, in most cases the background of single cross-over events was lower in PCR transformations; furthermore, false positives with no insertion at the target locus were very rare. Single cross-overs occurred overall at a similar frequency to double cross-overs across all experiments. Where single cross-over events occurred, the single cross-over consistently occurred with high preference (80–100% across all three loci) on the side homologous with the 3' sequence of the gene in the insertion locus (data not shown). We have no good explanation for this observation.

The high incidence of false positives (ampicillin resistant colonies with a wild type insertion locus) is surprising. It is unlikely to occur from residual uncut plasmid, because the plasmid replicates via the R6K *ori* and so cannot be maintained in the absence of a λ *pir* gene [48,49]. Therefore, we presume that the plasmid must be integrating elsewhere on the chromosome. Spontaneous mutation is also a possibility, although rates are too high for it to be the only possibility.

There were no clear differences between the different insertion loci, with the exception that false positives were never found when integrations were attempted at the *rbsAR* locus. This may be due to the small sample size used, or may be due to some unknown feature of the *rbsAR* insertion vector/locus.

Conclusions

Incorporation of large DNA fragments onto the *E. coli* genome at precise, well-defined loci is desirable both for metabolic engineering applications and for examining whole biological pathways. We have constructed a series

of knock-in/knock-out (KIKO) plasmid vectors designed to successfully achieve this in one step. We also address the need for multiple reliable insertion sites in one genome in order to facilitate complex engineering in *E. coli* and show that all three loci characterised can be used for integration in a single strain. The variation in expression level at the different loci serves to highlight the need for well characterised integration sites. We have shown the utility of this system in two laboratory strains, and we anticipate that it will be generally applicable across a range of strains.

The KIKO vectors have been used successfully within our laboratory for the integration of a variety of genes for production of industrially relevant products, including the polyhydroxybutyrate operon in *E. coli* W (*phbCBA* + selection marker, 5.8 kb total) and the methylerythritol pyrophosphate isoprenoid pathway in both W and K-12 (8 genes across three loci, including insert sizes of 7.3 kb at the *lacZ* locus and 4.0 kb at the *arsB* locus). Synthetic constructs for expressing multiple genes are typically synthesized commercially unless they are already available. Although the method described here involves linearization of the integration vector, we subsequently found that integration of the entire plasmid can occur at relatively high frequency. This may result from poor integration at the second homologous region, or from integration of residual non-linearised plasmid via host-encoded recombination functions (more likely in wild type unmodified strains that retain their native recombinase functions). We have found that either complete excision of the insert cassette by digestion at both ends (data not shown) or PCR amplification of the insertion cassette (Table 1), followed by gel excision of the fragment (in the former case) and/or DpnI digestion

to remove plasmid DNA (in the latter case), significantly increases the frequency of the desired recombination event. We have also used Gibson assembly [77] using the KIKO vectors as PCR templates to successfully construct insertion cassettes for integration. This approach shortens the strain construction process by two days and removes the vagaries of standard cut-and-paste ligation-mediated cloning, and is currently our preferred approach for production of complex insertion cassettes. Although we have not tested them specifically, it is also theoretically possible to use the KIKO vectors as templates to produce knock-out fragments as described by Datsenko & Wanner [23]; the universal priming site P2 is intact, but the P1 site is truncated by two base pairs, necessitating design of a slightly different primer for this application. The KIKO plasmids will be available from Addgene (<http://www.addgene.org/>).

Methods

Bacterial strains, media and growth conditions

Bacterial strains are shown in Table 2 and plasmids are shown in Table 3. R6K-based vectors were maintained in *E. coli* BW23474 [78]. *E. coli* W (NCIMB 8666, National Collection of Industrial, Food and Marine Bacteria, Aberdeen, UK) and MG1655 (*E. coli* Genetic Stock Centre, CGSC 7740) were used as targets for genomic integration. *E. coli* strains were grown in LB medium [79] supplemented with ampicillin (100 µg/mL), kanamycin (50 µg/mL), and/or chloramphenicol (40 µg/mL) as appropriate, unless otherwise stated. Sucrose utilization experiments were performed using MacConkey agar

(Difco, BD, North Ryde, NSW, Australia) supplemented with 1% sucrose or M9 medium [79] supplemented with 1 mg/L thiamine and 2% sucrose (M9S).

Molecular techniques

Standard molecular cloning procedures [79] were employed. Oligonucleotide primers used in this study are described in Table 4. PCR products and restriction endonuclease treated-plasmid DNA were purified using a MinElute PCR purification kit and QIAquick gel extraction kit (Qiagen, Doncaster, VIC, Australia), respectively.

Construction of the KIKO vectors

The KIKO vectors (detailed in Table 5) were constructed in two stages. In the first stage, the R6K origin and ampicillin resistance (*bla*, Ap^R) gene was amplified from pKD4 [23] using primers JSP64 and JSP65 and ligated to ~ 1 kb fragments amplified from *E. coli* W of *arsB* (primers JSP10 and JSP11), *lacZ* (primers JSP55 and JSP56), and *prbsAR* (primers JSP57 and JSP58) to generate the base vectors *parsB*-R6K, *placZ*-R6K, *prbsAR*-R6K respectively. A synthetic cassette that contained a multiple cloning site (MCS) and a chloramphenicol resistance gene (*catP*; Cm^R) flanked by FRT recombination sites [23] flanked by transcriptional terminators (hairpin loops) from the omega cassette [52] was synthesized in a plasmid backbone (DNA 2.0; CA, USA). In the second round of cloning, the synthetic cassette was excised from the synthesized plasmid using PmeI and subcloned into unique blunt sites within the base vectors (*arsB*, EcoRV; *lacZ*, EcoRV; *rbsAR*, XmnI) to generate the Cm^R KIKO vectors

Table 2 Strains used in this study

Strain	Properties	Reference/source
BW23474	$\Delta(\argF-lac)169$, $\Delta uidA4::pir-116$, <i>recA1</i> , <i>rpoS396(Am)</i> , <i>endA9(del-ins)::FRT</i> , <i>rph-1</i> , <i>hsdR514</i> , <i>rob-1</i> , <i>creC510</i>	[78] ¹
MG1655	λ -, <i>rph-1</i>	CGSC ¹
MG $\Delta arsB::xyn$	MG1655 <i>arsB::xynAcatP</i> ; Cm ^R	This study
MG $\Delta rbsAR::xyn$	MG1655 <i>rbsAR::xynAcatP</i> ; Cm ^R	This study
MG $\Delta lacZ::xyn$	MG1655 <i>lacZ::xynAcatP</i> ; Cm ^R	This study
MG $\Delta lacZ::xynFRT$	MG1655 <i>lacZ::xynA-FRT</i> ; Cm ^S	This study
MG $\Delta lacZ::xyn \Delta rbsAR::gfp-Cm$	MG1655 <i>lacZ::xynA-FRT rbsAR::gfpcatP</i> ; Cm ^R	This study
MG $\Delta lacZ::xyn \Delta rbsAR::gfp$	MG1655 $\Delta lacZ::xynA-FRT \Delta rbsA-rbsR::gfp-FRT$; Cm ^S	This study
MG $\Delta lacZ::xyn \Delta rbsAR::gfp \Delta arsB::csc-Cm$	MG1655 $\Delta lacZ::xynA-FRT \Delta rbsA-rbsR::gfp-FRT \Delta arsB::cscAKBcatP$; Cm ^R	This study
MG $\Delta lacZ::xyn \Delta rbsAR::gfp \Delta arsB::csc$	MG1655 $\Delta lacZ::xynA-FRT \Delta rbsA-rbsR::gfp-FRT \Delta arsB::cscAKB-FRT$; Cm ^S	This study
W	Wild type strain	NCIMB 8666 ²
W $\Delta arsB::xyn$	W <i>arsB::xynAcatP</i> ; Cm ^R	This study
W $\Delta lacZ::xyn$	W <i>lacZ::xynAcatP</i> ; Cm ^R	This study
W $\Delta rbsAR::xyn$	W <i>rbsAR::xynAcatP</i> ; Cm ^R	This study

¹Purchased from the *Coli Genetic Stock Centre* (CGSC).

²National Collection of Industrial Bacteria, Aberdeen, Scotland. This strain is also archived as ATCC9637 in the American Type Culture Collection.

Table 3 Plasmids used in this study

Plasmid	Properties	Reference/ source
pKD3	oriR6Kgamma <i>bla catP</i>	[23]
pKD4	oriR6Kgamma <i>bla aphA</i>	[23]
pKD46	repA101ts & oriR101 <i>P_{araB} exo, bet, gam araC bla</i>	[23]
pCP20	Flp recombinase expression plasmid	[51]
pNPDX2	PBTac2 (Boehringer) carrying modified <i>xynA</i> cDNA from <i>Neocallimastix patriciarum</i>	[80]
pBAV1K-T5-gfp	pBAV1K-T5 plasmid that carries <i>gfp</i> gene	[81]
pCSCX	pCR2.1 plasmid carrying <i>cscA, cscK</i> and <i>cscB</i> gene from <i>E. coli</i> W	[46]
pJAS01	Synthetic construct. Contains MCS & FRT- <i>catP</i> -FRT cassette flanked by hairpin loops.	This study
<i>parsB</i> -R6K	JSP64/JSP65 PCR product from pKD4 (PvuII; 1,495; R6K <i>ori</i> and <i>bla</i>) ΩJSP10/JSP11 PCR product (PmeI; 1,109; <i>arsB</i>)	This study
<i>placZ</i> -R6K	JSP64/JSP65 PCR product from pKD4 (PvuII; 1,495; R6K <i>ori</i> and <i>bla</i>) ΩJSP57/JSP58 PCR product (PmeI; 1,318; <i>lacZ</i>)	This study
<i>prbsAR</i> -R6K	JSP64/JSP65 PCR product from pKD4 (PvuII; 1,495; R6K <i>ori</i> and <i>bla</i>) ΩJSP55/JSP56 PCR product (PmeI; 1,034; <i>rbsAR</i>)	This study
pKIKOarsBCm	<i>parsB</i> -R6K (PmeI; 2,604; linear) ΩpJAS01 (PmeI; 1,143; MCS/ <i>catP</i> cassette)	This study
pKIKO <i>lacZ</i> Cm	<i>placZ</i> -R6K (PmeI; 2,813; linear) ΩpJAS01 (PmeI; 1,143; MCS/ <i>catP</i> cassette)	This study
pKIKOrbsARcm	<i>prbsAR</i> -R6K (PmeI; 2,529; linear) ΩpJAS01 (PmeI; 1,143; MCS/ <i>catP</i> cassette)	This study
pKIKOarsBKm	pKIKOarsBCm (XbaI; 2,839; <i>catP</i> drop-out) ΩpKD3 (XbaI; 930; <i>aphA</i>)	This study
pKIKO <i>lacZ</i> Km	pKIKO <i>lacZ</i> Cm (XbaI; 2,764; <i>catP</i> drop out) ΩpKD3 (XbaI; 930; <i>aphA</i>)	This study
pKIKOrbsARKm	pKIKOrbsARcm (XbaI; 3,111; <i>catP</i> drop out) ΩpKD3 (XbaI; 930; <i>aphA</i>)	This study
pKIKOarsBΩ <i>xyn</i> Cm	pKIKOarsBCm (PstI/SalI; 3,763) ΩpNPDX2 (PstI/SalI; 1,108; <i>xynA</i>)	This study
pKIKO <i>lacZ</i> Ω <i>xyn</i> Cm	pKIKO <i>lacZ</i> Cm (PstI/SalI; 3,688) ΩpNPDX2 (PstI/SalI; 1,108; <i>xynA</i>)	This study
pKIKOrbsARΩ <i>xyn</i> Cm	pKIKOrbsARcm (PstI/SalI; 4,035) ΩpNPDX2 (PstI/SalI; 1,108; <i>xynA</i>)	This study
pKIKOrbsARΩ <i>gfp</i> Cm	pKIKOrbsARcm (SacI/SpeI; 4,035) ΩpBAV1K-T5-gfp (SacI/SpeI; 1,849; <i>aphA</i> & <i>gfp</i>)	This study
pKIKOarsBΩ <i>cscAKB</i> Cm	pKIKOarsBCm (EcoRI; 3,771) ΩpCSCX (EcoRI; 4,064; <i>cscAKB</i>)	This study

pKIKOarsBCm [GenBank KC503965] pKIKO*lacZ*Cm [GenBank KC503966] and pKIKOrbsARcm [GenBank KC503967]. To generate KIKO vectors conferring kanamycin resistance (*aphA*; Km^R), the *catP* gene was replaced with an *aphA*-containing fragment from pKD3 [23] using *XbaI* sites. This produced the Km^R KIKO vectors pKIKOarsBKm [GenBank KC503968], pKIKO*lacZ*Km [GenBank KC503969] and pKIKOKm [GenBank KC503970]. These plasmids will be available from Addgene (<http://www.addgene.org/>).

Construction of xylanase KIKO vectors

To generate xylanase knock-in constructs, PCR products carrying the xylanase gene (*xynA*) under the control of the strong inducible *tac* promoter were amplified from the plasmid pNPDX2 [63] using the primers JSP148 and JSP149, and cloned into each of the Cm^R KIKO vectors using PstI and SalI to generate pKIKOarsBΩ*xyn*Cm, pKIKO*lacZ*Ω*xyn*Cm and pKIKOrbsARΩ*xyn*Cm. In these constructs, the *xynA* gene cassette is inserted in the opposite orientation relative to the natural expression of

the gene encoding the homologous arms. To generate a *gfp* knock-in construct, the *gfp* gene was digested from pBAV1K-T5-gfp [81] with SacI and SpeI and ligated into pKIKOrbsARcm at the cognate sites to produce pKIKOrbsARΩ*gfp*Cm. To generate a sucrose utilization gene knock-in construct, the *cscAKB* operon was amplified from pCSCX [46] using primers JSP250 and JSP251. The PCR product was digested with SpeI, purified, and ligated into SpeI-digested pKIKOarsBCm, generating pKIKOarsBΩ*cscAKB*. All ligations were transformed into rubidium chloride treated *E. coli* BW23474 chemically competent cells (λ pir⁺; see Table 2). For pKIKOrbsARΩ*gfp*Cm, the transformants were plated on LB agar supplemented with chloramphenicol and kanamycin. For pKIKOarsBΩ*cscAKB*, the transformants were plated on the LB agar supplemented with chloramphenicol; colonies were inoculated in 1 ml of M9S broth with chloramphenicol and were grown overnight at 37°C to confirm the suc⁺ phenotype. All of the putative recombinants were further screened using colony PCR and confirmed using Sanger sequencing of plasmid DNA. Sequencing was performed by Australian Genome

Table 4 Primers used in this study

Primer	Application	Sequence (5'-3')
JSP10	Amplification of <i>arsB</i> fragment (PmeI)	GGCGCAGCT <u>GTTTAAAC</u> GTCTGACCATCGTATTGG
JSP11	Amplification of <i>arsB</i> fragment (PmeI)	GGGTCAGCT <u>GTTTAAAC</u> GTCCCAATCGCAGCCAAT
JSP22	Screening oligo for KIKO vector inserts	TTCTGCGAAGTGATCTTCCG
JSP55	Amplification of <i>lacZ</i> fragment	GCGAG <u>TTTAAAC</u> CATTTTCCGTGACGTCTCGT
JSP56	Amplification of <i>lacZ</i> fragment (PmeI)	GCGAG <u>TTTAAAC</u> AAGACTGTTACCCATCGCGT
JSP57	Amplification of <i>rbsAR</i> fragment (PmeI)	GCGAG <u>TTTAAAC</u> AAGCGGCATTGTGTATATC
JSP58	Amplification of <i>rbsAR</i> fragment (PmeI)	GCGAG <u>TTTAAAC</u> CCCAGTTCATCTTTCGGTTG
JSP64	Amplification of R6K ori and <i>bla</i> (PvuII)	TAGGCAGCTGGGAGGATATCAAATGGACC
JSP65	Amplification of R6K ori and <i>bla</i> (PvuII)	CCCCAGCTGGATGCAGGTGGCACTTTTCG
JSP123	Confirmation of <i>arsB</i> genomic insertion	CAACCTGGCTCGACAAAAC
JSP124	Confirmation of <i>arsB</i> genomic insertion	GTGTCACAAACAGCACAGGC
JSP125	Confirmation of <i>rbsAR</i> genomic insertion	CCGAAGTGTAAAGTGCTC
JSP126	Confirmation of <i>rbsAR</i> genomic insertion	GCGTAAATCTAAGCCGAACC
JSP129	Confirmation of <i>lacZ</i> genomic insertion	GTCTGAATTTGACCTGAGCG
JSP130	Confirmation of <i>lacZ</i> genomic insertion	TCATACAGAAGTGGCGATCG
JSP148	Amplification of <i>xynA</i> (PstI)	AGGACTGCAGCGGAGCTTATCGACTGCACG
JSP149	Amplification of <i>xynA</i> (Sall)	AAACGCTGACTTAAACGAGGAGCGGAGAGG
cscB_F2	Screening oligo for <i>cscAKB</i> insertion	ATGGCACTGAATATCCATTCAGA
KO Test	Confirmation of insertion junction at Cm end	GGAGTGAATACCACGACGAT

Restriction enzyme sites included in primers are underlined, and the enzyme name is listed in the Application column.

Research Facility (The University of Queensland, St. Lucia, QLD, Australia).

Strain construction

Engineered strains are shown in Table 2. Gene knock-in was performed by homologous recombination using the arabinose-inducible λ Red recombination system as described previously [23] with minor modifications also described previously [46,72]. KIKO vectors containing xylanase genes were linearized with NotI and 250 ng of purified linear plasmid was electroporated into arabinose-induced W(pKD46) and MG1655(pKD46) electrocompetent cells. In parallel, the integration cassette was also PCR-amplified from each KIKO vector, purified and again 250 ng of product was used to transform cells. The

Table 5 List of kiko vectors, their characteristics and genbank accession numbers

KIKO vector	Target locus	Resistance gene	GenBank accession
pKIKOarsBCm	<i>arsB</i>	<i>catP</i> (Cm ^R)	KC503965
pKIKOlacZCm	<i>lacZ</i>	<i>catP</i> (Cm ^R)	KC503966
pKIKOrbsARcm	<i>rbsAR</i>	<i>catP</i> (Cm ^R)	KC503967
pKIKOarsBKm	<i>arsB</i>	<i>aphA</i> (Km ^R)	KC503968
pKIKOlacZKm	<i>lacZ</i>	<i>aphA</i> (Km ^R)	KC503969
pKIKOrbsARKm	<i>rbsAR</i>	<i>aphA</i> (Km ^R)	KC503970

The resistance gene listed is the gene inserted onto the chromosome after homologous recombination. Km^R kanamycin resistance, Cm^R chloramphenicol resistance.

transformation efficiency (competency) of the cells was determined using 120 ng of pACYC184 (New England Biolabs, Ipswich, MA, USA). Integration of the xylanase gene was confirmed by colony PCR using three separate PCRs: 1) a *xynA*-specific primer (JSP148) coupled with a chromosome-specific primer (JSP123 for the *arsB* locus, JSP129 for the *lacZ* locus and JSP125 for the *rbsAR* locus); 2) a *catP*-specific primer (KO test) coupled with a chromosome-specific primer (JSP124 for the *arsB* locus, JSP130 for the *lacZ* locus and JSP126 for the *rbsAR* locus); and 3) a PCR to screen across the whole insertion locus (JSP123 and JSP124, JSP129 and JSP130 and JSP125 and JSP126, respectively). The resulting W-derived strains were designated W Δ *arsB*::*xyn*, W Δ *lacZ*::*xyn*, W Δ *rbsAR*::*xyn*, and the resulting K-12(MG166)-derived strains were designated MG Δ *arsB*::*xyn*, MG Δ *rbsAR*::*xyn*, and MG Δ *lacZ*::*xyn*.

To generate a double knock-in with xylanase and GFP, the Cm^R resistance gene was first removed from *E. coli* MG Δ *lacZ*::*xyn* by flippase-mediated recombination as described previously [23,51]. The resulting strain (MG Δ *lacZ*::*xyn*FRT) was transformed with pKD46 and NotI-linearized pKIKOrbsAR Ω *gfp*Cm plasmid DNA was integrated into the *rbsA-rbsR* locus using the method described above. Transformants were selected on LB agar supplemented with chloramphenicol (25 μ g/ml) and kanamycin (25 μ g/ml). Integration of the *gfp* gene in the *rbsA-rbsR* locus was confirmed by colony PCR using

JSP125 and JSP126 primers. Expression of the *gfp* gene in the resulting strain *MGΔlacZ::xyn ΔrbsAR::gfp-Cm* was confirmed by colony fluorescence under UV light. The Cm^R resistance gene was then eliminated as described above to produce strain *MGΔlacZ::xyn ΔrbsAR::gfp*. To generate a triple knock-in, the sucrose utilization operon *cscAKB* from NotI-linearized *pKIKOarsBΩcscAKB* was integrated at the *arsB* locus using the method described above. Transformants were selected on LB agar supplemented with 25 μg/ml chloramphenicol and the *suc*⁺ phenotype confirmed using M9S medium supplemented with 25 μg/ml chloramphenicol. Integration of the *cscAKB* genes was verified by colony PCR using the JSP123 and *cscB_F2* primers [72].

Growth rate analysis

The growth rate of the recombinant strains were determined from 96-well microtitre plate cultures monitored using a FLUOStar Omega multi-detection microplate reader (BMG Labtech, Offenburg, Germany) as described previously [46] with the exception that LB medium was used instead of defined medium.

Xylanase assay

Strains were cultured in LB to an OD_{600} of 0.6-0.8 then induced for 3 h with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were harvested by centrifugation, washed and resuspended in 1 volume of PBS buffer, and lysed with PopCulture™ reagent (Novagen, Billerica, USA). The lysed cell culture was clarified by centrifugation and the xylanase activity measured using an EnzChek® Ultra Xylanase assay kit (Invitrogen, Eugene, USA). Fluorescence was measured on FLUOStar Omega multi-detection microplate reader (BMG Labtech, Offenburg, Germany) every 30 sec min for 30 min at room temperature using excitation at 360 nm and emission detection at 460 nm. Commercial xylanase from *Thermomyces lanuginosus* (Sigma, St Louis, USA) was used to develop a standard curve. Protein concentration was determined by the method of Bradford [82] using a BioRad Protein Assay kit (BioRad, Gladesville, Australia) and bovine serum albumin as a standard. Xylanase activity was expressed as change in relative fluorescence over time, adjusted for total protein present ($\Delta RFU/min/mg$ protein).

Statistical analysis

The xylanase activity and the growth rate data were analyzed using analysis of variance (ANOVA). All statistical analyses were performed using R software (<http://www.r-project.org/>).

Abbreviations

csc: Chromosomally-encoded sucrose catabolism; Cm^R : Chloramphenicol resistance; FRT: Flippase recombinase target sites; GFP: Green fluorescent

protein; HA: Homologous arm; KIKO: Knock-in/knock-out; Km^R : Kanamycin resistance; MCS: Multiple cloning site; QC: Quality control; RFU: Relative fluorescence units; *xynA*: Xylanase gene.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

This study was designed by JAS, CEV, LKN and MB. Plasmid vector generation, strain generation and strain characterization was performed by JAS and SS. The manuscript was written by CEV, JAS, and SS and approved by all authors prior to submission.

Acknowledgements

We thank Jessica McClintock and Sarah Bydder for technical assistance. This research was supported by a Queensland State Government grant under National and International Research Alliances Program. SS was supported by Skim Latihan Akademik IPTA (SLAI) from the Ministry of Higher Education, (MOHE) Malaysia and Universiti Putra Malaysia. CEV was supported by a Queensland State Government Smart Futures Fellowship.

Received: 18 January 2013 Accepted: 23 May 2013

Published: 24 June 2013

References

- Stephanopoulos GN, Aristidou AA, Nielsen J: *Metabolic Engineering*. San Diego: Academic Press; 1998.
- Archer C, Kim J, Jeong H, Park J, Vickers C, Lee S, Nielsen L: **The genome sequence of *E. coli* W ATCC 9637: comparative genome analysis and an improved genome-scale model of *E. coli***. *BMC Genomics* 2011, **12**:9.
- Vickers CE, Klein-Marcuschamer D, Krömer JO: **Examining the feasibility of bulk commodity production in *Escherichia coli***. *Biotechnol Lett* 2012, **34**:585.
- Blanch HW, Simmons BA, Klein-Marcuschamer D: **Biomass deconstruction to sugars**. *Biotechnol J* 2011, **6**:1086–1102.
- Nielsen LK: **Metabolic engineering: From retrofitting to green field**. *Nat Chem Biol* 2011, **7**:408–409.
- Cunningham D, Koepsel R, Ataai M, Domach M: **Factors affecting plasmid production in *Escherichia coli* from a resource allocation standpoint**. *Microb Cell Fact* 2009, **8**:27.
- Paulsson J, Ehrenberg M: **Trade-off between segregational stability and metabolic burden: A mathematical model of plasmid ColE1 replication control**. *J Mol Biol* 1998, **279**:73–88.
- Diaz Ricci JC, Hernandez ME: **Plasmid effects on *Escherichia coli* metabolism**. *Crit Rev Biotechnol* 2000, **20**:79–108.
- Wang ZJ, Li XA, Shao JJ, Wegryzn A, Wegryzn G: **Effects of the presence of ColE1 plasmid DNA in *Escherichia coli* on the host cell metabolism**. *Microb Cell Fact* 2006, **5**:34. doi:10.1186/1475-2859-5-34.
- Pedraza JM, van Oudenaarden A: **Noise propagation in gene networks**. *Science* 2005, **307**:1965–1969.
- Jones KL, Kim SW, Keasling JD: **Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria**. *Metab Eng* 2000, **2**:328–338.
- Marcellin E, Chen WY, Nielsen LK: **Understanding plasmid effect on hyaluronic acid molecular weight produced by *Streptococcus equi* subsp. *zooepidemicus***. *Metab Eng* 2010, **12**:62–69.
- Axe DD, Bailey JE: **Application of 31 P nuclear magnetic resonance spectroscopy to investigate plasmid effects on *Escherichia coli* metabolism**. *Biotechnol Lett* 1987, **9**:83–88.
- Lerner CG, Inouye M: **Low copy number plasmids for regulated low-level expression of cloned genes in *Escherichia coli* with blue/white insert screening capability**. *Nucleic Acids Res* 1990, **18**:4631.
- Payne MS, Picataggio SK, Kuang-Hsu A, Nair RV, Valle F, Soucaille P, Trimbur DE: *Promoter and plasmid system for genetic engineering*. E. I. du Pont de Nemours and Company; 2006.
- Yuan LZ, Rouvière PE, LaRossa RA, Suh W: **Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in *E. coli***. *Metab Eng* 2006, **8**:79–90.
- Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G: **Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli***. *Science* 2010, **330**:70–74.

18. Hayes F: Transposon-based strategies for microbial functional genomics and proteomics. *Annu Rev Genet* 2003, **37**:3–29.
19. Akhverdyan VZ, Gak ER, Tokmakova IL, Stoyanova NV, Yomantas YA, Mashko SV: Application of the bacteriophage Mu-driven system for the integration/amplification of target genes in the chromosomes of engineered Gram-negative bacteria—mini review. *Appl Microbiol Biotechnol* 2011, **91**:857–871.
20. Fineran PC, Petty NK, Salmond GPC: Transduction: Host DNA transfer by bacteriophages. In *Encyclopedia of microbiology*. Edited by Schaechter M. Oxford: Elsevier; 2009:666–679.
21. Haldimann A, Wanner BL: Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol* 2001, **183**:6384–6393.
22. Murray NE: The impact of phage lambda: from restriction to recombineering. *Biochem Soc Trans* 2006, **34**:203–207.
23. Datsenko KA, Wanner BL: One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 2000, **97**:6640–6645.
24. Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court D: An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci USA* 2000, **97**:5978–5983.
25. Sawitzke JA, Thomason LC, Costantino N, Bubunenko M, Datta S, Court D: Recombineering: In vivo genetic engineering in *E. coli*, *S. enterica*, and beyond. In *Methods in Enzymology Advanced Bacterial Genetics: Use of Transposons and Phage for Genomic Engineering*, Volume 421. Edited by Kelly TH. Stanley: Academic Press; 2007:171–199.
26. Murphy KC: Use of bacteriophage recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol* 2003, **198**:180.
27. Zhang Y, Buchholz F, Muylers JPP, Stewart AF: A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet* 1998, **20**:123–128.
28. Sharan SK, Thomason LC, Kuznetsov SG, Court DL: Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* 2009, **4**:206–223.
29. Benzinger R, Enquist LW, Skalka A: Transfection of *Escherichia coli* spheroplasts. V. Activity of *recBC* nuclease in *rec+* and *rec minus* spheroplasts measured with different forms of bacteriophage DNA. *J Virol* 1975, **15**:861–871.
30. Murphy KC, Campellone KG, Poteete AR: PCR-mediated gene replacement in *Escherichia coli*. *Gene* 2000, **246**:321–330.
31. Muylers JPP, Zhang YM, Benes V, Testa G, Ansoorge W, Stewart AF: Point mutation of bacterial artificial chromosomes by ET recombination. *EMBO Rep* 2000, **1**:239–243.
32. Thomason L, Court DL, Bubunenko M, Costantino N, Wilson H, Datta S, Oppenheim A: Recombineering: genetic engineering in bacteria using homologous recombination. *Current Protocols in Molecular Biology* 2007, **78**:1.16.11–1.16.24.
33. 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 I Red recombination. *Mol Biotechnol* 2006, **32**:43–53.
34. Meynial-Salles I, Cervin MA, Soucaille P: New tool for metabolic pathway engineering in *Escherichia coli*: One-step method to modulate expression of chromosomal genes. *Appl Environ Microbiol* 2005, **71**:2140–2144.
35. Kuhlman TE, Cox EC: Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res* 2010, **38**:e92.
36. Orford M, Nefedov M, Vadolas J, Zaibak F, Williamson R, Ioannou PA: Engineering EGFP reporter constructs into a 200 kb human β -globin BAC clone using GET Recombination. *Nucleic Acids Res* 2000, **28**:e84.
37. Pohl T, Uhlmann M, Kaufenstein M, Friedrich T: Lambda Red-mediated mutagenesis and efficient large scale affinity purification of the *Escherichia coli* NADH:ubiquinone oxidoreductase (complex I). *Biochemistry* 2007, **46**:10694–10702.
38. Yu BJ, Kang KH, Lee JH, Sung BH, Kim MS, Kim SC: Rapid and efficient construction of markerless deletions in the *Escherichia coli* genome. *Nucleic Acids Res* 2008, **36**:e84.
39. Ublinskaya AA, Samsonov VV, Mashko SV, Stoyanova NV: A PCR-free cloning method for the targeted ϕ 80 Int-mediated integration of any long DNA fragment, bracketed with meganuclease recognition sites, into the *Escherichia coli* chromosome. *J Microbiol Methods* 2012, **89**:167–173.
40. Li MZ, Eledge SJ: MAGIC, an in vivo genetic method for the rapid construction of recombinant DNA molecules. *Nat Genet* 2005, **37**:311–319.
41. Minaeva NI, Gak ER, Zimenkov DV, Skorokhodova AY, Biryukova IV, Mashko SV: Dual-In/Out strategy for genes integration into bacterial chromosome: a novel approach to step-by-step construction of plasmid-less marker-less recombinant *E. coli* strains with predesigned genome structure. *BMC Biotechnol* 2008, **8**:63. doi:10.1186/1472-6750-8-63.
42. Block DHS, Hussein R, Liang LW, Lim HN: Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res* 2012, **40**:8979–8992.
43. Carlin A, Shi W, Dey S, Rosen B: The *ars* operon of *Escherichia coli* confers arsenical and antimicrobial resistance. *J Bacteriol* 1995, **177**:981–986.
44. Lopilato JE, Garwin JL, Emr SD, Silhavy TJ, Beckwith JR: D-ribose metabolism in *Escherichia coli* K-12: genetics, regulation, and transport. *J Bacteriol* 1984, **158**:665–673.
45. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2006, **2**:2006–0008. doi:10.1038/msb4100050.
46. Bruschi M, Boyes S, Sugiarto H, Nielsen LK, Vickers CE: A transferrable sucrose utilization approach for non-sucrose-utilizing *Escherichia coli* strains. *Biotech Adv* 2011, **30**:1001–1010.
47. Kerkhoven R, van Enckevort FHJ, Boekhorst J, Molenaar D, Siezen RJ: Visualization for genomics: the Microbial Genome Viewer. *Bioinformatics* 2004, **20**:1812–1814.
48. Donnenberg MS, Kaper JB: Construction of an *eae* deletion mutant of enteropathogenic *Escherichia coli* by using a positive-selection suicide vector. *Infect Immun* 1991, **59**:4310–4317.
49. Philippe N, Alcaraz JP, Coursange E, Geiselmann J, Schneider D: Improvement of pCVD442, a suicide plasmid for gene allele exchange in bacteria. *Plasmid* 2004, **51**:246–255.
50. Martinez-Morales F, Borges AC, Martinez A, Shanmugam KT, Ingram LO: Chromosomal integration of heterologous DNA in *Escherichia coli* with precise removal of markers and replicons used during construction. *J Bacteriol* 1999, **181**:7143–7148.
51. Cherepanov PP, Wackernagel W: Gene disruption in *Escherichia coli*: Tc^R and Km^R cassettes with the option of FLP-catalyzed excision of the antibiotic-resistance determinant. *Gene* 1995, **158**:9–14.
52. Prentki P, Krisch HM: In vitro insertional mutagenesis with a selectable DNA fragment. *Gene* 1984, **29**:303–313.
53. Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, Ohtsubo E, Baba T, Wanner BL, Mori H, Horiuchi T: Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* 2006, **2**:2006–0007. doi:10.1038/msb4100049.
54. Soupene E, van Heeswijk WC, Plumbridge J, Stewart V, Bertenthal D, Lee H, Prasad G, Pally O, Charennoppakul P, Kustu S: Physiological studies of *Escherichia coli* strain MG1655: Growth defects and apparent cross-regulation of gene expression. *J Bacteriol* 2003, **185**:5611–5626.
55. Conrad T, Joyce A, Applebee MK, Barrett C, Xie B, Gao Y, Palsson B: Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biol* 2009, **10**:R118. doi:10.1186/gb-2009-10-10-r118.
56. Lee SY, Chang HN: High cell density cultivation of *Escherichia coli* W using sucrose as a carbon source. *Biotechnol Lett* 1993, **15**:971–974.
57. Camara JE, Skarstad K, Crooke E: Controlled initiation of chromosomal replication in *Escherichia coli* requires functional Hda protein. *J Bacteriol* 2003, **185**:3244–3248.
58. Sezonov G, Joseleau-Petit D, D'Ari R: *Escherichia coli* physiology in Luria-Bertani broth. *J Bacteriol* 2007, **189**:8746–8749.
59. Piper SE, Mitchell JE, Lee DJ, Busby SJ: A global view of *Escherichia coli* Rsd protein and its interactions. *Mol Biosyst* 2009, **5**:1943–1947.
60. Baev MV, Baev D, Radek AJ, Campbell JW: Growth of *Escherichia coli* MG1655 on LB medium: monitoring utilization of sugars, alcohols, and organic acids with transcriptional microarrays. *Appl Microbiol Biotechnol* 2006, **71**:310–316.
61. Volkmer B, Heinemann M: Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PLoS One* 2011, **6**:e23126.
62. Velagapudi VR, Wittmann C, Lengauer T, Talwar P, Heinze E: Metabolic screening of *Saccharomyces cerevisiae* single knockout strains reveals

- unexpected mobilization of metabolic potential. *Process Biochem* 2006, **41**:2170–2179.
63. Xue GP, Denman SE, Glassop D, Johnson JS, Dierens LM, Gobius KS, Aylward JH: Modification of a xylanase cDNA isolated from an anaerobic fungus *Neocallimastix patriciarum* for high-level expression in *Escherichia coli*. *J Biotechnol* 1995, **38**:269–277.
 64. Vickers CE, Xue GP, Gresshoff PM: A synthetic xylanase as a novel reporter in plants. *Plant Cell Rep* 2003, **22**:135–140.
 65. Rocha EPC: The replication-related organization of bacterial genomes. *Microbiology* 2004, **150**:1609–1627.
 66. Kuhlman TE, Cox EC: Gene location and DNA density determine transcription factor distributions in *Escherichia coli*. *Mol Syst Biol* 2012, **8**:610.
 67. Fijalkowska LJ, Jonczyk P, Tkaczyk MM, Bialoskorska M, Schaaper RM: Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. *Proc Natl Acad Sci USA* 1998, **95**:10020–10025.
 68. Klumpp S, Zhang Z, Hwa T: Growth rate-dependent global effects on gene expression in bacteria. *Cell* 2009, **139**:1366–1375.
 69. Lee J, Choi S, Park J, Vickers C, Nielsen L, Lee S: Development of sucrose-utilizing *Escherichia coli* K-12 strain by cloning β -fructofuranosidases and its application for L-threonine production. *Appl Microbiol Biotechnol* 2010, **88**:905–913.
 70. Sahin-Tóth M, Lengyel Z, Tsunekawa H: Cloning, sequencing, and expression of *cscA* invertase from *Escherichia coli* B-62. *Can J Microbiol* 1999, **45**:418–422.
 71. Tsunekawa H, Azuma S, Okabe M, Okamoto R, Aiba S: Acquisition of a sucrose utilization system in *Escherichia coli* K-12 derivatives and its application to industry. *Appl Environ Microbiol* 1992, **58**:2081–2088.
 72. Sabri S, Nielsen LK, Vickers CE: Molecular control of sucrose utilization in *Escherichia coli* W, an efficient sucrose-utilizing strain. *Appl Environ Microbiol* 2013, **79**:478–487.
 73. Peredelchuk MY, Bennett GN: A method for construction of *E. coli* strains with multiple DNA insertions in the chromosome. *Gene* 1997, **187**:231–238.
 74. Tolmachov O, Palaszewski I, Bigger B, Coutelle C: RecET driven chromosomal gene targeting to generate a RecA deficient *Escherichia coli* strain for Cre mediated production of minicircle DNA. *BMC Biotechnol* 2006, **6**:17.
 75. Arakawa H, Lodygin D, Buerstedde JM: Mutant loxP vectors for selectable marker recycle and conditional knock-outs. *BMC Biotechnol* 2001, **1**:7.
 76. Albert H, Dale EC, Lee E, Ow DW: Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J* 1995, **7**:649–659.
 77. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO: Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth* 2009, **6**:343–345.
 78. Haldemann A, Prahalad MK, Fisher SL, Kim S-K, Walsh CT, Wanner BL: Altered recognition mutants of the response regulator PhoB: A new genetic strategy for studying protein–protein interactions. *Proc Natl Acad Sci USA* 1996, **93**:14361–14366.
 79. Sambrook J, Russell DW: *Molecular Cloning: A Laboratory Manual*. 3rd edition. Cold Spring Harbour, NY: Cold Spring Harbor Laboratory Press; 2001.
 80. Xue GP, Johnson JS, Smyth DJ, Dierens LM, Wang X, Simpson GD, Gobius KS, Aylward JH: Temperature-regulated expression of the *tac/lacI* system for overproduction of a fungal xylanase in *Escherichia coli*. *Appl Microbiol Biotechnol* 1996, **45**:120–126.
 81. Bryksin AV, Matsumura I: Rational design of a plasmid origin that replicates efficiently in both gram-positive and gram-negative bacteria. *PLoS One* 2010, **5**:e13244.
 82. Bradford MM: A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976, **72**:248–254.

doi:10.1186/1475-2859-12-60

Cite this article as: Sabri et al.: Knock-in/Knock-out (KIKO) vectors for rapid integration of large DNA sequences, including whole metabolic pathways, onto the *Escherichia coli* chromosome at well-characterised loci. *Microbial Cell Factories* 2013 **12**:60.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

