

ARTICLE

Received 24 Feb 2013 | Accepted 23 Aug 2013 | Published 23 Sep 2013

DOI: 10.1038/ncomms3503

Implementation of stable and complex biological systems through recombinase-assisted genome engineering

Christine Nicole S. Santos^{1,†}, Drew D. Regitsky^{1,†} & Yasuo Yoshikuni¹

Evaluating the performance of engineered biological systems with high accuracy and precision is nearly impossible with the use of plasmids due to phenotypic noise generated by genetic instability and natural population dynamics. Minimizing this uncertainty therefore requires a paradigm shift towards engineering at the genomic level. Here, we introduce an advanced design principle for the stable instalment and implementation of complex biological systems through recombinase-assisted genome engineering (RAGE). We apply this concept to the development of a robust strain of *Escherichia coli* capable of producing ethanol directly from brown macroalgae. RAGE significantly expedites the optimal implementation of a 34 kb heterologous pathway for alginate metabolism based on genetic background, integration locus, copy number and compatibility with two other pathway modules (alginate degradation and ethanol production). The resulting strain achieves a ~40% higher titre than its plasmid-based counterpart and enables substantial improvements in titre (~330%) and productivity (~1,200%) after 50 generations.

¹Bio Architecture Lab, Inc., 604 Bancroft Way, Suite A, Berkeley, California 94710, USA. † Present addresses: Manus Biosynthesis, Inc., 790 Memorial Drive, Suite 102, Cambridge, Massachusetts 02139, USA (C.N.S.S.) and Calysta Energy Inc., 1140 O'Brien Drive Menlo Park, California 94025, USA (D.D.R.). Correspondence and requests for materials should be addressed to Y.Y. (email: yoshikuni@ba-lab.com).

In recent years, we have seen rising interest in harnessing the power of microbes for a variety of applications, including engineering pathways for renewable fuel and chemical production^{1–4}, understanding fundamental biological functions (biosensors, gene circuits and rewiring molecular scaffolds)^{5,6}, and studying multigene and genome-wide molecular genetics^{7,8}. In synthetic biology, most ventures proceed through four phases of strain engineering—design, construction, implementation (booting) and troubleshooting (debugging)⁷. The design of increasingly more complex biological systems has been greatly enhanced by the wide array of tools now available for controlling heterologous protein and pathway expression at both the transcriptional (promoter and messenger RNA) and translational (RBS) levels^{9–11}. Similarly, the construction of large and complex designs has become an almost routine process through the use of cheap, automated and sophisticated techniques for both *de novo* DNA synthesis^{12,13} and genetic parts assembly^{14–17}. Despite this unprecedented level of genetic control, however, most engineered systems still do not perform as expected when first implemented in a microbial host. Thus, an additional phase for troubleshooting and debugging is often required to generate genetic diversity and select for improvements in the desired biological function.

Due to the cumbersome nature of the implementation and debugging process, most synthetic biology endeavours still rely on plasmid-borne protein expression. However, while the ease of manipulating and transferring extrachromosomal DNA makes plasmids an attractive tool for cursory proof-of-concept studies, their use introduces severe inaccuracies when carefully evaluating the performance of engineered biological systems. Indeed, the combined metabolic burden of both plasmid propagation/maintenance^{18,19} and pathway overexpression often results in segregational or genetic population events promoting plasmid loss or allele inactivation²⁰. In addition, several studies have shown that cell-to-cell variability in copy number can be significant under different cellular growth states and conditions^{21,22}, a phenomenon that can lead to erratic strain performance. Clearly, the use of stable, integrated components is needed to be able to accurately and precisely assess phenotypic functionality and performance.

Unfortunately, the development of efficient, flexible and specific tools for genome engineering, particularly for manipulating and integrating large, complex schemes, is still in its infancy and has hampered progress in the field of synthetic biology. Homologous recombination of single or double-stranded DNA fragments (also called ‘recombineering’) works well for smaller genetic fragments but is cumbersome for larger ones due to both natural limitations in the construction or PCR amplification of long linear fragments and the considerably lower efficiency of integration for inserts greater than just a few kilobases^{23–25}. Using this technique, the genomic incorporation of a large multigene pathway would require several sequential rounds of integration and would be a time-consuming and laborious process. Phage integration methods based on recombination between *attP*–*attB* sites have higher integration efficiencies and can easily accommodate large plasmid-based fragments^{26,27}. However, they provide little flexibility over location, as integration is typically limited to existing *att* sites within the bacterial chromosome. Phage integration also does not allow one to control the specific fragments that are incorporated into the genome; the entire donor plasmid is assimilated, including extraneous genetic components such as the origin of replication. More recently, I-SceI recognition sites have been used to generate double strand DNA breaks in order to mediate site-specific integrations within the chromosome²⁸. However, this technique was only demonstrated for a maximum fragment size

of 7 kb and is additionally limited to only a single round of incorporation within the same host cell.

Here, we describe an advanced design paradigm for the stable and optimal implementation of complex biological systems through recombinase-assisted genome engineering (RAGE). Although RAGE offers broad utility for many areas of biology (that is, it is heavily used in genetics studies of higher organisms^{29,30}), the efficiency, flexibility, specificity and speed of recombination make it a disruptive technology for synthetic biology. A detailed schematic of its application to strain implementation is described in Fig. 1. First, the biological system of interest is constructed on a single copy plasmid or bacterial artificial chromosome (BAC) using various cloning or assembly techniques^{12–17} and then transformed into different hosts for proof-of-concept studies aimed at identifying genetic backgrounds that yield the desired phenotype. (Single copy plasmids or BACs are recommended at this step, as they can maintain large DNA fragments at copy numbers that are representative of genome-based expression.) Most strain engineering endeavours do not proceed beyond this step due to technical limitations inherent to genome engineering³¹; however, the use of RAGE not only allows for the stable instalment of large, complex biological systems, but, more importantly, also enables a full exploration of various parameters, including genetic background, integration locus/location and copy number. Due to the number of strain variants that may need to be generated, efficiency and simplicity are key assets of RAGE, particularly when the performance of the targeted biological system is dependent on other enzymes or genetic modules requiring parallel optimization^{32,33}. As an added benefit, a stably integrated strain is also much more amenable to additional iterations of strain engineering, including directed or adaptive evolutionary processes aimed at improving phenotype^{4,8,20}. We demonstrate the practical application of this method through the installation and optimization of a 34 kb heterologous gene cluster enabling alginate degradation and utilization in *E. coli*. The resulting strains are capable of producing ethanol from macroalgae-derived sugars for up to 50 generations at significantly higher titres and productivities over their plasmid-based counterparts.

Results

Chromosomal integration of large DNA fragments through RAGE. In this specific implementation of RAGE, we make use of the Cre recombinase and mutated *lox* site pairings^{29,30,34–36} to facilitate integration of large genetic fragments into a precise and predetermined location within the bacterial genome (Fig. 2a). In the first step of the process, the host or recipient strain is modified by introducing a *lox* site-flanked antibiotic marker, or ‘targeting cassette,’ at the location of the desired integration. Here, a chloramphenicol resistance gene (*cat*) flanked by *loxP* and *lox5171* sites³⁴ was integrated using standard λ -RED recombination protocols²³. In the second step, Cre recombinase expression is induced from the plasmid pJW168 (Lucigen) in order to direct the transfer of a similarly *lox*-flanked fragment and marker into the genome. This ‘integration cassette’ can be provided on a single-copy plasmid or BAC pre-transformed into the recipient strain (plasmid delivery) or transferred via phage transduction (phage delivery). For the latter option, P1vir phage lysates are prepared from a donor strain containing the genetic fragment (either on a single or multicopy plasmid or present within the genome) and subsequently used to infect a recipient strain induced for Cre recombinase expression. In the following examples, an FRT-flanked kanamycin resistance gene (*kan*) was included in the cassette to select for proper integration events.

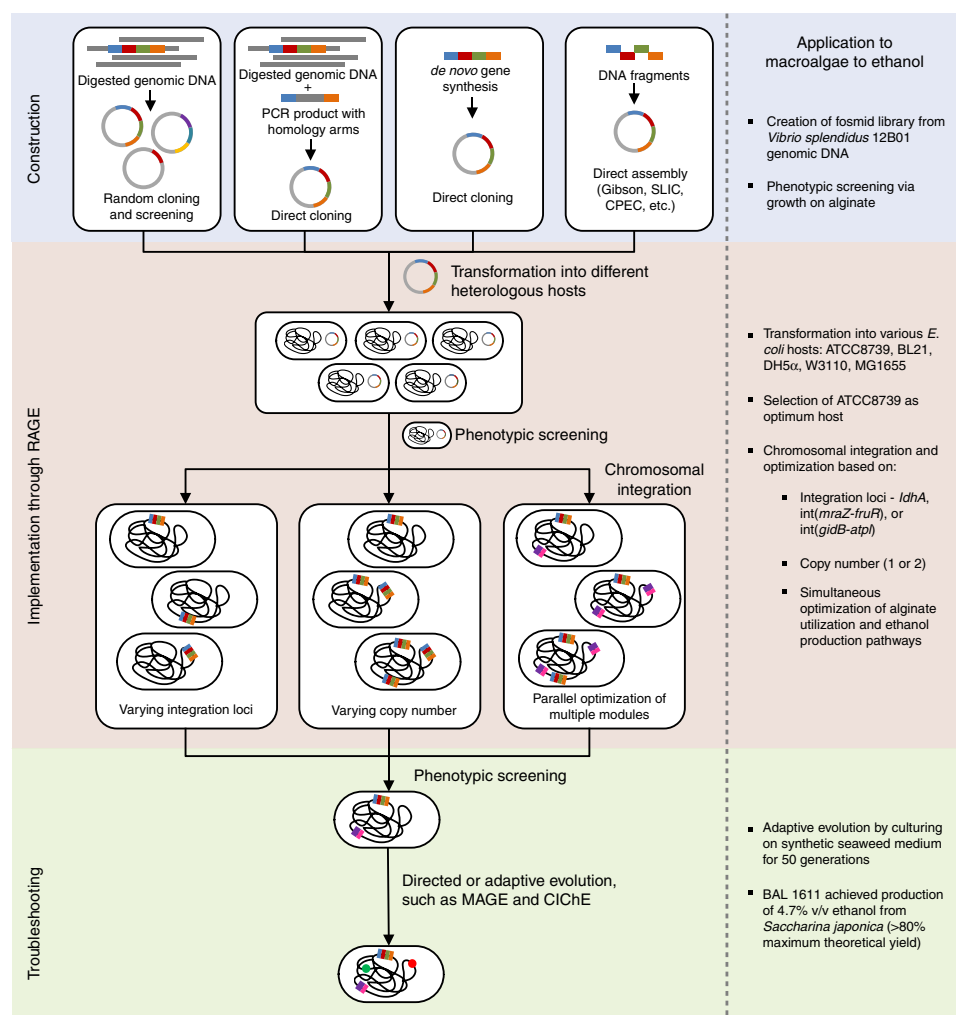


Figure 1 | Advanced design principle based on the optimal implementation of complex biological systems through RAGE. Top: Construction of biological systems (assembly of genetic parts): A plasmid containing genes responsible for a designed biological function is constructed from either genomic DNA, *de novo* synthesized DNA and/or DNA fragments through random cloning (library construction), direct cloning or DNA assembly methods^{12–17}. Middle: Implementation through RAGE (host strain selection and genome engineering): The expression vector is transformed into different heterologous hosts to identify the specific genetic backgrounds yielding the most desirable cellular phenotypes. Using RAGE, complex biological systems can be integrated into the bacterial genome to explore the effects of integration loci and copy number on phenotype. The efficiency of RAGE also allows for the parallel optimization of multiple genetic modules. Bottom: Troubleshooting desired biological functions (directed and adaptive evolution): The resulting strain can be subjected to additional optimization, including directed or adaptive evolutionary techniques aimed at increasing integrated copy numbers (chemically inducible chemical evolution, CICE²⁰) or fine-tuning biological function at the genomic level (multiplex automated genome engineering, MAGE⁸). Classical adaptive evolution approaches can also be applied to improve the desired phenotype. The panel on the right demonstrates the application of RAGE towards engineering macroalgae utilization in *E. coli* for the production of ethanol.

Excision of *kan* through expression of the FLP recombinase²³ allows for the generation of markerless integration strains capable of undergoing additional rounds of genetic modification.

Advantages of RAGE. The utility of RAGE was first demonstrated through the incorporation of a 34 kb alginate metabolic pathway (from pALG3.4, a modified version of pALG3 (ref. 37)) into the *ldhA* locus of BAL 1075 (Table 1) (Fig. 2b–d). Strains that have successfully undergone fragment integration into the targeted locus should exhibit both kanamycin resistance and chloramphenicol sensitivity (conferred by the loss of *cat* through cassette replacement), thus providing a simple initial screen for recombination accuracy (Fig. 2b). Indeed, colony PCR verification of the *ldhA* junctions revealed correct fragment placement in all 10 chloramphenicol-sensitive strains tested (Fig. 2c). Because

of the large size of the integration cassette, fragment integrity was tested through growth on alginate-derived medium rather than through resequencing of the entire genetic construct. Although innocuous mutations cannot be detected through this assay, insertions, deletions or point mutations that alter protein function should be reflected as a shift in the growth properties of the strains. As shown in Fig. 2d, however, five individual colonies exhibited similar growth profiles on 2% degraded alginate medium, indicating a low rate of mutation within the genetic fragment. (The likelihood of introducing point mutations remains low, as genetic fragments are incorporated directly from the donor material without the need for PCR amplification.) The five clones tested were obtained by both plasmid and phage delivery of the integration cassette, demonstrating that the use of either methodology leads to 100% cassette integrity and functionality after selecting for kanamycin resistance and chloramphenicol

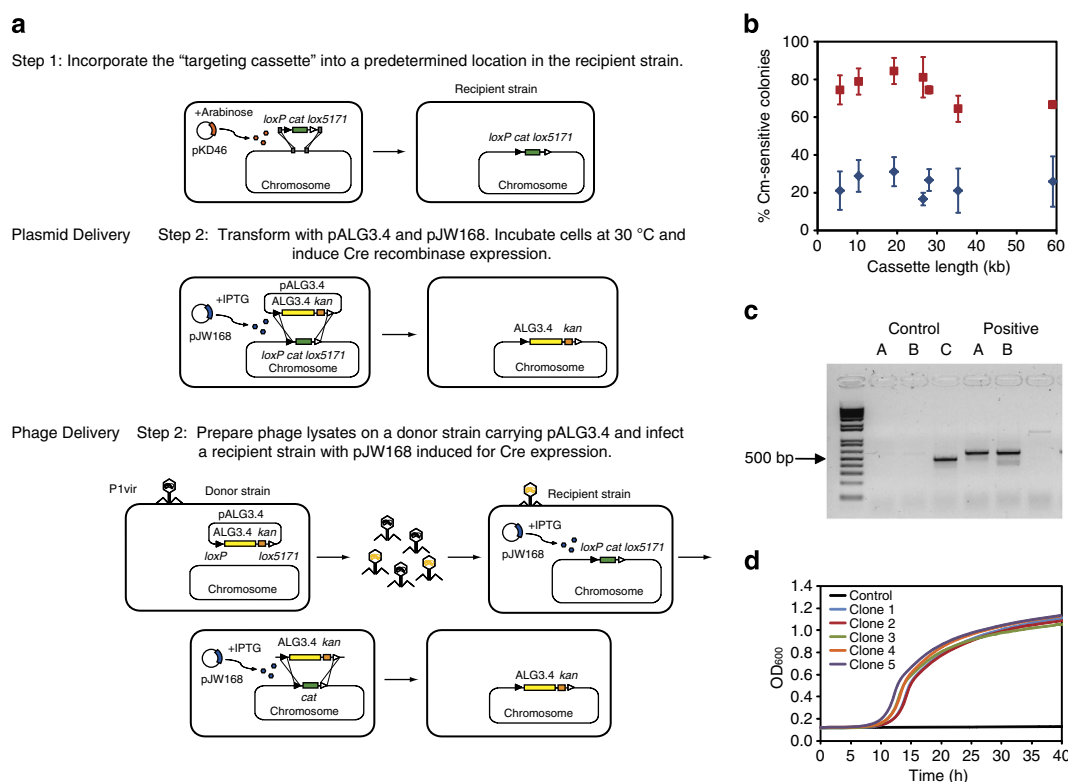


Figure 2 | Recombinase-assisted genome engineering (RAGE). RAGE is a genome engineering strategy that facilitates stable instalment and implementation of complex biological systems by recombinase-mediated cassette exchange. **(a)** In the first step, the *cat* gene flanked by two mutually exclusive *lox* sites ('targeting cassette') is incorporated into the bacterial genome through arabinose induction of the λ -RED recombination genes (on pKD46). With plasmid delivery, this strain is subsequently transformed with both pJW168 and the donor plasmid carrying the *lox*-flanked genetic fragment (in this case, pALG3.4), then grown at 30 °C with IPTG to induce Cre recombinase expression. With phage delivery, P1vir lysates are prepared from a recipient strain containing pALG3.4 and subsequently used to infect a recipient strain induced for Cre recombinase expression. In both cases, the temperature-sensitive plasmid pJW168 is lost following plating on kanamycin and growth at 37 °C. **(b)** Relationship between integration efficiency and cassette length. Efficiency is represented as the percent of kanamycin-resistant colonies that also exhibited sensitivity to chloramphenicol by either plasmid delivery (red squares) or phage delivery (blue diamonds). Error bars represent s.d. from at least three replicate platings. **(c)** Colony PCR verification of integrated strains across the *ldhA* junction. BAL 1075 *ldhA::loxP-cat-lox5171* was used as a negative control for the PCR reactions. Expected product sizes are as follows—A (left end verification): 588 bp; B (right end verification): 608 bp; C (*ldhA::loxP-cat-lox5171* cassette): no product expected if correct or 471 bp if integration failed. Because similar results were found for all 10 strains tested, only one representative set is shown. **(d)** Growth of five integration clones on 2% degraded alginate medium. BAL 1075 was used as a negative control for these growth assays. Clones 1–3 were derived through the plasmid delivery method, and clones 4–5 were obtained through phage delivery.

sensitivity. These results suggest that a simple antibiotic screen is sufficient for obtaining a proper integrant. After engineering *lox* sites flanking the donor region of interest, the entire integration procedure can be implemented and confirmed in under a week.

Homologous recombination-based techniques for genomic integration are widely used in bacterial engineering but often exhibit very low efficiencies, particularly for cassettes greater than just a few kilobases in length³⁴. As such, they often require a cumbersome validation process to select for a correct integrant. To determine the effect of cassette size on the performance of our method, we constructed several alternate versions of pALG3 containing partial or extended alginate pathways ranging from 5.6 kb to 59 kb in length (Table 2). Efficiencies of integration and correct cassette localization were then determined based on the percent of recovered (kanamycin-resistant) colonies exhibiting chloramphenicol sensitivity (although we note that selected colonies from each cassette length were additionally verified by performing colony PCR across the integration junctions). As shown in Fig. 2b, the phage delivery method showed no dependence on length with efficiencies ranging between 20–30% across all fragment sizes. Efficiencies with plasmid delivery were

much higher (65–85%) and showed only a slight drop for cassettes greater than 30 kb. Because the 59 kb fragment comprises an extended alginate pathway containing additional copies or orthologs of several alginate degradation enzymes, strains with a correctly integrated cassette should display enhanced growth properties over its 34 kb counterpart. Indeed, strains with the extended pathway displayed both higher growth rates ($\mu = 0.34$ per hour versus 0.30 per hour) and maximum cell densities ($OD_{600} = 1.6$ versus 1.1) when grown on 2% degraded alginate medium. Thus, this methodology can be utilized for the integration of cassette sizes up to at least 59 kb with both high accuracy and efficiencies.

Efficient and optimal implementation of biological systems. To demonstrate the potential of the advanced design principle, we applied it towards engineering a microbial platform for the conversion of brown macroalgae (brown seaweed) into ethanol (Fig. 1). Brown macroalgae possess several attractive features, which make them ideal next-generation feedstocks for biofuel and chemical production³⁷. Unlike sugarcane or corn, seaweed

Table 1 Strains used in this study.		
Strain	Description	Reference
BAL 1075	<i>E. coli</i> ATCC 8739 Δ ldhA Δ frd::Zm_adhB	This study
BAL 1303	<i>E. coli</i> ATCC 8739 Δ ldhA Δ frd::P ₆₂₅ -Zm_pdc-Zm_adhB Δ focA-pflB	(37)
BAL 1117	BAL 1075 pALG3-N455 + tSM0524 Aly	This study
BAL 1373	BAL 1303 pALG3-N455 + tSM0524 Aly	This study
BAL 1450	BAL 1303 int(mraZ-fruR)::ALG3.4	This study
BAL 1452	BAL 1075 int(mraZ-fruR)::ALG3.4 int(gidB-atpI)::ALG3.4	This study
BAL 1453	BAL 1075 int(mraZ-fruR)::ALG3.4	This study
BAL 1493	BAL 1303 int(mraZ-fruR)::ALG3.4 int(gidB-atpI)::ALG3.4	This study
BAL 1611	BAL 1450 Δ frd::N455 + tSM0524 Aly	(37)

Table 2 Constructed pALG plasmid versions and their corresponding lox-flanked cassette sizes.		
Original plasmid	Modified plasmid	Cassette size (kb)*
pALG3	pALG3.4	35.3
pALG3	pALG3.5	26.5
pALG3	pALG3.6	19.2
pALG3	pALG3.7	10.3
pALG3	pALG3.8	5.6
pALG4	pALG4.4	28.0
pALG7	pALG7.4	59.1

*Includes antibiotic selection marker.

cultivation does not require arable land, fertilizer or fresh water resources, thus minimizing its impact on both the environment and existing food supplies. The absence of lignin in seaweed also eliminates the requirement for expensive pretreatment operations to promote sugar release. Although *E. coli* readily utilizes two of the three main sugars in seaweed—mannitol and glucan (glucose polymers in the form of laminarin or cellulose)—no industrial strain to-date has the capability of metabolizing the third sugar component, alginate.

We have previously shown that efficient utilization of alginate requires a 34 kb heterologous gene cluster from *Vibrio splendidus* 12B01 present on the single-copy plasmid pALG3 (originally constructed by using recombineering techniques on the fosmid pALG1, which was isolated from a genomic library of *V. splendidus* 12B01)³⁷. To determine the optimal strain for this pathway, we first transformed pALG3 into five *E. coli* variants – ATCC 8739, BL21, DH5 α , K-12 (MG1655) and W3110. Growth on 2% degraded alginate was then used as a crude readout for pathway expression and functionality. Although the presence of pALG3 was able to impart growth on alginate for all five genotypes tested, ATCC 8739 exhibited a much faster growth rate and was therefore selected for subsequent integration studies (Fig. 3a).

Because chromosomal positioning has been shown to have a role in gene expression³⁸, we explored loci-dependent transcriptional variation by integrating our pathway into three different locations within the ATCC 8739-derived strain, BAL 1075 (Table 1). The following positions were selected based on their distance to the chromosomal origin of replication (*oriC*): (1) the intergenic region of *gidB* and *atpI* (0.1% away from *oriC*), (2) the intergenic region of *mraZ* and *fruR* (33.3% away from *oriC*) and (3) *ldhA* (99.5% away from the *oriC*). In this context,

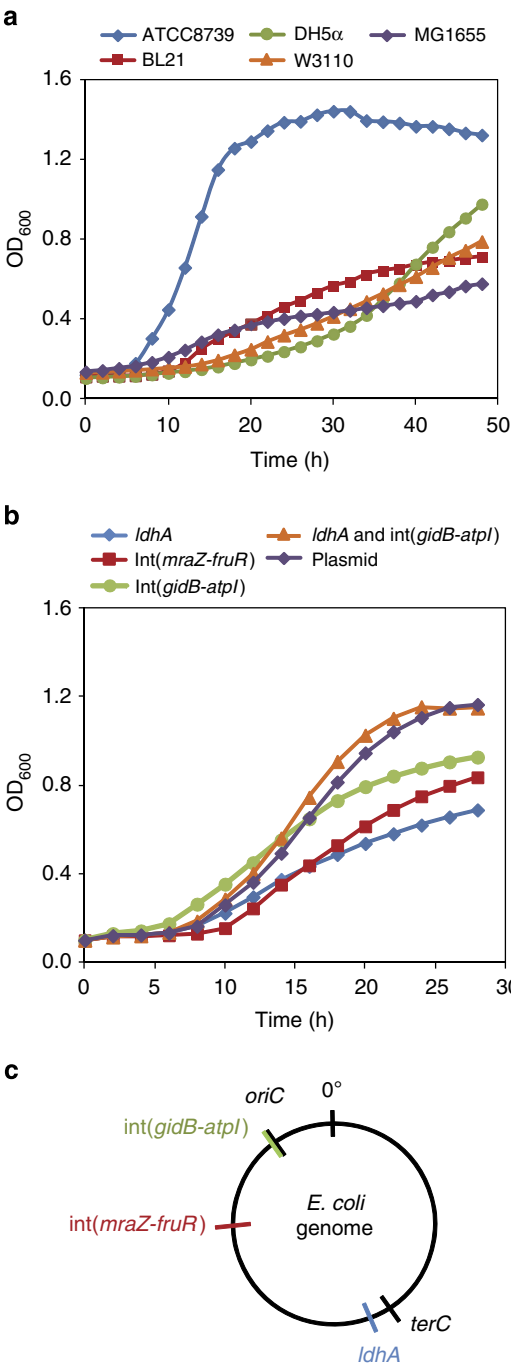


Figure 3 | Growth of strains on 2% degraded alginate. (a) pALG3 in five different *E. coli* strain backgrounds **(b)** Integration of ALG3.4 into different loci within *E. coli* ATCC8739's genome. Data points represent averages from at least three biological replicates. All standard deviations were less than 5% of the average, and error bars have been omitted for visual clarity. **(c)** Positioning of three integration loci relative to the *oriC* and the terminus of replication (*terC*) of the *E. coli* genome.

percent distance represents the fraction of the genome that is closer to the *oriC*. As chromosomal replication begins at *oriC*, genes that are located near this reference point are duplicated first and possess a higher 'effective copy number' within the cell³⁸. As seen in Fig. 3b, we found a clear negative correlation between growth and distance from *oriC*, confirming that the best expression can be obtained at locations closer to the origin of replication. Interestingly, the growth of even the best single-copy

integration (in the intergenic region of *gidB* and *atpI*) still lagged in comparison to a strain carrying the single-copy plasmid, pALG3. These differences were largely eliminated through the introduction of a second copy of the alginate pathway into the *ldhA* locus via phage transduction.

Parallel optimization of a multi-module microbial system. Although two integrated copies of the alginate metabolic pathway are clearly needed for robust growth on degraded alginate, phenotypes that divert cellular resources from biomass generation into alternate products may possess different expression level optima. As an example, we postulated that efficient production of ethanol from alginate may require balancing of the two pathway modules for alginate consumption and the conversion of pyruvate to ethanol. To determine the most effective coupling of alginate utilization and ethanol production, we constructed and tested the full matrix of single and double integrations for each independent module (Table 3). The alginate metabolic pathway from pALG3 was incorporated into the intergenic regions of *mraZ* and *fruR* (single copy) and *gidB* and *atpI* (double copy). To engineer robust

ethanol production, the homoethanol pathway encoded by the *Zymomonas* pyruvate decarboxylase (Pdc) and alcohol dehydrogenase B (AdhB) genes was integrated as a single copy (*frd* locus, controlled by the strong constitutive promoter P_{G25} from phage T5 in BAL 1303 (Table 1)) or as double copies (*frd/pta/focA-pflB* loci in BAL 1075 controlled by their native promoters). We also expressed a secreted alginate lyase system on the single-copy plasmid N455 + tSM0524 Aly to catalyse the extracellular depolymerization of alginate³⁷.

To evaluate their performance, strains were grown micro-aerobically in synthetic seaweed media containing 50 g l⁻¹ total sugars (alginate, mannitol and glucose at a ratio of 5:8:1) and assayed for ethanol production (maximum theoretical titre of ethanol is 25 g l⁻¹). In all cases, we observed lower ethanol titres (Fig. 4a) and productivities (Fig. 4b) in strains harbouring a plasmid-based alginate pathway relative to their integrated counterparts (*P* < 0.001 by Student's *t*-test), indicating that plasmid instability is an issue, even at these shorter time scales. (The only exception is BAL 1452, which seems to be impaired by the overexpression of enzymes from double chromosomal copies of both the ethanol and alginate pathways.) Although both BAL 1075 and BAL 1303 possess the ability to produce ethanol from glucose with nearly theoretical maximum yield, ethanol production is significantly reduced in BAL 1075 with an integrated alginate pathway, thus highlighting the importance of optimizing relative expression levels through empirical testing. The best overall performance was obtained from two strains containing a single integrated ethanol pathway and either one or two copies of the alginate pathway (BAL 1450 and BAL 1493, respectively), yielding titres of 21–22 g l⁻¹ ethanol (83–87% of the maximum theoretical yield) and productivities of 0.35 g l/h. Integration of the secreted alginate lyase system from N455 + tSM0524 Aly into BAL 1450 yielded BAL 1611, a fully integrated strain possessing an ~40% improvement in ethanol production over its plasmid-based counterpart BAL 1373 (Fig. 4).

Stable maintenance of phenotype over 50 generations. From an industrial context, strain implementation through RAGE is especially advantageous, as it allows for the extended maintenance of phenotypes, even in the absence of external selection pressures such as antibiotics. To assess the genetic and phenotypic stability of our engineered strains, BAL 1373, BAL 1450 and BAL 1611 were grown in synthetic seaweed media over 50 generations through 11 successive rounds of serial subculturing. Strains that retain their properties through 50 doublings are expected to remain stable from a seed inoculum all the way up to an industrial scale process (~150,000 l). Although plasmid-based

Table 3 Strains generated for the comparison of plasmid- and chromosomal-based expression for three independent genetic modules.			
BAL #	Alginate lyase (N455 + tSM0524 Aly)*	Alginate pathway (pALG3 or ALG3.4)*	Ethanol pathway (<i>pdc</i> , <i>adhB</i>)*
1373	Plasmid	Plasmid	Chromosome (1) <i>frd</i>
1117	Plasmid	Plasmid	Chromosome (2) <i>frd</i> , <i>pta</i> , <i>focA-pflB</i>
1450	Plasmid	Chromosome (1) <i>int(mraZ-fruR)</i>	Chromosome (1) <i>frd</i>
1453	Plasmid	Chromosome (1) <i>int(mraZ-fruR)</i>	Chromosome (2) <i>frd</i> , <i>pta</i> , <i>focA-pflB</i>
1493	Plasmid	Chromosome (2) <i>int(mraZ-fruR)</i> , <i>int(gidB-atpI)</i>	Chromosome (1) <i>frd</i>
1452	Plasmid	Chromosome (2) <i>int(mraZ-fruR)</i> , <i>int(gidB-atpI)</i>	Chromosome (2) <i>frd</i> , <i>pta</i> , <i>focA-pflB</i>
1611	Chromosome (1) <i>Frd</i>	Chromosome (1) <i>int(mraZ-fruR)</i>	Chromosome (1) <i>frd</i>

*Numbers in parenthesis indicate the number of copies integrated into the chromosome. Italics indicates locus or loci of integration.

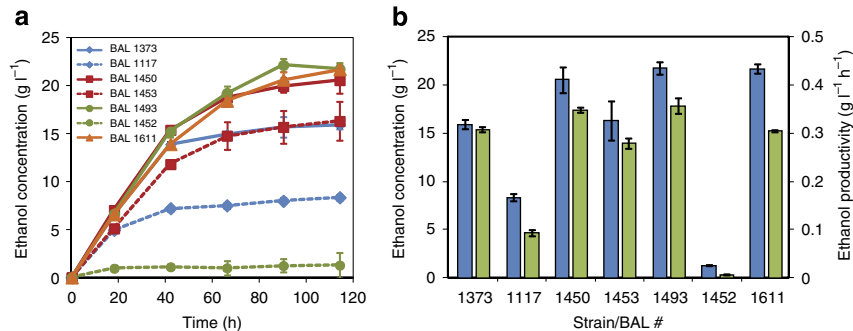


Figure 4 | Effect of plasmid- and chromosomal-based expression on ethanol production. Strain numbers correspond to Table 1. (a) Ethanol production profiles—Matching colours represent strains with identical alginate lyase and alginate pathway configurations with either a single integrated copy (solid line) or two integrated copies (dashed line) of the ethanol pathway. (b) Final ethanol titres (114 h, blue) and productivities (18–42 h, green). Error bars represent standard deviations from at least three biological replicates.

expression of the alginate pathway in BAL 1373 was already previously shown to be unfavourable (Fig. 4), extended culturing of this strain in synthetic seaweed media led to an additional twofold drop in final ethanol titre (Fig. 5a). In stark contrast, integration of the alginate pathway in both BAL 1450 and BAL 1611 led to stable phenotypes over 50 generations, with strains achieving even slightly higher titres and productivities than the original, uncultured variants ($P < 0.01$ by Student's *t*-test) (Fig. 5b,c). The performance of BAL 1611 beats its plasmid counterpart BAL 1373 by a significant margin ($\sim 330\%$ in titre and $\sim 1,200\%$ in productivity) despite the presence of strong selection pressures (antibiotic supplementation and carbon utilization for growth) to encourage plasmid maintenance.

Discussion

Genetic instability and subsequent loss of function can be severe in engineered strains even with the use of single copy, low-metabolic burden expression vectors and multiple routes for selection and plasmid retention. As such, evaluating the performance of engineered biological systems with high accuracy and precision is nearly impossible with the use of plasmid systems. As demonstrated here, RAGE encompasses several important features that make it an enabling technology for many applications within synthetic biology, including the construction and optimization of genetically stable strains. First, the use of a site-specific recombinase such as the Cre enzyme allows for a high efficiency of integration, even for large genetic fragments. Although the maximum size tested in this study was 59 kb, we expect this methodology to have the capacity to incorporate even larger pieces with minimal impact on efficiency. Phage delivery can mediate the insertion of pieces up to 90 kb in size, an upper bound imposed by the amount of DNA packaged into the head of a bacteriophage particle³⁹. Plasmid delivery has even less stringent limitations as it only requires stable maintenance of the genetic fragment on a plasmid or BAC, the latter of which has been routinely used for genetic fragments as large as 300 kb in size⁴⁰. Second, RAGE is versatile and flexible with respect to both integration locus (determined through placement of the targeting cassette) and mode of fragment delivery (phage or plasmid). Although our efficiency studies revealed a two- to fourfold higher rate of integration via (single copy) plasmid delivery (Fig. 2b), phage delivery allows for the insertion of fragments that reside within a chromosome or on a multicopy plasmid, thus expanding the array of acceptable donor material. A third important feature of this method is its ability to maintain cassette integrity during chromosomal incorporation, even for large constructs. Due to the size of our alginate utilization pathway, direct sequencing of the genetic cluster would have been both expensive and impractical. We therefore used growth rates on 2% degraded alginate as surrogate readout to verify cassette integrity and found very close agreement among several colonies that were tested (Fig. 2d). Indeed, the likelihood of introducing point mutations remains low, as genetic fragments are incorporated directly from the donor material without the need for PCR amplification.

Taken altogether, the efficiency, specificity and speed of RAGE pave the way for an advanced design paradigm for the stable and optimal implementation of complex biological systems, facilitating explorations based on several genetic parameters, including strain background, integration loci and copy number. The potential of this method is most palpable when working with long heterologous pathways, particularly in situations when cellular phenotype is dependent on two or more modules requiring parallel optimization. Although RAGE was only applied towards integration of a single module in this example, it is a straightforward exercise to pursue the integration of multiple

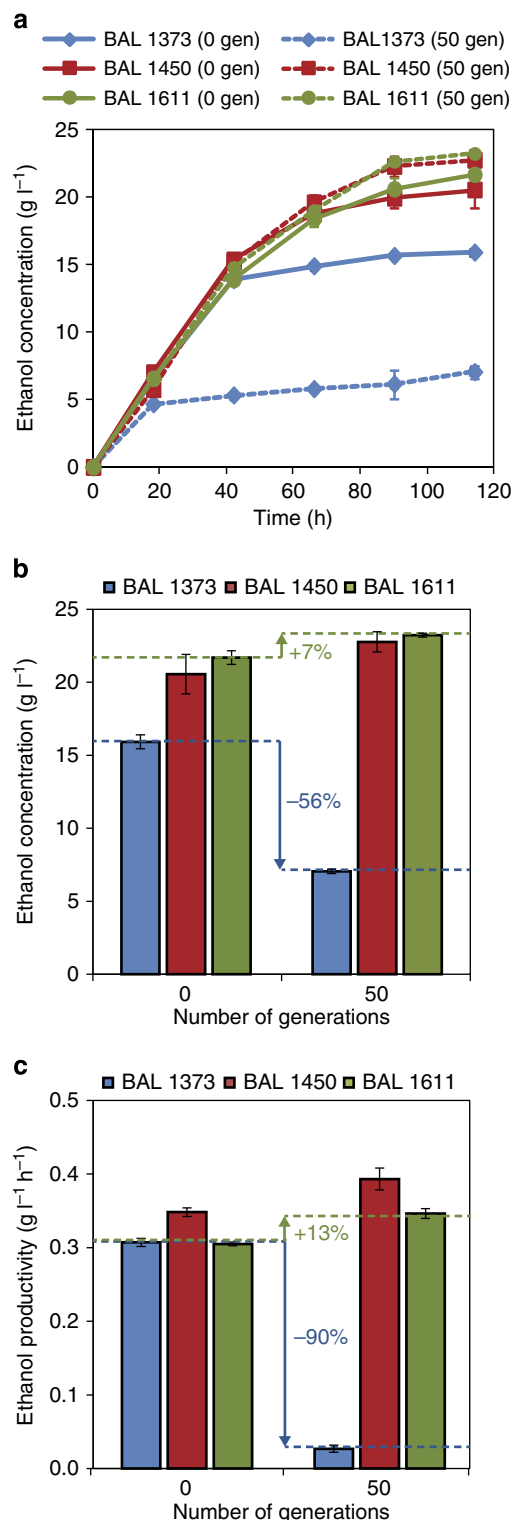


Figure 5 | Performance of select strains after 0 or 50 generations of culturing. Strain numbers correspond to Table 1. **(a)** Ethanol production profiles—Matching colours represent identical strains tested after 0 generations (solid line) or 50 generations (dashed line) of culturing. **(b)** Final ethanol titres (114 h) and **(c)** ethanol productivities (18–42 h). Error bars represent standard deviations from at least three biological replicates.

modules into a variety of backgrounds through the use of alternate mutually exclusive *lox* site pairings^{34–36}. Finally, coupling this technique with adaptive evolution or direct

genome editing^{4,8,20} can further optimize the performance of the designed biological functions of interest. Indeed, the recovery of better ethanol producers after 50 generations of culturing clearly demonstrates the potential of generating additional improvements in phenotype through adaptive evolution and screening. Given such advantages, we expect the application of RAGE and this implementation paradigm to significantly accelerate the exploration of many more avenues within synthetic biology.

As we have previously reported, the fully integrated strain BAL 1611 generated by this study has proven to be well-optimized and is capable of producing 4.7% v/v ethanol (over 80% of the maximum theoretical yield) directly from brown macroalgae with a yield of 0.281 g-ethanol/g-dry seaweed³⁷. This efficient utilization of macroalgae sugars coupled with its enhanced genetic stability make this strain a suitable microbial platform for the industrial-scale production of renewable fuels and chemicals from seaweed.

Methods

Cultivation conditions. All strain/plasmid manipulations were conducted on Luria-Bertani (LB) medium. When appropriate, antibiotics were added at the following concentrations: 25 µg ml⁻¹ chloramphenicol for maintenance of *cat*-containing plasmids or strains, 12.5 µg ml⁻¹ kanamycin for maintenance of *kan*-containing plasmids or strains and 100 µg ml⁻¹ ampicillin for maintenance of pKD46, pCP20 and pJW168. Primers used in this study are listed in Supplementary Table S1.

Modification of pALG plasmids for Cre-lox recombination. Two *lox* sites—*loxP* and *lox5171*—were introduced into pALG3 (ref. 37) through four rounds of sequential modifications with λ-RED recombination²³. pALG3.1 was constructed by amplifying *cat* from pCm-R6Kγ⁴¹ with primers CS009 Cm/Km hom sense and CS010 Cm anti-hom and transforming the resulting cassette into DH5α pALG3 pKD46 electrocompetent cells. This initial step was necessary to change the antibiotic resistance of pALG3 from kanamycin to chloramphenicol (*kan* to *cat*) in order to facilitate downstream integration steps. To insert a *loxP* site into pALG3.1 (to form pALG3.2), the *loxP::kan^{FRT}* cassette was amplified from pKD13 with primers CS037 hom-pKD13 sense and CS038 hom-*loxP*-pKD13 anti, followed by a second round of amplification with CS039 pKD13 sense 2 and CS040 *lox*-pKD13 anti 2. Excision of FRT-flanked *kan* was mediated by transformation with the FLP recombinase-expressing plasmid pCP20 (ref. 23) to form pALG3.3. In the final step of modification, a *lox5171* site was introduced into pALG3.3 by amplifying a *lox5171::kan^{FRT}* cassette from pKD13 with primers CS011 *lox5171*-pKD13 anti and CS013 hom1-pKD13 sense, followed by a second round of amplification with CS012 *lox5171*-pKD13 anti 2 and CS013 hom1-pKD13 sense. The cassette was subsequently transformed into DH5α pALG3.3 pKD46 to form pALG3.4, a plasmid containing a 35 kb fragment flanked by *loxP*, *lox5171* and *kan^{FRT}*.

Plasmids pALG3.5–8 were additionally constructed to provide integration cassettes of varying lengths, ranging from 5.6 kb to 26.5 kb. *lox5171::kan^{FRT}* cassettes for construction were amplified with CS011 *lox5171*-pKD13 and CS014-17 hom2-5 pKD13 sense, followed by a second round of amplification with CS012 *lox5171*-pKD13 anti 2 and the same set of second primers. Cassettes were subsequently transformed into DH5α pALG3.3 pKD46.

Two *lox* sites—*loxP* and *lox5171*—were introduced into pALG4 to yield pALG4.4 through a similar method as described above. CS037 hom-pKD13 sense-CS115 hom-*loxP*-pKD13 anti and CS039 pKD13 sense 2-CS116 *lox*-pKD13 anti 2 primer pairings were used to amplify the *loxP::kan^{FRT}* cassette from pKD13 for the construction of pALG4.2. Excision of FRT-flanked *kan* was mediated by transformation with pCP20 to form pALG4.3, and the *lox5171::kan^{FRT}* cassette was amplified from pKD13 with primer pairs CS011 *lox5171*-pKD13 anti-CS120 hom-pKD13 sense and CS012 *lox5171*-pKD13 anti 2-CS120 hom-pKD13 sense to form pALG4.4.

Two *lox* sites—*loxP* and *lox5171*—were introduced into pALG7 to yield pALG7.4. Construction of pALG7.8.2 and pALG7.8.3 proceeded through the same steps as for pALG3.2 and pALG3.3. The *lox5171::kan^{FRT}* cassette was amplified from pKD13 with primer pairs CS011 *lox5171*-pKD13 anti-CS113 hom-pKD13 sense and CS012 *lox5171*-pKD13 anti 2-CS113 hom-pKD13 sense to form pALG7.4.

Correct integration events were verified after each step by colony PCR and sequencing.

Construction of BAL1075. To achieve stable ethanol production, *pdh* and *adhB* derived from *Zymomonas mobilis* (*Zm_pdh* and *Zm_adhB*) and *pdh* derived from *Zymomonas palmae* (*Zp_pdh*) were integrated into the *E. coli* host ATCC8739 *ΔldhA*³⁷. Briefly, DNA fragments for *Zm_adhB* and the FRT-flanked kanamycin

cassette from pKD13 (*kan^{FRT}*) were first amplified from *Z. mobilis* genomic DNA and pKD13 using forward primers *adhB*-pKD13-F and pKD13-F and reverse primers *adhB*-R and *adhB*-pKD13R, respectively. Amplified fragments were gel purified and adjoined by PCR using the primer set pKD13-F and *adhB*-R, then treated with T4 DNA polynucleotide kinase and ligated with T4 DNA ligase to form pKD13-*adhB*. The *Zm_adhB-kan^{FRT}* cassette was amplified from pKD13-*adhB* using forward primer *frd*-*adhB*-F and reverse primer *frd* *adhB*-R and integrated into the *frd* gene locus of *E. coli* ATCC8739 *ΔldhA* via λ-RED recombination²³ to form ATCC8739 *ΔldhA Δfrd::Zm_adhB*.

A similar procedure was used to create pKD13-*pdh* using forward primers *pdh*-F and *pdh*-pKD13-F and reverse primers *pdh*-R and pKD13-R to amplify *Zm_pdh* and *kan^{FRT}* from *Z. mobilis* genomic DNA and pKD13, respectively. The *Zm_pdh-kan^{FRT}* cassette was then amplified from pKD13-*pdh* using forward primer *pta*-*pdh*-F and reverse primer *pta*-*pdh*-R and integrated into the *pta* gene locus of *E. coli* strain ATCC8739 *ΔldhA Δfrd::Zm_adhB* via λ-RED recombination to form ATCC8739 *ΔldhA Δfrd::Zm_adhB Δpta::Zm_pdh*.

pKD13-pflPr-*Zp_pdh-adhB-kan^{FRT}*-pflTm was created by amplifying DNA fragments for the *pfl* promoter (pflPr), *Zp_pdh*, *Zm_adhB*, *kan^{FRT}*, the *pfl* terminator (pflTm) and R6Kγ using forward primers pflPr-F, pflPr-*Zp_pdh*-F, *Zp_pdh*-*adhB*-F, *adhB*-FRT-Km-FRT-F, FRT-Km-FRT-pflTm-F and pflTm-R6Kγ-F and reverse primers *Zp_pdh*-*adhB*-R, *adhB*-FRT-Km-FRT-R, FRT-Km-FRT-pflTm-R, pflTm-R6Kγ-R and R6Kγ-R, respectively. Genomic DNA from *E. coli*, *Z. palmae*, and *Z. mobilis* and pKD13 were used as templates. The amplified fragments were gel purified and adjoined by PCR using the primer set pflPr-F and pflTm-R, then treated with T4 DNA polynucleotide kinase and ligated with T4 DNA ligase to form pKD13-pflPr-*Zp_pdh-adhB-kan^{FRT}*-pflTm. *Zp_pdh-adhB-kan^{FRT}* was amplified from this plasmid using forward primer pflPr-F and the reverse primer pflTm-R and integrated into the *focA-pflB* gene locus of *E. coli* ATCC8739 *ΔldhA Δfrd::Zm_adhB Δpta::Zm_pdh* via λ-RED recombination to form BAL1075 (ATCC8739 *ΔldhA Δfrd::Zm_adhB Δpta::Zm_pdh ΔfocA-pflB::Zp_pdh-Zm_adhB*). Excision of *kan^{FRT}* after all integration events was mediated by transformation with the FLP recombinase-expressing plasmid pCP20 (ref. 23).

Genomic insertion of *lox* targeting cassette. *Lox* sites were integrated into the *ldhA* locus of BAL 1075 (Table 1) using λ-RED recombination²³ and a chloramphenicol marker (*cat*). Briefly, *cat* was amplified from pCm-R6Kγ⁴¹ with primers CS001 *lox5171*-Cm sense, CS002 *loxP*-Cm anti and Phusion Hot-Start II DNA polymerase (New England BioLabs). Two distinct and mutually exclusive *lox* sites (*loxP*, *lox5171* (ref. 29)) were incorporated into the primer sequences to allow for double-crossover recombination of similarly *lox*-flanked fragments into the genome. In addition, these primers contained 28–29 bp homology with the *ldhA* region of *E. coli* ATCC 8739. The resulting *ldhA::loxP-cat-lox5171* cassette was then re-amplified with primers CS003 *lox*-Cm sense 2 and CS004 *lox*-Cm anti 2 to extend its *ldhA* homology region (for a total of 78 bp) and subsequently utilized for λ-RED recombination²³. *Lox* sites were integrated into the intergenic regions of *gidB* and *atpI* and *mraZ* and *fruR* through a similar method as above. The int(*gidB-atpI*):*loxP-cat-lox5171* integration cassette was amplified with CS095 *lox5171-gidB-atpI* sense and CS096 *loxP-gidB-atpI* anti, followed by a second round of amplification with CS097 *lox-gidB-atpI* sense 2 and CS098 *lox-gidB-atpI* anti 2. The int(*mraZ-fruR*):*loxP-cat-lox5171* cassette was amplified with CS105 *lox5171-mraZ-fruR* sense and CS106 *loxP-mraZ-fruR* anti, followed by a second round of amplification with CS107 *lox-mraZ-fruR* sense 2 and CS108 *lox-mraZ-fruR* sense 2. Colony selection was performed on LB-agar plates with 25 µg ml⁻¹ chloramphenicol, and correct integration events were verified by colony PCR and sequencing.

Cre-lox-mediated integration via plasmid delivery. To create BAL 1075 *ldhA::loxP-ALG3.4-lox5171*, BAL 1075 *ldhA::loxP-cat-lox5171* was transformed with pALG3.4 and pJW168 (Lucigen), a plasmid containing a temperature-sensitive replicon and an inducible Cre recombinase. After overnight growth in LB medium at 30 °C, 25 µl was used to inoculate 2.5 ml fresh LB with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 12.5 µg ml⁻¹ kanamycin. Cultures were grown for 3–6 h at 30 °C, then streaked out on LB-agar plates with kanamycin to isolate single colonies. After overnight growth at 37 °C, individual colonies were streaked out on LB-kanamycin and LB-chloramphenicol to identify chloramphenicol-sensitive colonies. These colonies were additionally verified for proper end integration by colony PCR using primer pairs CS005 *ldhA* *verif* sense-CS063 left *verif* anti and CS008 *ldhA* *verif* anti-CS064 right *verif* sense. Correct colonies were also streaked out on LB-ampicillin plates to verify loss of pJW168. A similar procedure was used to create strains BAL 1075 int(*gidB-atpI*):*loxP-ALG3.4-lox5171*, BAL 1075 int(*mraZ-fruR*):*loxP-ALG3.4-lox5171* (BAL 1453), and BAL 1303 int(*mraZ-fruR*):*loxP-ALG3.4-lox5171* (BAL 1450). Primers used for colony PCR are as follows: int(*gidB-atpI*)—CS063 left *verif* anti-CS099 *gidB-atpI* *verif* sense and CS064 right *verif* sense-CS100 *gidB-atpI* *verif* anti; int(*mraZ-fruR*)—CS064 right *verif* sense-CS109 *mraZ-fruR* *verif* sense and CS063 left *verif* anti-CS110 *mraZ-fruR* *verif* anti. Excision of FRT-flanked *kan* was mediated by transformation with the FLP recombinase-expressing plasmid pCP20.

To test the effect of cassette length on integration efficiency, the same procedure was repeated for pALG3.5–8. Efficiency was defined as the percent of colonies that

were found to be chloramphenicol sensitive (out of a total of about 30 colonies). In addition, two random chloramphenicol-sensitive colonies from each culture were chosen for colony PCR verification. Triplicate cultures and recombination experiments were performed for each plasmid version.

Cre-lox-mediated integration via phage delivery. BAL 1075 was transformed with pALG3.4 and used for the preparation of lysates from the P1vir bacteriophage^{39,42}. This lysate was subsequently used to infect an overnight culture of BAL 1075 *ldhA::loxP-cat-lox5171* pJW168 grown at 30 °C in LB medium with ampicillin and 1 mM IPTG. Following a 1-h infection, cells were plated on LB-agar plates with kanamycin to isolate single colonies. After overnight growth at 37 °C, individual colonies were streaked out on LB-kanamycin and LB-chloramphenicol to identify chloramphenicol-sensitive colonies. These colonies were additionally verified for proper end integration by colony PCR using the primer pairs listed above. Correct colonies were also streaked out on LB-ampicillin plates to verify loss of pJW168. Excision of FRT-flanked *kan* was mediated by transformation with the FLP recombinase-expressing plasmid pCP20.

To test the effect of cassette length on integration efficiency, the same procedure was repeated for pALG3.5–8. Efficiency was defined as the percent of colonies that were found to be chloramphenicol sensitive (out of a total of about 30 colonies). In addition, two random chloramphenicol-sensitive colonies from each culture were chosen for colony PCR verification. Triplicate lysate preparation and infection experiments were performed for each plasmid version.

Increasing gene copy number through P1vir phage transduction. Transfer of a second integrated copy of ALG3.4 (into BAL 1450 to generate BAL 1493 or into BAL 1453 to generate BAL 1452) was mediated by P1vir phage transduction^{39,42}. Proper integration was verified by colony PCR as described above. Excision of FRT-flanked *kan* was mediated by transformation with the FLP recombinase-expressing plasmid pCP20.

Alginate growth assays. Strains used for alginate growth assays were first grown overnight in LB medium at 30 °C. Hundred-microliters of cell culture was then washed and resuspended in an equal volume of 2% degraded alginate medium (M9 minimal medium with 2% alginate pre-degraded with 10 µg ml⁻¹ alginate lyase (Sigma) overnight at 30 °C), and 4 µl was used as an inoculum for 196 µl 2% degraded alginate medium. Alginate growth assays were performed at 30 °C, and cell density (OD₆₀₀) was monitored with a BioTek Synergy HT Multidetector microplate reader. All liquid cultivations were conducted with at least three biological replicates.

Ethanol fermentations. Ethanol fermentations were conducted in stirred bottles containing 100 ml M9 minimal media with 50 g l⁻¹ sugars (alginate, mannitol and glucose at a ratio of 5:8:1) and 5 g l⁻¹ LB. Cultures were inoculated to a starting cell density (OD₆₀₀) of ~0.6 and kept at 25 °C. Ethanol concentrations were quantified by high performance liquid chromatography (Shimadzu, Columbia, MD) equipped with an organic acid column (Phenomenex, Torrance, CA). Chromatography was operated at 60 °C using 5 mM H₂SO₄ as a mobile phase at a flow rate of 1 ml min⁻¹ (5 µl injection volume, 15 min isocratic method). Ethanol peaks were detected using a refractive index detector and compared with chemical standards.

References

1. Felnagle, E. A., Chaubey, A., Noey, E. L., Houk, K. N. & Liao, J. C. Engineering synthetic recursive pathways to generate non-natural small molecules. *Nat. Chem. Biol.* **8**, 518–526 (2012).
2. Lee, J. W. *et al.* Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat. Chem. Biol.* **8**, 536–546 (2012).
3. Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B. & Keasling, J. D. Microbial engineering for the production of advanced biofuels. *Nature* **488**, 320–328 (2012).
4. Santos, C. N. S. & Stephanopoulos, G. Combinatorial engineering of microbes for optimizing cellular phenotype. *Curr. Opin. Chem. Biol.* **12**, 168–176 (2008).
5. Tabor, J. J. *et al.* A synthetic genetic edge detection program. *Cell* **137**, 1272–1281 (2009).
6. Nandagopal, N. & Elowitz, M. B. Synthetic biology: integrated gene circuits. *Science* **333**, 1244–1248 (2011).
7. Carr, P. A. & Church, G. M. Genome engineering. *Nat. Biotechnol.* **27**, 1151–1162 (2009).
8. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
9. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc. Natl Acad. Sci. USA* **102**, 12678–12683 (2005).
10. Pfleger, B. F., Pitera, D. J., Smolke, C. D. & Keasling, J. D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* **24**, 1027–1032 (2006).
11. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
12. Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
13. Kodumal, S. J. *et al.* Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl Acad. Sci. USA* **101**, 15573–15578 (2004).
14. Fu, J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30**, 440–446 (2012).
15. Quan, J. & Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One* **4**, e6441 (2009).
16. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
17. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256 (2007).
18. Birnbaum, S. & Bailey, J. E. Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*. *Biotechnol. Bioeng.* **37**, 736–745 (1991).
19. Jones, K. L., Kim, S. W. & Keasling, J. D. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab. Eng.* **2**, 328–338 (2000).
20. Tyo, K. E., Ajikumar, P. K. & Stephanopoulos, G. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat. Biotechnol.* **27**, 760–765 (2009).
21. Bentley, W. E. & Quiroga, O. E. Investigation of subpopulation heterogeneity and plasmid stability in recombinant *Escherichia coli* via a simple segregated model. *Biotechnol. Bioeng.* **42**, 222–234 (1993).
22. Paulsson, J. & Ehrenberg, M. Noise in a minimal regulatory network: plasmid copy number control. *Q. Rev. Biophys.* **34**, 1–59 (2001).
23. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
24. Yu, D., Sawitzke, J. A., Ellis, H. & Court, D. L. Recombineering with overlapping single-stranded DNA oligonucleotides: testing a recombination intermediate. *Proc. Natl Acad. Sci. USA* **100**, 7207–7212 (2003).
25. Yu, D. *et al.* An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **97**, 5978–5983 (2000).
26. Haldimann, A. & Wanner, B. L. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.* **183**, 6384–6393 (2001).
27. Boyd, D., Weiss, D. S., Chen, J. C. & Beckwith, J. Towards single-copy gene expression systems making gene cloning physiologically relevant: lambda InCh, a simple *Escherichia coli* plasmid-chromosome shuttle system. *J. Bacteriol.* **182**, 842–847 (2000).
28. Kuhlman, T. E. & Cox, E. C. Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res.* **38**, e92 (2010).
29. Glaser, S., Anastassiadis, K. & Stewart, A. F. Current issues in mouse genome engineering. *Nat. Genet.* **37**, 1187–1193 (2005).
30. Louwerse, J. D. *et al.* Stable recombinase-mediated cassette exchange in *Arabidopsis* using *Agrobacterium tumefaciens*. *Plant. Physiol.* **145**, 1282–1293 (2007).
31. Cobb, R. E. & Zhao, H. Direct cloning of large genomic sequences. *Nat. Biotechnol.* **30**, 405–406 (2012).
32. Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70–74 (2010).
33. Santos, C. N. S., Koffas, M. & Stephanopoulos, G. Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab. Eng.* **13**, 392–400 (2011).
34. Lee, G. & Saito, I. Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene* **216**, 55–65 (1998).
35. Langer, S. J., Ghafoori, A. P., Byrd, M. & Leinwand, L. A genetic screen identifies novel non-compatible loxP sites. *Nucleic Acids Res.* **30**, 3067–3077 (2002).
36. Siegel, R. W., Jain, R. & Bradbury, A. Using an in vivo phagemid system to identify non-compatible loxP sequences. *FEBS Lett.* **499**, 147–153 (2001).
37. Wargacki, A. J. *et al.* An engineered microbial platform for direct biofuel production from brown macroalgae. *Science* **335**, 308–313 (2012).
38. Sousa, C., de Lorenzo, V. & Cebolla, A. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology*. **143**(Pt 6): 2071–2078 (1997).
39. Thomason, L. C., Costantino, N. & Court, D. L. *E. coli* genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.* **Chapter 1**, Unit 1 17 (2007).
40. Monaco, A. P. & Larin, Z. YACs, BACs, PACs and MACs: artificial chromosomes as research tools. *Trends Biotechnol.* **12**, 280–286 (1994).

41. Yoshikuni, Y., Wargacki, A. J. & Herman, A. Microbial systems for producing commodity chemicals. *US Patent 12/899,419* (2011).
42. Miller, J. H. *A short course in bacterial genetics: a laboratory manual and handbook for Escherichia coli and related bacteria* (Cold Spring Harbor Laboratory Press, Plainview, N.Y., 1992).

Acknowledgements

This work is supported by the DOE under Advanced Research Projects Agency–Energy (ARPA-E) award DE-AR0000006 and by the CORFO INNOVA CHILE (código 09CTEI-6866). Patents describing components of this work can be found under U.S. patent application nos. 12/245537, 12/361293, 12/899419, 61/427077 and 61/436173. This report was prepared as an account of work sponsored by an agency of the U.S. government. Neither the U.S. government nor any agency thereof, nor any of their employees, makes any warranty, express or implied or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer or otherwise does not necessarily constitute or imply its endorsement, recommendation or favoring by the U.S. government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. government or any agency thereof.

Author contributions

C.N.S.S. designed the experiments and performed all strain engineering and genetic manipulations. C.N.S.S. and D.R.R. characterized and evaluated strains. Y.Y. supervised all work. C.N.S.S. and Y.Y. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: Y.Y. is a co-inventor on a patent and a patent application entitled BIOFUEL PRODUCTION (US Patent 8198056, 8211689, 8318473, 8318464 and U.S. Patent Application 2009-0139134). All other authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Santos, C.N.S. *et al.* Implementation of stable and complex biological systems through recombinase-assisted genome engineering. *Nat. Commun.* 4:2503 doi: 10.1038/ncomms3503 (2013).