Title

Parallel Integration and Chromosomal Expansion of Metabolic Pathways

Journal

ACS Synthetic Biology, 7(11)

ISSN

2161-5063

Authors

Goyal, Garima

Costello, Zak

Alonso-Gutierrez, Jorge

et al.

Publication Date

2018-11-16

DOI

10.1021/acssynbio.8b00243

Peer reviewed

ACS SyntheticBiology
Research Article
Cite This: ACS Synth. Biol. XXXX, XXX, XXX−XXX
pubs.acs.org/synthbio

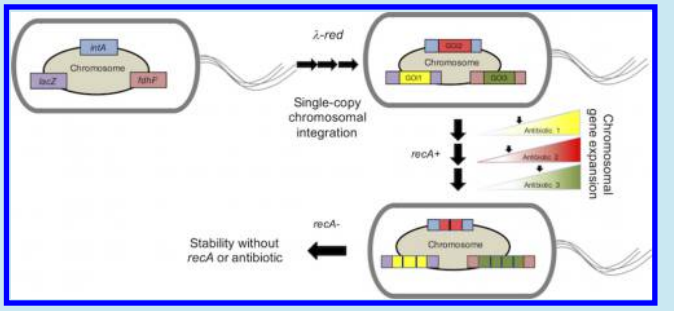
Parallel Integration and Chromosomal Expansion of Metabolic Pathways

Garima Goyal,[†,§,∥] Zak Costello,[‡,§,∥] Jorge Alonso-Gutierrez,[‡,∥] Aram Kang,[‡,∥] Taek Soon Lee,[‡,∥] Hector Garcia Martin,[‡,§,∥] and Nathan J. Hillson*[,†,§,∥]

[†]Technologies and [‡]Biofuels and Bioproducts Divisions, DOE Joint BioEnergy Institute, Emeryville, California 94608, United States
[§]DOE Agile BioFoundry, Emeryville, California 94608, United States
[∥]Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States

Supporting Information

ABSTRACT: Robust fermentation of biomass-derived sugars into bioproducts demands the reliable microbial expression of metabolic pathways. Plasmid-based expression systems may suffer from instability and result in highly variable titers, rates, and yields. An established mitigation approach, chemical induced chromosomal expansion (CIChE), expands a singly integrated pathway to plasmid-like copy numbers while maintaining stability in the absence of antibiotic selection pressure. Here, we report parallel integration and chromosomal expansion (PIACE), extensions to CIChE that enable independent expansions of pathway components across multiple loci, use suicide vectors to achieve high-efficiency site-specific integration of sequence-validated multigene components, and introduce a heat-curable plasmid to obviate *recA* deletion post pathway expansion. We applied PIACE to stabilize an isopentenol pathway across three loci in *E. coli* DH1 and then generate libraries of pathway component copy number variants to screen for improved titers. Polynomial regressor statistical modeling of the production screening data suggests that increasing copy numbers of all isopentenol pathway components would further improve titers.



KEYWORDS: *metabolic pathway, chromosomal integration, stabilization, copy number, statistical modeling, machine learning, optimization, isopentenol*

Proof-of-concept metabolic engineering and synthetic biology tool development efforts in *Escherichia coli* often favor replicating plasmid vectors over chromosomal genetic manipulations (see, for example, refs 1−4). This preference for plasmids is not surprising because it can be faster, easier, and less expensive to prototype metabolic pathways or genetic circuitry using replicating multicopy plasmids than through (generally single copy) chromosomal integration. However, replicating plasmid approaches can suffer from extensive genetic instability due to structural and allele distribution instability under high metabolic pressure.[5] Chromosomal integration is preferable (e.g., for industrial biomanufacturing economic and regulatory reasons) when seeking to minimize cell-to-cell variability and maintain genetic stability in the absence of antibiotic selection pressure.

There are multiple chromosomal integration methods available for *E. coli*,[6−8] although integrating large DNA fragments into desired chromosomal loci can be challenging, as recombination efficiency sharply drops with increasing DNA fragment size.[9,10] Representative approaches include phage-mediated site-specific recombination via conditional-replication integration and modular (CRIM) plasmids,[11,12] transposon-mediated gene transposition,[13,14] λ-red-mediated homologous recombination,[7,8,15−17] and knock-in/knockout vectors.[18]

Only a subset of these methods, however, enable the rapid and controllable chromosomal integration of multiple gene copies (toward parity with what replicating plasmids can provide). Examples within this subset include CRIM (see above) as well as chemically inducible chromosomal evolution (CIChE)[19] and chromosomal integration of genes with multiple copies (CIGMC).[20] CIGMC uses the FLP recombinase to achieve variable (colony-to-colony) numbers of chromosomal insertions in a single step. While CIGMC has been shown to achieve up to 15 integrated copies of a gene, average copy numbers were in the range of 2−4, with higher copy numbers occurring at very low frequencies. In contrast, CIChE monotonically controls gene copy number through antibiotic titration and has been shown to achieve up to 60 gene copies, which is comparable to multicopy plasmids.

Of these established methods, we chose the CIChE approach (as a good trade-off between process control and labor/time requirements) for our efforts to genetically stabilize

Received: June 8, 2018
Published: October 23, 2018

DOI: 10.1021/acssynbio.8b00243
ACS Synth. Biol. XXXX, XXX, XXX−XXX

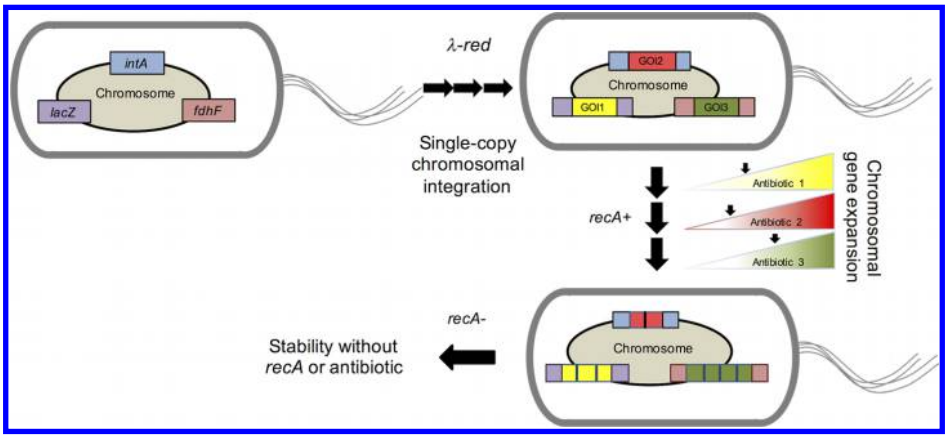


Figure 1. PIACE schematic. Three gene cassettes of interest (GOI1, GOI2, and GOI3) are sequentially integrated (one copy each) at corresponding *E. coli* genomic loci (*lacZ*, *intA*, and *fdhF*) via λ-red-mediated suicide-vector homologous recombination. The introduction of *recA* via plasmid transformation facilitates the orthogonal copy number expansion of each gene cassette via sequential antibiotic (gentamycin, chloramphenicol, and spectinomycin) titrations that tunably select for each gene cassette-associated marker (GentR, CmR, and SpecR). The removal of *recA* via temperature-sensitive plasmid-curing post chromosomal expansion promotes gene cassette copy number stability in the absence of continued antibiotic selection pressure.

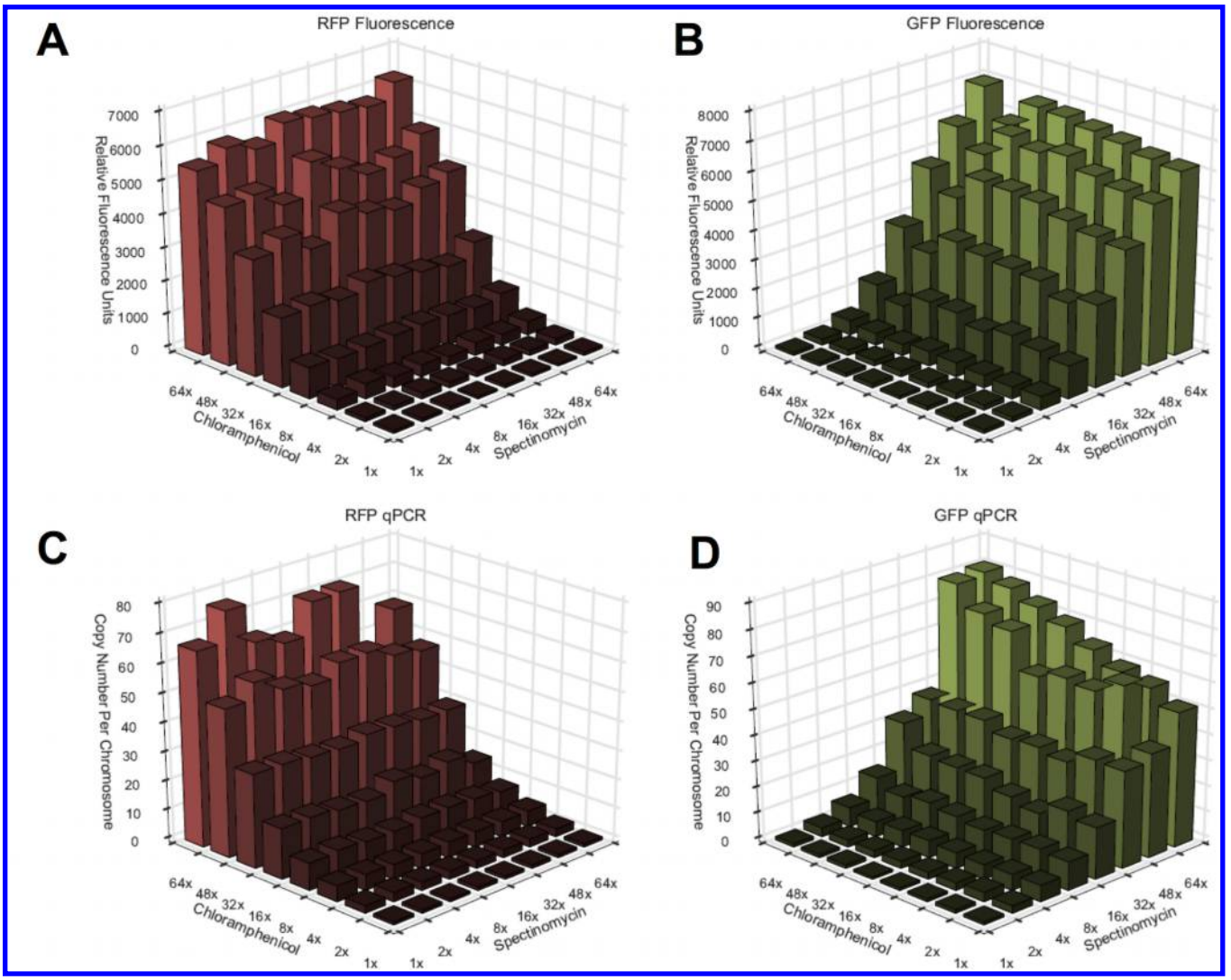


Figure 2. Pilot experiment expanding *rfp* (CmR) and *gfp* (SpecR) gene cassettes at two loci (*intA* and *fdhF*). An 8 × 8 matrix of chloramphenicol and spectinomycin titration conditions results in different extents of *rfp* and *gfp* expansion, which were quantitatively assessed by fluorescence and qPCR measurements. (A) RFP fluorescence. (B) GFP fluorescence. (C) *rfp* qPCR. (D) *gfp* qPCR.

DOI: 10.1021/acssynbio.8b00243
ACS Synth. Biol. XXXX, XXX, XXX−XXX

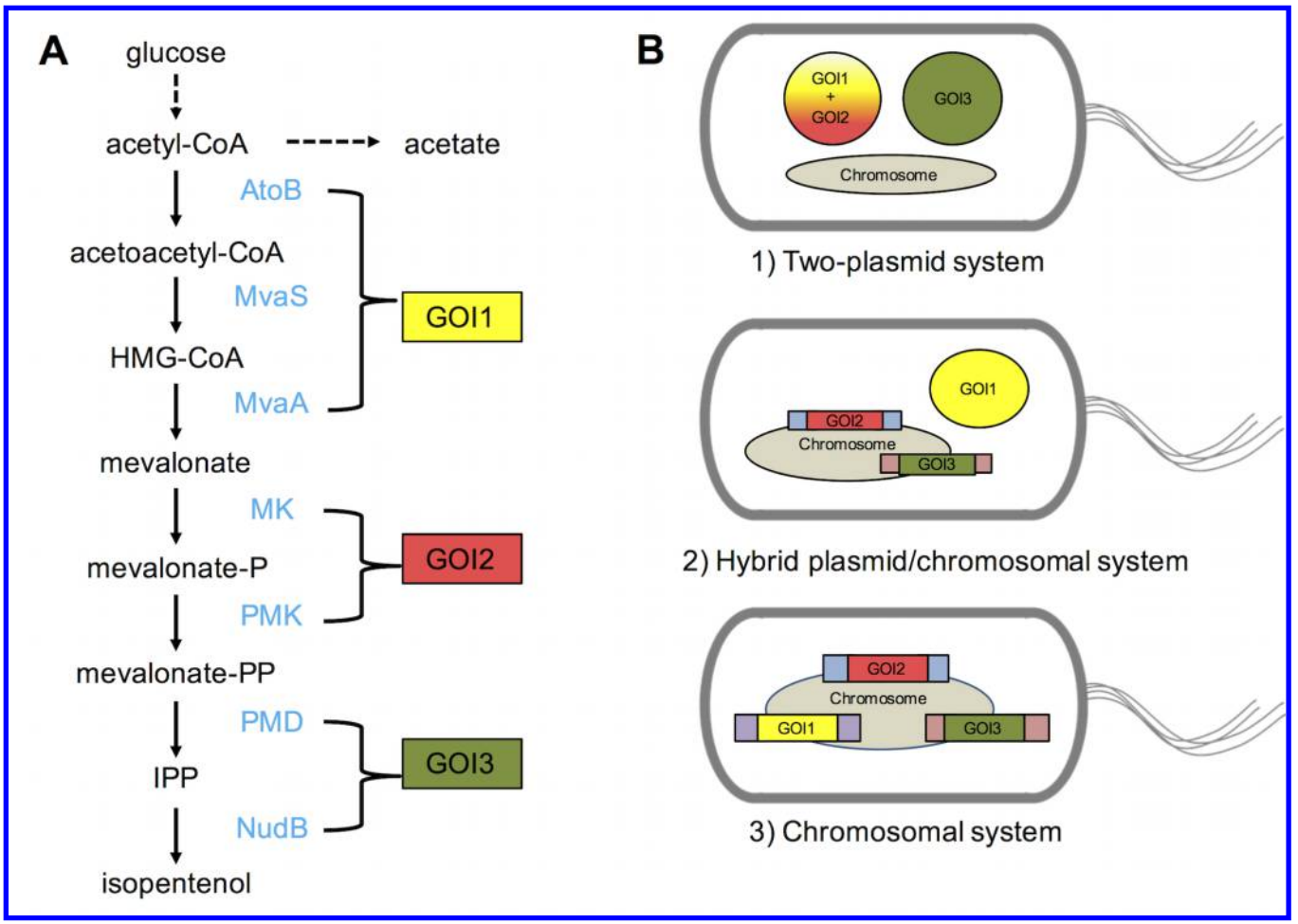


Figure 3. Schematic of three configurations evaluated for expressing a metabolic pathway to isopentenol. (A) An implementation of a glucose to isopentenol pathway that includes the enzymes AtoB, MvaS, and MvaA (herein collectively referred to GOI1); MK and PMK (GOI2); and PMD and NudB (GOI3). (B) Schematic of three expression configurations: (1) two-plasmid system with GOI1+GO2 and GOI3 split between p15A- and pBR322-derivative medium-copy plasmids, respectively; (2) hybrid plasmid/chromosomal integration system with GOI1 residing in a p15A-derivative medium copy plasmid and GOI2 and GOI3 integrated into the *intA* and *fdhF* loci, respectively; and (3) chromosomal integration system with GOI1, GOI2, and GOI3 integrated into the *intA*, *fdhF*, and *lacZ* loci, respectively.

a multicopy plasmid-borne isopentenol metabolic pathway.[21−24] During the process of applying CIChE to the isopentenol pathway, we identified several opportunities for further extending the methodology. One such opportunity that still remains to be pursued, for example, would be to take inspiration from Chen et al.[25] and use a recombinase to remove superfluous (e.g., vector backbone components) sequences prior to the CIChE expansion process. Here, we report our development of these extensions to CIChE, namely parallel integration and chromosomal expansion (PIACE), and our application of PIACE to the chromosomal stabilization and titer optimization of isopentenol production in *E. coli* DH1.

RESULTS AND DISCUSSION

PIACE Strategy. The PIACE strategy is summarized in Figure 1. The approach begins with a *recA*-negative base strain of *E. coli* (e.g., *E. coli* DH1). This base strain is transformed with the heat-curable helper plasmid pKD46 (Supplementary Figure S1A), which provides arabinose-inducible λ-red homologous recombination functionality. Gene cassettes GOI1, GOI2, and GOI3 are sequentially integrated (one copy each) at corresponding *E. coli* genomic loci (*lacZ*, *intA*, and *fdhF*) via λ-red-mediated suicide-vector (R6K origin) homologous (1 kb targeting sequences) recombination. The introduction of the heat-curable helper plasmid pRedi2RecA (Supplementary Figure S1B) provides rhamnose-inducible *recA* homologous recombination functionality. Under *recA* expression conditions, the copy numbers of gene cassettes GOI1, GOI2, and GOI3 can be orthogonally and tunably expanded (via the previously established CIChE mechanism) by sequential antibiotic (gentamycin, chloramphenicol, and spectinomycin, respectively) titration. Heat-curing pRedi2RecA post expansion promotes chromosomal gene cassette copy number stability in the absence of continued antibiotic selection pressure.

***rfp* and *gfp* Chromosomal Integration and Expansion.** As a preliminary PIACE proof of concept, *rfp* and *gfp* were selected as genes of interest (for slots GOI2 and GOI3 in Figure 1). Following the approach shown in Figure 1, suicide vectors pPIACE_Cm_RFP and pPIACE_Sp_GFP (Supplementary Figures S1C and D) were sequentially integrated at *intA* and *fdhF* loci, respectively, and integration at these loci was confirmed by colony PCR. This resulting strain (1R1G_pRedi2RecA_DH1, Supplementary Table S1), containing one copy each of *rfp* and *gfp*, was then exposed (in the presence of *recA*) to stepwise parallel increases in chloram-

DOI: 10.1021/acssynbio.8b00243
ACS Synth. Biol. XXXX, XXX, XXX−XXX

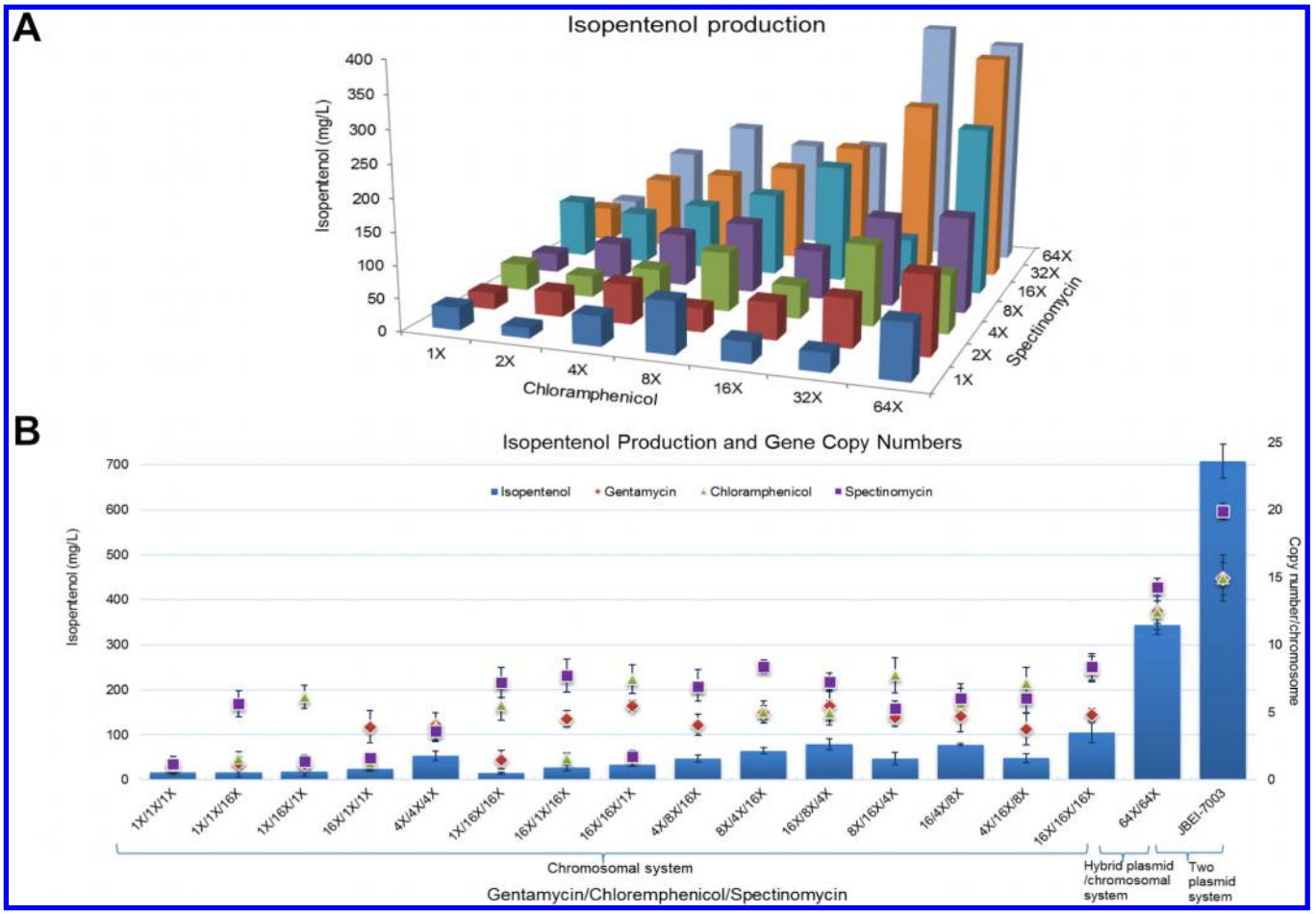


Figure 4. Isopentenol production and gene copy numbers for the three expression configurations. (A) For the hybrid plasmid/chromosomal configuration, isopentenol production for a 7 × 7 matrix of chloramphenicol and spectinomycin titration conditions. (B) For the chromosomal configuration, 15 strains subsampled from a 5 × 5 × 5 matrix of gentamycin, chloramphenicol, and spectinomycin titration conditions compared with the 64X/64X hybrid plasmid/chromosomal and two-plasmid configurations. Copy numbers of GOI1 (AtoB, MvaS, and MvaA; GentR), GOI2 (PMK and MK; CmR), and GOI3 (PMD and NudB; SpecR) were assessed by qPCR.

phenicol and spectinomycin titers to an 8 × 8 matrix of final antibiotic concentrations (see Methods). We have not yet systematically tested different expansion strategies (in terms of serial or parallel sequencing and stepwise increments of antibiotic titrations), which could be done (not only for *rfp* and *gfp* here, but also for GOI1−3, below). For each of these 64 resulting strains (1R1G−8R8G), Figures 2A and B show the resulting RFP and GFP relative fluorescence units (microplate reader), and Figures 2C and D show the *rfp* and *gfp* copy numbers (qPCR). For the most part, RFP and GFP relative fluorescence units along with *rfp* and *gfp* copy numbers increase monotonically (and orthogonally) with chloramphenicol and spectinomycin titers, respectively. Up to 66 and 80 copies of *rfp* and *gfp*, respectively, were obtained in strain 8R8G. Two individual colonies were picked post expansion for each expansion condition, and the presented data in Figure 2 consist of the average (of the two colonies) values for fluorescence and copy numbers. A given colony's RFP or GFP copy number varied on average by roughly 15% or less from the average value, about 4−5 copies. A given colony's fluorescence values varied on average by roughly 5% from the average value, about a 140 (RFP) or 110 (GFP) relative units difference between the two colonies.

Stability of *rfp* and *gfp* Copy Numbers Post Expansion. We then assayed the stability of a subset of these *rfp* and *gfp* gene expansions post *recA* removal (via heat-curing pRedi2RecA, as in Figure 1) in the absence of antibiotics. We began with 10 mL starting cultures (with antibiotics) of three strains (low-copy *rfp/gfp* (1R1G−1X/1X), medium-copy *rfp/gfp* (4R4G−8X/8X), and high-copy *rfp/gfp* (8R8G−64X/64X)) and then performed three cycles of 1000-fold (serial) dilution into fresh LB media (lacking antibiotics) followed by growth overnight at 37 °C. For each of the three initial strains, RFP and GFP relative fluorescence units along with *rfp* and *gfp* copy numbers were measured for the starting cultures and after each cycle of dilution and outgrowth (Supplementary Figure S2). Even the least stable (highest *rfp/gfp* copy) initial strain (8R8G−64X/64X) retained more than 71 and 75% of its original copies of *rfp* and *gfp*, respectively, despite a billion-fold (volumetric) outgrowth in the absence of antibiotics.

Isopentenol Pathway Integration. Following the *rfp/gfp* proof of concept experiments, we applied PIACE to a glucose to isopentenol metabolic pathway. We divided this isopentenol pathway into three segments (Figure 3A), namely the enzymatic steps carried out by AtoB, MvaS, and MvaA (collectively referred to GOI1); MK and PMK (GOI2); and PMD and NudB (GOI3). We compared isopentenol titers and gene copy numbers for three isopentenol pathway expression configurations (Figure 3B): (1) a two-plasmid system, (2) a

DOI: 10.1021/acssynbio.8b00243
ACS Synth. Biol. XXXX, XXX, XXX−XXX

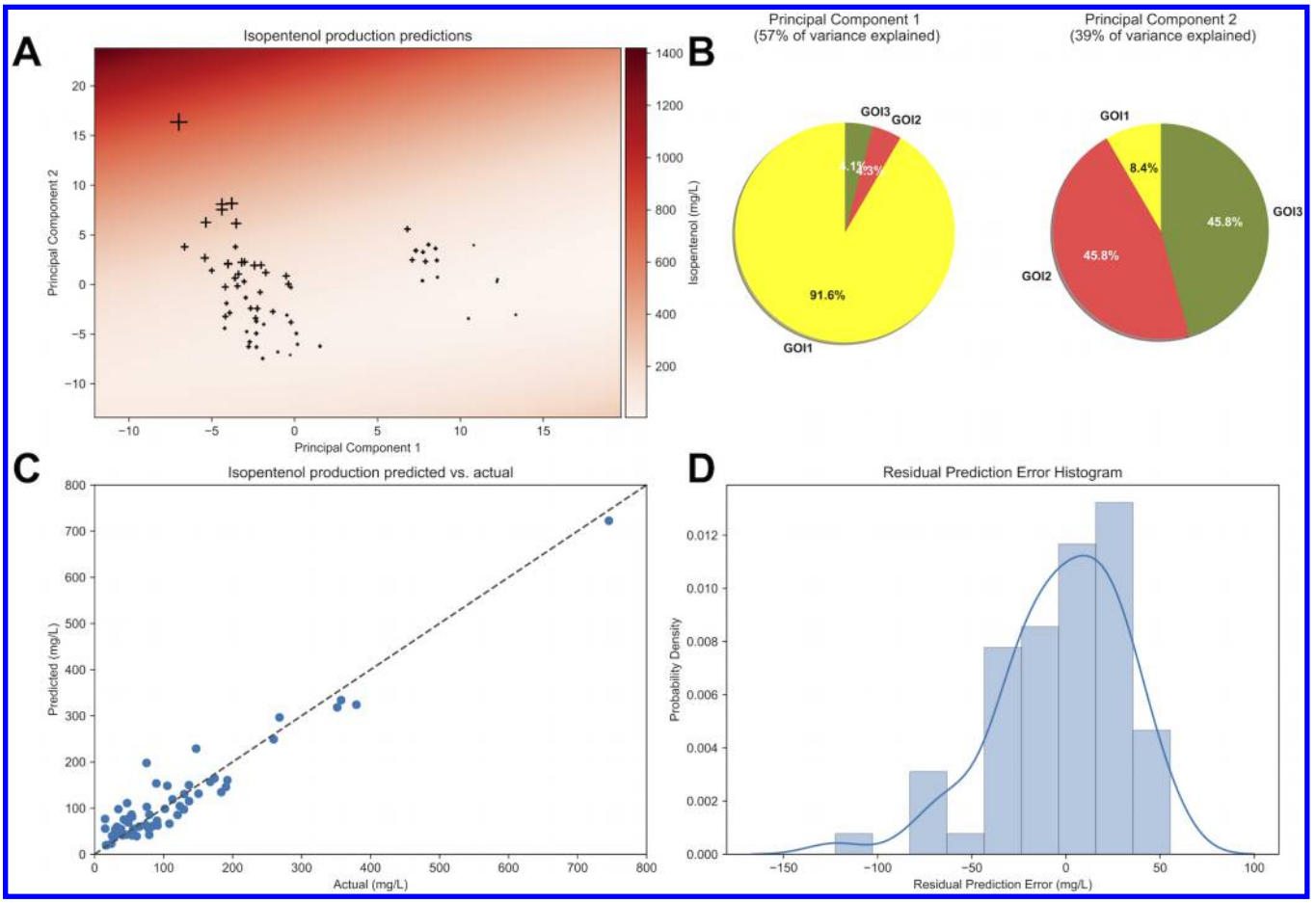


Figure 5. Polynomial regressor modeling of isopentenol production data. (A) Predicted isopentenol production (darker red indicates higher isopentenol titer) plotted against the first two principal components of gene copy number measurements. Cross-hatch marks (+) correspond to isopentenol/copy number measurements with larger cross-hatch marks indicating higher isopentenol titers. The gradient toward the darkest red region at upper left corresponds to increasing copy numbers of all isopentenol pathway genes. (B) Decomposition of the first two principal components of gene copy number measurements. (C) Cross-validated model predictions plotted against actual measurements. Dashed line coincides with flawless predictions. (D) The residuals of the prediction errors in panel C are illustrated in a histogram which shows that, within one standard deviation, all points have less than 34.1 mg/L isopentenol prediction error.

hybrid plasmid/chromosomal system, and (3) a chromosomal system.

The two-plasmid system is a previously published strain (JBEI-7003)[24] that contains two plasmids, namely (1) pMevTsa-PMKco-MKco, which has a p15A origin and contains GOI1 and GOI2 under the inducible expression control of a lacUV5 promoter, and (2) pTrc99A-NudB-PMD, which has a pBR322 origin and contains GOI3 under the inducible expression control of a pTrc promoter (Supplemental Figure S3).

The foundation of the hybrid plasmid/chromosomal system is an *E. coli* DH1 strain (pMevTop_Kan_DH1) that contains plasmid pMev_AtoB-mvaS-mvaA (Supplementary Figure S1E), which has a p15A origin and contains GOI1 under the control of a lacUV5 promoter. Into this strain, pMevTop_Kan_DH1, GOI2, and GOI3 are sequentially integrated into the *intA* and *fdhF* chromosomal loci (via suicide vectors pPIACE_Cm_PMK-MK and pPIACE_Sp_NudB-PMD, Supplementary Figures S1F and G), respectively yielding (post heat-curing pKD46 and introducing pRedi2RecA) the hybrid plasmid/chromosomal system base strain 1p1N_pRedi2RecA_DH1, which is then competent (under *recA* expression) for orthogonal GOI2 and GOI3 copy number expansions. It could be noted that we could have alternatively selected GOI2 or GOI3 to retain in the plasmid instead of GOI1. In a manner analogous to that described above for a matrix of *rfp*/*gfp* expansions, we sequentially expanded GOI2 and then GOI3 in a 7 × 7 matrix of chloramphenicol and spectinomycin titrations. These resulting 49 strains were then heat-cured of pRedi2RecA (resulting in strains 1P1N through 7P7N).

The chromosomal system is *E. coli* DH1, into which GOI1, GOI2, and GOI3 are sequentially integrated into the *lacZ*, *intA*, and *fdhF* chromosomal loci (via suicide vectors pPIACE_Gent_AtoB-mvaS-mvaA, pPIACE_Cm_PMK-MK, and pPIACE_Sp_NudB-PMD; Supplementary Figures S1F–H), respectively yielding (post heat-curing pKD46 and introducing pRedi2RecA) the chromosomal system base strain 1A1P1N_pRedi2RecA_DH1, which is then competent (under *recA* expression) for orthogonal GOI1, GOI2, and GOI3 copy number expansions. In a manner similar to that described immediately above for the hybrid plasmid/chromosomal system, we sequentially expanded GOI1, and then GOI2, and finally GOI3 in a 5 × 5 × 5 matrix of gentamycin, chloramphenicol, and spectinomycin titrations. A subset of 15 of these 125 total strains was selected for further

DOI: 10.1021/acssynbio.8b00243
ACS Synth. Biol. XXXX, XXX, XXX−XXX

characterization (post heat-curing pRedi2RecA, resulting in strains 1A1P1N through 5A5P5N).

The three pathway expression configurations, while not appearing to differ in their cellular growth phenotypes (all achieving similar ODs at the stationary phase time point in which isopentenol production was assessed), produce different isopentenol titers (Figure 4) and consist of different copy numbers of GOI1, GOI2, and GOI3 (Figure 4B, Supplemental Figure S4). The maximum isopentenol titers achieved for the three expression configurations were: (1) two-plasmid system (strain JBEI-7003), 708 mg/L; (2) hybrid plasmid/chromosomal system, 344 mg/L (strain 7P7N); and (3) chromosomal system, 105 mg/L (strain 5A5P5N). The maximum copy numbers we achieved for GOI1, GOI2, and GOI3 across the three different configurations were: (1) two-plasmid system: 15 GOI1, 15 GOI2, and 20 GOI3 (strain JBEI-7003); (2) hybrid plasmid/chromosomal system: 15 GOI1, 13 GOI2, and 12 GOI3 (strain 7P7N); and (3) chromosomal system: 5 GOI1, 8 GOI2, and 8 GOI3 (strain 5A5P5N). While error bars (standard deviation of three replicate colonies) are presented in Figure 4B, for the hybrid plasmid/chromosomal systems (with the exception of strain 7P7N, all those strains for which data are shown in Figures 4A and S4), two individual colonies were picked post expansion for each expansion condition, and the presented data consist of the average (of the two colonies) values for copy number and isopentenol production. In terms of copy number variation for the hybrid plasmid/chromosomal systems, a given colony's GOI1−3 copy numbers varied on average by roughly 15% or less from the average value, about 1−2 copy difference between the two colonies for GOI1−3. For the hybrid plasmid/chromosomal system, a given colony's isopentenol production comparably varied on average by roughly 15% from the average value, about a 30 mg/L difference between the two colonies.

Machine Learning for the Prediction of Isopentenol Titers from Gene Copy Numbers. Because there are no available mechanistic models to predict how GOI1−3 copy numbers will affect isopentenol titers, we opted for a statistical approach. The performance of three (namely support vector regressor, random forest regressor, and polynomial regressor) machine learning models was assessed for the prediction of isopentenol titers, given GOI1−3 copy numbers. Each model's fit, cross validation, and error residuals are shown in Figure 5 and in Supplemental Figure S5. To evaluate the extents of putative model overfitting, one data point (strain 7P7N) was withheld during model development. The root-mean-square residual prediction errors of the models (lower is better) are 80.5, 58.9, and 34.1 mg/L isopentenol for the support vector regressor, random forest regressor, and polynomial regressor models, respectively. While these models are qualitatively similar (each suggests that to increase isopentenol titers, GOI1−3 copy numbers should all be increased beyond that present in the training set), the polynomial regressor (Figure 4) significantly outperforms the others (Supplemental Figure S5) in terms of its prediction accuracy. For strain 7P7N (the data point withheld during model development), which achieved 344 mg/L isopentenol, the support vector regressor and random forest regressor do not perform very well, predicting 211.5 and 276.0 mg/L of production, respectively. In comparison, the polynomial regressor has a more accurate prediction of 292.4 mg/L isopentenol.

Stability of Isopentenol Titers and GOI1−3 Copy Numbers across Expression Configurations. A stability assay was performed to test the stability of isopentenol titers and GOI1−3 copy numbers across the two-plasmid (strain JBEI-7003), hybrid plasmid/chromosomal (strains 1P1N, 4P4N, and 7P7N), and chromosomal (strains 1A1P1N, 3A3P3N, and 5A5P5N) pathway expression configurations. Analogous to that performed above to assess the stability of *rfp*/*gfp* strains, these isopentenol-producing strains were repetitively diluted in fresh LB media without antibiotics 1000-fold and grown overnight for three cycles total. The resulting progression of isopentenol titers and GOI1−3 copy numbers are shown in Figure 6. The relative isopentenol titer and GOI1−3 copy number stability of the chromosomal expression configuration outperforms the hybrid plasmid/chromosomal and two-plasmid systems, although the starting isopentenol titer and GOI1−3 copy numbers are lowest for the chromosomal system.

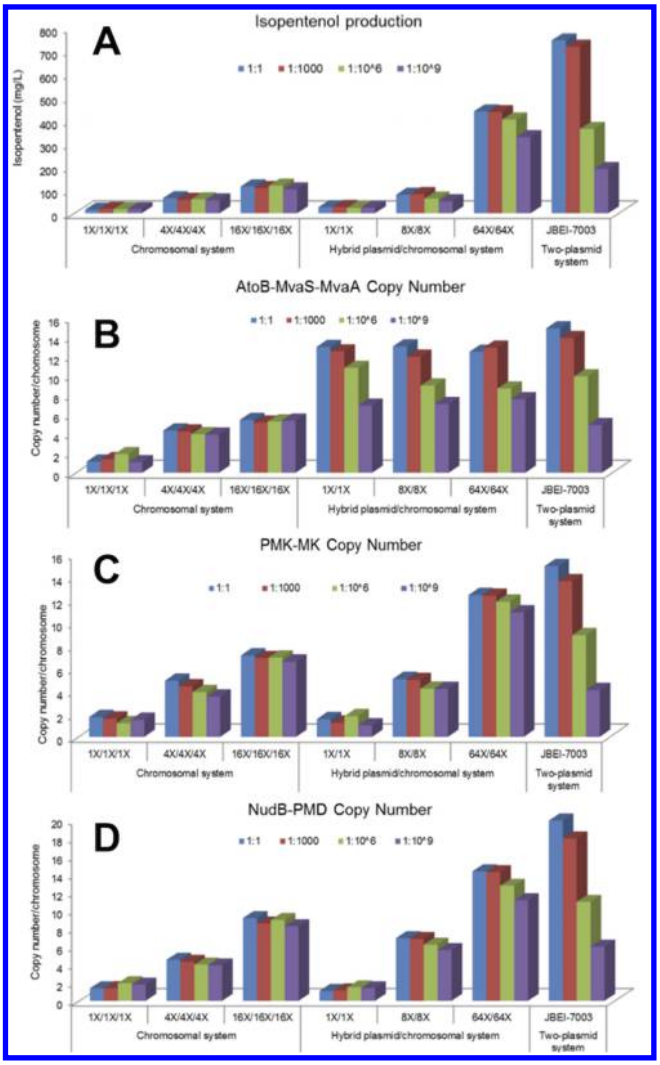


Figure 6. Comparison of isopentenol production and gene copy number stability across three variations of the chromosomal configuration, three variations of the hybrid plasmid/chromosomal configuration, and the two-plasmid configuration. Ten milliliter cultures were repetitively (three cycles total) diluted 1000-fold in fresh LB media (lacking antibiotics) and grown overnight at 37 °C. (A) Isopentenol production. (B) GOI1 (AtoB, MvaS, and MvaA), (C) GOI2 (PMK and MK), and (D) GOI3 (PMD and NudB) copy numbers, as assessed by qPCR.

DOI: 10.1021/acssynbio.8b00243
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Discussion. In this study, we developed an approach for the PIACE of metabolic pathways (Figure 1), which extends the CIChE method[19] in several key ways. Following proof-of-concept experiments with *gfp*/*rfp* (Figure 2), we applied PIACE to the integration of genes encoding the mevalonate metabolic pathway to isopentenol, divided across three gene cassettes, into three loci within the *E. coli* DH1 chromosome (Figure 3). Following gene cassette copy number expansion (under *recA* expression and tuned antibiotic concentration conditions), we achieved a three-dimensional grid sampling of GOI1, GOI2, and GOI3 cassette copy number variants, assessed these variants for isopentenol production (Figure 4), and applied statistical modeling to make predictions as to which additional gene cassette copy number variants should be pursued in the future to further improve isopentenol titers (Figure 5). Finally, we assessed the stabilities of isopentenol production and gene cassette copy numbers in the absence of antibiotic pressure, for the fully chromosomal configuration in comparison with hybrid chromosomal/plasmid and two-plasmid configurations (Figure 6). In this study, we exclusively focused on increasing isopentenol production through gene copy number variation. We did not, for example, explore alternative promoters, RBSs, enzyme orthologs, operon structures (e.g., gene order, genes per operon, genes in same operon), or terminators. These additional engineering approaches are complementary and could be combined with PIACE in the future.

We sought to extend CIChE to overcome several of its limitations, namely (1) low integration efficiencies for multigene cassettes, (2) that only a single locus per strain was used for gene integration, and (3) the requirement to delete *recA* post gene cassette copy number expansion. To increase the integration efficiency of multigene cassettes of interest rather than introduce a gene cassette through the transformation of linear DNA, we opted to use supercoiled suicide plasmid vectors. These suicide vectors contain a R6K replicon, which can only replicate in strains expressing the π protein[26,27] (such as *E. coli* PIR1 but not *E. coli* DH1). This approach significantly increased integration efficiencies and our ability to integrate multigene cassettes (data not shown). While the original Tyo et al. CIChE publication[19] does speculate that other integration methods would also suffice, our comparison here was made with respect to the integration method actually used in Tyo et al. Three different loci (*lacZ*, *intA*, and *fdhF*) in *E. coli* DH1 were selected for integrating GOI1, GOI2, and GOI3, respectively, all into the same strain. Dividing a metabolic pathway across multiple loci enables each portion (e.g., GOI1, GOI2, and GOI3) to have its copy number expanded independently of the others, providing more fine-grained control for optimizing pathway performance, as compared with integration at a single locus, in which the copy numbers of all portions of the pathway must be coarsely expanded together in lock-step. Rather than starting with a *recA* positive strain of *E. coli* and having to delete *recA* post gene expansion, we opted to start with a *recA* negative strain (*E. coli* DH1) and transiently induce *recA* expression from a heat-sensitive plasmid (pSC101 origin of replication) during the expansion process and then heat-cure this helper plasmid post gene expansion. This methodological distinction from CIChE is especially important when generating libraries of copy number variants because introducing the helper plasmid (and for *recA* positive *E. coli* backgrounds, deleting *recA*) once to the base strain prior to expansion and then heat-curing the plasmid from each derivative strain variant post expansion is far faster and requires less effort than deleting *recA* from each individual expanded strain variant. Phage transduction could possibly be an equally quick and easy method for removing *recA* (from the genome) post CIChE expansion, although this would necessitate the use of an additional selectable marker with which to indicate and select for positive transductants of the *recA* deletion.

The PIACE method continues to offer the best features of CIChE, such as enhanced pathway stability (even in the absence of selection pressure) and titratable pathway copy numbers. As reported in the CIChE publication[19] and supported here in Figure 6, chromosomally integrated pathways, in comparison with plasmid-borne metabolic pathways (especially for those pathways invoking high metabolic demands in the absence of antibiotics), retain pathway copy numbers better, which results in more sustained levels of target metabolite production over time. While it may be easier to initially obtain high gene copy numbers through the use of multicopy plasmid systems, it has been shown that a six-copy chromosomal gene integration produced more L-serine than the same gene harbored in medium copy plasmid.[20] Besides copy number stability, the ordered segregation (following DNA replication) of multicopy chromosomal integrations makes it more difficult for pathway-attenuating mutations to take over a cellular population, compared with the random segregation that is characteristic of some multicopy plasmid systems.[19] These stability advantages of chromosomally integrated vs plasmid-borne pathways are of key importance in the context of industrial metabolite production, in which cost (and regulatory) drivers lead to large-scale (many cellular generations) and antibiotic-free fermentations. Beyond stability, the ability to fine-tune copy number (through selective pressure titration) offers advantages over some other gene integration methods, such as CIGMC,[20] that result in variable (colony-to-colony) copy numbers. While methods like PIACE and CIChE require more steps than CIGMC to integrate multiple gene copies into the chromosome, they do not require additional molecular biology (beyond the creation of the base strain) or colony screening to generate strain variant libraries with monotonically increasing gene copy numbers.

There are at least three major opportunities for further improving PIACE, namely (1) detune selectable marker expression to boost maximal achievable copy numbers, (2) develop alternative selectable markers, and (3) use a recombinase to remove superfluous (e.g., vector backbone components) sequences prior to the expansion process.[25] As shown in Figure 5, statistical modeling suggests that increasing GOI1−3 copy numbers would result in higher isopentenol titers (despite the potential concern that the coupled expansion of superfluous vector backbone component sequences would adversely affect performance). (As an aside, we expect that, as higher isopentenol titers are achieved, the interplay of GOI1−3 copy numbers with host metabolism will become more complex, and the relationship will still be accessible through this statistical technique.) As shown in Figure 4, however, we were unable to achieve more than 10−15 copies of GOI1−3. This is in contrast with *gfp*/*rfp* that could be expanded to about 80 copies (Figure 2). Because there are roughly the same maximal copy numbers of each GOI (15 GOI1, 12 GOI2, and 13 GOI3) in the maximally expanded hybrid expression configuration strain 7P7N, this

DOI: 10.1021/acssynbio.8b00243
ACS Synth. Biol. XXXX, XXX, XXX−XXX

places some doubt that having selected GOI2 or GOI3 rather than GOI1 as the plasmid-borne GOI would have enabled access to a significantly more diverse space of gene copy numbers. To attain higher copy numbers for larger gene cassettes (e.g., GOI1−3), we speculate that it may be necessary/possible to detune (i.e., reduce the strength of) the promoters/RBSs governing the expression of the selectable marker (e.g., the antibiotic resistance gene) to decrease the effective amount of selectable activity per copy and thus increase the total number of gene copies required to withstand a given amount of selective pressure. Second, our PIACE developments to date have exclusively used antibiotic resistance genes for selectable markers. It would be of great utility to develop alternative titratable selectable markers, such as auxotrophic markers, or perhaps even screenable fluorescent markers, so that the resulting strains do not bear antibiotic resistance genes, which may prove problematic in industrial contexts from an economic or regulatory standpoint. It would additionally be possible to use *fabI*/triclosan as one of the selectable marker/pressure pairs;[25] however, Because PIACE requires multiple orthogonal markers, using *fabI* would only partially address the replacement of antibiotic resistance markers. We speculate that selectively eliminating antibiotic resistance genes post expansion (e.g., flipping them out via a recombinase) could be possible but that this would need to be done in such as manner as not to eliminate/regress the expanded gene copies of the pathway of interest at the same time.

Beyond the expression of *gfp*/*rfp* or the isopentenol metabolic pathway, PIACE could be applied to other metabolic or regulatory pathways of interest and in organisms besides *E. coli*. Whereas applying PIACE to another pathway is straightforward (e.g., swap out GOI1−3 contents), implementing PIACE in another organism may prove challenging. An important PIACE prerequisite (shared in common with CIChE) is switchable (e.g., conditionally inducible) homologous recombination activity, given the reversible nature of its mechanism for expanding gene copy number. Recombinase activity should be conditionally switched on to enable selectable gene duplication/expansion but should then (post expansion) be switched off to prevent regressive copy number elimination/reduction (at least in the absence of selective pressure). It would be a technical challenge, then, to implement PIACE in organisms without endogenous (or heterologously expressed) recombinase activity or (in the absence of selective pressure) in organisms with chronic endogenous recombinase activity that cannot be switched off. The latter issue may not pose too great an obstacle in the case of persistent auxotrophic selection pressure, which could be cost- and regulatory-effective (in contrast with antibiotic selection pressure) even at industrial scales. We look forward to the application of PIACE (or an improved derivative thereof) to other pathways and to its implementation in organisms beyond *E. coli*.

METHODS

Materials. For PCR, Q5 high-fidelity 2X master mix and Taq polymerase from New England Biolabs (Ipswich, MA) were used. For qPCR, SYBR green master mix was purchased from Bio-Rad (Hercules, CA). All DNA oligos were purchased from Integrated DNA Technologies (Coralville, IA). DNA assembly reagents were purchased from New England Biolabs. Unless otherwise specified, all chemicals, solvents, and media components were purchased and used without modification from Sigma-Aldrich (St. Louis, MO), Fisher Scientific (Pittsburgh, PA), or VWR (West Chester, PA). *E. coli* strains DH10B and PIR1 (Invitrogen, Carlsbad, CA) were used for plasmid construction, and DH1 (ATCC) was used for homologous recombination and production experiments. Miniprep and gel extraction kits were purchased from Qiagen (Hilden, Germany). Genomic DNA extraction kit was purchased from Promega (Madison, WI). Echo 550 liquid handler supplies were purchased from Labcyte (San Jose, CA). Supplies and reagents for the Nimbus gel extraction robot were purchased from Coastal Genomics (Burnaby, British Columbia, Canada).

DNA Design/Visualization/Protocol Generation Software. DNA software tools, including DeviceEditor, j5, ICE, and VectorEditor were used for designing and visualizing DNA constructs and for designing the DNA assembly protocols used for their construction.[28−30]

Plasmid and Strain Construction and Availability. *E. coli* DH10B was used as the cloning host for the plasmids pRedi2RecA and pMevTop_AtoB_mvaS_mvaA. *E. coli* PIR1 was used as the cloning host for all the suicide vectors constructed for homologous recombination. DNA oligos and templates, PCR conditions, and the DNA assembly method type used to construct each plasmid are listed in Supplementary Table S2. DNA fragments were amplified using Q5 hot-start high-fidelity 2X master mix (NEB): 50 μL PCR reactions consisted of 5 μL (5 μM) of each forward and reverse primer, 10 μL of template (5 ng/μL), 25 μL of Q5 high fidelity 2X master mix, and 5 μL of deionized water. The following touchdown PCR thermocycling conditions were used: 98 °C for 30 s, then 10 cycles of 98 °C for 10 s, annealing at specified temperature for 30 s with a decrease in annealing temperature of 0.5 °C per cycle, 72 °C for 20 s/kb; then 25 cycles of 98 °C for 10 s, annealing at the specified temperature for 30 s, 72 °C for 20 s/kb; and final extension at 72 °C for 2 min. Following PCR amplification, residual (methylated) DNA template in each PCR reaction was DpnI-digested and purified using a Qiagen gel extraction kit or Nimbus Size Selection, as specified. Gibson and CPEC assembly methods were used for constructing the plasmids as described elsewhere.[31,32] All suicide vectors have a R6K origin of replication, which can be propagated only in a PIR1 strain background, as well as a 1 kb homologous sequence targeting their respective locus for chromosomal integration. Chloramphenicol (pPIACE_Cm_RFP and pPIACE_Cm_PMK-MK), spectinomycin (pPIACE_Cm_GFP and pPIACE_Sp_NudB-PMD), and gentamycin (pPIACE_Gent_AtoB-mvaS-mvaA) resistance genes were chosen as selectable markers due to their ability to titrate antibiotic resistance roughly linearly (as opposed to an all-or-nothing threshold) with copy number expansion. Homologous recombination was performed in freshly prepared *E. coli* DH1 electrocompetent cells, as described elsewhere.[18] Transformations were otherwise performed in chemically competent cells.[33] Integration of each gene cassette was confirmed by colony PCR and Sanger sequencing (of its colony PCR product, using the corresponding colony PCR primers). Twenty microliters of colony PCR reaction consisted of 1 μL (5 μM) of each forward and reverse primer, 1 μL of template (boiled cell culture or colony in water), 0.1 μL of Taq DNA polymerase (NEB), 0.4 μL of dNTPs, 2 μL of Taq polymerase buffer, and 14.5 μL of deionized water. The following thermocycler conditions were used: 95 °C for 30 s,

DOI: 10.1021/acssynbio.8b00243
ACS Synth. Biol. XXXX, XXX, XXX−XXX

and then 30 cycles of 95 °C for 10 s, annealing at 60 °C for 30 s, 72 °C for 1 min/kb, and final extension at 72 °C for 5 min. The complete list of DNA oligos used for colony PCR is provided in Supplementary Table S3. Strains and plasmids developed for this study (Supplementary Table S1), along with annotated DNA sequences were deposited in the public instance of the Joint BioEnergy Institute Registry (https://public-registry.jbei.org/folders/354) and are physically available from the authors and/or addgene (https://www.addgene.org) upon reasonable request.

Chromosomal Integration, Gene Copy Number Expansion, and Curation of Helper Plasmids. Suicide vectors were integrated at their respective chromosomal loci using helper plasmid pKD46.[18] Strains containing pKD46 were cultured from a single colony in LB media with 100 μg/mL ampicillin at 30 °C and 200 rpm to an $OD_{600\ nm}$ of 0.25−0.3. Arabinose was then added (to 10 mM) to the cell cultures to induce the expression of λ-red proteins, and the cells were then grown for another 30 min (OD of 0.5−0.6). Electrocompetent cells were then prepared, and cells were transformed with the corresponding suicide vector. Following electroporation, cells were grown out in 1 mL of SOC media (no antibiotic) and incubated at 37 °C for 3 h. Cells were then plated on LB agar plates with the addition of the appropriate antibiotic overnight at 37 °C. Suicide vector integrations were verified by colony PCR using oligos listed in Supplementary Table S3 and then by Sanger sequencing of the amplicons. pKD46 was cured from the verified integrants through outgrowth at 37 °C, isolation of colonies on LB agar (no ampicillin), and subsequent counter-screening of colonies on LB agar with ampicillin (no growth indicating the absence of pKD46).

Following suicide vector integration(s), gene copy (CIChE) expansion was carried out using helper plasmid pREDI2RecA. For the RFP/GFP experiments and hybrid plasmid/chromosomal systems, expansions were carried out in parallel with stepwise simultaneous antibiotic (chloramphenicol and spectinomycin) increases (concentrations doubling per step). For the fully chromosomal system, we observed that simultaneous titration increases of all three antibiotics significantly impaired cell growth. As such, for the fully chromosomal system, expansions were carried out sequentially starting with gentamycin followed by chloramphenicol and spectinomycin. Strains were grown in 5 mL of LB medium supplemented with 100 mM L-rhamnose (inducing the expression of *recA*) and increasing concentrations of antibiotics (one or more of gentamycin, starting concentration 12.5 mg/L; chloramphenicol, starting concentration 15 mg/L; and spectinomycin, starting concentration 25 mg/L) in 15 mL culture tubes at 30 °C until stationary phase, as previously described.[19] Fifty microliters of culture was then inoculated into 5 mL of fresh LB media with double the concentration of one or more of antibiotics (see above), and the process was repeated as desired or until the gene expansions were saturated (no further growth observed). pREDI2RecA was cured from all the strains following gene copy (CIChE) expansions. Strains were grown at 42 °C and shaken at 200 rpm overnight in liquid LB media. Thereafter, 5−10 μL of each strain was plated on LB agar, and single colonies were isolated and replica streaked on LB agar with or without ampicillin, with no growth on LB with ampicillin indicating the loss of pRedi2RecA.

Growth Conditions and Isopentenol Production. Cultures of *E. coli* DH1 harboring heat-sensitive auxiliary plasmids pRedi2RecA or pKD46 were grown overnight in LB medium containing ampicillin (100 mg/L final concentration) at 30 °C and shaken at 200 rpm in rotary shakers. Sequential integration experiments were performed with appropriate antibiotic in LB media at 37 °C, with one or more of the antibiotics gentamycin, chloramphenicol, and/or spectinomycin provided at final concentrations of 12.5, 15, and 25 mg/L, respectively. Expansion experiments were performed in LB media induced with 100 mM L-rhamnose at 30 °C and shaken at 200 rpm in rotary shakers. Production assays were performed in EZ-Rich defined medium (Teknova, Hollister, CA) containing 1% glucose. Briefly, starter cultures were used to inoculate 10 mL of EZ-Rich medium in a culture tube to an OD_{600} of 0.1. Production cultures were grown in rotary shakers (200 rpm) at 37 °C to an OD_{600} of 0.6 and induced with 500 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Following induction, cultures were moved to 30 °C for the duration of the assay. Samples were taken after 48 h for isopentenol analysis by GC-FID as described previously.[34] All cultures were grown in 10 mL medium volumes in culture tubes.

Fluorescence and Gene Copy Number Measurement. Red fluorescence proteins and green fluorescence proteins were measured using a Tecan F200 Pro (Männedorf, Switzerland) microplate reader. Measurement of RFP fluorescence was performed with excitation/emission at 590/635 nm. Measurement of GFP fluorescence was performed with excitation/emission at 485 nm/530 nm. Captured data were transferred directly into an Excel file for direct data analysis. Fluorescence values were corrected for background fluorescence (as a function of OD_{600}) of the negative control (wild-type DH1). qPCR analysis was used to measure gene copy numbers. A Bio-Rad CFX96 (Hercules, CA) was used for qPCR analysis. Quantitative PCR reactions were set up by mixing 10 μL of 2X SYBR green super mix, 2 μL of genomic DNA (50 ng), 1 μL of each DNA oligo (10 μM), and nuclease-free water to a total reaction volume of 20 μL. Cycling conditions used were one cycle of initial denaturation and enzyme activation at 95 °C for 3 min, 40 cycles of denaturing at 95 °C for 15 s, annealing at 55 °C for 30 s, and extension at 72 °C for 30 s, and one cycle of a melting curve at 55−95 °C (in 0.5 °C increments) for 30 s. A complete list of DNA oligos used for qPCR is provided in Supplementary Table S4.

Stability Tests. Stability tests were performed on chromosomal, hybrid plasmid/chromosomal, and two-plasmid isopentenol-producing strains, and on chromosomal RFP/GFP-expressing strains. Strains were grown overnight in LB media with antibiotics and in the absence of *recA* at 37 °C in 10 mL cultures. Cultures were diluted 1000-fold into fresh LB media without antibiotics and this process was repeated for three cycles. Glycerol stocks were maintained for each cycle. At the end of the experiment, RFP/GFP strains were tested for stability by measuring RFP/GFP fluorescence by microplate reader and *rfp*/*gfp* copy number by qPCR. Isopentenol-producing strains were tested for stability by measuring isopentenol production by GC-FID and copy number of each gene cassette by qPCR.

Machine Learning. Three regression models were applied to the prediction of isopentenol titer from the number of plasmid-borne or chromosomally integrated copies of three gene cassettes (GOI1−3). Sixty-five strains with varying copy numbers for each gene cassette were used as a training set (as indicated in Supplementary Table S1). Polynomial, random forest, and support vector regressors were applied to this training set. All regressors were implemented in python using

DOI: 10.1021/acssynbio.8b00243
ACS Synth. Biol. XXXX, XXX, XXX−XXX

the scikit-learn library,[35] and a jupyter notebook[36] is provided as Supplementary File S1 (and is also freely available from https://github.com/JBEI/PIACE), from which the analyses can be reproduced, and in which the specific implementation details of these models are described. Each model's performance was evaluated using 10-fold cross validation and the comparison of error residuals of the resulting predictions. One data point was held back (strain 7P7N, as indicated in Supplementary Table S1) during model development and was tested against fitted models to protect against/assess the extent of overfitting the data.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00243.

Plasmid and chromosomal maps, fluorescence and qPCR measurements, support vector and random forest regressor modeling, table of plasmids and strains, PCR amplification and DNA assembly reactions, colony PCR DNA oligos, and qPCR DNA oligos (PDF)

Jupyter notebook (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: njhillson@lbl.gov.

ORCID

Hector Garcia Martin: 0000-0002-4556-9685

Nathan J. Hillson: 0000-0002-9169-3978

Author Contributions

G.G., J.A.G., T.S.L., and N.J.H. conceived the project. G.G., Z.C., J.A.G., A.K., and N.J.H. designed the experiments. G.G. and Z.C. performed the experiments. G.G., Z.C., and N.J.H. analyzed the results. T.S.L, H.G.M., and N.J.H. supervised the work. G.G., Z.C., T.S.L, H.G.M., and N.J.H. wrote the manuscript. All authors have given approval of the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Michael Lee and Eric Steen for originally bringing the CIChE method to our attention and for their early suggestions of putative improvements to the CIChE method, including the use of BglBrick-compatible vectors. The authors thank John Wayne Thorne for early work towards extending the CIChE method. This work was part of the DOE Joint BioEnergy Institute (https://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, and was part of the Agile BioFoundry (http://agilebiofoundry.org) supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, through Contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed 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. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

■ ABBREVIATIONS

CIChE, chemical induced chromosomal expansion; CIGMC, chromosomal integration of gene(s) with multiple copies; CmR, chloramphenicol-resistance gene cassette; CRIM, conditional-replication, integration and modular; FLP, flippase; GC-FID, gas chromatography with flame ionization detector; GentR, gentamycin-resistance gene cassette; GOI, gene of interest; IPTG, isopropyl β-D-1-thiogalactopyranoside; LB, lysogeny broth; PHB, poly-3-hydroxybutyrate; OD, optical density; PIACE, parallel integration and chromosomal expansion; qPCR, quantitative PCR; SpecR, spectinomycin-resistance gene cassette

■ REFERENCES

(1) Juminaga, D., Baidoo, E. E., Redding-Johanson, A. M., Batth, T. S., Burd, H., Mukhopadhyay, A., Petzold, C. J., and Keasling, J. D. (2012) Modular engineering of L-tyrosine production in Escherichia coli. *Appl. Environ. Microbiol.* *78* (1), 89−98.

(2) Jiang, M., and Pfeifer, B. A. (2013) Metabolic and pathway engineering to influence native and altered erythromycin production through E. coli. *Metab. Eng.* *19*, 42−9.

(3) Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U. S. A.* *102* (36), 12678−83.

(4) Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* *10* (4), 354−60.

(5) Friehs, K. (2004) Plasmid copy number and plasmid stability. *Adv. Biochem. Eng./Biotechnol.* *86*, 47−82.

(6) Bassalo, M. C., Garst, A. D., Halweg-Edwards, A. L., Grau, W. C., Domaille, D. W., Mutalik, V. K., Arkin, A. P., and Gill, R. T. (2016) Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. *ACS Synth. Biol.* *5* (7), 561−8.

(7) Baumgartner, F., Seitz, L., Sprenger, G. A., and Albermann, C. (2013) Construction of Escherichia coli strains with chromosomally integrated expression cassettes for the synthesis of 2′-fucosyllactose. *Microb. Cell Fact.* *12*, 40.

(8) Martinez-Morales, F., Borges, A. C., Martinez, A., Shanmugam, K. T., and Ingram, L. O. (1999) Chromosomal integration of heterologous DNA in Escherichia coli with precise removal of markers and replicons used during construction. *J. Bacteriol.* *181* (22), 7143−7148.

(9) Ublinskaya, A. A., Samsonov, V. V., Mashko, S. V., and Stoynova, N. V. (2012) A PCR-free cloning method for the targeted phi80 Int-mediated integration of any long DNA fragment, bracketed with meganuclease recognition sites, into the Escherichia coli chromosome. *J. Microbiol. Methods* *89* (3), 167−73.

(10) Kuhlman, T. E., and Cox, E. C. (2010) Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res.* *38* (6), e92.

(11) Haldimann, A., and Wanner, B. L. (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.* *183* (21), 6384−93.

(12) Sauer, B. (1994) Site-specific recombination: developments and applications. *Curr. Opin. Biotechnol.* *5* (5), 521−7.

DOI: 10.1021/acssynbio.8b00243
ACS Synth. Biol. XXXX, XXX, XXX−XXX

(13) Mizuuchi, K. (1992) Transpositional recombination: mechanistic insights from studies of mu and other elements. *Annu. Rev. Biochem.* *61*, 1011−51.
(14) Wei, X. X., Shi, Z. Y., Li, Z. J., Cai, L., Wu, Q., and Chen, G. Q. (2010) A mini-Mu transposon-based method for multiple DNA fragment integration into bacterial genomes. *Appl. Microbiol. Biotechnol.* *87* (4), 1533−41.
(15) Song, C. W., and Lee, S. Y. (2013) Rapid one-step inactivation of single or multiple genes in Escherichia coli. *Biotechnol. J.* *8* (7), 776−84.
(16) Yang, J., Sun, B., Huang, H., Jiang, Y., Diao, L., Chen, B., Xu, C., Wang, X., Liu, J., Jiang, W., and Yang, S. (2014) High-efficiency scarless genetic modification in Escherichia coli by using lambda red recombination and I-SceI cleavage. *Appl. Environ. Microbiol.* *80* (13), 3826−34.
(17) Song, J., Dong, H., Ma, C., Zhao, B., and Shang, G. (2010) Construction and functional characterization of an integrative form lambda Red recombineering Escherichia coli strain. *FEMS Microbiol. Lett.* *309* (2), 178−183.
(18) Sabri, S., Steen, J. A., Bongers, M., Nielsen, L. K., and Vickers, C. E. (2013) 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. *Microb. Cell Fact.* *12*, 60.
(19) Tyo, K. E., Ajikumar, P. K., and Stephanopoulos, G. (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat. Biotechnol.* *27* (8), 760−5.
(20) Gu, P., Yang, F., Su, T., Wang, Q., Liang, Q., and Qi, Q. (2015) A rapid and reliable strategy for chromosomal integration of gene(s) with multiple copies. *Sci. Rep.* *5*, 9684.
(21) Liu, H., Wang, Y., Tang, Q., Kong, W., Chung, W.-J., and Lu, T. (2014) MEP pathway-mediated isopentenol production in metabolically engineered Escherichia coli. *Microb. Cell Fact.* *13* (1), 135.
(22) Atsumi, S., and Liao, J. C. (2008) Metabolic engineering for advanced biofuels production from Escherichia coli. *Curr. Opin. Biotechnol.* *19* (5), 414−419.
(23) Peralta-Yahya, P. P., Zhang, F., Del Cardayre, S. B., and Keasling, J. D. (2012) Microbial engineering for the production of advanced biofuels. *Nature* *488* (7411), 320.
(24) George, K. W., Chen, A., Jain, A., Batth, T. S., Baidoo, E. E., Wang, G., Adams, P. D., Petzold, C. J., Keasling, J. D., and Lee, T. S. (2014) Correlation analysis of targeted proteins and metabolites to assess and engineer microbial isopentenol production. *Biotechnol. Bioeng.* *111* (8), 1648−58.
(25) Chen, Y.-Y., Shen, H.-J., Cui, Y.-Y., Chen, S.-G., Weng, Z.-M., Zhao, M., and Liu, J.-Z. (2013) Chromosomal evolution of Escherichia coli for the efficient production of lycopene. *BMC Biotechnol.* *13* (1), 6.
(26) Kolter, R., Inuzuka, M., and Helinski, D. R. (1978) Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. *Cell* *15* (4), 1199−1208.
(27) Metcalf, W. W., Jiang, W., and Wanner, B. L. (1994) Use of the rep technique for allele replacement to construct new Escherichia coli hosts for maintenance of R6Kλ origin plasmids at different copy numbers. *Gene* *138* (1), 1−7.
(28) Chen, J., Densmore, D., Ham, T. S., Keasling, J. D., and Hillson, N. J. (2012) DeviceEditor visual biological CAD canvas. *J. Biol. Eng.* *6* (1), 1.
(29) Hillson, N. J., Rosengarten, R. D., and Keasling, J. D. (2012) j5 DNA assembly design automation software. *ACS Synth. Biol.* *1* (1), 14−21.
(30) Ham, T. S., Dmytriv, Z., Plahar, H., Chen, J., Hillson, N. J., and Keasling, J. D. (2012) Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. *Nucleic Acids Res.* *40* (18), e141.
(31) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* *6* (5), 343−345.
(32) Quan, J., and Tian, J. (2009) Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One* *4* (7), e6441.
(33) Hanahan, D., Jessee, J., and Bloom, F. R. (1991) Plasmid transformation of Escherichia coli and other bacteria. In *Methods in Enzymology*, Vol. *204*, pp 63−113, Elsevier.
(34) Chou, H. H., and Keasling, J. D. (2012) Synthetic pathway for production of five-carbon alcohols from isopentenyl diphosphate. *Appl. Environ. Microbiol.* *78* (22), 7849−55.
(35) Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* *12*, 2825−2830.
(36) Pérez, F., and Granger, B. E. (2007) IPython: a system for interactive scientific computing. *Computing in Science & Engineering* *9* (3), 1.

DOI: 10.1021/acssynbio.8b00243
ACS Synth. Biol. XXXX, XXX, XXX−XXX