ACS SyntheticBiology

pubs.acs.org/synthbio Research Article

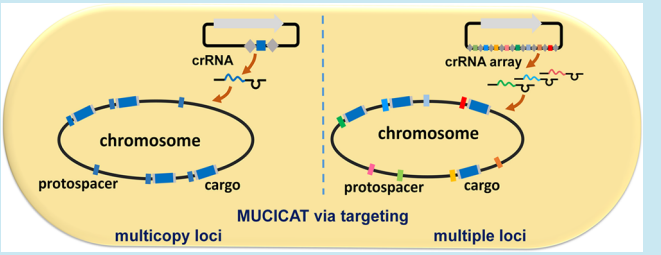
Multicopy Chromosomal Integration Using CRISPR-Associated Transposases

Yiwen Zhang, Xiaoman Sun, Qingzhuo Wang, Jiaqi Xu, Feng Dong, Siqi Yang, Jiawei Yang, Zixu Zhang, Yuan Qian, Jun Chen, Jiao Zhang, Yingmiao Liu, Rongsheng Tao, Yu Jiang, Junjie Yang,* and Sheng Yang*

Cite This: https://dx.doi.org/10.1021/acssynbio.0c00073 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Controlling the copy number of gene expression cassettes is an important strategy to engineer bacterial cells into high-efficiency biocatalysts. Current strategies mostly use plasmid vectors, but multicopy plasmids are often genetically unstable, and their copy numbers cannot be precisely controlled. The integration of expression cassettes into a bacterial chromosome has advantages, but iterative integration is laborious, and it is challenging to obtain a library with varied gene doses for phenotype characterization. Here, we demonstrated that multicopy chromosomal integration using CRISPR-associated transposases (MUCICAT) can be achieved by designing a crRNA to target multicopy loci or a crRNA array to target multiple loci in the *Escherichia coli* genome. Within 5 days without selection pressure, *E. coli* strains carrying cargos with successively increasing copy numbers (up to 10) were obtained. Recombinant MUCICAT *E. coli* containing genomic multicopy glucose dehydrogenase expression cassettes showed 2.6-fold increased expression of this important industrial enzyme compared to *E. coli* harboring the conventional protein-expressing plasmid pET24a. Successful extension of MUCICAT to *Tatumella citrea* further demonstrated that MUCICAT may be generally applied to many bacterial species.



KEYWORDS: *multicopy, chromosomal integration, CRISPR-associated transposase, insertion sequence, CRISPR array*

Downloaded via SHANGHAI NORMAL UNIV on July 10, 2020 at 02:30:36 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

Next-generation sequencing and low-cost DNA synthesis technology provide massive genetic resources together with fast and inexpensive access to gene acquisition, which offers near-infinite possibilities for the design and creation of biocatalysts.[1] Optimizing the genetic dose of a specific gene for efficient expression is a key step in the engineering of high-performance biocatalysts.[2] However, commonly used methods relying on plasmids are generally genetically unstable.[3,4] It is challenging to accurately control the dose of the gene expression cassette due to the limited plasmid replicons and the varied external factors that affect the copy number of a plasmid.[5,6] Chromosomal integration methods provide precise gene dosage control and recombinant genetic stability. However, even with *E. coli*, which is one of the most genetically tractable microorganisms, at least 2 days were needed for one copy of chromosomal integration, making it difficult to quickly obtain recombinant descendants that have successively varied copy numbers of gene expression cassettes.[7]

Screening for high-copy tandem repeat gene expression cassettes with elevated chloramphenicol concentrations accelerates this process.[8] However, this approach leaves chloramphenicol-resistant markers on the *E. coli* chromosome, which limits its industrial application.[9] The bacteriophage Mu-driven transposition system, in principle, allows for the integration/amplification of target genes in the chromosomes without the aid of antibiotic concentration gradients, but the integration sites are random.[10]

Although the Cas nucleases cleave multiple loci on the chromosome, it remains difficult to obtain a multiplex-edited genome with more than 3 sites simultaneously[7] due to limitations in homologous recombination efficiency. Therefore, an editable, fast, marker-free, site-directed chromosomal multicopy integration system will be a powerful tool for metabolic engineering and synthetic biology. Recently discovered transposon-encoded CRISPR−Cas systems,[11−14] including CRISPR-associated transposases (CAST)[15] and insert transposable elements by guide RNA-assisted targeting (INTEGRATE),[16,17] enable genomic site-specific DNA integration independent of homologous recombination. CAST is the subtype V−K of Tn7-like transposons from

Received: February 9, 2020
Published: June 17, 2020

© XXXX American Chemical Society

Scytonema hofmanni, which contains TniQ, TnsBC, and a single-protein CRISPR-Cas effector, Cas12k. INTEGRATE is a Tn6677 transposon from *Vibrio cholerae* that contains TnsABC, TniQ, and a CRISPR−Cas effector complex, Cas678. These systems may be capable of being developed into multicopy chromosomal integration by CRISPR-associated transposases (MUCICAT), which we designed and demonstrated in the present work (Figure 1).

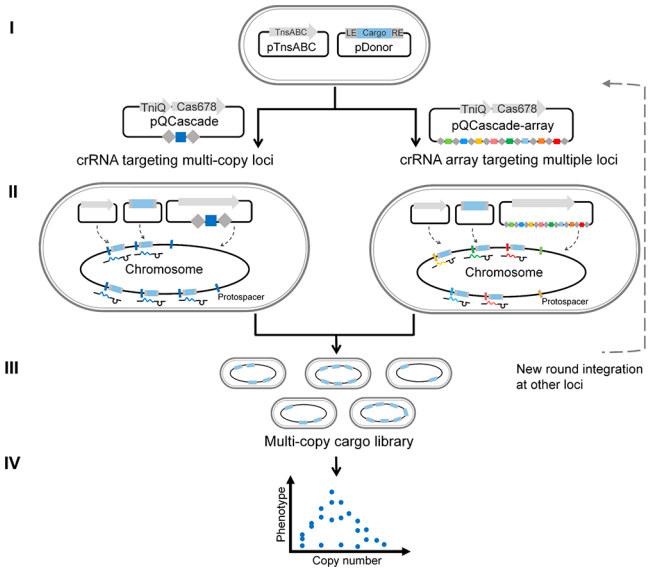


Figure 1. Schematic illustration of multicopy chromosomal integration by CRISPR-associated transposases (MUCICAT). (I) Transfer of plasmids pTnsABC, pDonor, pQCascade, or pQCascade-array into the recipient strain. pQCascade carries one CRISPR array targeting multicopy loci of the chromosome. pQCascade-array carries a series of CRISPR arrays targeting multiple different loci of the chromosome. Each CRISPR array contains two repeats (gray diamonds) and a single spacer (rectangle of blue or other colors). (II) IPTG induces the expression of transposition-associated enzymes to initiate transposition. The cargo gene is integrated into the corresponding loci of the chromosome guided by crRNAs. Rounds of transfer increase the integration ratio of the cargo gene. (III) A plasmid-free multicopy cargo library is generated after transposition-associated plasmid curing. By altering the crRNA in a new round of transposition experiments that target other loci, the copy number range was further increased. (IV) Screening strains containing different copy numbers of cargo for optimized enzyme yield during fermentation.

RESULTS

MUCICAT *via* Targeting Multicopy Loci. There are many naturally existing multicopy sequences in bacteria, such as Insertion Sequences (ISs).[18−20] We speculated that using crRNA targeting IS loci would result in cargo integration *via* CRISPR-associated transposases into all targeting sites at once or in a battery of strains having various copies of integrations that can be readily characterized in subsequent work. On the basis of the ShCAST system, we first designed crRNA-IS1C to target the IS1 loci in *E. coli* Pir1. Twenty-five randomly picked *E. coli* Pir1 colonies after transposition underwent junction PCR using one external primer (located in IS1) and one internal primer (located in cargo gene). The results showed that 15 colonies had at least one cargo integration in the monitored orientation (Supplementary Figure S1). Two positive colonies in Supplementary Figure S1 were selected randomly for genome sequencing, and site 3 and site 4 of IS1 loci were detected as integrated in a total of 5 IS1 loci (Supplementary Figure S2). On the basis of the genome information on the Pir1 strain, we designed specific primer pairs to detect the integration site in more colonies. After restreaking to yield pure single colonies, colony PCR using the external primer pairs of IS1 sites showed that only 3 colonies contained one copy cargo integrated at site 4 (Supplementary Figure S3). We speculated that the integration at site 3 was not detected because only a small proportion of cells achieved cargo integration at that site and those cells were not picked up during single colony purification. The transformant with one integration at site 4 was restreaked for 8 transfers. We failed to detect additional integration at the other four genomic IS1 sites (Supplementary Figure S4). We speculated that the relatively low transposition efficiency led to a high probability of selecting cells that had not undergone transposition in the mixed colonies for subsequent passage, which resulted in the failure to obtain pure multicopy strains.

The INTEGRATE system enabled DNA cargo insertion into the *lacZ* locus of *E. coli* BL21 (DE3) guided by crRNA-3[16] (Supplementary Figure S5a). To investigate whether integration also occurred on the other *lacZ* locus at the λDE3 prophage on the chromosome, we verified both sites using colony PCR and found that 10 of the 16 colonies had cargo integrated at both *lacZ* loci (Supplementary Figure S5b), which indicates that INTEGRATE simultaneously targeted two identical sequences. To measure the potential of the integration efficiency, we replaced crRNA-3 with the crRNA-IS186 to target 5-copy IS186 loci on the genome. Colony PCR showed that all 14 randomly picked colonies had at least 3 copies of cargo integration, and 11 colonies had integration at all 5 IS186 loci (Supplementary Figure S5c). Encouraged by these results, we designed crRNA-IS1 to target 28 copies of IS1 loci in the *E. coli* BL21(DE3) (Figure 2a). Three of the 11 colonies showed at least 4 copies of cargo integration after a single round of transposition experiment. To test the effect of longer transposition time, we transferred these colonies by restreaking them on the transposition-inducing agar-plates for colony out-growth and found that one of the colonies accumulated at least 10 copies of cargo integration at IS1 loci in the third round of transfer. Further transferring increased the portion of the colonies with higher cargo copies (Figure 2b). To assess the transposition efficiency in *E. coli* MG1655, we chose *E. coli* MG1655 Δ*endA* Δ*recA* (DE3)[21] as a test strain (Addgene bacterial strain #37854) to target IS186 (3 copies), IS3 (5 copies), and IS1 (8 copies). One of the 11 colonies achieved cargo gene integration at two IS1 loci (Supplementary Figure S5d). Only single copy cargo integration was detected at the IS3 (Supplementary Figure S5e) and no cargo integration was detected at the IS186 site. Regarding the difference in transposition efficiency between MG1655Δ*endA* Δ*recA* (DE3) and BL21 (DE3), we speculate that the deletion of *recA* or *endA* affected the transposition event.[22] The reason also may be that the protein expression advantage of the BL21 strain made the expression of transposase more abundant than in MG1655, which made the transposition efficiency higher. This result suggests that CRISPR-associated transposases need to be fine-tuned in each specific strain for optimized MUCICAT efficiency.

MUCICAT Using a crRNA Array. With the positive results above, we investigated whether MUCICAT would be achieved

https://dx.doi.org/10.1021/acssynbio.0c00073
ACS Synth. Biol. XXXX, XXX, XXX−XXX

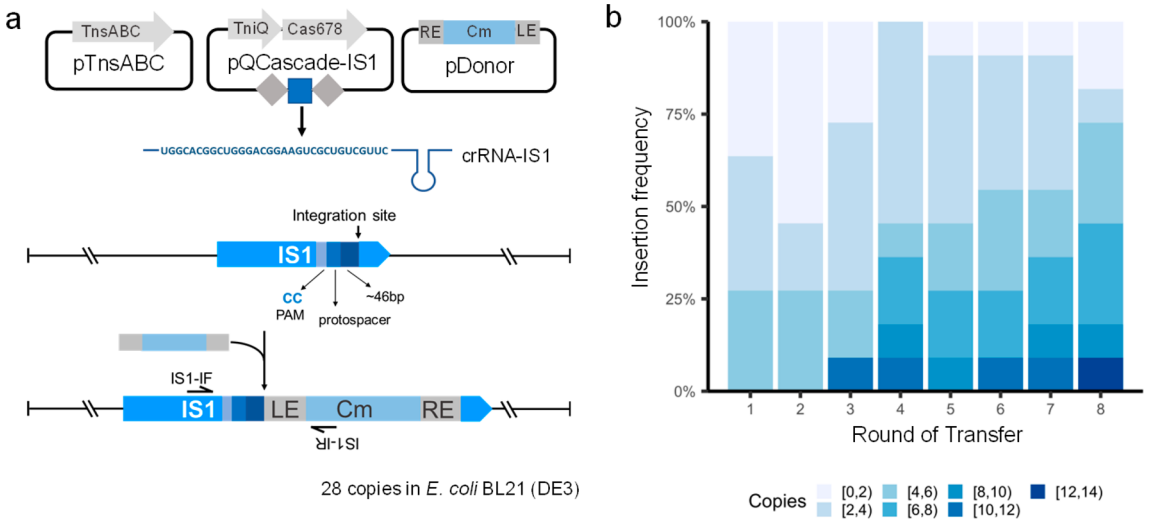


Figure 2. Targeting IS1 sequences enables multiple-copy genomic integration of a cargo gene. (a) Schematic illustration of transposition at genomic IS1 loci guided by crRNA-IS1 in the INTEGRATE system, including three transposition-associated plasmids, the sequence of crRNA-IS1, potential insertion at IS1 loci, and quantitative PCR (qPCR) primer pair IS1-IF and IS1-IR (external-internal). crRNA-IS1 targets 28 of 29 copies of IS1 in the chromosome of *E. coli* BL21 (DE3). Because the insertion orientation of the cargo gene in the INTEGRATE system is uncertain, only insertion in the RL direction was detected using the qPCR primer pair above. (b) The copy-number of cargo gene at targeted genomic sites increased with successive rounds of transfer to new transposition inducing agar plates. The copy numbers were determined using the primer pair in (a) and qPCR. In transposition experiments, cells grown in the first transposition-inducing plates containing 1 mM IPTG were defined as Transfer 1 and new cells generated by streaking single colonies onto new inducing plates were defined as Transfer rounds 2, 3, ..., *n*. The copy-numbers of the cargo gene increases with more rounds of transfer. The darkness of the color gradient indicates the copy-number. All data were measured using qPCR.

by targeting multiple different loci on the chromosome. We designed an 8-loci targeting crRNA array in which 8 spacers that were validated in the INTEGRATE system[16] (Figure 3a) were used. 40% of the colonies had one copy integration, and 20% of the colonies contained 2 copies after a single round of transposition. After 7-round transfers, 65% of colonies had all 8 loci inserted with the cargo (Figure 3c). Targeting sites 4 and 5 appeared to have higher integration frequencies than the other sites (Figure 3d). Genome sequencing showed that the cargo was inserted into the 8 targeting sites as expected, with an on-target rate higher than 99% according to NGS analysis of the mixed colonies of strain BL21(DE3) harboring pTnsABC, pDonor, and pQCascade-array1 after the 2-round and 4-round transfers (Figure 3e). The final strain CmA-8A with insertions at all 8 sites was verified using genomic sequencing (Figure 3e), and other strains had multiple copies of cargo integration during the transposition experiments (Supplementary Figure S11). Therefore, the crRNA array may be applied to insertionally disrupt multiple genes and achieve multicopy integration of a gene of interest simultaneously.

We also assembled the 8 crRNAs validated in the ShCAST system into a crRNA array. After transposition experiments in *E. coli* Pir1, 2 of 8 colonies showed integration bands in 5 of 8 loci *via* junction PCR (Supplementary Figure S6). After restreaking for pure single colonies, 6 colonies were detected. Colony PCR by external primer pairs outside cargo did not show any integrated bands (Supplementary Figure S7). These results indicate the low multiple integration frequency of ShCAST, and only a small portion of cells in the colonies was integrated with the cargo. We speculated that the above results may be due to the low efficiency or low expression of transposition-associated enzymes in *E. coli* Pir1.

A One-Step Plasmid-Curing System for MUCICAT. In order to terminate the process of MUCICAT at an appropriate time, we designed a plasmid pCutamp, which carries a SpCas9 cassette and a rhamnose-inducing sgRNA that simultaneously targets the ampR promoter on the backbone of all three plasmids, *i.e.*, pDonor, pQCascade, and pTnsABC. After the transformation of pCutamp, 3 h of recovery with the addition of 10 mM rhamnose, three plasmids in MUCICAT were cleaved with a plasmid-curing efficiency up to 98% (Figure 4b). pCutamp carries a SacB cassette as a negative selection marker, so plasmid-free colonies may be obtained by subsequently curing pCutamp on the sucrose-contained agar plate (Figure 4a). After each transfer, some colonies were selected randomly to perform a plasmid-curing step to obtain a multicopy integration strain library that varied in copy numbers and integration sites (Figure 1).

Application of MUCICAT to Rapidly Generate Genomic Multicopy Library. Glucose dehydrogenase (GDH, EC:1.1.1.47) is one of the most significant enzymes in industrial application because of its function of regenerating NAD(P)H,[23−25] and it is currently commercially produced using the pET plasmid system in China.[26] We used MUCICAT and replaced the cargo on pDonor with the T7lac promoter-controlled GDH expression cassette (T7-*gdh*) to perform transposition and generate a multicopy integration library of T7-*gdh* within 8 loci (Figure 5a). To examine whether even higher copy numbers may be achieved from a strain that already contained 8 T7-*gdh* insertions at the array 1 sites *via* targeting 5 additional genomic sites at IS186 loci, we replaced pQCascade-array1 with pQCascade-IS186. GDH-producing strains with 9- and 10-copy of T7-*gdh* were obtained successfully (Supplementary Figure S8). After curing all plasmids, strains were selected randomly from the library. The 9-copy and 10-copy constructs above were fermented, and the enzyme activity was monitored for each strain. The GDH activities of 15 of the 49 strains tested were significantly higher

https://dx.doi.org/10.1021/acssynbio.0c00073
ACS Synth. Biol. XXXX, XXX, XXX−XXX

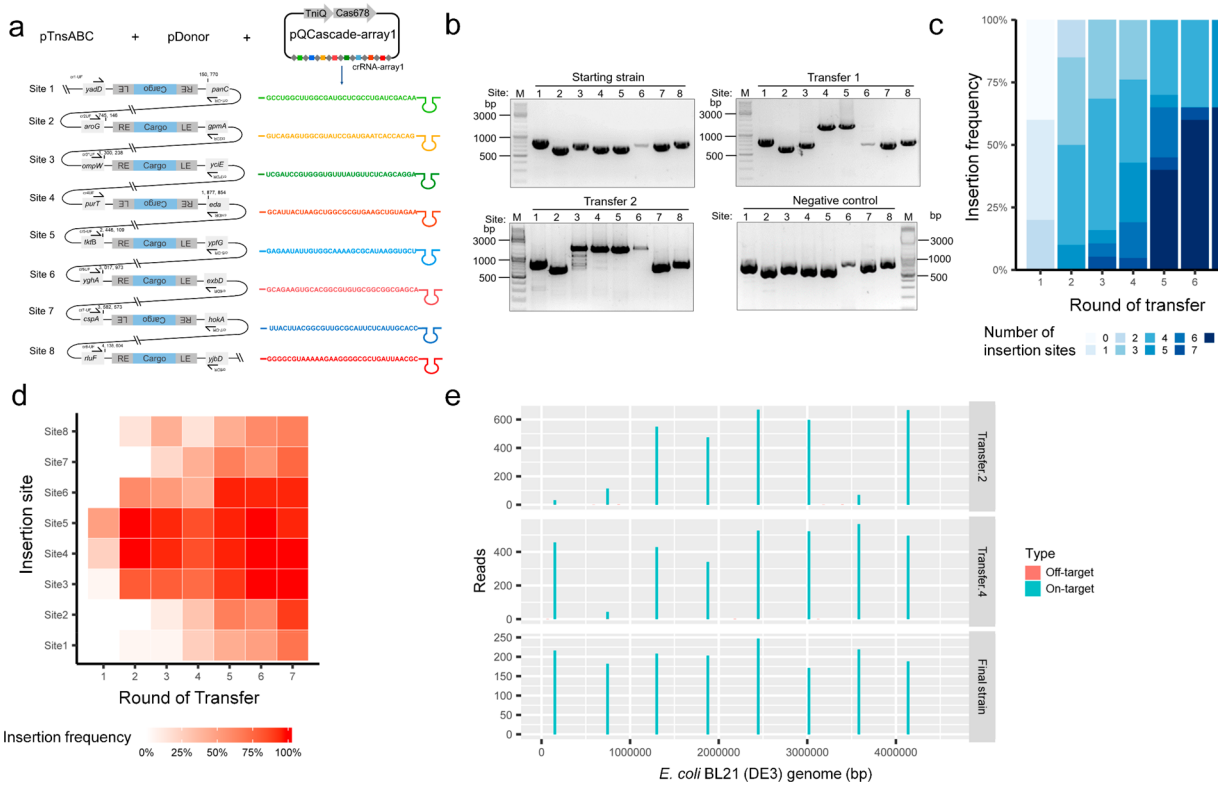


Figure 3. Targeting 8 different sites with a crRNA array to achieve multicopy integration of a cargo gene. (a) Schematic illustration of inserting the target gene into multiple genomic loci using crRNA array1 and the INTEGRATE system, including potential transposition products at 8 genomic sites (left), PCR primer pairs for verifying (arrows), and sequences of each spacer (right). Numbers no the top of the genes indicate the genomic positions of the 3′ end of each spacer. (b) PCR analysis of 8 genomic sites for cargo insertion during transferring. Unintegrated sites PCR products: ~700 bp; Integrated sites PCR products: ~1700 bp. *E. coli* BL21(DE3)/pTnsABC, pDonor, pQCascade-array1 grown in plates without IPTG was defined as the starting strain, and cells initially grown on IPTG-containing plates were defined as Transfer 1. Cells streaked from the initial IPTG-containing plates to fresh IPTG-containing agar for another round of transposition was defined as Transfer 2. *E. coli* BL21(DE3) harboring pTnsABC, pDonor, and pQCascade_entry served as negative controls. (c) Genomic insertion of cargo gene during strains transferring. Genomic insertion was determined *via* colony PCR and the insertion at the 8 genomic sites increased with successive rounds of transfer. (d) Analysis of integration site distribution guided by crRNA-array1 and determined from the PCR data. Primer pairs in (b), (c), and (d) are shown in (a). (e) Genomic sequencing confirmed that the copy-number of the inserted gene increased with successive rounds of transferring. Samples analyzed include mixed colonies of BL21(DE3) harboring pTnsABC, pDonor and pQCascade-array1 in the second Transfer (Transfer 2), the fourth Transfer (Transfer 4), and a pure single colony strain ArCm-8A (Final strain) that had all 8 sites inserted as verified using colony PCR.

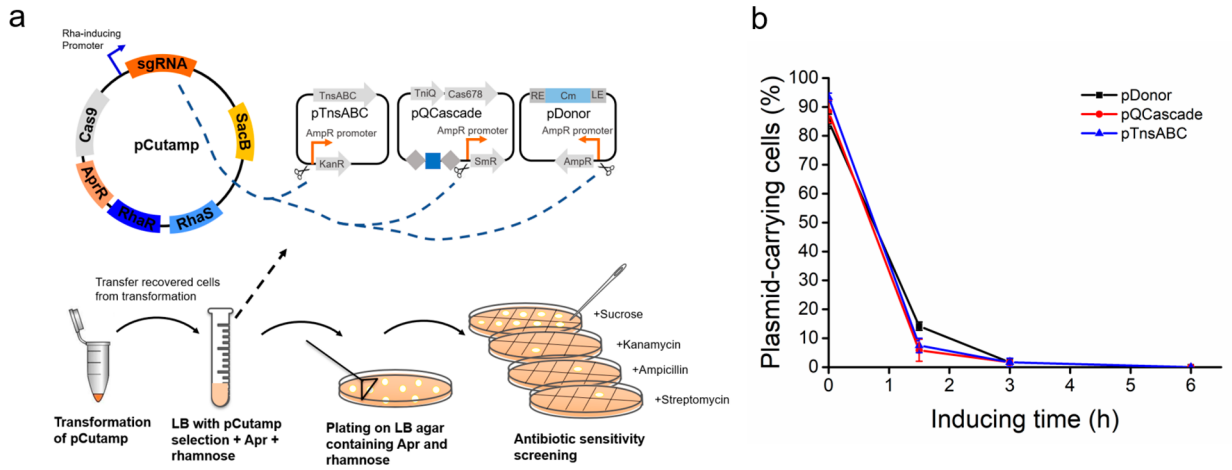


Figure 4. Plasmid pCutamp expressed Cas9 and sgRNA for MUCICAT plasmid-curing. (a) SgRNA was induced by rhamnose and targeted the 20-nt sequence in AmpR promoter of pDonor, pQCascade, and pTnsABC plasmid. Because there is no template, the three plasmids above cannot be repaired after cutting by Cas9. Strains that needed to cure plasmids were transformed with pCutamp *via* electroporation. (b) To obtain a high plasmid-curing efficiency with a shorter inducing time, different inducing times were tested in rhamnose-containing LB medium. When the inducing time was 3 h, the efficiency of plasmid-curing was approximately 98.33%, and 140 colonies were replated for verification at every time point.

https://dx.doi.org/10.1021/acssynbio.0c00073
ACS Synth. Biol. XXXX, XXX, XXX−XXX

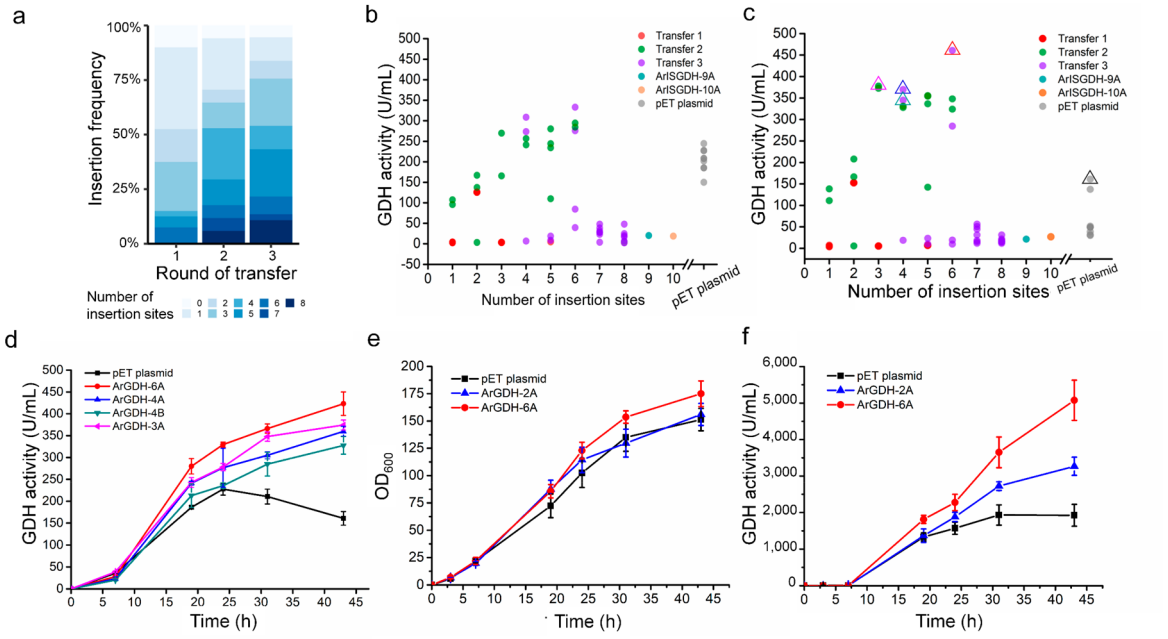


Figure 5. Comparison of GDH activity between strains constructed by MUCICAT and a strain carrying the pET plasmid. (a) Successive increase in genomic T7-*gdh* cassette insertion sites with a series of transfer for new rounds of induced transposition. The T7-*gdh* cassette was targeted at 8 different sites of the chromosome in *E. coli* BL21(DE3) guided by crRNA array1; the insertion sites on the chromosome were verified using colony PCR. (b,c) Enzyme activity analysis of strains with genomic insertion of T7-*gdh* at different sites and a strain carrying the pET24a-GDH plasmid (gray) in flasks at 24 h (b) and 43 h (c). Strains from the first transfer (red), the second transfer (green), and the third transfer (violet) and strains carrying the pET24a-GDH (gray), harboring 9-site insertion (blue), and 10-site insertion (orange) engineered by targeting the IS186 loci after prior transposition at 8 loci (array1) were included. Strains from Transfers 1−3 were selected randomly from different transfer plates. (d) Flask-level fermentation of strains with high enzyme activity in (c) (triangle boxes). GDH activities of strains with different numbers of sites insertions of T7-*gdh*, including strain ArGDHA-6A (red triangle in (b)), strain ArGDH-4A (blue triangle in (b)), ArGDH-4B (green triangle in (b)), and ArGDH-3A (pink triangle in (b)), and a strain harboring pET24a-GDH (pET plasmid). The corresponding antibiotic (kanamycin) was added during fermentation of the pET24a-GDH containing strain in (b) and (c). (e,f) The growth (e) and production (f) of strain ArGDH-2A, strain ArGDH-6A, and the strain carrying plasmid pET24a-GDH in fed-batch fermentation. Data are shown as the means ± s.d. for $n = 3$ biologically independent samples.

than the strain carrying the pET24a-GDH plasmid in flask culture at 43 h (Figure 5c). Because the GDH enzyme activity of the same copy of the strain was quite different, we speculated that different insertion positions would affect GDH enzyme activity. Therefore, the insertion positions of different strains were counted (Supplementary Table S4). No correlation was found between GDH enzyme activity and genomic positions, which may require different single copy GDH strains with different integration sites to further accurately determine. These 15 strains had 2−6 sites of insertion of T7-*gdh* (Figure 5c). The GDH activities from 17 strains carrying 7−10 sites of insertion were substantially lower than the pET24a-GDH control strain (Figure 5b,c). We suspected that strains with extremely low enzyme activity were no-longer able to produce functional T7 RNA polymerase, which is essential for MUCICAT-associated enzymes and GDH expression. To investigate this possibility, the pET24a-GFP plasmid was transformed into three nonproductive strains (with T7-*gdh* insertion occurred at 1, 3, and 5 sites) selected from strains above. None of the transformants from these 3 strains exhibited fluorescence after induction (Supplementary Figure S9). Genomic sequencing revealed that the T7 RNA polymerase gene of ArGDH-5D and ArGDH-1D was inactivated by the off-target insertion of a *T7-gdh* cargo (Supplementary Figure S11) and the T7 RNA polymerase gene of GDHA-3D harbors a nonsense mutation (Gln232Stop) (Supplementary Table S5). This high ratio of T7 RNA polymerase inactivation during MUCICAT suggests that the GDH expression produced a heavy metabolic burden and induced cellular adaptive evolution. The design of MUCICAT with induction conditions that differ from the target gene cassette may reduce the proportion of non-productive cells. Further, we selected a portion of strains that carry 7−10 sites of insertion to sequence the gene encoding T7 RNA polymerase and found there was no mutation. The results of fluorescence intensity after transforming pET24a-GFP into these strains also proved that the T7 system could work (Supplementary Figure S14). After transforming with the plasmid pET24a-GDH, the enzyme activities of the strains carrying 7−10 sites of insertion were recovered to the level that similar to that of the control strain BL21(DE3) (Supplementary Figure S15).

A 6-genomic-sites insertion strain, ArGDH-6A, showed the highest enzyme production, and several well-performing strains in Figure 5b were further benchmarked with the pET24a-GDH strain. The GDH activity of the strain carrying the pET plasmid began to decrease from 24 h in the middle of fermentation, and the strains containing genomic multiple integrations remained stable and increased steadily in the end phase of fermentation (Figure 5d). In the scale-up fed-batch fermentation, the growth of the MUCICAT strains and the plasmid-carrying strain was almost consistent (Figure 5e), but the GDH activity of plasmid-carrying strain peaked at 31 h and remained nearly unchanged thereafter. The enzyme activity of

https://dx.doi.org/10.1021/acssynbio.0c00073
ACS Synth. Biol. XXXX, XXX, XXX−XXX

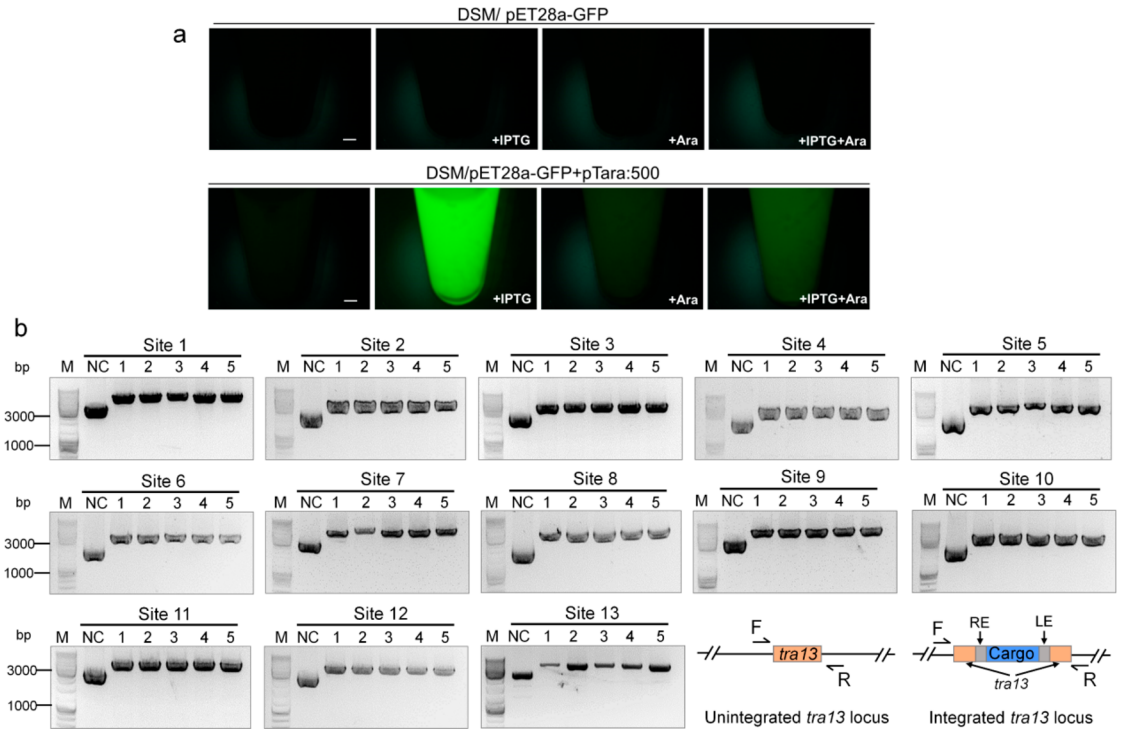


Figure 6. Application of the MUCICAT system in *Tatumella citrea* DSM 13699. (a) The T7 transcription system worked in *Tatumella citrea* DSM 13699. Plasmid pET28a-GFP, which carries the GFP-encoding gene controlled by the IPTG-inducing T7 promoter, and the plasmid pTara:500,[29] which carries an arabinose (Ara)-inducing T7 RNA polymerase, were transformed into *Tatumella citrea* (*T. citrea*) DSM 13699. After induction by IPTG, green fluorescence was observed under a fluorescence microscope, which confirmed the expression of GFP and the function of the T7 transcription system in *T. citrea* DSM 13699. In *T. citrea* DSM 13699, without the induction of Ara, the leakage expression of T7 RNA polymerase was adequate to make the system work. (b) crRNA-tra13 was designed to target the consensus sequence of the transposases encoding gene (13 copies, the genomic positions are shown in Supplementary Table S3) in *T. citrea* DSM 13699 chromosome using the INTEGRATE system. After 2 rounds of transfer, 5 colonies were randomly selected for integration verification using colony PCR, which showed that all 13 sites in 5 colonies were inserted. Integrated *tra13* locus: ∼3700 bp; Unintegrated *tra13* locus: ∼2700 bp. 1−5: strains 1−5.

the MUCICAT strain ArGDH-6A reached 1.68-fold of the plasmid-carrying strain at 31 h and further increased to 2.3-fold at 43 h (Figure 5f), which indicates that a better yield may be achieved from expression cassettes integrated in the chromosome than genes harbored in a multicopy plasmid.

Application of MUCICAT in *Tatumella citrea* for Multicopy Integration. To evaluate the possibility of a broader applicability of MUCICAT, *Tatumella citrea* (*T. citrea*) DSM 13699,[27] another member of *Enterobacteriaceae*, was selected. *T. citrea* is used in the bioindustry for the synthesis of the vitamin C precursor 2-keto-L-gulonic acid, and it has broad application prospects.[28] The multicopy integration tool of *T. citrea* will promote the development and utilization of this species. First, *T. citrea* DSM 13699 (NCBI Reference Sequence: NZ_CP015579.1) was transformed with a T7 RNA polymerase-expressing plasmid (pTara:500)[29] and a pET28a-GFP plasmid. Strong fluorescence was observed after induction, which demonstrated that the T7 system worked (Figure 6a). The cargo gene, a GFP-encoding gene, was targeted to 13 identical genomic loci by the crRNA-tra13 (PAM and spacer sequence see Supplementary Table S3). After plasmid-curing of mixed colonies on the second-round streaking transfer plate, colony PCR analysis of 5 colonies selected randomly showed that the cargo gene was integrated at all 13 sites (Figure 6b). Genome sequencing data verified that all 13 sites were inserted (Supplementary Figure S10). We speculated that the higher integration efficiency in *T. citrea* may due to the binding efficiency of crRNA-tra13 or the better expression of transposases in its growth temperature, 30 °C. In conclusion, MUCICAT may be broadly applied to other bacterial species.

DISCUSSION

On the basis of the INTEGRATE system, which uses a crRNA that targets multiple identical sequences, such as IS1 or IS186, or a crRNA array targeting diverse sequences, a multiplex chromosomal insertion library with copy numbers of up to 10 was obtained in 5 days. Compared with the requirement of 2 days per round of CRISPR-Cas-based homologous recombination, MUCICAT had a significant advantage in the manipulation speed. Compared with CIChE,[3] MUCICAT does not rely on selection pressure and does not leave selection markers on the chromosome. MUCICAT was highly efficient in *T. citrea* with a T7 RNA polymerase expressing plasmid, which suggests that this system is broadly applicable.

The gene dose of the glucose dehydrogenase expression cassette was optimized using MUCICAT. The GDH activity of a MUCICAT strain having a 6-genomic-site insertion of T7-*gdh* was 2.63-fold of the pET24a-GDH-carrying reference strain. The copy number of the pET plasmid is approximately 15−20.[30,31] Although 10, 5, or 1 copy may be obtained by changing the replicon to p15A,[32] pSC101,[33] or use of F plasmid, more precise control of 2−4, 6−9, and 11−14 copies cannot be achieved using plasmid systems. MUCICAT could reach 10 copies of cargo by targeting IS1 loci in BL21 (DE3) with 3 rounds of transfer, and more transfers resulted in higher

https://dx.doi.org/10.1021/acssynbio.0c00073
ACS Synth. Biol. XXXX, XXX, XXX−XXX

copy numbers up to 16 (Supplementary Figure S12). The pET series of plasmids, and their compatible plasmid, may be combined to achieve a wider range of copy numbers. However, MUCICAT can also combine crRNA arrays to target more loci. Therefore, at least in the field of recombinant protein expression by *E. coli*, MUCICAT may be used as an upgrade alternative of the pET plasmid vectors. The additional steps are pCutamp transformation to cure all plasmids, and several additional transfers to obtain a wider range of copy numbers of genes of interest.

From our experimental results, the efficiency of the CAST system in *E. coli* Pir1 strains was not sufficiently high to easily select positive colonies with multiple copies of genomic insertion. Therefore, we replaced the previous constitutive lac promoter of this system with the stronger T7lac promoter to increase the expression of the guide RNA and transposition-associated enzymes. To extend this system to other *E. coli* strains except for Pir1, the λ pir protein-dependent replicon R6k of plasmid pDonor_ShCAST_kanR may be replaced with the temperature-sensitive replicon pSC101.[15] By leveraging the updated CAST system above in *E. coli* BL21(DE3), two-copy genomic insertion was achieved by targeting 29-copy IS1 loci after 3 rounds of transfer (Supplementary Figure S13). This result means that the CAST system promises to be a genomic multicopy integration tool, but it requires more fine-tuning.

A key deficiency of the current version of MUCICAT is that when the cargo gene is T7lac-inducing, the same inducer (IPTG) is used for the transposon and cargo gene expression simultaneously. In this case, cells that undergo regulation system mutation tend to inactivate the transposition system and the desired protein-producing system. Because the metabolic burden of protein production in these cells was reduced due to the inactivation, these cells generally have higher fitness than their non-inactivated siblings and these cells are prone to be enriched during transposition. For example, we encountered the mutation of T7 RNA polymerase in our work. Cells with regulation system mutations dramatically overtake the cells that still retain the ability of transposition and production, which makes the proportion of the latter drop sharply. This reduction leads to cells that are capable of transposing may be acquired only using high-throughput screening methods, especially after a long period of induced protein expression and transposition.

Developing a new version of MUCICAT is proposed to avoid this negative selection caused by the metabolic burden of desired protein production toward the transposition system. Arabinose, rhamnose, anhydrotetracycline, and a thermal induction system may be tried to control MUCICAT to separate the regulation between the transposition system and the desired protein expression system, which is promising in quickly establishing an engineered bacterial library with higher and wider copy numbers of gene expression cassette(s). Because the transposition efficiency might largely depend on the expression doses of transposase and crRNA, we speculated that after replacing the T7 system with other induction systems, the induction intensity of the protein and crRNA that affect the integration efficiency must to be fine-tuned.

The other aspect that may be improved is the use of orthogonal CRISPR-associated transposases for different cargos, and orthogonal dose combinations of multiple gene expression cassettes may be achieved. Therefore, MUCICAT may play a greater role in the field of metabolic engineering, and a combination of various gene disruptions and insertions may be generated in one transformation. The low transposition efficiency of ShCAST is most likely attributed to its weak promoter lac, and MUCICAT may be achieved by replacing it with a T7 or another higher-strength promoter. These further modifications may make MUCICAT a convenient and broadly applicable genetic tool for the stable and optimized expression of proteins from candidate expression cassettes integrated into the chromosome of producing strains.

MATERIALS AND METHODS

Plasmid Construction. All plasmids used in this study were listed in Supplementary Table S1, and a subset of them are available from Addgene (ID: 140622−140632) and Molecular Cloud (ID: 101225−101235).

Plasmids construction based on INTEGRATE system: plasmids pTnsABC, pDonor, pQCascade_entry, and pQCascade_crRNA-3 were synthesized by GenScript according to sequences reported in Klompe's research.[34] crRNA-9 to crRNA-16 in Klompe's research were selected to assemble crRNA array1, which was synthesized and cloned into pQCascade_entry by GenScript, yielding pQCascade-array1. The plasmid pQCascade_entry was digested by BsaI and inserted a 32-nt spacer of crRNA-IS1 by GoldenGate,[26] yielding pQCascade-IS1. The same method was performed to construct pQCascade-IS3 and pQCascade-IS186. To generate pDonor-GDH, T7-*gdh* gene[24] was ligated to pDonor, which was digested prior by NcoI and XbaI. The same method was performed to construct pDonor-GFP.

Plasmids construction based on ShCAST system: 8 crRNAs in Strecker's research[15] were selected to assemble crRNA array2, which was synthesized and cloned into pHelper_ShCAST (Addgene plasmid #127922) by GenScript, yielding pHelper-array2. The plasmid pHelper_ShCAST_sgRNA (Addgene plasmid #127921) was digested by LguI and inserted a 24-nt spacer of crRNA-IS1C by GoldenGate,[26] yielding pHelper-IS1. Using the primers in Supplementary Table S1, the promoters in pHelper-IS1 were replaced with T7lac promoter by PCR and Gibson assembly, yielding pHelper-T7IS1.

Plasmids were cloned and propagated in *E. coli* Top10 (Thermo Fisher Scientific), purified using Miniprep Kits (Qiagen), and verified by Sanger sequencing (GENEWIZ). Gibson assembly Cloning Kit was from NEB.

Transposition Experiments in *E. coli*. Transposition experiments of INTEGRATE system comprise the transformation and induction of pDonor, pTnsABC, and pQCascade (or its variant). Chemically competent *E. coli* BL21 (DE3) cells were first cotransformed with pDonor and pTnsABC, and transformants were screened by double antibiotic LB-agar plates containing 100 μg/mL ampicillin and 50 μg/mL kanamycin. The resulting strains were made chemically competent using standard methods, followed by introducing pQCascade through heat shock. After 1 h recovery in fresh LB medium at 37 °C, cells were plated on triple antibiotic LB-agar plates containing 100 μg/mL ampicillin, 50 μg/mL kanamycin, and 50 μg/mL streptomycin. After a 16-h growth at 37 °C, hundreds of colonies were scraped from the plates, and a portion was resuspended in fresh LB medium. These cells were then replated on LB-agar plates containing triple-antibiotics and 0.1 mM IPTG to induce transposition related protein expression and grown at 37 °C for another 16 h. After biofilms were formed, scrape them off and resuspended in LB medium. Cells were rediluted appropriately

https://dx.doi.org/10.1021/acssynbio.0c00073
ACS Synth. Biol. XXXX, XXX, XXX−XXX

and plated on triple antibiotic LB-agar plates containing 1 mM IPTG (*i.e.*, inducing plates), and grown overnight at 37 °C. For clonal out-growth and integration rate improvement, we further tried to transfer colonies by restreaking as follows. Colonies grown on the inducing plates above would be defined as Tansfer1, and a portion of these colonies were streaked onto new triple-antibiotic LB-agar plates containing 1 mM IPTG and grown overnight at 37 °C to generate a new transfer, *i.e.*, Transfer 2, continuous transferring produced Transfer *n* (*n* = 3, 4, ...). Colony PCR or qPCR was then performed to identify copy number.

Transposition experiments of CAST system: The plasmids pDonor_ShCAST_kanR (Addgene plasmid #127924) and pHelper-IS1 were transformed to Pir1 cells, recovered in fresh LB medium at 37 °C for 1 h, plated on plates containing 100 μg/mL ampicillin and 50 μg/mL kanamycin and grown overnight at 37 °C. The colony transfer method was consistent with the one above, except on ampicillin and kanamycin-containing plates. Colony PCR or qPCR was performed after overnight growth at 30 °C. Integrative copy numbers were analyzed by colony PCR or qPCR.

Transposition Experiments of *Tatumella citrea*. The transposition experiments of *Tatumella citrea* using INTEGRATE system included transformation and induction of pTara:500 (Addgene plasmid #60717),[29] pDonor-GFP, pTnsABC, and pQCascade-tra13. Electrically competent DSM13699 cells were cotransformed with pTara:500, pQCascade -tra13, and pTnsABC, and plated on triple antibiotic LB-agar plates containing 12.5 μg/mL chloramphenicol, 50 μg/mL kanamycin, and 50 μg/mL streptomycin. Then, colonies were inoculated into LB medium to prepare electrically competent cells, and which were transformed with pDonor-GFP by electroporation. Cells were recovered in fresh LB medium at 30 °C for 1 h and then plated onto antibiotic plates containing 12.5 μg/mL chloramphenicol, 100 μg/mL ampicillin, 50 μg/mL kanamycin, and 50 μg/mL streptomycin. After a 20-h growth at 30 °C, colonies were induced on quadruple antibiotic plates supplemented with 0.1 mM IPTG for another 16 h at 30 °C until biofilms formed. A portion of cells were resuspended and rediluted in fresh LB medium and then plated on quadruple antibiotic LB-agar plates supplemented with 0.3 mM IPTG, and grown overnight at 30 °C. These colonies were defined as the first generation of mixed bacteria. The subsequent transfer was performed by scraping, suspending, and diluting these mixed colonies and plating them onto fresh quadruple-antibiotic plates supplemented with 0.3 mM IPTG, grown overnight at 30 °C. Colony PCR was performed to analyze copy numbers.

Transposition Efficiency Analysis by Colony PCR and Sanger Sequencing. Colony PCR products were generated with Taq DNA Polymerase using Green Taq Mix (Vazyme Biotech). The colonies on the plate were individually picked up into 10 μL of H_2O and resuspended as a template. Ten μL reactions contained 5 μL Green Taq Mix, 0.5 μM primers, 0.5 μL template, and H_2O. Thermal cycling conditions for colony PCR were as follows: 1 cycle, 95 °C, 5 min; 26 cycles, 95 °C, 30 s, 57 °C, 30 s, 72 °C, 70 s; 4 °C hold. Primer pairs were designed about 300 bp upstream and 300 bp downstream of the insertion sites so that all possible integration orientations could be detected (Supplementary Table S2). PCR products were resolved by 1% agarose gel electrophoresis and stained by SYBR Safe (Thermo Scientific). Negative control samples were always analyzed in parallel with the experimental samples to identify nonspecific bands. PCR products were sent to GENEWIZ for isolation and analysis.

Transposition Efficiency Analysis by qPCR. All genomic DNA (gDNA) extraction was carried out with the HiPure Bacterial DNA Kit (Magen). Specific primers for qPCR were designed as one located upstream of the insertion site, the other located inside the cargo gene, and the whole amplified fragment is about 220 bp. An *E. coli* reference gene (*gapA*) was chosen for standardization and a pair of primers were designed (Supplementary Table S2). The qPCR reaction (20 μL) contained 10 μL of SsoAdvanced Universal SYBR Green Supermix (BioRad), 20 μM primers and 10 ng genomic extraction, and H_2O. Reactions were performed in 96-well white PCR plates (BioRad) and measured on a CFX384 real-time PCR detection system (BioRad) using the following thermal cycling parameters: 1 cycle, 95 °C, 2 min; 40 cycles, 95 °C, 10 s, 59 °C, 20 s, 72 °C, 20 s; terminal melt-curve analysis (59−95 °C in 0.5 °C/5 s increments).

We first prepared gDNA extraction of BL21(DE3) strain harboring pDonor, pTnsABC, and pQCascade-entry as negative controls. To measure the copy-number of cargos inserted into IS1 sites accurately, we prepared an MG1655 strain (MIS1A (1)) with one copy in one IS1 site for standard sample, which had been identified by colony PCR, and the qPCR primers of IS1 could also be applied in MIS1A (1). In our final qPCR experiments, each biological sample was analyzed by two parallel reactions: one containing the primer pair of the *E. coli* reference gene and the other containing the primer pair of the R/L integration direction. Copy-numbers for all samples were calculated by Pfaffl Method. All the measurements shown in the text and figures were from three separate biological replicates.

pCutamp Construction and Plasmid Curing Experiments. pCutamp (Supplementary Table S1) was constructed by inserting a rhamnose-inducing promoter controlling sgRNA that targeting pDonor, pTnsABC, and pQCascade into the plasmid pCasSac (stored in our laboratory). The spacer of sgRNA was the consensus sequence 5′-TGCTTCAATA-ATATTGAAAAAGG-3′ of the AmpR promoter in the three plasmids. Rhamnose induced transcription of sgRNA, which guided *Streptococcus pyogenes* Cas9 to target and cut the consensus sequences. The targeted plasmids could not be repaired since no homologous template was provided. Strains harboring three plasmids above were transformed with pCutamp by electroporation (1.85 kV, 200 Ohm, 25 μF) and recovered for 1 h in 1 mL SOC medium at 37 °C. To improve the plasmid-curing efficiency, recovered cells were transferred to 4 mL LB medium containing 50 μg/mL apramycin and 10 mM rhamnose, shaking for 3 h at 37 °C. The inducing time was optimized, as shown in Supplementary Figure S8. Then 100 μL of cultures were plated on LB-agar plates containing 50 μg/mL apramycin and 10 mM rhamnose, being incubated overnight at 37 °C.

pCutamp also contains the SacB gene, encoding levansurase that catalyzes the hydrolysis of sucrose and the polymerization of fructan, such that *E. coli* carrying pCutamp cannot grow in plates containing 10 g/L of sucrose. Colonies grown on plates containing apramycin and rhamnose were selected and replated on four kinds of LB-agar plates containing 100 μg/mL ampicillin, 50 μg/mL kanamycin, 50 μg/mL streptomycin, and 10 g/L sucrose, respectively. The colonies that could not grow on the plates containing ampicillin, kanamycin, and streptomycin were considered to have eliminated the pDonor,

https://dx.doi.org/10.1021/acssynbio.0c00073
ACS Synth. Biol. XXXX, XXX, XXX−XXX

pTnsABC, and pQCascade. Then the corresponding colonies grown on the sucrose plate would be replated on nonselective LB agar plates and apramycin-contained plates to verify the curing of pCutamp. Colonies that could not grow in apramycin-containing plates were plasmid-free multicopy integrated strains.

In CAST system, pCutamp can only target the series of pHelper_ShCAST_sgRNA plasmids, such as pHelper-IS1, but not pDonor_ShCAST_kanR. To stop transposition and obtain more genetic stable strains, we use pCutamp for pHelper-IS1 curing by the same protocol above, except for changing of corresponding antibiotics. The same method above was used for pCutamp curing.

Fermentation and GDH Activity Analysis. Flask fermentation was performed as follows: *E. coli* BL21(DE3) integrated with multicopy-GDH cassette or harboring pET24a-GDH were grown in 3 mL of LB liquid medium overnight at 37 °C. Then, 300 μL of cultures above were inoculated into 30 mL of Terrific Broth (TB) medium in 250 mL shake flasks, respectively. Cells were cultured at 37 °C to OD_{600} = 2−4, at which point IPTG was added to a final concentration of 0.3 mM, and fermentation was allowed to continue at 28 °C for additional 40 h.

The GDH expression strains that had been cultured at 37 °C, 220 rpm for 16 h were inoculated into a 500 mL shake flask containing 100 mL of TB medium, and cultured at 37 °C, 220 rpm for 5 h. In the scale-up fermentation experiments, the above 100 mL culture solution was all inoculated into a 7.5-L fermenter containing 4 L of TY medium[26] at 37 °C, 800 rpm, and aerated at 2.5 vvm. Agitation is steadily increased to a maximum of 800 rpm to reach 20% dissolved oxygen (DO). When the DO was further increased to more than 60%, 400 g/L glycerol, 50 g/L yeast extract, and 25 g/L tryptone were added into the medium at an aeration rate of 2.5 vvm, 800 rpm. When the OD_{600} reached 20, 0.2 mM IPTG was added to the medium and incubate at 28 °C for 43 h.

To measure GDH activity, 1 mL of fermentation culture was collected by centrifuging at 1500*g* for 10 min at 4 °C. The culture supernatant was discarded, and cells were resuspended using 1 mL of 100 mM Tris-HCl (pH 7.5). The cell suspensions were lysed by sonication on ice: 200 W, working 3 s, interval 5 s for 40 times, the resulting liquid was the crude enzyme solution. The reaction mixture contained 100 mM glucose, 100 mM Tris-HCl (pH 7.5), 2 mM NAD^+, and an 8000-fold diluted crude enzyme solution, and the enzyme activity was assayed at 25 °C. Following the increase of the absorbance at OD_{340}, the enzyme activity was calculated by the formula $U/mL = [\Delta A/min] \times [1/\varepsilon] \times 8000$ ($\varepsilon = A/cL$ = 6.402 mL/(μmol × cm)).[35]

Next-Generation Sequencing (NGS) and Data Analysis. Paired-end libraries for whole-genome sequencing were prepared using Nextera DNA Flex Library Prep Kit or TruSeq DNA PCR-Free Library Prep Kits. At least 7 000 000 read-pairs were obtained from each sample, for a total length of 2.1 Gb (450-fold average coverage of the *E. coli* genome), using Illumina Novaseq systems (performed by Novogene China, Genewiz China or CHGC Shanghai China).

To determine the integration sites and their distribution for a given sample from the whole genome sequencing (WGS) data, the following steps were performed using the split-read mapping or discordant read-pair mapping approaches[36] (using custom Bash scripts).

Split-read mapping approach: First, the reads from WGS that can be mapped to transposon end sequence were picked up to reduce the amount of data using BWA mem.[37] Thereafter, reads from WGS data were filtered based on the requirement that one segment maps to some location in the reference genome and the remaining segment maps to the transposon end sequence with 20 bp perfectly matching transposon end sequence, using LAST.[38] Finally, The picked-up reads were mapped to the reference genome using BWA bwasw[37] for identifying the exact insertion location and determine the integration site distributions.

Discordant read-pair mapping approach: The read-pairs from WGS that only one end can be mapped to transposon end sequence were considered discordant and were picked up using BWA mem.[37] The reads unmapped to transposon end sequence were then mapped to the reference genome using BWA bwasw for identifying the insertion locations.

To determine whether new mutations were introduced during the operation, SNPs/Indels were detected using the snippy (https://github.com/tseemann/snippy) by mapping WGS reads to the reference genome.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.0c00073.

Figure S1: Integration of cargo into 5-copy IS1 loci in *E. coli* Pir1 strain based on the ShCAST system; Figure S2: Schematic of genome sequencing of integrated colonies; Figure S3: Targeting 5-copy IS1 loci in *E. coli* Pir1 by the ShCAST system; Figure S4: Genomic insertion at 5-copy IS1 loci in *E. coli* Pir1 using the ShCAST system after 8 rounds of transfer; Figure S5: INTEGRATE system enables multicopy cargo integration on *E. coli* BL21(DE3) chromosome; Figure S6: Integration of cargo gene guided by crRNA-array2 using the ShCAST system in *E. coli* Pir1; Figure S7: Integration of cargo gene guided by crRNA-array2 using the ShCAST system in *E. coli* Pir1 after restreaking and plasmid-curing; Figure S8: Replacing crRNA-array1 with crRNA-IS186 to achieve more than 8 copies on the chromosome; Figure S9: Verifying the function of the T7 transcription system in nonproductive cells by transforming pET24a-GFP; Figure S10: Genomic analysis of *T. citrea* strain Tc-2 that has insertions in all 13 tra13 loci; Figure S11: Genomic schematics of 17 important strains in this study and sequences flanked in the off-target sites; Figure S12: Confirmation of the copy-number of pure strain IS1Cm-16A using site-specific PCR; Figure S13: Targeting 29-copy IS1 loci in *E. coli* BL21(DE3) using the updated ShCAST system; Figure S14: Verifying the function of the T7 transcription system in strains that carry 7−10 sites of genomic GDH-insertion by transforming pET24a-GFP; Figure S15: The growth, enzyme activity, and protein expression of strains that carry 7−10 sites of genomic GDH-insertion and without or with the plasmid pET24a-GDH; Table S1: Strains and plasmids used in this study; Table S2: Primers used in this study; Table S3: Spacers and target sites used in this study; Table S4: Distribution of insertion sites of GDH strains with different copy numbers; Table S5: List of SNPs found

https://dx.doi.org/10.1021/acssynbio.0c00073
ACS Synth. Biol. XXXX, XXX, XXX−XXX

between genomes of the *E. coli* BL21(DE3) (CP001509) and ArGDH-3D strain (PDF)

AUTHOR INFORMATION

Corresponding Authors

Junjie Yang − *Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China*; Email: yangjunjie@sibs.ac.cn

Sheng Yang − *Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China; Huzhou Center of Industrial Biotechnology, Shanghai Institutes for Biological Sciences, Huzhou 313000, China*; orcid.org/0000-0003-3742-9989; Email: syang@sibs.ac.cn

Authors

Yiwen Zhang − *Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China; University of Chinese Academy of Sciences, Beijing 100049, China*

Xiaoman Sun − *School of Food Science and Pharmaceutical Engineering, Nanjing Normal University, Nanjing 210046, China*

Qingzhuo Wang − *Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China; University of Chinese Academy of Sciences, Beijing 100049, China*

Jiaqi Xu − *Institute of Bioengineering, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China*

Feng Dong − *Huzhou Center of Industrial Biotechnology, Shanghai Institutes for Biological Sciences, Huzhou 313000, China*

Siqi Yang − *Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China; University of Chinese Academy of Sciences, Beijing 100049, China*

Jiawei Yang − *School of Food Science and Pharmaceutical Engineering, Nanjing Normal University, Nanjing 210046, China*

Zixu Zhang − *School of Food Science and Pharmaceutical Engineering, Nanjing Normal University, Nanjing 210046, China*

Yuan Qian − *Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China*

Jun Chen − *Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China*

Jiao Zhang − *Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China; University of Chinese Academy of Sciences, Beijing 100049, China*

Yingmiao Liu − *Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China*

Rongsheng Tao − *Huzhou Center of Industrial Biotechnology, Shanghai Institutes for Biological Sciences, Huzhou 313000, China*

Yu Jiang − *Shanghai Taoyusheng Biotechnology Co., Ltd, Shanghai 201203, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.0c00073

Author Contributions

Y.W.Z., J.J.Y., and S.Y. designed the overall method and approach. Y.J., J.J.Y., and S.Y. supervised the research. Y.W.Z. developed and performed transferring method and made the strain library. X.M.S., Z.X.Z., J.W.Y., and R.S.T. performed the fermentation experiments. J.Q.X., F.D., and Y.Q. performed experiments in ShCAST system. Q.Z.W. and J.Z performed experiments in *T. citrea*. S.Q.Y. and J.C. performed PCR and qPCR experiments. Y.M.L. was responsible for constructing plasmid pCutamp. Y.W.Z., J.W.Y., and S.Y. wrote the manuscript. All authors read and commented the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (Number: 21825804, 31921006), the National Science & Technology Major Project "Key New Drug Creation and Manufacturing Program", China (Number: 2018ZX09711002), and the Natural Science Foundation of Shanghai (18ZR1446800, 18ZR1446500).

REFERENCES

(1) Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., and Robins, K. (2012) Engineering the third wave of biocatalysis. *Nature 485* (7397), 185−194.

(2) Nakamura, T., Koma, D., Oshima, M., Hoshino, H., Ohmoto, T., and Uegaki, K. (2018) Application of chromosomal gene insertion into Escherichia coli for expression of recombinant proteins. *J. Biosci Bioeng. 126* (2), 266−272.

(3) 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.

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

(5) Corchero, J. L., and Villaverde, A. (1998) Plasmid maintenance in Escherichia coli recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnol. Bioeng. 58* (6), 625−32.

(6) Lin-Chao, S., and Bremer, H. (1986) Effect of the bacterial growth rate on replication control of plasmid pBR322 in Escherichia coli. *Mol. Gen. Genet. 203* (1), 143−9.

(7) Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., and Yang, S. (2015) Multigene editing in the Escherichia coli genome using the CRISPR-Cas9 system. *Appl. Environ. Microbiol. 81*, 2506.

(8) Tyo, K. E. J., Ajikumar, P. K., and Stephanopoulos, G. (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat. Biotechnol. 27* (8), 760−U115.

(9) Vandermeulen, G., Marie, C., Scherman, D., and Preat, V. (2011) New generation of plasmid backbones devoid of antibiotic resistance marker for gene therapy trials. *Mol. Ther. 19* (11), 1942−9.

(10) Akhverdyan, V. Z., Gak, E. R., Tokmakova, I. L., Stoynova, N. V., Yomantas, Y. A. V., and Mashko, S. V. (2011) Application of the bacteriophage Mu-driven system for the integration/amplification of target genes in the chromosomes of engineered Gram-negative bacteria-mini review. *Appl. Microbiol. Biotechnol. 91* (4), 857−871.

(11) Koonin, E. V., and Makarova, K. S. (2017) Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back. *Genome Biol. Evol. 9* (10), 2812−2825.

(12) Krupovic, M., Beguin, P., and Koonin, E. V. (2017) Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol. 38*, 36−43.

(13) Peters, J. E., Makarova, K. S., Shmakov, S., and Koonin, E. V. (2017) Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl. Acad. Sci. U. S. A. 114* (35), E7358−E7366.
(14) Peters, J. E. (2019) Targeted transposition with Tn7 elements: safe sites, mobile plasmids, CRISPR/Cas and beyond. *Mol. Microbiol. 112*, 1635.
(15) Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., and Zhang, F. (2019) RNA-guided DNA insertion with CRISPR-associated transposases. *Science 365* (6448), 48−53.
(16) Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., and Sternberg, S. H. (2019) Transposon-encoded CRISPR−Cas systems direct RNA-guided DNA integration. *Nature 571* (7764), 219−225.
(17) Halpin-Healy, T. S., Klompe, S. E., Sternberg, S. H., and Fernandez, I. S. (2020) Structural basis of DNA targeting by a transposon-encoded CRISPR-Cas system. *Nature 577*, 271.
(18) Saedler, H., and Heiss, B. (1973) Multiple copies of the insertion-DNA sequences IS1 and IS2 in the chromosome of E. coli K-12. *Mol. Gen. Genet. 122* (3), 267−77.
(19) Sousa, A., Bourgard, C., Wahl, L. M., and Gordo, I. (2013) Rates of transposition in Escherichia coli. *Biol. Lett. 9* (6), 20130838.
(20) Siguier, P., Gourbeyre, E., and Chandler, M. (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev. 38* (5), 865−91.
(21) Tseng, H. C., Harwell, C. L., Martin, C. H., and Prather, K. L. (2010) Biosynthesis of chiral 3-hydroxyvalerate from single propionate-unrelated carbon sources in metabolically engineered E. coli. *Microb. Cell Fact. 9*, 96.
(22) Tietze, E., Tschepe, H., Golubev, A. V., and Smirnov, G. B. (1988) [Characteristics of transposition of Tn7-like elements determining resistance to trimethoprim and streptomycin]. *Mol. Genet. Mikrobiol. Virusol.* No. 10, 10−17.
(23) Wong, C. H., and Drueckhammer, D. G. (1985) Enzymatic-Synthesis of Chiral Hydroxy Compounds Using Immobilized Glucose-Dehydrogenase from Bacillus-Cereus for Nad(P)H Regeneration. *Nat. Biotechnol. 3* (7), 649−651.
(24) Shaw, N. M., Robins, K. T., and Kiener, A. (2003) Lonza: 20 years of biotransformations. *Adv. Synth. Catal. 345* (4), 425−435.
(25) Jin, L. Q., Xu, W., Yang, B., Liu, Z. Q., and Zheng, Y. G. (2019) Efficient Biosynthesis of Xylitol from Xylose by Coexpression of Xylose Reductase and Glucose Dehydrogenase in Escherichia coli. *Appl. Biochem. Biotechnol. 187* (4), 1143−1157.
(26) Tao, R., Jiang, Y., Zhu, F., and Yang, S. (2014) A one-pot system for production of L-2-aminobutyric acid from L-threonine by L-threonine deaminase and a NADH-regeneration system based on L-leucine dehydrogenase and formate dehydrogenase. *Biotechnol. Lett. 36* (4), 835−41.
(27) Brady, C. L., Venter, S. N., Cleenwerck, I., Vandemeulebroecke, K., De Vos, P., and Coutinho, T. A. (2010) Transfer of Pantoea citrea, Pantoea punctata and Pantoea terrea to the genus Tatumella emend. as Tatumella citrea comb. nov., Tatumella punctata comb. nov. and Tatumella terrea comb. nov. and description of Tatumella morbirosei sp. nov. *Int. J. Syst. Evol. Microbiol. 60* (3), 484−494.
(28) Grindley, J. F., Payton, M. A., van de Pol, H., and Hardy, K. G. (1988) Conversion of Glucose to 2-Keto-l-Gulonate, an Intermediate in l-Ascorbate Synthesis, by a Recombinant Strain of Erwinia citreus. *Appl. Environ. Microbiol. 54* (7), 1770−5.
(29) Shis, D. L., and Bennett, M. R. (2013) Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc. Natl. Acad. Sci. U. S. A. 110* (13), 5028−33.
(30) Balbas, P., Soberon, X., Bolivar, F., and Rodriguez, R. L. (1988) The plasmid, pBR322. *Biotechnology. 10*, 5−41.
(31) Balbas, P., Soberon, X., Merino, E., Zurita, M., Lomeli, H., Valle, F., Flores, N., and Bolivar, F. (1986) Plasmid vector pBR322 and its special-purpose derivatives–a review. *Gene 50* (1−3), 3−40.
(32) Chang, A. C., and Cohen, S. N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol. 134* (3), 1141−56.
(33) Gamas, P., Burger, A. C., Churchward, G., Caro, L., Galas, D., and Chandler, M. (1986) Replication of pSC101: effects of mutations in the E. coli DNA binding protein IHF. *Mol. Gen. Genet. 204* (1), 85−9.
(34) Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., and Sternberg, S. H. (2019) Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature 571* (7764), 219−225.
(35) Bucher, T., Krell, H., and Lusch, G. (1974) Molar extinction coefficients of NADH and NADPH at Hg spectral lines. *Z. Klin. Chem. Klin. Biochem. 12* (5), 239−240.
(36) Ewing, A. D. (2015) Transposable element detection from whole genome sequence data. *Mobile DNA 6*, 24.
(37) Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics 25* (14), 1754−60.
(38) Kielbasa, S. M., Wan, R., Sato, K., Horton, P., and Frith, M. C. (2011) Adaptive seeds tame genomic sequence comparison. *Genome Res. 21* (3), 487−93.

https://dx.doi.org/10.1021/acssynbio.0c00073
ACS Synth. Biol. XXXX, XXX, XXX−XXX