A New Era of Genome Integration—Simply Cut and Paste!

Zihe Liu,‡§ Youyun Liang,‡§ Ee Lui Ang,†‡§ and Huimin Zhao*†‡§

†Metabolic Engineering Research Laboratory, Science and Engineering Institutes, Agency for Science, Technology and Research, Singapore
‡Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

ABSTRACT: Genome integration is a powerful tool in both basic and applied biological research. However, traditional genome integration, which is typically mediated by homologous recombination, has been constrained by low efficiencies and limited host range. In recent years, the emergence of homing endonucleases and programmable nucleases has greatly enhanced integration efficiencies and allowed alternative integration mechanisms such as nonhomologous end joining and microhomology-mediated end joining, enabling integration in hosts deficient in homologous recombination. In this review, we will highlight recent advances and breakthroughs in genome integration methods made possible by programmable nucleases, and their new applications in synthetic biology and metabolic engineering.

KEYWORDS: genome integration, programmable nucleases, synthetic biology, metabolic engineering

Genome integration, which involves the precise and stable insertion of functional DNA parts into the chromosome of a host organism, is of central importance in both basic and applied biological research. It allows in-frame tagging of gene of interest (GOI) and facilitates the generation of reporter cell lines to address key biological questions. Moreover, compared to plasmid expression systems, integrating heterologous genes or biochemical pathways into chromosomes of host organisms reduces metabolic burden, allows precise regulation of metabolic fluxes, and provides genetic stability. The resulting recombinant organism is more suitable for industrial applications, especially in long-term fed-batch or chemostat cultures without costly antibiotic selection.

However, genome integration is traditionally associated with low efficiencies even with selection markers. In addition, integration capabilities (efficiencies, integrated copy numbers, accuracy, etc.) often decrease rapidly when integrating larger or multiple cassettes. Some endeavors to increase genome integration efficiencies include optimization of integration sites, elevation of selection pressure, and most notably—generation of double stranded breaks (DSBs) in the chromosomes by homing endonucleases (HEs) and programmable nucleases.

The discovery of HEs and programmable nucleases has been a game changer in terms of genome integration capabilities. Integration efficiencies can be increased by several orders of magnitude when DSBs are generated in the chromosomes. Some endeavors to increase genome integration efficiencies include optimization of integration sites, elevation of selection pressure, and most notably—generation of double stranded breaks (DSBs) in the chromosomes by homing endonucleases (HEs) and programmable nucleases.

Moreover, as they can only cut their cognate DNA target sites, a tedious workflow involving the preintegration of HE target site is required. More recently, programmable endonucleases, such as Zinc Finger Nucleases (ZFN), Transcription Activator-Like Effector Nucleases (TALEN), Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated system (CRISPR/Cas9), have been employed for genome integration. As in-depth comparisons of different nuclease systems are covered in several recent reviews, they will not be addressed here. The high integration efficiencies and ease of use provided by programmable nucleases enable marker-less integration, multiplicity integration of large DNA cassettes, in vivo assembly of large synthetic pathways, and precise integration in a wide range of cell types and organisms. Moreover, the generation of DSBs allows various repair mechanisms to be used to achieve unprecedented genome integration capabilities and host range.

This review describes recent advances in genome integration methods made possible by programmable nucleases and the application of these methods in synthetic biology and metabolic engineering. Integration via common DNA repair mechanisms including homologous recombination (HR), nonhomologous end joining (NHEJ), and microhomology-mediated end joining (MMEJ) will be addressed. Other integration methods that have benefited significantly from the use of programmable nucleases, such as single-stranded oligodeoxyribonucleotides integration and recombination will also be covered. Finally, selected applications in which genome
integrations go beyond single gene integration will be highlighted.

**KEY ELEMENTS IN GENOME INTEGRATION**

**Selection Methods for Genome Integration.** In genome integration, selection markers have routinely been used to ensure successful integration events. Selection systems for genome integration are mostly based on antibiotic or auxotrophic marker systems and are well-characterized in many model organisms.32–34 Recently, a variety of new selection systems in less characterized organisms were developed, including the thymidine kinase gene (tdk) which confers sensitivity to S-fluoro-2′-deoxyuridine in *Thermoanaerobacter ethanethiolicus*,35 the hypoxanthine phosphoribosyl transferase gene (*hpt*) which confers sensitivity to 8-azaguanine in *Clostridium thermocellum*,36 the a CoA-ligase gene (*acsA*) for acrylate counter-selection in cyanobacterium *Synechococcus*,37 two genes encoding orotidine phosphoribosyltransferase (*pyrE*) and orotidine 5-phosphate decarboxylase (*pyrF*) which convert S-fluorourotic acid to S-fluoro-UMP in *Staphylococcus*,38 and the pregallidimer protease gene (*gdnP*) which converts pregallidimer to gallidimer in *Staphylococcus*.39

While selection markers used to be instrumental to successful genome integration, it has become dispensable with the advent of programmable nucleases. In nuclease-assisted integrations, successful integration of donor DNA at the target sites protects chromosomal DNA from being cut while unmodified target sites are cleaved leading to growth arrest; thereby providing a convenient selection platform for successful integration, and paving the way for markerless integration.40,41

**Effects of Different Integration Loci.** Generally, the integration locus is selected based on whether it will support high gene expression without causing any significant growth impact. It has been shown in multiple studies that in the absence of nucleases, integration efficiencies and expression levels of integrated genes vary greatly with different integration loci even in the same organism. Mikkelsen *et al.* evaluated 14 yeast integration loci for genome integration using the lacZ gene as a reporter. A 2.5-fold difference of the specific β-galactosidase activity was reported between different loci. Moreover, 3 out of 14 loci showed significant fitness defects to the cell growth, with only 46%, 58% and 81% of the reference strain’s growth rates, respectively.42 Fang *et al.* characterized 14 different loci in yeast *Saccharomyces cerevisiae*, including URA3, MET15, LEU2, TRP1, seven Ty elements, and one Ty3 element.8 The levels of expression at different loci were comparable with approximately 50% variation and were around 60% of plasmid-based expression using a CEN/ARS vector. Flagfeldt *et al.* characterized *lacZ* expression under control of two strong promoters at 20 loci in *S. cerevisiae*: URA3, SPB1/PPN1, PDC6, and 17 different solo long terminal repeats (LTRs).33 Up to an 8-fold variation was reported with gene integration at the LTR locations, and the results were consistent for both promoters. It was postulated that the observed differences were due to epigenetic effects with sites closest to the telomeres possessing reduced expression levels.

With the adoption of programmable nucleases, an added factor to consider in addition to the above-mentioned epigenetic effects is differences in cleavage efficiency at different integration loci and potential off-target effects at different sites. For instance, with the widely used CRISPR/Cas9 system, studies have shown that cleavage efficiencies and off-target effects varied widely with different loci and different guide RNAs.44,45 To improve genome editing outcome and predictability, several groups have developed design rules and algorithms to enhance cleavage efficiencies.46–48

**Effects of Different Double-Stranded Break Mechanisms.** Since there are multiple mechanisms to generate DSBs, it is desirable to compare genome integration efficiencies associated with these mechanisms. Lee *et al.* tested both I-SceI and CRISPR/Cas9 on their effects on genome integration.49 In the study, CRISPR/Cas9-assisted DSBs increased the integration efficiency by 2-fold compared to unassisted integration, whereas a 5-fold increase was achieved using I-SceI. However, as the comparison was only done for a single locus, we cannot conclude that I-SceI is a superior nuclease to CRISPR/Cas9 for genome integration. In another study, comparison of TALEN and CRISPR/Cas9-mediated integration revealed no clear winner with each outperforming the other under different integration strategies.49 In addition to the relative integration efficiencies, other considerations that might be more pertinent when selecting the appropriate nuclease system include ease of application and cellular toxicities.

**Effects of the Inset Size and the Size of the Genomic Region to Be Replaced.** In studies investigating integration with simultaneous multiple knockouts, it was found that replacement efficiencies generally decreased with increasing sizes of the integration insets and the genomic regions to be replaced.50,51 Pyne *et al.* used CRISPR/Cas9 together with the ζ-Red recombination system to test the integration efficiency in *Escherichia coli*.50 In this work, the integration efficiencies of different sized cassettes (550, 560, 1264, 1756, 2492 and 3000 bp), all with the same 40 bp homology arms, decreased from 47% to 1%; and the replacement efficiencies for genomic regions with different sizes (8, 818, 2428, 5123, 9590, 11068 and 19378 bp), all with a 560 bp lacZ reporter gene and 40 bp homology arms, decreased from 47% to 3%. In all cases, it is
notable that almost no integration were observed without the CRISPR/Cas9 plasmid, thus highlighting the benefits of nuclease-assisted integrations.

■ INTEGRATION VIA HOMOLOGOUS RECOMBINATION

Mechanism of HR. HR is the most studied and widely used DSBs repair mechanism for genome integrations. Generally, DSBs first undergo end resection to yield duplexes with 3'-ended single stranded tails (Figure 1A). Then, after gap-repair DNA synthesis and ligation, the DSBs are repaired at Holiday junctions in either crossover or noncrossover mode. HR is an error-free mechanism that repairs damaged loci with donor templates containing homology arms.59,62 Methods to increase the efficiency of HR-mediated genome integration are discussed below.

Suppression of NHEJ. In some organisms, the frequency of HR is less than that of NHEJ,53 which might result in random and erroneous integration of only part of the construct.54 This problem can be partially addressed by suppression of host NHEJ, or by increasing the HR efficiency with long homology arms or HR enhancers.55,56 Verbeke et al. disrupted the ku70 gene required for the NHEJ pathway in Yarrowia lipolytica57 and reported a 100-fold increase of the accuracy compared to the reference strain. Chu et al. reported that in both human and mouse cell lines, suppression of NHEJ-associated proteins Ku70 and DNA ligase IV increased the integration efficiency by 4 to 5-fold, and coexpression of adenovirus E1B55K and E4orf6 ligase IV-degrading proteins together with Cas9 improved the genome integration efficiency by 8-fold.58

Extension of the Homology Arms. While ~50 bp homology arms are sufficient for most single gene integrations in S. cerevisiae, 500 to 5000 bp homology arms might be required in nonconventional yeast or other organisms.22,29 For example, Ishii et al. reported that the integration efficiency in human cells correlates well with the length of homology arms.60 The integration efficiency of a puromycin-resistant gene using ~4 kb homology arms was more than twice of the one using ~1 kb homology arms.

Design of Donor DNA. It has been suggested that the slow dissociation of Cas9 (~6h) from cleaved double-stranded DNA impeded HR-mediated repair and reduced HR efficiencies.61 Further mechanistic studies revealed that the 3' end of the strand noncomplementary to the single-guide RNA (sgRNA) is released earlier than the complementary DNA. The authors exploited the asymmetric release by using defined single-stranded DNA (ssDNA) donors that were complementary to the released strand (known as target-strand donor).63 With this strategy, HR efficiencies of up to 60% could be obtained in human cells, and it is 2.6-fold higher than using the complementary nontarget ssDNA donor.

Use of a Chemical HR Enhancer. Through screening a 10 000-compound library, Jayathilaka et al. identified a small molecule RS-1 that can stimulate the human HR protein Rad51.64 Later on, Song et al. reported that HR enhancer RS-1 increased the knock-in efficiency by 2 to 5-fold at different loci through in vitro studies, whereas NHEJ inhibitor SCR7 had minor effects. RS-1 was also applied in mammalian cells and showed multifold improvement on knock-in efficiencies.65

Tuning of the Antibiotic Concentrations. Shi et al. reported a 6-fold increase of fatty acid ethyl esters production through integration in yeast delta-sequences.7 The clones were identified through selection using increased antibiotic concentra-

trations after repeated transformations. Yuan et al. reported an in vivo assembly method that took advantage of yeast delta-integration coupled with antibiotic selection.35 The method was demonstrated through simultaneous integration of a five-gene isobutanol pathway together with an eight-gene mevalonate pathway.

■ INTEGRATION VIA NONHOMOLOGOUS END JOINING

Mechanism of NHEJ. NHEJ is present in all eukaryotes and many prokaryotes, and is critical for preserving genome integrity. In eukaryotes, it is initiated through the binding of the heterodimer Ku80-Ku70 which further recruits other end-processing proteins to initiate cofactor Xrcc4-dependent ligation by ligase IV (Figure 1B).66 In prokaryotes, a different set of proteins mediates a similar process involving end protection, processing, and ligation.66 As its name suggests, NHEJ does not require a homologous template and has been regarded as an error-prone process due to the indels and even translocations that are observed at the sites of repair. As such, NHEJ is traditionally employed for generating frame-shifting knockouts. However, recent studies have revealed that error-free NHEJ events can occur at high frequencies of up to 100%, with end-processing and deletions only arising in the absence of directly ligatable DNA ends.30,65

NHEJ Applications. This paradigm shift has motivated the utilization of NHEJ for genome integrations. In a typical workflow, DSBs are generated by chosen nucleases in the presence of linear DNA donors which can be directly transfected, transformed, or introduced through intracellular digestion of circular plasmids. Advantages of NHEJ-integration include higher efficiencies and greater ease of donor construction, as long homologous ends are not required.66,67 The usefulness of NHEJ-mediated integration is most apparent in cell types such as embryonic stem cells, induced pluripotent stem cells, zebrafish embryos, diploid frog Xenopus tropicalis embryos, and Chinese hamster ovary (CHO) cells, where HR is nonexistent or extremely inefficient.53,67–70 Another major advantage is the ease of integrating large DNA constructs which can be challenging with traditional HR-based approaches. Studies comparing integration efficiencies of 15 kb DNA constructs found that NHEJ integration efficiencies were higher than HR in all the tested cell lines.68,69 However, a more recent study found that HR can be the dominant mode of repair in some instances and the ratio of HR versus NHEJ repair frequencies was dependent on gene locus, nuclease platform and cell type.71 Aside from routine genome integration, NHEJ-integration has also recently been adopted for synthetic biology applications where it enabled the rapid integration-mediated modification of pseudorabies virus (PRV) by simple cotransfection of PRVs genome with Cas9 complex and customized DNA cassettes.72

Increasing Precision and Efficiency of NHEJ-Mediated Integration. Although NHEJ-based integration holds a lot of promise, it is not without limitations. Unlike HR-based approaches, NHEJ-integrations are not multiplexable as there are no homologies to direct donor DNAs to their specific loci and orientations. Furthermore, while some of the small deletions inherent to NHEJ might be tolerated in integrations at noncoding regions, these deletions impede precise modifications such as in-frame tagging of coding regions.

To reduce NHEJ-associated errors, two novel strategies were investigated in a recent study. The first approach involved
fusion of a destabilization domain to Cas9 to facilitate dissolution of the tightly bound sgRNA-Cas9 complex from cleaved DNA (Figure 2). This served to expedite NHEJ before significant degradation to the donor DNA could occur. With this approach, integration efficiency was improved and precise integration was successfully enhanced from 40% to 60%. In a second related strategy, protection of donor DNA ends with phosphorothioate bonds also minimized undesired end-processing of the donor DNA and improved the efficiency of precise ligations.

**INTEGRATION VIA MICROHOMOLOGY-MEDIATED END JOINING**

**Mechanism of MMEJ.** MMEJ is a DSB repair mechanism that uses microhomologous sequences (5–25 bp). The process is initiated by end-resection to expose ssDNAs, followed by annealing of short microhomologous regions, flap trimming, fill-in, and ligation (Figure 1C). Unlike NHEJ, it is independent of Ku-80 and Xrc4. MMEJ is traditionally known as a mutagenic mode of repair as it can utilize microhomologies distant from the DSBs, resulting in deletions at the repair junctions. However, with emerging studies, it is now known that MMEJ-mediated repair can be surprisingly accurate in the presence of compatible ends.

**MMEJ Applications.** Similar to NHEJ-mediated integrations, donor construction for MMEJ is relatively straightforward as the microhomologies of 5–25 bp are readily introduced by polymerase chain reaction, bypassing the long workflow typically required for attaching long homologous arms. As a strong advocate of MMEJ-mediated integrations, Yamamoto’s laboratory developed a series of CRIS-PITCH (Precise Integration into Target Chromosome) and TAL-PITCH vectors for CRISPR/Cas9 and TALEN assisted integration in silkworms, frogs, frog embryos, zebrafish, CHO cells, HEK293T cells, and HeLa cells. Others have also used MMEJ-integration for knock-ins in a range of organisms such as Aspergillus fumigatus, and Leishmania donovani in which genome integration is challenging due to nonexistent or inefficient HR. To simplify the workflow, all-in-one vectors encoding Cas9, multiple sgRNAs, and integration cassettes have also been developed to facilitate single-step, efficient integration of up to seven expression cassettes in their unique integration loci. From these studies, MMEJ-mediated integrations appear to be a nice complement to HR- and NHEJ-based methods, offering multiplexability, simplified workflow, good efficiencies, and precision in many instances.

**Increasing Precision and Efficiency of MMEJ-Mediated Integration.** One challenge in MMEJ-mediated integration is the occurrence of competing NHEJ repairs even with donor DNA containing microhomologies, resulting in unpredictable modification outcomes. To overcome this, a ku80-deletion strain of *A. fumigatus* was used for MMEJ-mediated insertion of GFP-protein tags. In this study, it was found that NHEJ was completely abolished in the absence of Ku80, whereas MMEJ-integration of insertion cassettes bearing microhomologous ends was highly efficient and precise. In a separate study, ligase 3 and ligase 4 were deleted separately to investigate their roles in both NHEJ- and MMEJ-mediated integration in zebrafish. The deletion of ligase 4 inhibited NHEJ, significantly increasing MMEJ efficiency whereas deletion of ligase 3 greatly reduced MMEJ. Taken together, these results suggest that ligase 3 but not ligase 4 is essential for MMEJ in zebrafish.

**INTEGRATION VIA ALTERNATIVE MECHANISMS**

Integration via Single-Stranded Oligodeoxyribonucleotides (ssODNs). In addition to double-stranded genome editing DNA templates, single-stranded oligodeoxyribonucleotides (ssODNs) have also been employed for homology-directed repairs in mammalian cells. While the invasion mechanism of ssODNs integration is still not fully elucidated, it has been postulated that ssODNs are incorporated at replication forks possibly as “pseudo-Okazaki-fragments”. Although ssODN-based integrations were reported more than a decade ago, significant enhancements in integration efficiencies from 1% to 30% were only achieved recently with the introduction of DSBs. This could be attributed to the deleterious nature of DSBs, providing selection pressure against unmodified DNA sequences. Numerous advantages for ssODNs donors over double-stranded DNA (dsDNA) have been proposed. These include ease of synthesis, more efficient integration and greater accuracy. Recently, with the aid of programmable nucleases, the applications of ssODNs have been expanded from integration of small elements such as tags and restriction sites to integration of large DNA fragments. For example, by combining CRISPR/Cas9 with ssODNs, Yoshimi et al. reported the integration of a 5.5 kb CAG-GFP vector into the rat zygotes Rosa26 locus with two 80-bp ssODNs, the replacement of the rat Sirpa gene with a 200 kb bacterial artificial chromosome (BAC) containing the human Sirpα.
locus, and the replacement of the rat 58 kb Cyp2d cluster with a 6.2 kb human CYP2D6 gene.86

Integration via Nuclease-Assisted Recombination. The most widely used recombination system is the phage lambda-derived λ-Red recombinase or “recombineering” system which comprises of Gam, Exo, and Beta. Gam protects linear dsDNA from degradation, Exo chews back dsDNA to reveal complementary ssDNA, while Beta binds to exposed ssDNA to facilitate annealing.88 Recombineering only requires short homologies of 35 bp or less and is also readily implemented with ssDNAs. However, the low efficiency of recombineering (<1%) necessitates the incorporation of selection markers in dsDNA donors and the screening of large number of clones in the case of ssDNA donors. In addition, recombineering is more suitable for small genetic modifications and has limited utility for integration of large DNA fragments. To overcome this limitation, the HE I-ScI and programmable nucleases such as CRISPR/Cas9 have been coupled with λ-Red recombination to increase integration efficiencies by selecting against unmodified DNA sequences.12,41,89 With this approach, the scarless integration of up to 9 kb in E. coli has been reported.90

Another common class of recombinases is site-specific recombinases (SSRs) such as tyrosine recombinases and serine recombinases. Tyrosine and serine recombinases such as Cre, Flp and Fc31 integrate promote recombination in specific loci between particular short sequences (about 12 to 34 bp).91,92 Integration mediated by SSRs does not provoke error-prone DNA repair processes, and is thus generally with high accuracy.93 However, the nonprogrammable DNA target specificity limits the application of these enzymes. To overcome this limitation and increase the number of the sites amenable to targeted recombination, the hyper-activated catalytic domains of small serine recombinases have been fused to programmable zinc finger, TALE and CRISPR/Cas DNA-binding proteins.93–95 Using this approach, the scarless integration of EGFP in three safe-harbor loci in human HEK293T cells has been reported.93

■ CHOOSING THE RIGHT INTEGRATION PLATFORM

As described in the earlier sections, the emergence of HEs and programmable nucleases has greatly expanded the array of integration platforms available. Some of the key considerations when selecting the appropriate integration platform include the host and cell type, size of insets, simplicity of workflow, target efficiency and accuracy. Generally speaking, HR mediated genome integration strategies are always with a newly studied organism with less-characterized genetic background. If they prove to be low efficiency, other genome integration methods discussed in this review can be tested. To facilitate the selection of the right integration platform, the requirement, host range and performance of each integration method has been summarized in Table 1.

<table>
<thead>
<tr>
<th>Integration Method</th>
<th>Homology Requirement</th>
<th>Donor Size</th>
<th>Multiplexable</th>
<th>Cell Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR</td>
<td>Varies from 40 bp to 5000 bp in different hosts93–95</td>
<td>≤24 kb86,90</td>
<td>Yes</td>
<td>Various cells and organisms including bacteria (Lactobacillus acidophilus, Lactobacillus gasseri, Bacillus subtilis), yeast (Pichia pastoris, Hansenula polymorpha, Yarrowia lipolytica, Saccharomyces cerevisiae), fungi (Ashbya gossypii, Glomereula cingulate), plants (Arabidopsis thaliana, Physcomitrella patens), zebrash, mice, human HEK293 cells, human pre-B leukemia cell line Nalm-6, and stem cell59,109,115</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Homology not required76,77</td>
<td>≤15 kb68,69</td>
<td>No</td>
<td>Various cells and organisms including human embryonic stem cells, induced pluripotent stem cells, zebrash, embryos, frog embryos, and Chinese hamster ovary (CHO) cells76,77</td>
</tr>
<tr>
<td>MMEJ</td>
<td>Varies from 5 bp to 25 bp76,77</td>
<td>≤50 kb117</td>
<td>Yes</td>
<td>Various cells and organisms including silkworms, frogs, frog embryos, zebrafish, CHO cells, HEK293T cells, HeLa cells, Aspergillus fumigatus, and Leishmania donovani74–79</td>
</tr>
<tr>
<td>ssODNs</td>
<td>25–80 bp83,106</td>
<td>≤200 kb86</td>
<td>Yes</td>
<td>Various cells and organisms including rat zygotes, mouse embryos, zebrafish, and HEK293T cells53,67,85,86</td>
</tr>
<tr>
<td>Recombination</td>
<td>12–35 bp89,91</td>
<td>≤9 kb90</td>
<td>Yes</td>
<td>Various cells and organisms including E. coli and HEK293T cells12,41,89,95</td>
</tr>
</tbody>
</table>

■ STRAIN AND PATHWAY ENGINEERING THROUGH NUCLEASE-ASSISTED GENOME INTEGRATION METHODS

Besides the direct genome integration of genes or pathways, the enhanced efficiencies, multiplexability, and expanded capabilities conferred by HEs and programmable nucleases, have greatly accelerated complex and stable strain and pathway engineering efforts, which would otherwise be tedious and difficult to achieve by traditional means.86,97 Some of these advanced applications include in vivo assembly of large synthetic pathways, multiplicity integration of large DNA cassettes, and library integration and screening.

In Vivo Assembly and Integration of Large Biochemical Pathways. Prior to the development of nuclease-assisted genome integrations, in vivo assembly of biochemical pathways was mainly conducted using plasmid based systems.89 With the rapid advancements in genome integration methods, it is now possible to simultaneously assemble and integrate biochemical pathways into the genome. Jakoučiūnas et al. developed a CRISPR/Cas9-mediated system, called CasSEMBlR (Cas9-facilitated multi loci integration of assembled DNA parts), that allows marker-free, multi loci, in vivo assembly and integration of pathways with various phenotypes.99 The authors reported creation of an orange carotenoid production strain from 15 DNA parts into three loci with 31% efficiency and construction of a tyrosine production strain from 10 parts into two integration loci with an average efficiency of 58%, together with two gene knockouts through one transformation.98 Similarly, CrEdit (CRISPR/Cas9 mediated genome Editing) has been used to reconstruct the β-carotene pathway by simultaneously integrating individual pathway genes into three separate integration loci.100 In this work, 84% efficiency was achieved for the three-gene integration without any selection. With these approaches, strain engineering is greatly simplified, obviating the need for traditional multistep integrations of individual pathway genes and deletions of competing enzymes.

Multicopy Integration of Large Pathways. Traditionally, multicopy integrations were challenging and were often met with low efficiencies and low copy numbers, especially when integrating large constructs. With the enhanced efficiencies conferred by HEs and programmable nucleases, multicopy integration is now readily achieved. Kuhlman and Cox reported a two-step genome integration method in E. coli...
by coupling I-SceI with λ-Red recombineering. The method was demonstrated through integration of the entire lac operon (∼9 kb) into four predefined locations. Shi et al. developed a single-step, markerless method by coupling the CRISPR/Cas9 system with yeast delta-site integration. As a proof of concept, the authors reported an unprecedented 18-copy genome integration of a 24 kb recombinant pathway. The ability to integrate multiple copies of the pathway genes simultaneously removes the need for repeated integrations and greatly accelerates strain engineering efforts.

**Library Integration and Screening.** Rather than the traditional workflow of screening plasmid libraries and integrating the final shortlisted constructs, the vastly improved integration efficiencies have enabled the direct integration of libraries of genetic elements. Ryan et al. developed quantitative gene assembly and DNA library integration in _S. cerevisiae_ using CRISPR/Cas9. The authors demonstrated integration of a nourseothricin-resistance gene from three parts with 85% efficiency in a diploid strain and 70% efficiency in a polyploid strain. Moreover, a DNA library of PCR generated variants of a cellobiose utilization pathway was assembled in vivo and integrated into a diploid yeast, which led to 10-fold increase of the cellobiose uptake rate.

**REFERENCES**


mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res.* 38, e152.


