Molecular Systems Biology 9; Article number 641; doi:10.1038/msb.2012.66
Citation: *Molecular Systems Biology* **9**:641
www.molecularsystemsbiology.com

OPEN ACCESS

molecular systems biology

REVIEW

Genome-scale engineering for systems and synthetic biology

Kevin M Esvelt[1],* and Harris H Wang[1,2,3],*

[1] Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA and [2] Department of Systems Biology, Harvard Medical School, Boston, MA, USA
[3]Present address: Department of Systems Biology, Columbia University Medical Center, 701 West 168th Street, Room 1308-B, New York, NY 10032, USA
* Corresponding authors. HH Wang or KM Esvelt, Wyss Institute for Biologically Inspired Engineering, Harvard University, 3 Blackfan Circle, Boston, MA 02115, USA. Tel.: +1 617 955 9575; Fax: +1 617 432 7828;
E-mail: hw2429@columbia.edu or Tel.: +1 857 919 3375;
Fax: +1 617 432 7828; E-mail: kevin.esvelt@wyss.harvard.edu

Received 21.9.12; accepted 16.12.12

Genome-modification technologies enable the rational engineering and perturbation of biological systems. Historically, these methods have been limited to gene insertions or mutations at random or at a few pre-defined locations across the genome. The handful of methods capable of targeted gene editing suffered from low efficiencies, significant labor costs, or both. Recent advances have dramatically expanded our ability to engineer cells in a directed and combinatorial manner. Here, we review current technologies and methodologies for genome-scale engineering, discuss the prospects for extending efficient genome modification to new hosts, and explore the implications of continued advances toward the development of flexibly programmable chasses, novel biochemistries, and safer organismal and ecological engineering.

Molecular Systems Biology **9**: 641; published online 22 January 2013; doi:10.1038/msb.2012.66
Subject Categories: synthetic biology
Keywords: directed evolution; genome engineering; metabolic engineering; synthesis; synthetic chassis

Introduction

The phrase ‘genome-scale engineering’ invokes a future in which organisms are custom designed to serve humanity. Yet humans have sculpted the genomes of domesticated plants and animals for generations. Darwin’s contemporary William Youatt described selective breeding as ‘that which enables the agriculturalist, not only to modify the character of his flock, but to change it altogether. It is the magician’s wand, by means of which he may summon into life whatever form and mold he pleases’ (Youatt, 1837). Selective breeding has transformed aurochs into Holsteins, wolves into Chihuahuas and Great Danes, and teosinte into maize. All of these examples involved genomic changes at a scale dwarfing any attempted through rational design. Understanding why genomes have been more readily shaped by evolutionary principles than conventional design-based approaches is important for current and future genome engineering endeavors.

Engineering is a human enterprise consisting of iterative cycles of design, construction, and testing. Optimizing this iterative process involves balancing the relative time, costs, and expected benefits gained at each phase. However, rationally designing and building a genome to produce the desired phenotype has proven exceedingly difficult. Designing organisms to specification requires accurately predicting phenotype from genotype, a complex problem that is worsened by our incomplete knowledge of biomolecule production, degradation, and interaction rates. Moreover, the computational resources required to run bottom-up molecular-level simulations are daunting even for simpler systems (Karr *et al*, 2012; Koch, 2012). Nevertheless, models have been useful for generating new hypotheses and targeting promising areas for engineering. Yet, even with the best *in silico* predictions, we are still limited by our ability to construct the designed genome. More than any other factor, the absence of molecular tools for manipulating genomic sequences has forced us to rely on selective breeding and evolutionary optimization (Conrad *et al*, 2011) rather than rational genome design.

Recent breakthroughs in genomics and genome editing have promised a greater role for rational design in biological engineering (Figure 1), offering new opportunities for systems and synthetic biologists aiming to reverse-engineer naturally evolved systems and to build new systems. In particular, advances in high-throughput DNA sequencing and large-scale biomolecular modeling of metabolic and signaling networks represent two important new frontiers that aid genome-scale engineering. Over the last few years, thousands of bacterial genomes have been sequenced from a wide variety of natural species and numerous laboratory-generated strains (Pagani *et al*, 2012). These efforts have illuminated many essential features of the core genome (Lukjancenko *et al*, 2010), the extent and importance of genetic heterogeneity across populations (Avery, 2006), the ubiquity of horizontal gene transfer (Smillie *et al*, 2011), and the evolution and selection of functional genetic elements (David and Alm, 2011). At the same time, new computational tools have used the flood of data to model metabolic processes and signaling networks across the entire cell, generating many new testable hypotheses (Lewis *et al*, 2012). Most importantly, emerging advances in *de novo* synthesis and *in vivo* gene targeting allow empirical validation of these model-driven hypotheses. By building and testing synthetic variants of biological systems, we have a unique opportunity to decipher the constraints imposed by the complexity of evolved systems and develop strategies

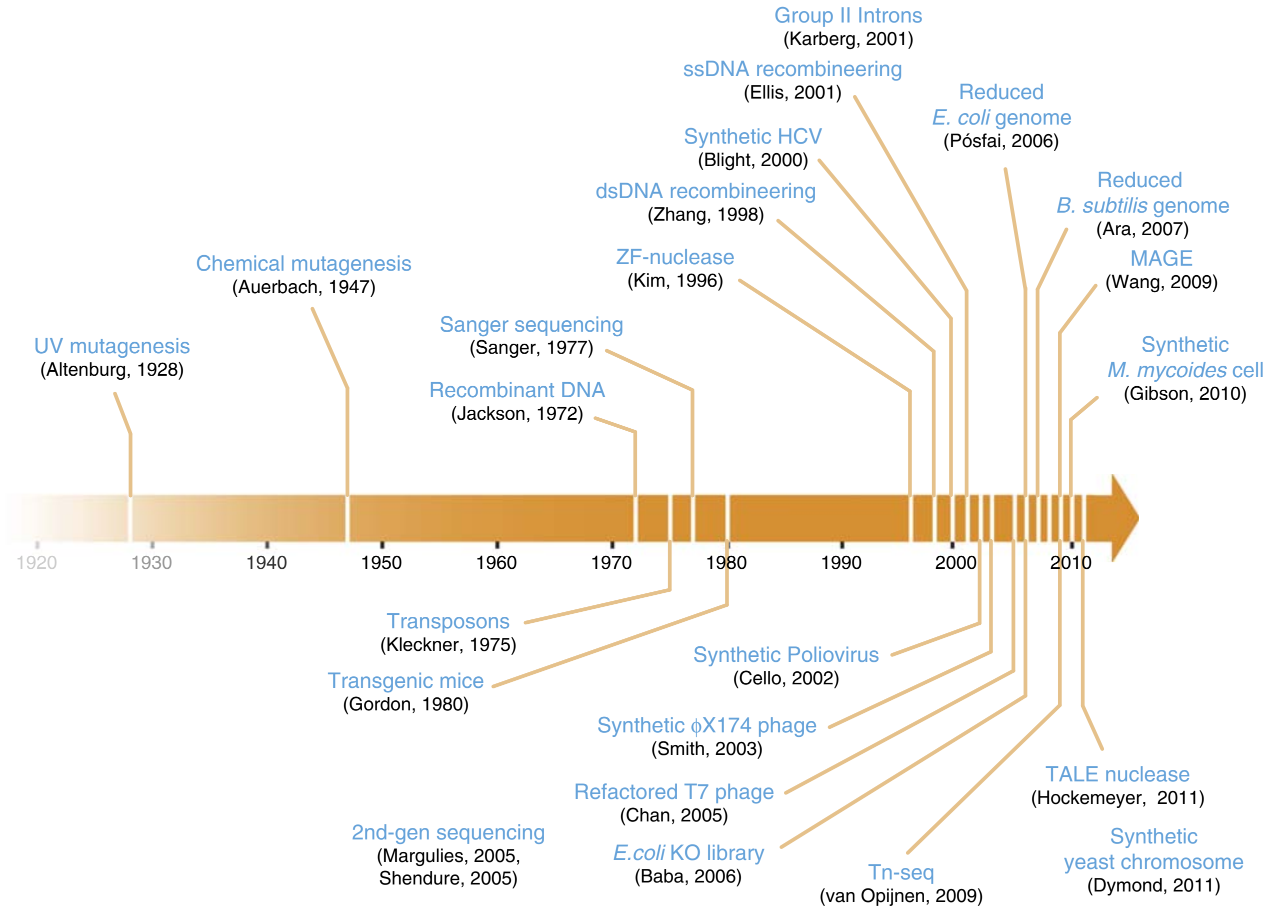

Figure 1 A historical timeline of selected advances leading to genome-scale engineering.

for engineering living systems more conducive to quantitative modeling and rational design.

Here we review recent technologies that empower design-based genome engineering approaches, identify potential bottlenecks, discuss strengths and limitations of strategies employing rational design versus evolution, and consider future applications of genome-scale engineering. We advocate a synergistic engineering strategy that adopts the best aspects of rational genome design and evolutionary optimization.

What is genome-scale engineering?

Genome engineering is the art of constructing a genotype that gives rise to a desired phenotype, a challenge whose difficulty is influenced by the scale of genomic alteration required. One measure of scale is the number of changes that must be made to an existing genome to produce the desired phenotype. In some cases, this may require editing only one gene, a task that is clearly not genome scale. The same is true for a library of single-gene variants and even a complete collection of single-gene knockouts (Giaever *et al*, 2002; Baba *et al*, 2006), as each genome has only a single change. We define genome-scale engineering to be any endeavor involving sequence modifications to at least two distinct regions of a genome. In what follows, we will mainly focus on technologies potentially capable of modifying large fractions of a single genome.

Genome-scale engineering allows us to experimentally probe deep biological questions such as essentiality (Koonin, 2000), epistasis (Chou *et al*, 2011; Khan *et al*, 2011), encoding (Itzkovitz and Alon, 2007), evolvability (Tokuriki and Tawfik, 2009; Wagner and Zhang, 2011; Hill and Zhang, 2012), and robustness (Bershtein *et al*, 2006). At the same time, we aim to rationally build useful organisms that cannot be easily generated by harnessing evolution alone. Such endeavors require foundational tools in design, modeling, construction, and testing that extend from individual cells to populations of organisms (Figure 2). Iterations of design, model, build, and test phases are likely to be more important as the scale of the endeavor increases because biological complexity can grow exponentially. Below, we describe key features of these phases in genome-scale engineering, outline current capabilities, and suggest opportunities for improvement.

Genome designs and models

Design is a set of specifications intended to achieve a dedicated objective under various constraints. Biological designs are those that describe the underlying blueprint of living organisms, built upon the information encoded in genes across the

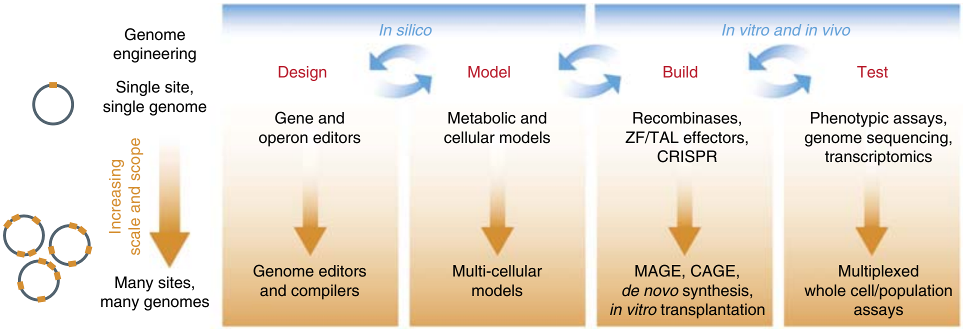
Genome engineering
Single site, single genome
Increasing scale and scope
Many sites, many genomes
In silico
In vitro and in vivo
Design
Model
Build
Test
Gene and operon editors
Metabolic and cellular models
Recombinases, ZF/TAL effectors, CRISPR
Phenotypic assays, genome sequencing, transcriptomics
Genome editors and compilers
Multi-cellular models
MAGE, CAGE, *de novo* synthesis, *in vitro* transplantation
Multiplexed whole cell/population assays

Figure 2 Foundational genome engineering tools and approaches are needed to extend single site genetic perturbations of a single genome to multiple changes across many genomes.

genome. As the focus of biological engineering shifts from individual genes to entire genomes, there is a growing need for more sophisticated genome design tools to assist such large-scale engineering endeavors. Recordkeeping software is essential for tracking numerous modifications designed and generated across libraries of genomes. Traditional gene editors such as Vector NTI and SeqBuilder are largely inadequate for such purposes. However, new design tools and software suites such as J5 (Hillson *et al*, 2012), Clotho (Xia *et al*, 2011), and Genome Compiler (http://www.genomecompiler.com/) provide better data management and user interfaces for the design of large operons and whole genomes.

Although recordkeeping is important, it is only one aspect of design, which must carefully define the experimental objective and triage candidate implementations according to likely failure modes. The complexity of biological systems often renders effective design a challenge. Fortunately, computational models can provide a useful guiding framework. Constraint-based reconstruction and analysis (COBRA) models such as flux-balance analysis have served as excellent predictive tools improve designs. These models generally rely on steady-state analysis of metabolic flux to determine useful genomic targets that optimize a desired phenotype *in silico*. Although a detailed discussion of such models is beyond the scope of this review, COBRA-based approaches have been reviewed extensively elsewhere (Lewis *et al*, 2012).

Whereas specialized metabolic models have been used for some years, Karr *et al* (2012) recently published the first complete virtual model of a cell, *M. genitalium*. At only ~525 genes, *M. genitalium* is one of the smallest genomes known. Nevertheless, its phenotype is determined by the interaction of so many molecular components that it cannot be accurately modeled using any single method. To surmount this problem, Karr *et al* (2012) partitioned *Mycoplasma* into 28 distinct modules, modeled each using the most appropriate representation, and integrated the results to describe the entire cell. Analysis of unexpected behaviors on the part of the resulting virtual cell led to novel hypotheses concerning emergent controls on cellular behavior and identification of promiscuous enzyme activities capable of compensating for the lost genes. Despite these successes, accurate genotype-to-phenotype predictions of multiple genomic perturbations are still challenging due to biological complexity, large combinatorial variations, and computational limitations. Nonetheless, these examples demonstrate the power and utility of predictive models in understanding cellular behavior and identifying promising biological designs.

A complementary alternative to *in silico* prediction is direct experimental perturbation to identify potential targets and failure modes. Recent breakthroughs combining large-scale mutagenesis with DNA sequencing have contributed significantly to improved genomic designs. Hutchison *et al* (1999) showed that sequencing transposon-generated libraries of mutants can be used to systematically identify essential genes within the Mycoplasma genome. More recent approaches have employed next-generation sequencing, including Insertion Sequencing (IN-Seq) (Goodman *et al*, 2011), transposon sequencing (Tn-seq) (van Opijnen *et al*, 2009), high-throughput insertion tracking by deep sequencing (Wong *et al*, 2011), and transposon-directed insertion-site sequencing (Eckert *et al*, 2011). IN-Seq, for example, involves the generation of libraries by random insertion of a Himar1 transposon containing a modified inverted repeat (IR) sequence. This IR is also recognized by the Type IIS restriction enzyme *Mme*I, which cuts the DNA 17 bases outside of its recognition site. When digested *in vitro*, genomic DNA carrying transposons harboring *Mme*I sites will generate fragments that include an extra 16–17 bp of genomic DNA, allowing high-throughput sequencing to pinpoint the locations of all insertions. By enabling researchers to compare the abundance of individual mutants in the library before and after an experimental perturbation, Tn-seq techniques enable multiplexed functional analysis of entire genomes. Every gene essential for the survival of a species can be identified in a single experiment that simultaneously rank-orders all nonessential 'accessory' genes by their relative importance to organismal fitness under the conditions of interest. Other approaches such as global transcription machinery engineering (Alper and Stephanopoulos, 2007) and genome-scale profiling of barcoded mutant libraries (Warner *et al*, 2010) can have a similar role in informing design. Expansion and broader adoption of these methods to guide genome-scale design is needed for both single-cell and multicellular organisms.

An expanding toolbox for genome construction and manipulation

A wide variety of tools for targeted gene disruption and transgenesis are currently available (Figure 3). These tools

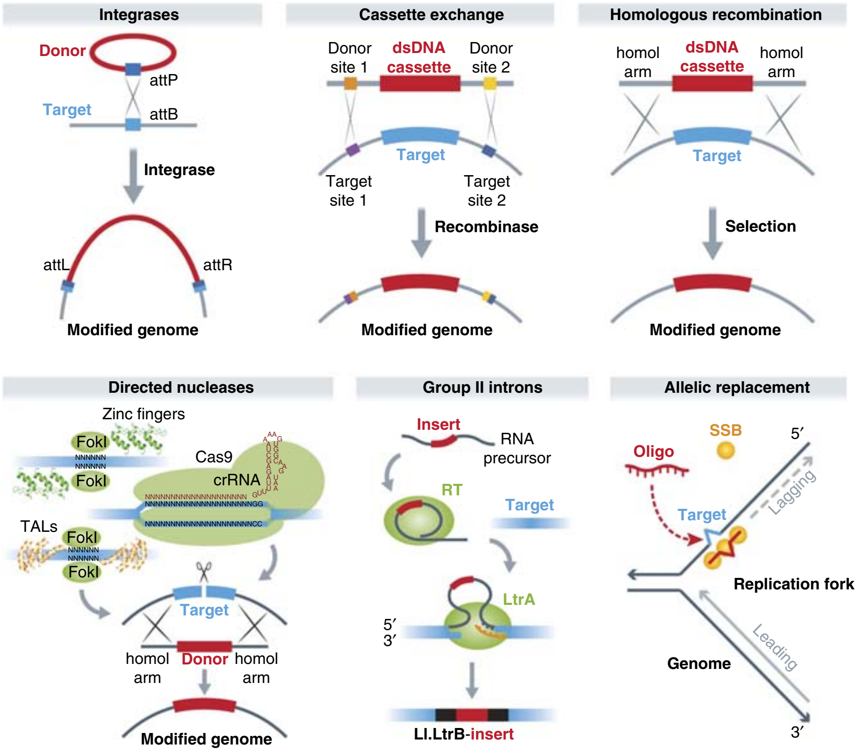

Figure 3 Mechanisms of various targeted genome-modification tools. Integrases can insert a circular donor construct into a recognition site on the genome. Recombinase-mediated cassette exchange (RMCE) involves the replacement of a target sequence flanked with recognition sites with a donor cassette flanked by compatible sites. Homologous recombination using double-stranded DNA cassettes enables programmable target replacement using RecA or RecET-like machinery, which can be stimulated via site-specific cleavage using zinc-finger, CRISPR, or TAL nucleases. Group II introns and insertional elements can be designed to insert into site-specific genome targets. Oligo-mediated allelic replacement incorporates short oligonucleotides into the lagging strand of replicating DNA, which are then resolved upon subsequent cell divisions to inherit the designed mutation.

Table I Comparison of various targeted genome engineering methods

Class	System	Programmable	Efficiency	Multiplexable	Examples	Organism
Protein site-specificity	Integrase	Limited	High	No	λ-int, ϕC31	B, P, M
	Recombinase	Limited	High	No	Cre, Flp, RMCE	B, P, M
	Zinc-finger	Moderate	Variable	Maybe	ZF-FokI, ZF-Tn3	B, P, M
	TAL effector	Moderate/High	Variable	Maybe	TAL-FokI	B, P, M
Nucleic acid site-specificity	Retro-transposon	Moderate	Variable	Maybe	Ll.LtrB intron	B
	dsDNA Homol. recomb.	High	Low	No	λ-Red, RecET	(B, M)
	ssDNA Homol. recomb.	High	High	Yes	Redβ, RecT	(B, M)
	CRISPR/Cas	High	Variable	Maybe	Cas9	B, M
De novo synthesis	Variable	High	N/R	N/R	*M. mycoides*	B, M

Abbreviations: B, bacteria; M, metazoan; N/R, not relevant; P, plants; (), has not worked in all species tested.

vary considerably in their targeting efficiency, ease of retargeting, and effectiveness across a variety of different organisms (Table I). We focus on those with the greatest potential to enable large-scale changes to single or multiple genomes by replacing large contiguous sequences or modifying numerous smaller sites serially or in parallel.

Targeted genome engineering

Recombinases

Because delivering large genetic constructs into many cell types is difficult, highly efficient methods of recombining the host genome with an introduced construct are useful for applications requiring large amounts of foreign DNA or the replacement of many contiguous genes with modified or synthetic variants. Recombinases are DNA-binding enzymes that catalyze highly specific and efficient DNA splicing reactions between two sites. Early experiments with phage-derived recombinases irreversibly incorporated circular constructs containing the phage attP site into the attB site of the host genome normally utilized by the phage (Mizuuchi and Mizuuchi, 1980). Later work demonstrated that these ‘integrases’ can perform a similar role in a wide variety of species if the appropriate attB or attP target site is inserted into the genome by other means (Kilby *et al*, 1993), or may alternatively utilize ‘pseudo-att’ sites native to the genome at somewhat lower efficiency (Thyagarajan *et al*, 2001). Cre recombinase, originally from phage P1 (Sternberg *et al*, 1981), is the gold standard for efficient recombination of target sites across a wide variety of species. However, its comparative promiscuity leads to toxicity in some eukaryotes, leading to the development of Flp recombinase as an alternative (Turan *et al*, 2011). Unlike integrases, Cre and Flp are reversible enzymes that normally recombine two identical recognition sites to invert or excise the intervening sequence, but they can be made irreversible by utilizing ‘poisoned’ half-sites that generate an inactive site upon recombination (Schlake and Bode, 1994; Albert *et al*, 1995).

In the context of genome-scale engineering, recombinases are most useful for efficiently inserting large DNA constructs into the genome. By flanking an endogenous sequence with orthogonal recognition sites from two different recombinases or two orthogonal sites recognized by the same recombinase, the sequence may be replaced by a synthetic donor construct containing compatible sites (Schlake and Bode, 1994; Missirlis *et al*, 2006; Sheren *et al*, 2007). With three pairs of orthogonal sites, this technique could conceivably be used to iteratively insert large cassettes into the genomes of many different organisms *ad infinitum* (Turan *et al*, 2011; Obayashi *et al*, 2012). Unfortunately, recombinases require pre-existing recognition sites, which must be introduced to the target site by another method. Although directed evolution methods have yielded recombinases capable of recognizing alternative sites (Buchholz and Stewart, 2001; Sarkar *et al*, 2007), such approaches are presently too laborious for most laboratories. New methods of performing directed evolution may relax this limitation (Esvelt *et al*, 2011). A promising design-based alternative involves replacing the native DNA-binding domain with an exogenous domain that can be more easily engineered to target a sequence of interest (Akopian *et al*, 2003). Although the resulting chimeric enzymes are highly specific, they are currently inefficient compared with natural recombinases (Gordley *et al*, 2009). It is likely that extensive directed evolution will be required to render the catalytic domain suitable for retargeting by replacement of the DNA-binding domain.

Zinc-finger nucleases and TAL effector nucleases

Targeted genome engineering requires a means of specifically recognizing the sequence of each site to be modified. Zinc-fingers (ZFs) and TAL (transcription activator-like) effectors are a class of versatile and programmable DNA-binding proteins that have enabled effector proteins, including DNA-modifying enzymes, to be targeted to specific sequences in a variety of organisms. ZFs are stackable motifs of ∼30 amino acids that recognize approximately three base pairs of DNA with varying specificity. Although ZFs recognizing each triplet cannot be simply stacked to reliably recognize longer sequences (Ramirez *et al*, 2008), a variety of design (Sander *et al*, 2011b) and selection-based (Maeder *et al*, 2009) methods are capable of generating specific DNA binders. Unfortunately, custom ZFs remain relatively difficult and expensive to obtain for the typical laboratory. DNA recognition by TAL effector domains is more straightforward, with each 34-aa TAL motif recognizing a single basepair through contacts with amino acids 12 and 13, known as the repeat variable di-residue (RVD) (Boch *et al*, 2009). Unlike ZFs, TAL effectors are readily stacked to recognize long sequences. Although the assembly of TALs is complicated by their larger size and abundant repeat regions, a number of recently described approaches have the potential to overcome these challenges (Weber *et al*, 2011; Briggs *et al*, 2012; Reyon *et al*, 2012).

ZF and TAL nucleases (ZFNs and TALENs) are created by coupling a ZF or TAL DNA-binding domain to the nonspecific nuclease domain of the *Fok*I restriction enzyme. When two monomers bind to adjacent sites, their FokI domains dimerize and catalyze DNA cleavage, causing a double-strand break (DSB) (Kim *et al*, 1996). DSBs are most commonly repaired by homologous recombination (HR) or non-homologous end-joining (NHEJ). ZFN cleavage followed by HR with a donor sequence containing homologous flanking regions leads to insertion of the donor sequence at efficiencies of ∼1–15% (Urnov *et al*, 2005), while ZFN cleavage followed by error-prone NHEJ results in gene disruption from small deletions or insertions, typically at somewhat higher efficiencies (Urnov *et al*, 2010). Targeted gene editing using ZFNs has been demonstrated in a variety of cell types, including flies (Bibikova *et al*, 2003), worms (Wood *et al*, 2011), sea urchins (Ochiai *et al*, 2010), zebrafish (Ekker, 2008), silkworms (Takasu *et al*, 2010), frogs (Young *et al*, 2011), plants (Cai *et al*, 2009; Osakabe *et al*, 2010; Zhang *et al*, 2010), and numerous mammals (Urnov *et al*, 2005; Geurts *et al*, 2009; Hauschild *et al*, 2011). Nuclease activity can be toxic in some cell types, possibly due to off-target activity, but this problem can be mitigated by utilizing less toxic ‘nickase’ variants that cut only one strand (Kim *et al*, 2012; Ramirez *et al*, 2012). Customized ZFNs are commercially available, although at a significant cost. TAL effector nucleases (TALENs) can more readily target a variety of sequences by virtue of their more flexible RVD-based recognition. Although newer and less thoroughly studied, TALENs appear to have fewer off-target effects and lower toxicity than corresponding ZFNs (Mussolino *et al*, 2011). Design tools are freely available (Doyle *et al*, 2012) with predicted viable cleavage sites every 35 basepairs in mammalian genomes on an average (Cermak *et al*, 2011). Their primary weakness is the difficulty of assembling and delivering such large and repeat-prone sequences. TALENs have been successfully applied in numerous organisms including yeast (Li *et al*, 2011), flies (Liu *et al*, 2012), zebrafish (Sander *et al*, 2011a), plants (Li *et al*, 2012), rats (Tesson *et al*, 2011), and human cells (Hockemeyer *et al*, 2011) with gene disruption efficiencies of up to 25% (Miller *et al*, 2011).

Group II intron retrotransposition

Certain group II introns are selfish genetic elements that undergo genomic transposition through an RNA intermediate. Because targeting is determined primarily by base-pairing interactions with the intron RNA, these site-specific retrotransposons can be retargeted to accomplish both gene disruption and gene insertion. The commercially available Targetron system harnesses a retrotransposon capable of inserting up to 1.8 kb into the genome (Karberg *et al*, 2001).

Intron retrotransposition efficiencies vary from 1–80% depending on the site and species (Perutka *et al*, 2004). Sequences suitable for insertion are found every few hundred bases on average, permitting most genes to be disrupted. Moreover, the system is active in a wide variety of microbes, providing genetic manipulation of species that cannot be modified using other methods (Yao and Lambowitz, 2007). Notably, insertions of recombinase recognition sites may permit subsequent recombinase-mediated cassette exchange. Targeting efficiency may be high enough to permit multiplex modifications, though this has yet to be demonstrated. Interestingly, group II introns can also be used to generate DSBs (Karberg *et al*, 2001), suggesting a potential use in promoting HR if they can be engineered or evolved to function efficiently in eukaryotes.

Recombineering

Recombineering (or recombinogenic-engineering) uses a phage-derived HR pathway to recombine a donor DNA strand with a homologous sequence in the bacterial host. Given sufficient regions of flanking homology (>500 bp), endogenous HR, which is usually mediated by the RecA/Rad51 pathway, is capable of integrating sequences into the genome of almost any cell. However, low efficiency of the native HR machinery limits the use of this technique without efficient DNA delivery and selection. Recombineering is an improved approach that utilizes phage proteins (RecET, λ-Red) to dramatically increase HR frequencies across the entire genome (Zhang *et al*, 1998; Datsenko and Wanner, 2000; Yu *et al*, 2000). In *E. coli*, HR by λ-Red is RecA-independent and instead relies on three proteins: Exo, Beta, and Gam (Muyrers *et al*, 2000; Yu *et al*, 2000). Exo is a $5' \rightarrow 3'$ exonuclease that digests linear double-stranded DNA (dsDNA), leaving $3'$ single-stranded intermediates that then act as substrates for subsequent recombination (Maresca *et al*, 2010). Beta is a single-stranded DNA (ssDNA)-binding protein that facilitates recombination via hybridization of the linear fragment to its genomic complement. Gam acts to inhibit RecBCD activity *in vivo* to prevent the degradation of foreign linear dsDNA fragments. Although recombineering still requires a selection step, the λ-Red-like system will function with as few as 40 bp of homology flanking double-stranded donor DNA fragments of up to several kilobases in length, a limit imposed by a combination of the transformation and recombination efficiencies. Thus, simple PCR amplification of a selectable cassette (typically an antibiotic resistance or metabolic gene), with primers containing flanking homologous sequences to the target site, enables limited rewriting of any region of the genome (Sharan *et al*, 2009). A recent combinatorial example of this technique, Trackable Multiplex Recombineering, used primers derived from DNA microarrays to generate pools of barcoded dsDNA cassettes that can target different sites across the genome (Warner *et al*, 2010). Short ssDNA can also be used in recombineering, a process which requires only the λ-Beta protein. We discuss the utility of such approaches for multiplexed recombineering in the next section. Although recombineering systems have been developed for several model bacteria (van Kessel and Hatfull, 2007; Swingle *et al*, 2010a; van Pijkeren and Britton, 2012), more work is needed to expand the methodology to other organisms. A search for λ-Red-like enzymes derived from phages and viruses that infect other organisms is ongoing (Datta *et al*, 2008).

RNA-guided CRISPR nucleases

The nucleic acid-targeted CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) system has great potential for genome modification in many organisms. CRISPR systems defend bacteria and archaea from invading phage and plasmids by RNA-directed degradation of DNA (Wiedenheft *et al*, 2012). In Type II CRISPR systems, the Cas9 protein locates DNA ‘protospacer’ sequences homologous to the ‘spacer’ sequence in a guiding CRISPR RNA (crRNA) and checking for sufficient RNA–DNA base pairing (Jinek *et al*, 2012). Upon identifying a matching sequence that also contains an appropriate protospacer-adjacent motif (PAM), the enzyme cleaves both DNA strands ~ 3 bp from the start of the PAM, causing a DSB (Gasiunas *et al*, 2012). PAM sequences are quite short (NGG (Deltcheva *et al*, 2011), NGGNG (Horvath *et al*, 2008), NNAGAAW (Deveau *et al*, 2008), and NAAR (van der Ploeg, 2009) to date), permitting most sequences to be targeted. At least 12 bp of perfect homology, in addition to the PAM, appears to be necessary for CRISPR endonuclease activity (Deveau *et al*, 2008; Sapranauskas *et al*, 2011; Jinek *et al*, 2012; Mali *et al*, 2013; Cong *et al*, 2013). In bacterial CRISPR loci, the spacer regions of crRNAs are normally flanked by direct repeats of similar size that are critical for recognition and processing by Cas9 and RNaseIII (Deltcheva *et al*, 2011), but synthetic mimics of the mature crRNA function equally well *in vitro* (Jinek *et al*, 2012).

We and others have recently demonstrated that Cas9 can be used to engineer mammalian genomes (Mali *et al*, 2013; Cong *et al*, 2013). Cas9 can be directed to cleave any sequence with a compatible PAM—in these cases NGG—by expressing a chimeric RNA mimic (Mali *et al*, 2013) or a spacer array together with the tracrRNA required for processing (Cong *et al*, 2013). Gene modification via DSB-stimulated HR is accomplished by simply expressing Cas9 and a cassette that generates a RNA with a spacer matching the target sequence in the desired cell. Targeting two adjacent sites effectively deleted the intervening region, demonstrating limited but multiplexed gene disruption capabilities. Knocking out one of the two Cas9 nuclease domains converted the enzyme into a nickase capable of stimulating HR with comparable efficiency while reducing the frequency of NHEJ. Importantly, both gene disruption and HR rates appear to be comparable to or greater than those achieved with ZFNs and TALENs targeting the same loci.

Interestingly, sustained Cas9 activity might be used to simultaneously promote HR while selecting against cells retaining the target region, potentially obviating the need for positive selection markers. This approach would be feasible in genomes engineered to constitutively express Cas9, which could be subsequently edited by simply delivering the appropriate donor cassette and crRNA. Further development of CRISPR-mediated genome engineering technologies should focus on increasing the specificity beyond the current 12bp + NGG sequence, which would likely lead to some unintended off-target cutting, and on enabling genomic sequences with alternative PAMs to be targeted. Due to its significantly greater ease of use, Cas9-mediated gene targeting represents a new and promising genome editing approach, especially in mammalian systems.

Multiplexed genome engineering

The ability to edit single genes is an important step toward engineering whole genomes. The explosion of modifications achieved with ZFNs and TALENs are particularly striking given the dearth of prior alternatives for most multicellular organisms. Still, the sheer size of even the smallest bacterial genomes renders serial modification of limited utility for truly genome-scale engineering endeavors. Efficient methods enabling multiplex genome editing are urgently needed.

Unfortunately, techniques that generate DSBs to catalyze homology-directed repair may be difficult to multiplex due to the toxicity of multiple simultaneous breaks and the high rate of NHEJ, which could easily lead to unintended rearrangements. High-efficiency ZF or TAL effector recombinases represent one potential alternative, although quickly generating large numbers of ZFs or TALs presents an additional challenge. Another option might involve fusing a nuclease-inactivated Cas9 protein to the catalytic domain of a recombinase, although retaining function could prove to be difficult. Group II introns may be multiplexable for gene disruption, but they leave unavoidable scar sites, are limited to small cargo capacities, and have not been demonstrated to work efficiently in eukaryotes. Meanwhile, the low efficiency of double-stranded λ-Red-mediated recombineering also limits its use for multiplexed genome-scale engineering. However, λ-Red-like proteins also facilitate recombination of smaller ssDNA fragments. On the basis of prior work (Ellis *et al*, 2001), we recently described an approached known as Multiplex Automated Genome Engineering (MAGE) that utilizes short ssDNA oligonucleotides (oligos) instead of dsDNA cassettes to mediate targeted genome modification (Wang and Church, 2011; Wang *et al*, 2009). Specifically, oligos that are complementary to the lagging strand of replicating genomes are incorporated into the daughter genome at high efficiency, presumably by mimicking Okazaki fragments at the replication fork (Yu *et al*, 2003). Oligos that target the leading strand appear to have >50-fold lower incorporation efficiency.

MAGE can precisely engineer any site in the genome by simply introducing an oligo matching the desired sequence. Oligos ranging from 30 to 100 bases are efficiently integrated as long as there are sufficient homology arms to facilitate ssDNA annealing to the target (Ellis *et al*, 2001). At the center of the oligo, new sequences can be designed (up to 30 bases along a 90-base oligo) and introduced into the genome as a heteroduplex, which is resolved into fully mutated alleles during subsequent rounds of cell division. In *E. coli*, oligo incorporation is increased >1000-fold by the ssDNA-binding protein λ-Beta. Removal of the endogenous mismatch repair machinery (e.g., Δ*mutS*) (Costantino and Court, 2003) or evasion of mismatch repair through modified bases (Wang *et al*, 2011) can significantly increase the efficiency of oligo incorporation to levels >30% per viable progeny (Wang *et al*, 2009). Use of a co-selectable marker can further increase the efficiency to >70% (Carr *et al*, 2012; Wang *et al*, 2012b).

Several factors make the oligo-mediated MAGE approach particularly attractive for genome-scale engineering. First, the transformation efficiency of short oligos is high compared with plasmids or dsDNA cassettes, thereby allowing large pools of oligos with different genomic targets to simultaneously enter the cell and undergo incorporation. Because not all oligos are incorporated in every cell, combinations of mutations are generated through this process. With incorporation efficiencies above 70%, cells containing >10 targeted mutations can be isolated after a single transformation (Lajoie *et al*, 2012) by simply screening 100 colonies with multiplex allele-specific PCR (Wang and Church, 2011). Second, the protocol can be iteratively repeated on a population of cells with only 2–3 h of recovery growth needed between cycles. Iterative cycling enables further multiplexing and enrichment of mutants that are otherwise found at low frequencies in the population, which can be automated (Wang *et al*, 2009). Third, oligos can be easily and cheaply synthesized using commercial vendors and used directly in MAGE reactions without the need for further processing, in contrast to dsDNA cassettes which require additional steps of PCR amplification and purification. Furthermore, high-density DNA microarrays can serve as potential sources of large pools of unique DNA sequences to extend multiplexed genome-scale engineering. Finally, oligo-mediated genome engineering approaches such as MAGE will likely function in a variety of organisms by virtue of mechanistic simplicity. To date, oligo-mediated allelic replacement has been demonstrated in Gram-negative bacteria (Swingle *et al*, 2010b), Gram-positive bacteria (van Pijkeren and Britton, 2012), and mammalian cells (Rios *et al*, 2012).

Semi-synthetic and synthetic genomes

Since the chemical synthesis of the first gene in 1972 (Agarwal *et al*, 1972), the cost of DNA synthesis has precipitously decreased as the throughput has soared, enabling construction and assembly of genes and genomes *de novo* (Carr and Church, 2009). Individual gene-sized DNA fragments are readily synthesized commercially and assembled into larger operons (Kodumal *et al*, 2004; Tian *et al*, 2009). Efforts to build phage (Chan *et al*, 2005) and viral genomes (Blight *et al*, 2000; Cello *et al*, 2002), chromosomal arms of *S. cerevisiae* (Dymond *et al*, 2011), and, most impressively, the entire genome of *M. mycoides* (Gibson *et al*, 2008) have been described. New technologies enabling oligonucleotide synthesis on DNA microarrays continue to reduce the cost and increase the throughput for building synthetic genes and genomes (Tian *et al*, 2004; Kosuri *et al*, 2010; Quan *et al*, 2011).

The question of when it is best to adopt an editing, semi-synthetic, or synthetic approach to genome engineering hinges on the reliability of design. Without the ability to accurately evaluate large numbers of potential designs *in silico*, we must build and test them empirically. Currently, large-scale *de novo* synthesis of a genome requires a significantly greater level of resources and effort than directly editing an existing genome. Consequently, a genome editing approach may be optimal when generating genomes with a moderate degree of specified changes (i.e., $\leqslant 100$ s of changes, < 100 bp each), as is required for tuning regulatory networks (Wang *et al*, 2009,2012b) or altering protein sequences (Wang *et al*, 2012a). A *de novo* synthesis approach is more likely to be appropriate for larger-scale alterations such as codon optimization (Welch *et al*, 2009) or refactoring (Chan *et al*, 2005; Temme *et al*, 2012) that are recalcitrant to genome editing technologies.

Building an entire synthetic genome can be difficult to troubleshoot, costly, and prone to failure. An illustrative example of such issues was observed during the construction of the synthetic 1.1 Mb *M. mycoides* genome, when a single basepair deletion in the essential gene *dnaA* prevented the generation of a viable cell (Gibson *et al*, 2010). Only when different synthetic pieces were swapped with natural sequences did the researchers identify the source of the error, highlighting the importance of direct testing. Underlying design flaws may be even more difficult to assess as they may impact the cell physiology in non-linear and epistatic ways. Thus, step-wise construction and testing of progressively modified intermediates will be a crucial approach for most genome-scale engineering efforts until the failure rate of engineered biological designs can be reduced to acceptable levels. Consequently, methods capable of rapidly assembling and exchanging individually synthesized and separately tested genome fragments will be needed. Current examples include *in vitro* enzymatic assembly methods (Li and Elledge, 2007; Engler *et al*, 2008; Gibson *et al*, 2009; Zhang *et al*, 2012) and Conjugative Assembly Genome Engineering *in vivo* (Isaacs *et al*, 2011) (Figure 4). Recent studies have already described instances of cloned or hybrid genomes constructed by transformation or assembly of a donor genome into a recipient cell that retains its own genome. While the *Bacillus-Synechocystis* hybrid-genome (Itaya *et al*, 2005) and the *S. cerevisiae* clone containing a copy of the *A. laidlawii genome* (Karas *et al*, 2012) have yet to yield useful new phenotypes, they do illustrate cellular robustness to large-scale genomic insertions. Studies that evaluate the effects of swapping or refactoring essential operons will provide information more directly relevant to evaluating the feasibility of new designs. More generally, developments that further combine synthetic, semi-synthetic, and hybrid approaches will lead to deeper understanding of the limits of rational design and optimization for engineered biological systems.

Testing and validation of engineered genomes

Empirical testing and validation of modified and synthesized genomes is necessary to determine whether the design goals have been met. Applying high-throughput sequencing to confirm that a constructed genome matches its intended sequence is one such crucial test. Although viability and growth are also essential phenotypic tests, most design objectives require validation of function through other indirect assays. Moreover, many genome construction approaches result in libraries of different variants that require systematic curation to identify and isolate the best genomes from the rest of the population. Typical assays can be divided into low-throughput and high-throughput screens, which identify variants from populations of limited size (up to $\sim 10^5$), and high-throughput selections, which enable the isolation of variants from much larger populations (Figure 5). For example, validating a constructed genome sequence by high-throughput sequencing is a form of low-throughput screen, while a viability assay testing the ability to survive and replicate under specific conditions is a selection. In both cases, the stringency of the assay is crucial, as constructs that do not generate the desired phenotype but still pass the screen or selection can lead to substantial delays and wasted effort. Selections are considerably more powerful when it is possible to generate large libraries of variants, as testing more variants increases the likelihood of finding ones with the desired phenotype.

Unfortunately, many desirable phenotypes cannot be directly selected, including small-molecule biosynthesis and other traits that are among the most frequent targets for biological engineering. Low-throughput screens can generally perform much more detailed phenotypic measurements by employing microscopy, transcriptomics, proteomics, or metabolomics to interrogate biological function at the cellular level. As our ability to build large libraries of genome variants grows, methods to increase the scale and throughput of such phenotypic measurements toward high-throughput selections will be urgently needed to isolate and validate engineered genomes.

Genome-scale metabolic engineering

The application of genome-scale approaches to metabolic engineering provides an excellent example of an integrated platform that showcases the synthesis of rational design, computational modeling, and multiplexed construction and testing to tackle real-world biological engineering challenges. Numerous studies have used metabolic engineering to modify microbes to produce industrially relevant biochemicals and biofuels such as ethanol (Ingram *et al*, 1998) and higher alcohols (Atsumi *et al*, 2008), fatty acids (Steen *et al*, 2010), amino acids (Leuchtenberger *et al*, 2005), shikimate precursors (Bongaerts *et al*, 2001), terpenoids (Martin *et al*, 2003), polyketides (McDaniel *et al*, 1999; Pfeifer *et al*, 2001), and polymer precursors (e.g., 1,4-butanediol (Yim *et al*, 2011)). A great example of genome-scale metabolic engineering is Dupont's near-decade long optimization of *E. coli* for bioproduction of 1,3-propanediol (Nakamura and Whited, 2003). The industrially optimized strain required up to 26 genomic changes including insertions, deletions, and regulatory modifications. Recent advances in constraint-based modeling (Lewis *et al*, 2012) have enabled *in silico* prediction of genomic targets whose perturbation may enhance strain performance or product yield. These computational predictions are ripe for experimental validation using new genome engineering tools. For example, OptKnock (Burgard *et al*, 2003), a computational tool that uses bi-level metabolic flux optimization to predict the phenotype of gene knockout combinations, has been used to improve microbial production of lactic acid (Fong *et al*, 2005). Deleting different combinations of four identified genes (*adhE, pta, pfk, glk*) in *E. coli* significantly improved secretion of the desired product. Similarly, Alper *et al* (2005) described a set of strains generated

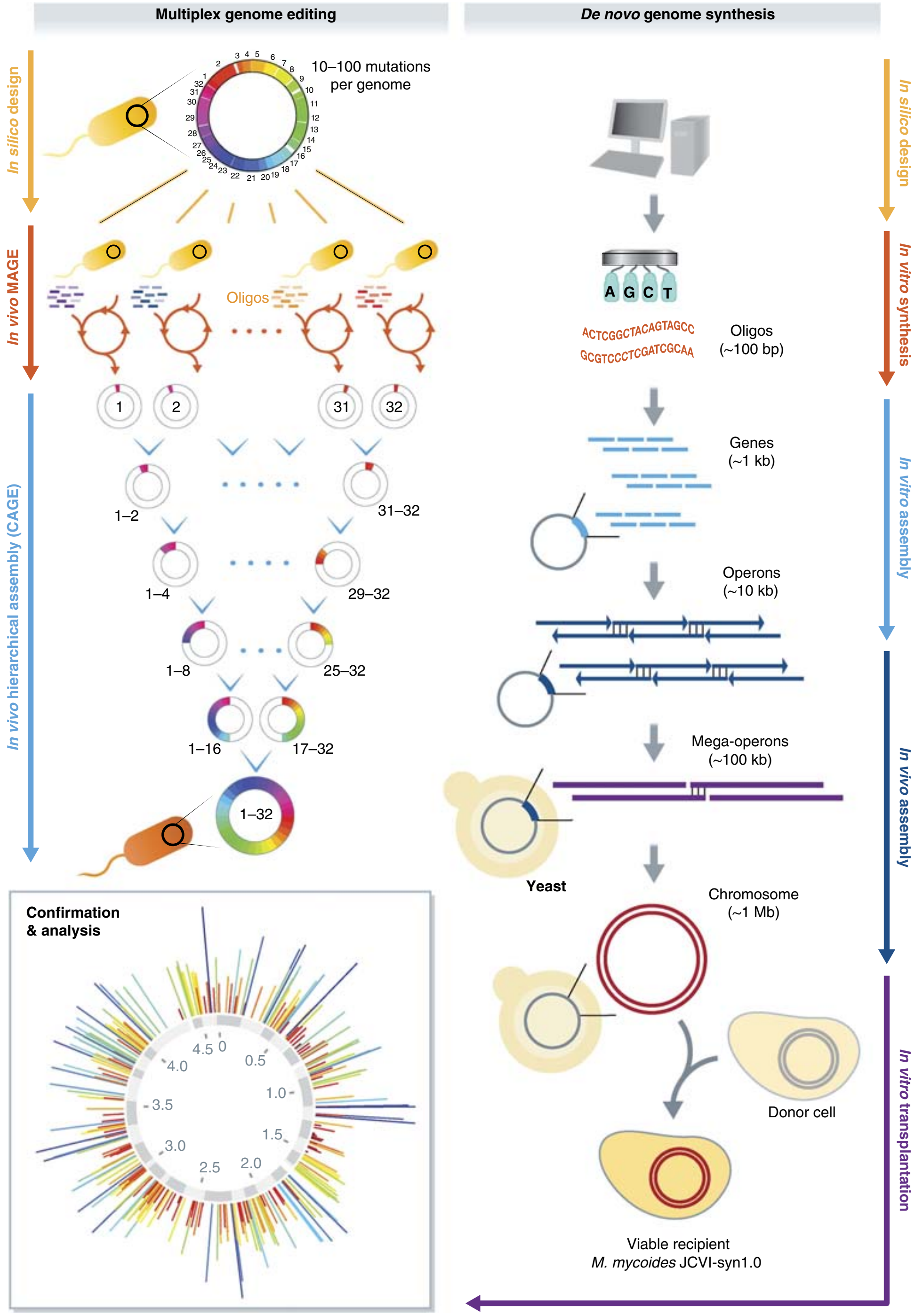

Figure 4 Strategies for genome-scale engineering by multiplexed *in vivo* editing or *de novo* synthesis. Multiplex genome editing enables construction of new genomes via living intermediates for rapid design and test cycles. *De novo* genome synthesis can build synthetic designs that are drastically different from natural genomes.

through model-driven combinatorial gene deletions of seven genomic targets that exhibited improved lycopene production by up to 8.5-fold. More recently, Xu *et al* (2011) described the use of genome-scale metabolic network modeling to generate genetic modifications that enhanced production of the useful precursor malonyl-CoA. Knockout and overexpression genotypes in up to nine genes were generated combinatorially, with some strains containing up to five modifications (triple knockout, double overexpression).

Although these few studies suggest the promising potential of higher-order mutants to access phenotypes needed to meet challenging design goals, the experimental difficulty of constructing such mutants has limited their use. The recent development of multiplex genome-scale engineering tools

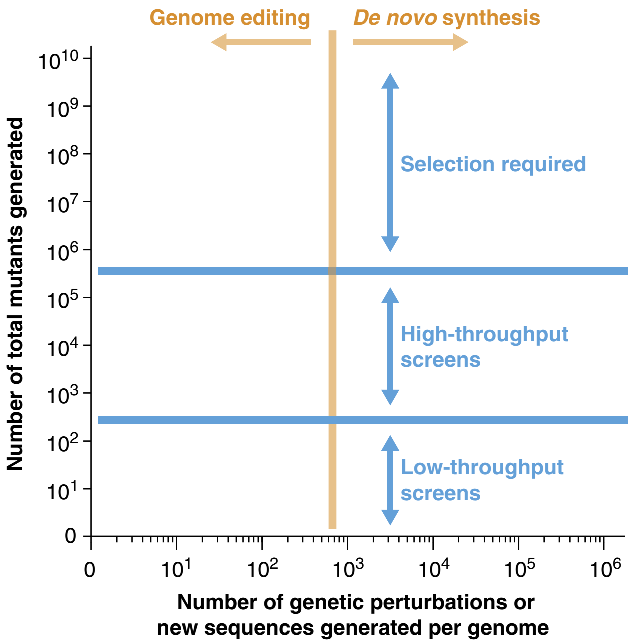

Figure 5 Current approaches to genome-scale engineering for building (editing versus synthesis) and testing (screening versus selection) genomes, when considering number of genetic perturbations against total number of mutants generated.

such as MAGE has dramatically reduced the time required to generate combinatorial libraries of targeted mutations. We have shown that combinatorial exploration of both translation efficiency and gene deletions in up to 24 genes can yield useful combinations of genomic modifications for production of lycopene (Wang *et al*, 2009). More recently, the MAGE approach was extended to build a combinatorial library of genomic variants that contained synthetic T7 promoters in up to 12 genes involved in aromatic amino-acid biosynthesis (Wang *et al*, 2012b). The combination of improved metabolic models and new techniques enabling combinatorial exploration and selection of specific genetic perturbations will substantially accelerate metabolic engineering (Sandoval *et al*, 2012).

Organismic genome engineering

When it comes to ease of designing, constructing, and testing genomes, not all organisms are created equal. Some have smaller genomes and unicellular lifestyles, while others have larger genomes and undergo complex multicellular development, both of which render genome design and modeling difficult. Some have many more tools available for genome editing, while others are burdened with polyploid genomes that increase the difficulty of constructing and testing new designs. Some organismal phenotypes can be readily measured, while others are subtle and hard to quantify. Most importantly, some replicate in mere minutes and are readily grown in large numbers, while others require years of labor-intensive care to reach adulthood. The advent of new technologies for genome design, construction, and testing have compensated for some of these differences, but accentuated the impact of others.

Dairy cows are classic examples of slow-growing, expensive, multicellular organisms that nonetheless have a large industry invested in their improvement. While cows have been modified through evolutionary engineering since antiquity, their slow growth and large diploid genomes render them recalcitrant to targeted variant construction and testing. Furthermore, *in silico* predictive models of mammals do not exist. Nevertheless, milk production has quadrupled over the last 60 years because the industry rigorously measured outputs and applied extremely strong selection in the form of artificial insemination (Funk, 2006). For decades, top bulls have routinely sired tens of thousands of offspring, efficiently transmitting only the best genes to the next generation—a purely blind evolutionary search, but the most effective strategy available given the constraints of the organism at the time. Thanks to high-throughput sequencing, it is now possible to design strategies to accelerate the rate of improvement. Although we are far from understanding the mechanistic basis of milk production, recent genotyping sequencing efforts have begun to identify the chromosomal regions and individual genes favored by the past few decades of selection (Larkin *et al*, 2012). The industry is now implementing rationally designed generations-long strategies to hasten the combination of known beneficial alleles into single genomes using selective breeding and perhaps, eventually, targeted genome editing.

Microbes are the mirror image of domesticated animals in almost every way. Unknown in antiquity due to their microscopic size, they tend toward small haploid genomes that can be grown quickly and in large numbers. Combined with a powerful selection, these traits permit swift evolutionary engineering, as first demonstrated by W.H. Dallinger's nineteenth-century-directed evolution of microbial thermal tolerance from 18°C to an astonishing 70°C over 7 years (Dallinger, 1887). A dearth of screening and selection technologies impeded further microbial engineering until the latter half of the twentieth century, but the subsequent explosion of such methods has rendered microbes—which combines rapid growth, large population sizes, and powerful selections—the organisms of choice for directed evolution studies. We recently demonstrated that even smaller and faster-replicating genomes can further accelerate and even automate evolutionary engineering (Esvelt *et al*, 2011). Our system harnesses filamentous phages, which require only minutes to replicate in host *E. coli* cells, to optimize phage-carried exogenous genes in a handful of days without researcher intervention. Compounding their growth advantage is the fact that microbes and phages are also ideal subjects for biological design, modeling, targeted genome editing, and genome synthesis, all of which can focus subsequent evolutionary searches on the regions of sequence space most likely to encode desirable phenotypes. Alternatively, these methods can compensate for the lack of a powerful selection that precludes evolution. Future technologies will ideally extend some of the advantages enjoyed by model organisms, such as *E. coli* and *S. cerevisiae* to other organisms, enabling more genome engineering endeavors to combine model-driven targeted manipulation with the best growth and selection paradigm available to the target organism.

Toward a flexibly programmable biological chassis

One of the overarching goals of genome-scale engineering is to develop insights and rules that govern biological design. Unfortunately, most biological systems are riddled with remnants of historically contingent evolutionary events—a complex, highly heterogeneous state woefully unsuitable for precise and rational engineering. Rational genome design would be greatly facilitated by the construction of an underlying biological 'chassis' that is simple, predictable, and programmable. From that foundation, we can begin to build more complex systems that expand the repertoire of biochemical capabilities and controllable parameters. Furthermore, the chassis organism must contain mechanisms ensuring safe and controlled propagation, with strong barriers preventing unintended release into the environment and mechanisms that genetically isolate it from other organisms. The chassis should also contain obvious and permanent genetic signatures of its synthetic origins for surveillance of its use and misuse. Here we outline several classes of capabilities that should serve as a framework for a flexibly programmable biological chassis (Figure 6). A combination of current and future genome engineering technologies will be needed to construct such an engineered system.

Reducing biological complexity

The difficulties inherent in designing living systems arise from the vast number of cellular components and the sheer complexity of their evolutionarily optimized network of interactions. Simulating large numbers of heterogeneously interacting molecules requires evaluating the probability and magnitude of all possible interactions between non-identical components, a task that would be computationally beyond us even if we had perfect knowledge of every interaction (Koch, 2012). We still do not understand the function of almost 20% of the ~4000 genes found in *E. coli* (Keseler *et al*, 2011). Given that biological complexity is one of the most significant barriers to rational genome design, we should aim to build a simplified microbial cell. Not only would such a cell serve as an improved chassis for future engineering, the act of constructing such a genome will transform our understanding of the factors contributing to the performance, evolvability, and robustness of cellular systems in general.

Single-gene deletion experiments (Giaever *et al*, 2002) suggest that a significant number of all genes are redundant, with only ~300 being individually essential (Feher *et al*, 2007). The first step toward a simplified cellular chassis is to reduce the genome to a functionally useful set of genes. Several groups have embarked upon endeavors to eliminate all nonessential genes, starting with *E. coli* (Hashimoto *et al*, 2005; Posfai *et al*, 2006), *B. subtilis* (Ara *et al*, 2007), and *S. pombe* (Giga-Hama *et al*, 2007). It is important to keep in mind that whether a gene is essential depends on the environmental conditions. Therefore, we define a set of useful traits for a biological chassis as (1) fast growing in minimal media with glucose, (2) capable of fermentation, (3) amenable to genetic manipulation, and (4) minimally sufficient such that removal of any additional gene negatively affects the other three stated considerations. A cell containing a set of genes that satisfy the above criteria is said to have a core or minimal chassis. Although a viable *E. coli* genome with 20% fewer genes has already been engineered (Posfai *et al*, 2006), it is likely that a reduction of 50% is achievable for the core chassis. Even though smaller genomes and simpler transcriptome do exist (e.g., *Mycoplasma pneumonia* (Guell *et al*, 2009)), our core chassis will be much more useful for biological engineering because it will not suffer from slow growth or depend upon additional exogenous metabolites. Moreover, engineering our

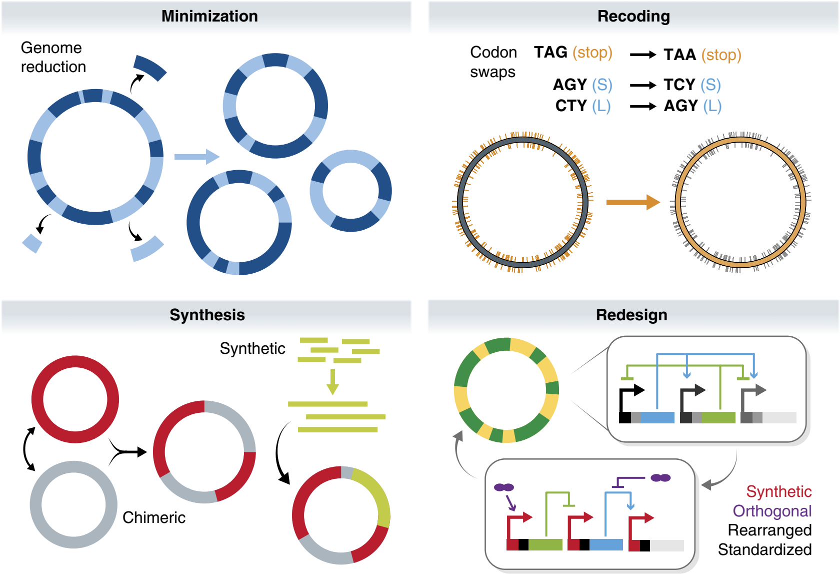

Figure 6 Toward the construction of a flexibly programmable chassis. Genome minimization reduces biological complexity and redundancy. Whole-genome codon remapping enables orthogonal information encoding and expansion of the genetic code. *De novo* genome synthesis and reconstitution from natural genomes enables creation of semi-synthetic and chimeric genomes with new and hybrid features. Whole-genome redesign and rewiring of regulatory systems enable new synthetic circuitries that are easier to design and model.

chassis could consolidate related genes into modular, functionally similar operons to facilitate future engineering.

With far fewer components and exponentially fewer possible interactions, a cell with a core chassis will be much more amenable to *in silico* modeling than wild-type *E. coli* or even *M. genitalium* (Karr *et al*, 2012). Still, its remaining components will interact in many more ways than we would prefer, and not all of them are understood. This might be remedied by reducing the number of regulatory interactions, ideally by replacing endogenous regulatory elements with well-defined orthogonal equivalents. Temme *et al* (2012) implemented this concept by 'refactoring' the nitrogen fixation cluster to remove all native gene regulation. Refactoring an operon involves removing all non-coding DNA, nonessential genes, and transcription factors, replacing essential genes with computer-designed synthetic genes recoded to eliminate internal regulatory sites, and adding synthetic regulation. Extending this approach to the entire core genome will be an immense challenge, as each replacement must be optimized with synthetic components. On the other hand, cellular growth and survival is a powerful and readily applicable selection, enabling libraries of synthetic or rewired regulatory elements to be quickly selected and sequenced to identify the best performers (Isalan *et al*, 2008). Minimizing the total number of orthogonal regulatory elements and compensating for changes in the expression of previously refactored operons caused by adding additional binding sites are likely to be the most challenging aspects of the project. Adding additional but well-defined levels of regulation such as orthogonal 16S ribosomes (Rackham and Chin, 2005), synthetic ZF transcription factors (Khalil *et al*, 2012), or orthogonal RNA-based translational repressors (Isaacs *et al*, 2004) may be necessary to increase growth to acceptable levels while minimizing the total number of components.

A final challenge concerns the effects of natural selection on our simplified genome. We expect our rationally designed synthetic chassis to be suboptimal, in that simple growth in glucose media may lead to accumulation of beneficial mutations. Careful tracking of these beneficial mutations as they occur will simplify the task of decoding the newly created interactions and reveal important design flaws in our *in silico* models. Only by understanding and attempting to compensate for these new interactions will we learn how to further simplify and optimize the performance of our engineered system.

Orthogonal information encoding

A frequent objection to the use of genetically modified organisms is the possibility of unintended consequences arising from accidental release. Improved methods for biological containment would reduce such risks while raising public awareness of beneficial genome engineering research. One such containment strategy is the development of a chassis that utilizes an orthogonal genetic code (Isaacs *et al*, 2011). The canonical encoding scheme maps 64 possible codons to 20 corresponding amino acids and three stop signals. Except for a few known organisms (Knight *et al*, 2001), the genetic code is the single most well-preserved property in all of biology and thought to be irreversibly fixed in its current configuration as a result of 'the frozen accident' (Crick, 1968). A codon-swapped organism might have codons that are normally assigned to leucine instead encode arginine. Although the resulting protein sequence would not change, the encoded nucleotide sequences would be quite different in a recoded organism compared with the wild type. Achieving this goal would involve not only recoding of all genes in the new genomic chassis, but would also require minor alterations to the anticodon sequences of tRNAs to accommodate different codon swaps. A combination of genome synthesis and engineering will be needed to realize such an endeavor.

More importantly, a radically recoded chassis would be unable to productively exchange genetic material with other organisms in the environment. When transferred into a wild-type cell, recoded genes from a swapped-codon chassis will generate meaningless proteins due to mistranslation from reassigned codons. Conversely, natural genes will not function in the swapped-codon chassis, preventing our synthetic genome from becoming contaminated with wild toxins, pathogenicity elements, or antibiotic resistance genes. Indeed, genetic isolation from all other domains of life will also confer broad immunity to natural viruses, a significant advantage for the industrial-scale production of biochemicals. However, the recoded chassis may still interact with the physical environment and with other organisms indirectly via nutritional exchange and space competition. These aspects present opportunities for further rational engineering. Finally, recoded organisms will contain many genomic signatures of their synthetic origin, allowing easy identification and surveillance of their origin, make, and purpose in comparison to natural variants.

Expanded biochemical repertoire

With the exception of post-translational modifications in higher-level organisms, the amino-acid repertoire of cells is mostly confined to the canonical 20 amino acids. Unnatural amino acids have been successfully incorporated into proteins using several strategies involving orthogonally evolved tRNA and tRNA synthetases (Hendrickson *et al*, 2004; Xie and Schultz, 2005), but this approach has been hampered by lower efficiencies of incorporation due to competition with existing codon recognition factors (Young *et al*, 2010). Expanding the repertoire of possible amino acids that the cell can use to build proteins is a powerful capability that will be readily available to any recoded chassis. Unnatural amino acids will dramatically expand the biochemical repertoire of cells by enabling new chemistries that are inaccessible to natural systems (Liu and Schultz, 2010). Whole-genome recoding can readily free up codons by reducing the degeneracy of the current codon mapping. New amino acids can be assigned to 'free codons' as long as the existing proteins are recoded with the synonymous codons to retain the amino-acid sequence. A similar event occurred when a handful of organisms began to encode the 21st amino acid, selenocysteine, with the TGA codon that functions as a STOP codon in other forms of life (Forchhammer *et al*, 1989). Although eliminating significant numbers of rare sense codons may be challenging, the prospect of engineering a flexible chassis with the ability to encode multiple unnatural amino acids and access phenotypes unavailable to natural organisms is worth the attempt.

Toward engineering the pan-genome

Thus far, we have only considered methods for engineering individual genomes in the laboratory. Similar and related techniques might be adapted to modify most or all of the individual genomes that together constitute a single species: the pan-genome. There are important safety and ecological considerations to assess before attempting any such project. Nevertheless, the environmental impact of human activity has already effected vast changes across the genomes of a large fraction of species all across Earth. It may be worth considering approaches that might correct such problems and accomplish desirable changes in a more benign manner. For example, we might spread a modification conferring drought resistance through the many local cultivars of a crop plant, with each cultivar retaining its local adaptations and genetic diversity. Such an approach would likely be superior in yield and lower in ecological impact to one in which all such variants are replaced with monocultures cloned from a single laboratory-modified plant. Similarly, human disease vectors such as mosquitoes might be engineered to resist pathogen transmission, which would be considerably cheaper and more ecologically friendly than heavy insecticide use. Several genome engineering tools might be used to address these challenges. Targeting the wild-type locus with nucleases would catalyze DSB repair using the transgenic cassette as a template, effectively converting all heterozygotes to homozygotes. Conceptually similar 'gene-drives' have proven effective in the laboratory (Windbichler *et al*, 2011). Alternatively, a site-specific recombinase targeted to the wild-type locus could exchange the ends of homologous chromosomes, moving the desired modification to the formerly wild-type chromosome and leaving behind a toxin rendering the donor chromosome sterile. Unlike other methods, this approach could be limited to a finite number of 'jumps' by placing a limited number of recombination sites and toxins on the initial donor, thereby improving our control over the spread of the engineered genetic element. Meanwhile, traits might be driven through microbiomes by combining horizontal gene transfer mechanisms with transposon- or retrotransposon-mediated gene insertion. Further advances in these areas scaled to the ecosystem level (Mee and Wang, 2012) may extend our genome engineering capabilities across the pan-genome, although we emphasize that ecological and safety considerations should be thoroughly assessed before such technologies are deployed.

Concluding remarks

Recent technological advances have overcome many of the limitations and bottlenecks that have constrained genome-scale engineering. The exponential decrease in cost of DNA sequencing has dramatically accelerated forward genomics while enabling sequence confirmation of synthesized and edited genomes. New methods are bringing down the cost of DNA synthesis at an even faster rate. Emerging technologies for gene insertion, multiplex editing, and large fragment assembly have dramatically expanded our capabilities in certain model organisms, but further enhancements and extension to other organisms and across species will be needed to extend our engineering capabilities to the ecological level. Similarly, improved *in silico* modeling capabilities are urgently needed to guide rational genome design and synergize productively with evolutionary optimization. Finally, we suggest that the construction of a flexibly programmable biological chassis may serve as a foundation and standard for synthetic biology. These and other ambitious endeavors will continue to challenge our capabilities as genome engineers and our competence as biological designers.

Acknowledgements

KME acknowledges funding from the Wyss Institute Technology Development Fellowship. HHW acknowledges funding from the Wyss Institute Technology Development Fellowship and the National Institutes of Health Director's Early Independence Award (Grant 1DP5OD009172-01). We thank N. Lewis, S. Kosuri, and G. Church for insightful discussions and critical reading of this manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Agarwal KL, Yamazaki A, Cashion PJ, Khorana HG (1972) Chemical synthesis of polynucleotides. *Angew Chem Int Ed Engl* **11:** 451–459

Akopian A, He J, Boocock MR, Stark WM (2003) Chimeric recombinases with designed DNA sequence recognition. *Proc Natl Acad Sci USA* **100:** 8688–8691

Albert H, Dale EC, Lee E, Ow DW (1995) Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J* **7:** 649–659

Alper H, Miyaoku K, Stephanopoulos G (2005) Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* **23:** 612–616

Alper H, Stephanopoulos G (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng* **9:** 258–267

Ara K, Ozaki K, Nakamura K, Yamane K, Sekiguchi J, Ogasawara N (2007) Bacillus minimum genome factory: effective utilization of microbial genome information. *Biotechnol Appl Biochem* **46:** 169–178

Atsumi S, Hanai T, Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451:** 86–89

Avery SV (2006) Microbial cell individuality and the underlying sources of heterogeneity. *Nat Rev Microbiol* **4:** 577–587

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2:** 2006.0008

Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444:** 929–932

Bibikova M, Beumer K, Trautman JK, Carroll D (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* **300:** 764

Blight KJ, Kolykhalov AA, Rice CM (2000) Efficient initiation of HCV RNA replication in cell culture. *Science* **290:** 1972–1974

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326:** 1509–1512

Bongaerts J, Kramer M, Muller U, Raeven L, Wubbolts M (2001) Metabolic engineering for microbial production of aromatic amino acids and derived compounds. *Metab Eng* **3:** 289–300

Briggs AW, Rios X, Chari R, Yang L, Zhang F, Mali P, Church GM (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res* **40:** e117

Buchholz F, Stewart AF (2001) Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat Biotechnol* **19:** 1047–1052

Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* **84:** 647–657

Cai CQ, Doyon Y, Ainley WM, Miller JC, Dekelver RC, Moehle EA, Rock JM, Lee YL, Garrison R, Schulenberg L, Blue R, Worden A, Baker L, Faraji F, Zhang L, Holmes MC, Rebar EJ, Collingwood TN, Rubin-Wilson B, Gregory PD *et al* (2009) Targeted transgene integration in plant cells using designed zinc finger nucleases. *Plant Mol Biol* **69:** 699–709

Carr PA, Church GM (2009) Genome engineering. *Nat Biotechnol* **27:** 1151–1162

Carr PA, Wang HH, Sterling B, Isaacs FJ, Lajoie MJ, Xu G, Church GM, Jacobson JM (2012) Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res* **40:** e132

Cello J, Paul AV, Wimmer E (2002) Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297:** 1016–1018

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39:** e82

Chan LY, Kosuri S, Endy D (2005) Refactoring bacteriophage T7. *Mol Syst Biol* **1:** 2005.0018

Chou HH, Chiu HC, Delaney NF, Segre D, Marx CJ (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332:** 1190–1192

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* (e-pub ahead of print 3 January 2013; doi:10.1126/science.1231143)

Conrad TM, Lewis NE, Palsson BO (2011) Microbial laboratory evolution in the era of genome-scale science. *Mol Syst Biol* **7:** 509

Costantino N, Court DL (2003) Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci USA* **100:** 15748–15753

Crick FH (1968) The origin of the genetic code. *J Mol Biol* **38:** 367–379

Dallinger WH (1887) The President’s address. *J R Microsc Soc* **7:** 185–199

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA* **97:** 6640–6645

Datta S, Costantino N, Zhou X, Court DL (2008) Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc Natl Acad Sci USA* **105:** 1626–1631

David LA, Alm EJ (2011) Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* **469:** 93–96

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471:** 602–607

Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* **190:** 1390–1400

Doyle EL, Booher NJ, Standage DS, Voytas DF, Brendel VP, Vandyk JK, Bogdanove AJ (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res* **40:** W117–W122

Dymond JS, Richardson SM, Coombes CE, Babatz T, Muller H, Annaluru N, Blake WJ, Schwerzmann JW, Dai J, Lindstrom DL, Boeke AC, Gottschling DE, Chandrasegaran S, Bader JS, Boeke JD (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477:** 471–476

Eckert SE, Dziva F, Chaudhuri RR, Langridge GC, Turner DJ, Pickard DJ, Maskell DJ, Thomson NR, Stevens MP (2011) Retrospective application of transposon-directed insertion site sequencing to a library of signature-tagged mini-Tn5Km2 mutants of Escherichia coli O157:H7 screened in cattle. *J Bacteriol* **193:** 1771–1776

Ekker SC (2008) Zinc finger-based knockout punches for zebrafish genes. *Zebrafish* **5:** 121–123

Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci USA* **98:** 6742–6746

Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3:** e3647

Esvelt KM, Carlson JC, Liu DR (2011) A system for the continuous directed evolution of biomolecules. *Nature* **472:** 499–503

Feher T, Papp B, Pal C, Posfai G (2007) Systematic genome reductions: theoretical and experimental approaches. *Chem Rev* **107:** 3498–3513

Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO (2005) *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* **91:** 643–648

Forchhammer K, Leinfelder W, Bock A (1989) Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. *Nature* **342:** 453–456

Funk DA (2006) Major advances in globalization and consolidation of the artificial insemination industry. *J Dairy Sci* **89:** 1362–1368

Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* **109:** E2579–E2586

Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, Choi VM, Jenkins SS, Wood A, Cui X, Meng X, Vincent A, Lam S, Michalkiewicz M, Schilling R, Foeckler J, Kalloway S, Weiler H, Menoret S, Anegon I, Davis GD *et al* (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325:** 433

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A *et al* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418:** 387–391

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison 3rd CA, Smith HO (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319:** 1215–1220

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH *et al* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329:** 52–56

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison 3rd CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6:** 343–345

Giga-Hama Y, Tohda H, Takegawa K, Kumagai H (2007) Schizosaccharomyces pombe minimum genome factory. *Biotechnol Appl Biochem* **46:** 147–155

Goodman AL, Wu M, Gordon JI (2011) Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nat Protoc* **6:** 1969–1980

Gordley RM, Gersbach CA, Barbas 3rd CF (2009) Synthesis of programmable integrases. *Proc Natl Acad Sci USA* **106:** 5053–5058

Guell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kuhner S, Rode M, Suyama M, Schmidt S, Gavin AC, Bork P, Serrano L (2009) Transcriptome complexity in a genome-reduced bacterium. *Science* **326:** 1268–1271

Hashimoto M, Ichimura T, Mizoguchi H, Tanaka K, Fujimitsu K, Keyamura K, Ote T, Yamakawa T, Yamazaki Y, Mori H, Katayama T, Kato J (2005) Cell size and nucleoid organization of engineered Escherichia coli cells with a reduced genome. *Mol Microbiol* **55:** 137–149

Hauschild J, Petersen B, Santiago Y, Queisser AL, Carnwath JW, Lucas-Hahn A, Zhang L, Meng X, Gregory PD, Schwinzer R, Cost GJ, Niemann H (2011) Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. *Proc Natl Acad Sci USA* **108:** 12013–12017

Hendrickson TL, de Crecy-Lagard V, Schimmel P (2004) Incorporation of nonnatural amino acids into proteins. *Annu Rev Biochem* **73:** 147–176

Hill WG, Zhang XS (2012) Assessing pleiotropy and its evolutionary consequences: pleiotropy is not necessarily limited, nor need it hinder the evolution of complexity. *Nat Rev Genet* **13:** 296

Hillson NJ, Rosengarten RD, Keasling JD (2012) j5 DNA assembly design automation software. *ACS Synth Biol* **1:** 14–21

Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC, Zeitler B, Cherone JM, Meng X, Hinkley SJ, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29:** 731–734

Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* **190:** 1401–1412

Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* **286:** 2165–2169

Ingram LO, Gomez PF, Lai X, Moniruzzaman M, Wood BE, Yomano LP, York SW (1998) Metabolic engineering of bacteria for ethanol production. *Biotechnol Bioeng* **58:** 204–214

Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333:** 348–353

Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* **22:** 841–847

Isalan M, Lemerle C, Michalodimitrakis K, Horn C, Beltrao P, Raineri E, Garriga-Canut M, Serrano L (2008) Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452:** 840–845

Itaya M, Tsuge K, Koizumi M, Fujita K (2005) Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. *Proc Natl Acad Sci USA* **102:** 15971–15976

Itzkovitz S, Alon U (2007) The genetic code is nearly optimal for allowing additional information within protein-coding sequences. *Genome Res* **17:** 405–412

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337:** 816–821

Karas BJ, Tagwerker C, Yonemoto IT, Hutchison CA, Smith HO (2012) Cloning the *Acholeplasma laidlawii* PG-8A genome in *Saccharomyces cerevisiae* as a yeast centromeric plasmid. *ACS Synth Biol* **1:** 22–28

Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM (2001) Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat Biotechnol* **19:** 1162–1167

Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival Jr B, Assad-Garcia N, Glass JI, Covert MW (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* **150:** 389–401

Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muniz-Rascado L, Bonavides-Martinez C, Paley S, Krummenacker M, Altman T, Kaipa P, Spaulding A, Pacheco J, Latendresse M, Fulcher C, Sarker M, Shearer AG, Mackie A, Paulsen I, Gunsalus RP *et al* (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res* **39:** D583–D590

Khalil AS, Lu TK, Bashor CJ, Ramirez CL, Pyenson NC, Joung JK, Collins JJ (2012) A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150:** 647–658

Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332:** 1193–1196

Kilby NJ, Snaith MR, Murray JA (1993) Site-specific recombinases: tools for genome engineering. *Trends Genet* **9:** 413–421

Kim E, Kim S, Kim DH, Choi BS, Choi IY, Kim JS (2012) Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res* **22:** 1327–1333

Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci USA* **93:** 1156–1160

Knight RD, Freeland SJ, Landweber LF (2001) Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet* **2:** 49–58

Koch C (2012) Systems biology. Modular biological complexity. *Science* **337:** 531–532

Kodumal SJ, Patel KG, Reid R, Menzella HG, Welch M, Santi DV (2004) Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc Natl Acad Sci USA* **101:** 15573–15578

Koonin EV (2000) How many genes can make a cell: the minimal-gene-set concept. *Annu Rev Genomics Hum Genet* **1:** 99–116

Kosuri S, Eroshenko N, Leproust EM, Super M, Way J, Li JB, Church GM (2010) Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat Biotechnol* **28:** 1295–1299

Lajoie MJ, Gregg CJ, Mosberg JA, Washington GC, Church GM (2012) Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Res* **40:** e170

Larkin DM, Daetwyler HD, Hernandez AG, Wright CL, Hetrick LA, Boucek L, Bachman SL, Band MR, Akraiko TV, Cohen-Zinder M, Thimmapuram J, Macleod IM, Harkins TT, McCague JE, Goddard ME, Hayes BJ, Lewin HA (2012) Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. *Proc Natl Acad Sci USA* **109:** 7693–7698

Leuchtenberger W, Huthmacher K, Drauz K (2005) Biotechnological production of amino acids and derivatives: current status and prospects. *Appl Microbiol Biotechnol* **69:** 1–8

Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* **10:** 291–305

Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods* **4:** 251–256

Li T, Huang S, Zhao X, Wright DA, Carpenter S, Spalding MH, Weeks DP, Yang B (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res* **39:** 6315–6325

Li T, Liu B, Spalding MH, Weeks DP, Yang B (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol* **30:** 390–392

Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. *Annu Rev Biochem* **79:** 413–444

Liu J, Li C, Yu Z, Huang P, Wu H, Wei C, Zhu N, Shen Y, Chen Y, Zhang B, Deng WM, Jiao R (2012) Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. *J Genet Genomics* **39:** 209–215

Lukjancenko O, Wassenaar TM, Ussery DW (2010) Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol* **60:** 708–720

Maeder ML, Thibodeau-Beganny S, Sander JD, Voytas DF, Joung JK (2009) Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nat Protoc* **4:** 1471–1501

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* (e-pub ahead of print 3 January 2013; doi:10.1126/science.1232033)

Maresca M, Erler A, Fu J, Friedrich A, Zhang Y, Stewart AF (2010) Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol Biol* **11:** 54

Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat Biotechnol* **21:** 796–802

McDaniel R, Thamchaipenet A, Gustafsson C, Fu H, Betlach M, Ashley G (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc Natl Acad Sci USA* **96:** 1846–1851

Mee MT, Wang HH (2012) Engineering ecosystems and synthetic ecologies. *Mol Biosyst* **8:** 2470–2483

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD, Rebar EJ (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29:** 143–148

Missirlis PI, Smailus DE, Holt RA (2006) A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer region in Cre-mediated recombination. *BMC Genomics* **7:** 73

Mizuuchi M, Mizuuchi K (1980) Integrative recombination of bacteriophage lambda: extent of the DNA sequence involved in attachment site function. *Proc Natl Acad Sci USA* **77:** 3220–3224

Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* **39:** 9283–9293

Muyrers JP, Zhang Y, Buchholz F, Stewart AF (2000) RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev* **14:** 1971–1982

Nakamura CE, Whited GM (2003) Metabolic engineering for the microbial production of 1,3-propanediol. *Curr Opin Biotechnol* **14:** 454–459

Obayashi H, Kawabe Y, Makitsubo H, Watanabe R, Kameyama Y, Huang S, Takenouchi Y, Ito A, Kamihira M (2012) Accumulative gene integration into a pre-determined site using Cre/loxP. *J Biosci Bioeng* **113:** 381–388

Ochiai H, Fujita K, Suzuki K, Nishikawa M, Shibata T, Sakamoto N, Yamamoto T (2010) Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. *Genes Cells* **15:** 875–885

Osakabe K, Osakabe Y, Toki S (2010) Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. *Proc Natl Acad Sci USA* **107:** 12034–12039

Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC (2012) The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* **40:** D571–D579

Perutka J, Wang W, Goerlitz D, Lambowitz AM (2004) Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J Mol Biol* **336:** 421–439

Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291:** 1790–1792

Posfai G, Plunkett 3rd G, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR (2006) Emergent properties of reduced-genome *Escherichia coli*. *Science* **312:** 1044–1046

Quan J, Saaem I, Tang N, Ma S, Negre N, Gong H, White KP, Tian J (2011) Parallel on-chip gene synthesis and application to optimization of protein expression. *Nat Biotechnol* **29:** 449–452

Rackham O, Chin JW (2005) A network of orthogonal ribosome x mRNA pairs. *Nat Chem Biol* **1:** 159–166

Ramirez CL, Certo MT, Mussolino C, Goodwin MJ, Cradick TJ, McCaffrey AP, Cathomen T, Scharenberg AM, Joung JK (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res* **40:** 5560–5568

Ramirez CL, Foley JE, Wright DA, Muller-Lerch F, Rahman SH, Cornu TI, Winfrey RJ, Sander JD, Fu F, Townsend JA, Cathomen T, Voytas DF, Joung JK (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods* **5:** 374–375

Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* **30:** 460–465

Rios X, Briggs AW, Christodoulou D, Gorham JM, Seidman JG, Church GM (2012) Stable gene targeting in human cells using single-strand oligonucleotides with modified bases. *PLoS One* **7:** e36697

Sander JD, Cade L, Khayter C, Reyon D, Peterson RT, Joung JK, Yeh JR (2011a) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol* **29:** 697–698

Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, Curtin SJ, Blackburn JS, Thibodeau-Beganny S, Qi Y, Pierick CJ, Hoffman E, Maeder ML, Khayter C, Reyon D, Dobbs D, Langenau DM, Stupar RM, Giraldez AJ, Voytas DF *et al* (2011b) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods* **8:** 67–69

Sandoval NR, Kim JY, Glebes TY, Reeder PJ, Aucoin HR, Warner JR, Gill RT (2012) Strategy for directing combinatorial genome engineering in Escherichia coli. *Proc Natl Acad Sci USA* **109:** 10540–10545

Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* **39:** 9275–9282

Sarkar I, Hauber I, Hauber J, Buchholz F (2007) HIV-1 proviral DNA excision using an evolved recombinase. *Science* **316:** 1912–1915

Schlake T, Bode J (1994) Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* **33:** 12746–12751

Sharan SK, Thomason LC, Kuznetsov SG, Court DL (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* **4:** 206–223

Sheren J, Langer SJ, Leinwand LA (2007) A randomized library approach to identifying functional lox site domains for the Cre recombinase. *Nucleic Acids Res* **35:** 5464–5473

Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480:** 241–244

Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, Del Cardayre SB, Keasling JD (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463:** 559–562

Sternberg N, Hamilton D, Hoess R (1981) Bacteriophage P1 site-specific recombination. II. Recombination between loxP and the bacterial chromosome. *J Mol Biol* **150:** 487–507

Swingle B, Bao Z, Markel E, Chambers A, Cartinhour S (2010a) Recombineering using RecTE from *Pseudomonas* syringae. *Appl Environ Microbiol* **76:** 4960–4968

Swingle B, Markel E, Costantino N, Bubunenko MG, Cartinhour S, Court DL (2010b) Oligonucleotide recombination in Gram-negative bacteria. *Mol Microbiol* **75:** 138–148

Takasu Y, Kobayashi I, Beumer K, Uchino K, Sezutsu H, Sajwan S, Carroll D, Tamura T, Zurovec M (2010) Targeted mutagenesis in the silkworm Bombyx mori using zinc finger nuclease mRNA injection. *Insect Biochem Mol Biol* **40:** 759–765

Temme K, Zhao D, Voigt CA (2012) Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc Natl Acad Sci USA* **109:** 7085–7090

Tesson L, Usal C, Menoret S, Leung E, Niles BJ, Remy S, Santiago Y, Vincent AI, Meng X, Zhang L, Gregory PD, Anegon I, Cost GJ (2011) Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol* **29:** 695–696

Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol Cell Biol* **21:** 3926–3934

Tian J, Gong H, Sheng N, Zhou X, Gulari E, Gao X, Church G (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432:** 1050–1054

Tian J, Ma K, Saaem I (2009) Advancing high-throughput gene synthesis technology. *Mol Biosyst* **5:** 714–722

Tokuriki N, Tawfik DS (2009) Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* **19:** 596–604

Turan S, Galla M, Ernst E, Qiao J, Voelkel C, Schiedlmeier B, Zehe C, Bode J (2011) Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. *J Mol Biol* **407:** 193–221

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435:** 646–651

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **11:** 636–646

van der Ploeg JR (2009) Analysis of CRISPR in Streptococcus mutans suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology* **155:** 1966–1976

van Kessel JC, Hatfull GF (2007) Recombineering in Mycobacterium tuberculosis. *Nat Methods* **4:** 147–152

van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6:** 767–772

van Pijkeren JP, Britton RA (2012) High efficiency recombineering in lactic acid bacteria. *Nucleic Acids Res* **40:** e76

Wagner GP, Zhang J (2011) The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat Rev Genet* **12:** 204–213

Wang HH, Church GM (2011) Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods Enzymol* **498:** 409–426

Wang HH, Huang PY, Xu G, Haas W, Marblestone A, Li J, Gygi S, Forster A, Jewett MC, Church GM (2012a) Multiplexed in vivo His-tagging of enzyme pathways for in vitro single-pot multi-enzyme catalysis. *ACS Synth Biol* **1:** 43–52

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460:** 894–898

Wang HH, Kim H, Cong L, Jeong J, Bang D, Church GM (2012b) Genome-scale promoter engineering by coselection MAGE. *Nat Methods* **9:** 591–593

Wang HH, Xu G, Vonner AJ, Church G (2011) Modified bases enable high-efficiency oligonucleotide-mediated allelic replacement via mismatch repair evasion. *Nucleic Acids Res* **39:** 7336–7347

Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LB, Gill RT (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol* **28:** 856–862

Weber E, Gruetzner R, Werner S, Engler C, Marillonnet S (2011) Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One* **6:** e19722

Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, Gustafsson C (2009) Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS One* **4:** e7002

Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482:** 331–338

Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, Hovde BT, Baker D, Monnat Jr RJ, Burt A, Crisanti A (2011) A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* **473:** 212–215

Wong SM, Gawronski JD, Lapointe D, Akerley BJ (2011) High-throughput insertion tracking by deep sequencing for the analysis of bacterial pathogens. *Methods Mol Biol* **733:** 209–222

Wood AJ, Lo TW, Zeitler B, Pickle CS, Ralston EJ, Lee AH, Amora R, Miller JC, Leung E, Meng X, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Meyer BJ (2011) Targeted genome editing across species using ZFNs and TALENs. *Science* **333:** 307

Xia B, Bhatia S, Bubenheim B, Dadgar M, Densmore D, Anderson JC (2011) Developer's and user's guide to Clotho v2.0 A software platform for the creation of synthetic biological systems. *Methods Enzymol* **498:** 97–135

Xie J, Schultz PG (2005) Adding amino acids to the genetic repertoire. *Curr Opin Chem Biol* **9:** 548–554

Xu P, Ranganathan S, Fowler ZL, Maranas CD, Koffas MA (2011) Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab Eng* **13:** 578–587

Yao J, Lambowitz AM (2007) Gene targeting in gram-negative bacteria by use of a mobile group II intron ("Targetron") expressed from a broad-host-range vector. *Appl Environ Microbiol* **73:** 2735–2743

Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, Estadilla J, Teisan S, Schreyer HB, Andrae S, Yang TH, Lee SY, Burk MJ, Van Dien S (2011) Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat Chem Biol* **7:** 445–452

Youatt W (1837) *Sheep: their Breeds, Management, and diseases. To which is Added the Mountain Shepherd's Manual*. London: Baldwin and Cradock

Young JJ, Cherone JM, Doyon Y, Ankoudinova I, Faraji FM, Lee AH, Ngo C, Guschin DY, Paschon DE, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Harland RM, Zeitler B (2011) Efficient targeted gene disruption in the soma and germ line of the frog *Xenopus tropicalis* using engineered zinc-finger nucleases. *Proc Natl Acad Sci USA* **108:** 7052–7057

Young TS, Ahmad I, Yin JA, Schultz PG (2010) An enhanced system for unnatural amino acid mutagenesis in E. coli. *J Mol Biol* **395:** 361–374

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci USA* **97:** 5978–5983

Yu D, Sawitzke JA, Ellis H, Court DL (2003) Recombineering with overlapping single-stranded DNA oligonucleotides: testing a recombination intermediate. *Proc Natl Acad Sci USA* **100:** 7207–7212

Zhang F, Maeder ML, Unger-Wallace E, Hoshaw JP, Reyon D, Christian M, Li X, Pierick CJ, Dobbs D, Peterson T, Joung JK, Voytas DF (2010) High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. *Proc Natl Acad Sci USA* **107:** 12028–12033

Zhang Y, Buchholz F, Muyrers JP, Stewart AF (1998) A new logic for DNA engineering using recombination in Escherichia coli. *Nat Genet* **20:** 123–128

Zhang Y, Werling U, Edelmann W (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res* **40:** e55