ANNUAL REVIEWS

Annual Review of Biochemistry

Synthetic Genomes

Weimin Zhang,[1] Leslie A. Mitchell,[1] Joel S. Bader,[2] and Jef D. Boeke[1,3]

[1]Institute for Systems Genetics and Department of Biochemistry and Molecular Pharmacology, New York University Langone Health, New York, NY 10016, USA; email: weimin.zhang@nyulangone.org, leslie.mitchell@nyulangone.org, jef.boeke@nyulangone.org

[2]Department of Biomedical Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA; email: joel.bader@jhu.edu

[3]Department of Biomedical Engineering, New York University Tandon School of Engineering, New York, NY 11201, USA

Annu. Rev. Biochem. 2020. 89:77–101

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

https://doi.org/10.1146/annurev-biochem-013118-110704

Keywords

DNA synthesis, genome design, genome editing, Sc2.0, ethics

Abstract

DNA synthesis technology has progressed to the point that it is now practical to synthesize entire genomes. Quite a variety of methods have been developed, first to synthesize single genes but ultimately to massively edit or write from scratch entire genomes. Synthetic genomes can essentially be clones of native sequences, but this approach does not teach us much new biology. The ability to endow genomes with novel properties offers special promise for addressing questions not easily approachable with conventional gene-at-a-time methods. These include questions about evolution and about how genomes are fundamentally wired informationally, metabolically, and genetically. The techniques and technologies relating to how to design, build, and deliver big DNA at the genome scale are reviewed here. A fuller understanding of these principles may someday lead to the ability to truly design genomes from scratch.

Contents

INTRODUCTION

DNA is the carrier of genetic information for all living organisms. In the 1950s, researchers determined the structure of DNA and described the flow of genetic information, now commonly known as the central dogma of molecular biology; since then, the study of genetics has greatly accelerated. While proteins and RNAs direct virtually all cellular activities, it is the DNA that provides the template for RNA transcription and coding information for protein translation and enables transmission of genetic information across generations. Thus, studying how DNA works in different organisms represents one way to probe the nature of life.

The ability to read DNA sequence was first established in the 1970s, and sequencing throughput scaled up massively with the emergence of next-generation sequencing technologies. Approximately 15,000 species in the tree of life were completely or partially sequenced as of April 2018 (1). It is now feasible to determine the sequence of some genomes in a matter of hours. The next great challenge is to decipher with base-pair resolution how genome sequences work to enable all aspects of cellular function and life.

Studies have been conducted from many perspectives to address the basic questions about genome function. From the synthetic biology perspective, an important demonstration of understanding something is to build it from scratch, as famously stated on Richard Feynman's blackboard. Chemical synthesis of genes and genomes started in the 1970s, when researchers built a 77-base-pair (bp) yeast tRNA gene (2). Technology advances in oligonucleotide synthesis, DNA assembly, and genome delivery have revolutionized the field of synthetic genomics. In the past two decades, not only have more synthetic genomes been completed but the size and complexity of synthetic genome projects have increased (**Figure 1**). In this review, we highlight progress in recent synthetic genome projects through the lens of the design-build-test-learn paradigm and discuss the current status and potential future directions of this field.

DESIGN OF SYNTHETIC GENOMES

Like drawing the blueprint of a building prior to construction, designing a synthetic genome is the first task of genome synthesis, and it is a major task that requires major consideration. Genome

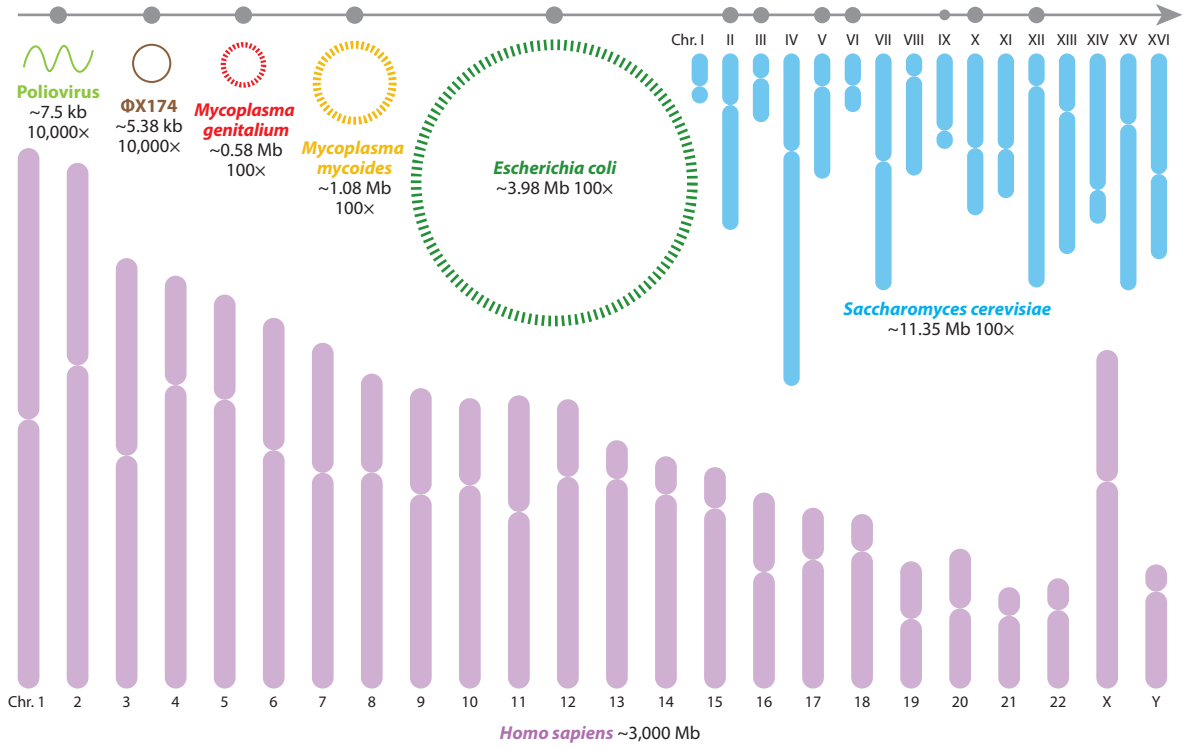

Figure 1

An illustration of six de novo genome synthesis projects and the wild-type human genome (not to scale). Poliovirus complementary DNA is presented by a light green line showing single-stranded linear DNA; synthetic φX174 is presented by a brown ring showing single-stranded circular DNA; synthetic *Mycoplasma genitalium*, *Mycoplasma mycoides*, and *Escherichia coli* genomes are presented by red, yellow, and green dashed circles showing double-stranded circular DNAs; the human and yeast chromosomes are denoted by two connected bars that represent chromosome arms consisting of double-stranded linear DNA. The 10,000× and 100× labels represent scale-up factors required to match the indicated human genome size. Abbreviations: Chr., chromosome; kb, kilobase; Mb, megabase. Figure adapted with permission from Reference 148.

designs have to match the capability of genome synthesis or genome engineering to ensure successful completion. Historically, with the growing capability of genome synthesis, researchers designed genomes from simple to complex. In general, previous genome designs have fallen into four categories, as described below. Here we specify some studies that represent each design category.

Distinguishing Synthetic and Native Genomes

Among early genome synthesis studies, designing a sequence nearly identical to the wild type was prioritized to minimize chances of failure (2). Here, distinguishing the synthetic product from its natural counterpart is an important issue, and several different strategies have been documented. Cello et al. (3) altered a total of 27 nucleotides in designing a synthetic poliovirus genome sequence [~7.5 kilobases (kb)] to introduce detectable changes by restriction digestion and sequencing. Gibson et al. (4) inserted five watermarks, 48–143-bp sequences used to encode tracking information into DNA, at intergenic positions that were known to tolerate transposon insertion. Smith et al. (5) used an entirely different approach and designed a φX174 bacteriophage genome sequence that matched a published reference sequence for which they did not have physical DNA

as a template (6). In all cases, the goal of the design was to make a minimal number of changes to a known, functional genome sequence in the hopes that the synthetic version would function in cells. This minimized the risk that the cost and effort of assembling these genomes would be wasted as a consequence of the designed sequence not working upon delivery to host cells. The projects were also important demonstrations that the nucleotide sequence was sufficient for function, without the need for unobserved chemical modifications or undetected additional vital components, perhaps serving as the ultimate death knell to vitalism.

Refactoring Native Genomes

A more aggressive strategy for synthetic genome design is termed refactoring and is particularly notable in viral and other compact genomes that have been evolutionarily selected for small size, a process sometimes referred to as streamlining. For example, the genome of T7 bacteriophage was partially redesigned to define one or more genetic elements as individual parts, which otherwise overlap with one another in the wild-type genome (7). The main advantage to disentangling genetic elements into stand-alone parts is that different modules can be assembled and tested individually to dissect their specific functionalities. Since the design maintains the same genetic features as the native genome, the refactored T7 genome should ideally maintain properties similar to those of the wild type while being easier to manipulate (7). The refactoring approach also has important applications for genetic pathway design to obtain better-controlled systems (8, 9).

Reducing Redundancy in Native Genomes

Are all DNA elements in a genome necessary? We would hypothesize the answer to be no, as some DNA elements are clearly redundant and others, such as transposable elements, can have deleterious effects in the short term, even though they may provide a strong selective advantage to a species in terms of adaptability in the long term (10). Additionally, cells may not need some genetic features when grown under stress-free or nutrient-rich conditions. *Escherichia coli*, a model organism for basic research and industrial applications, has been used to test these hypotheses. Researchers across laboratories reduced the size of the *E. coli* genome by 7% to 29.7% using a series of different designs (11–14), and the multiple-deletion series strains generated by Blattner's group were commercialized by Scarab Genomics. Claims were made regarding improved fitness and performance under certain conditions. But in some cases, the smaller synthetic genomes resulted in impaired strain fitness (14). One important outcome of this work was a test of gene essentiality in the *E. coli* genome, which was mapped in 1997 (15). Nevertheless, these *E. coli* genome reduction studies verified the hypothesis that native genomes could be simplified by following considered genome reduction designs.

Making Designer Alterations and Adding Elements to Endow New Functions

Eukaryotic genomes are more complicated than prokaryotic genomes due to their larger size and greater complexity of genome regulation and architecture. Designing such synthetic genomes requires deep consideration of many aspects, including basic biology, industrial potential, and biosafety. The Synthetic Yeast Genome Project (Sc2.0) is the first eukaryotic genome synthesis project in the world. The ultimate goal of Sc2.0 is to build a synthetic yeast genome that powers wild-type fitness while increasing genome versatility to probe new biological questions regarding gene content, genome structure–function relationships, and evolution. Therefore, the design

of Sc2.0 was a notable departure from the goals of previous genome synthesis projects in which the primary intention was to minimize the number of changes during design. Rather, during the design of Sc2.0, dramatic changes were made to the known *Saccharomyces cerevisiae* genome sequence to design a genome that can teach us new biology, the overarching theme of the project (16). First, retrotransposons, subtelomeric repeated sequences, other repeated genes such as *COS* and *PAU* that are presumed to be nonessential, and tRNA genes were removed or relocated to minimize sources of genome instability. Second, pre-mRNA and pre-tRNA introns were removed to eventually allow the following questions to be posed: Can all such introns be removed from the genome? Assuming that is possible, is it then possible to delete all or part of the splicing machinery from the genome? Third, stop codon swaps (a modest rewrite of the genetic code) and loxPsym site insertion [the basis for genome SCRaMbLE (17); see the section titled Insights from Synthetic Genomes] can endow the synthetic genome with new potential functions, allowing future genome modifications to provide new insights. Restriction site modifications facilitated the assembly of synthetic chromosomes. To distinguish synthetic products from native chromosomes, a watermarking system called PCRTags introduced synonymous changes into the nucleotide sequence within open reading frames to enable a polymerase chain reaction (PCR)-based assay for synthetic content. The PCRTags facilitate analysis of intermediates in the assembly of synthetic chromosomes to ensure not only that the synthetic sequences are incorporated but—an equally important practical consideration—that the native sequences are eliminated. The resulting synthetic genome has an ~8% size reduction and 1.1 megabases (Mb) of sequence alterations relative to the native genome (16). The design of Sc2.0 was a critical part of the experiment, and embarking on a systematic redesign makes the assumption that we have sufficient information about the yeast genome and biology in general to introduce massive changes, for example by deleting entire classes of genetic elements, but still retain high fitness once the genome is built. Importantly, the process used to build up the synthetic genome was segmental swapping of 30–60 kb at a time, which allowed for early tests of risky strategic elements of the design.

All genome synthesis projects to date have started with a known reference sequence and imposed some type of redesign, whether minor or major. An important prerequisite is to have an excellent understanding of the entire sequence of a genome in the form of a complete reference sequence. Additionally, having an accurate and comprehensive genome annotation is equally if not more critical and builds on the shoulders of communities that have spent decades dissecting the biology of an organism by more traditional methods. Efforts have been made to identify the functional elements in different genomes through both computational and experimental approaches. Experimentally, transposon-based mutagenesis methods (18, 19) and systematic gene knockouts (20, 21) have been used to identify gene essentiality in both bacteria and yeast. Similar studies were performed in several other species (22, 23). Computationally, genome comparison across subspecies is one way to identify gene essentiality. DNA elements that are present in one subspecies but are absent in other subspecies are more likely to be deletable, at least in stress-free, nutrient-rich conditions (13).

While genome characterization is important for designing a fully functional genome, even the most optimized genome design can still miss key unknown features that affect the complex biological system of a cell. We have termed these design bugs (24). Design bugs become apparent only during the in vivo phase of genome synthesis—for example, synonymous codon substitution, which affects growth fitness in *E. coli* (25), or watermark incorporation, which alters mRNA secondary structure (24) or introduces ectopic transcription binding sites (26) in yeast (see the section titled Insights from Synthetic Genomes). The mechanisms underlying these bugs might reveal unappreciated aspects of codon usage bias, translation efficiency, ribosome binding, or a multitude of other biological unknowns (27–29). As such, debugging, while operationally

frustrating and difficult, may represent one of the biggest contributions of the Sc2.0 project, as this is where current assumptions in our understanding of genome biology are challenged and, hopefully, new insights into biology are obtained.

Perhaps the ultimate alteration of biological systems is to fundamentally expand life's alphabet by expanding the number of bases from four to more. Important progress in this area has been made by the Romesberg and Benner groups. The Benner team (30) has recently reported expanding the alphabet by four, by strictly adhering to the known hydrogen bonding, geometric, and electrochemical features of natural base pairs to systematically explore otherwise conventional Watson–Crick base-pairing modalities. The resulting hachimoji set of eight nucleotides [and its precursor system, artificially expanded genetic information system (AEGIS)] is able to form conventional B-DNA structures in double-stranded DNA, even when the non-GATC bases are present in runs of contiguous base pairs. DNA and RNA polymerases that can incorporate the appropriate triphosphates in vitro have been identified, opening the possibility of deploying this expanded alphabet in vivo. However, the major challenge to such deployment inside living cells relates to engineering the entire suite of macromolecules required to efficiently import the synthetic precursors of the required nucleotides and convert them to triphosphates inside living cells. Impressive in vitro achievements with this system include the demonstration of functional RNA aptamers containing these nucleotides (30, 31) and the development of a large series of diagnostic tests based on PCR that use these orthogonal base pairs to dramatically enhance specificity (32).

An alternative approach, pioneered by Romesberg and colleagues (33), exploited a pair of modified nucleotides that interact hydrophobically rather than by hydrogen bonding. Here, too, engineering of polymerases was required to efficiently incorporate these unnatural bases. This approach, combined with innovative deployment of a nucleoside triphosphate transporter gene borrowed from a diatom genome, led to the impressive demonstration of not only the stable and persistent replication of DNA containing these bases but also their transcription into functional mRNA and tRNA, culminating in decoding and translation of a codon containing a synthetic base inside living bacteria (33). The novel base pair has an unusual structure in DNA in which the bases are slipped relative to each other rather than aligned as in Watson–Crick base pairing. This limitation means that unlike in the case of hachimoji DNA, it is likely to be difficult to make functional runs of such nucleotides, as their impact on DNA structure is less likely to permit proper decoding and stability properties.

BUILDING SYNTHETIC GENOMES FROM SCRATCH

Synthetic genomes are built from the bottom up using single-stranded oligonucleotides as starting material. Over the past two decades, automating and miniaturizing oligonucleotide synthesis has enabled significant cost reduction to meet the increased demand of the DNA synthesis market. However, synthesizing oligonucleotides longer than 200 nucleotides with high fidelity is still challenging due to the repetitive yield problem, which results in part from imperfect nucleotide chemical purity and chemical damage occurring during synthesis: even if each chemical step is 99% effective, only 13% of 200-mer molecules will be correct. Therefore, to produce longer DNA sequences to specification, single-stranded oligos are stitched together via a variety of methods into longer double-stranded DNA pieces, and the larger constructs are eventually assembled to genome size. In some cases, the synthetic DNA needs to be built in a different host to achieve higher assembly efficiency. In these cases, technologies for genome transplantation may be needed. This section reviews the methods of oligonucleotide synthesis, genome assembly, and genome delivery, discussing the strengths and weaknesses of each method.

Oligonucleotide Synthesis

After discovering the structure of DNA in the 1950s, researchers started trying to chemically join deoxyribonucleotides in vitro. The synthesis of dithymidine dinucleotide pioneered the field of in vitro DNA synthesis (34). In the 1980s, synthetic phosphoramidite chemistry was developed and subsequently automated and has since been used for more than 30 years (35). High-throughput and automated oligo synthesis technologies result in higher oligonucleotide production per day, particularly when they use microarrays that allow tens of thousands of oligos to be made in parallel in a pooled format (36). However, given the low yield and high error rates of array-based DNA synthesis, some obstacles had to be overcome for widespread use (37). Details of oligonucleotide synthesis methods have been reviewed elsewhere (38, 39).

Due to technical barriers, the chemical synthesis approach is unable to infinitely increase product length with high fidelity, making it challenging to meet the growing demand for DNA synthesis of products that can encode pathways or even of single genes. Also, hazardous organic solvents are used during synthesis. In contrast, enzyme-based single-stranded DNA synthesis methods have the potential to polymerize oligonucleotides faster and more accurately. Among all known DNA and RNA synthesis enzymes, terminal deoxynucleotidyl transferase (TdT) is dedicated to DNA synthesis in a template-independent manner, since it uses single-stranded DNA as an initiator and polymerizes nucleotides by adding deoxynucleotide triphosphates (dNTPs) stepwise to the 3′-OH group of the initiator, unlike other polymerases, which obligatorily use a double-stranded primer–template DNA complex. TdT is a member of the X family of DNA polymerases and works mostly in vertebrate immune systems to introduce a diversity of antigen receptors during replication because of its low fidelity (40). Bollum (41, 42) first isolated and characterized TdT in the 1960s. Studies show that TdT can incorporate multiple fluorescent nucleotides and catalyze DNA polymerization up to 8 kb on a surface (43), indicating that it has great potential for chip-based DNA synthesis. It is also capable of incorporating various nucleotide derivatives to endow more functionalities to DNA (44), such as nuclease resistance (45). Currently there are still many barriers to commercializing TdT for DNA synthesis, one of which is its template independence, which leads to random incorporation of nucleotides. Thus, dNTPs must be added one at a time and at very high purity. Efforts from both academic institutions and companies may lead to the development of efficient and effective enzymatic DNA synthesis technologies in the future.

Synthetic Genome Assembly

Typically, a gene or functional unit consists of hundreds to thousands of base pairs, a length scale well outside the range of oligonucleotide synthesis—not to mention chromosomes or genomes. Researchers have developed various protocols to assemble multiple oligos into designed sequences. Early gene synthesis relied on short oligonucleotides and T4 polynucleotide ligase. Following the chemical synthesis of two 12-mers, a 77-mer yeast alanine tRNA gene was synthesized by joining 17 segments varying in length from 5 to 20 nucleotides (2). Later approaches, which relied on DNA polymerases, include polymerase chain assembly (PCA) and PCR, which could assemble several kilobases of DNA from tens to hundreds of overlapped oligos in single-step assembly (46). By applying PCA and conventional cloning, Cello et al. (3) hierarchically built the synthetic genome of poliovirus. Methodological optimizations such as oligonucleotide gel purification and 55°C ligation prior to PCA have dramatically increased the assembly length and efficiency of PCA (5). Non-PCA-based gene synthesis methods have also emerged and have been reviewed elsewhere (47). Notably, synthetic DNA fragments can be directly assembled in yeast using overlapping oligonucleotides (48). A large number of biotechnology companies have taken

Table 1 Price filtering rules and the reasons for them in gene synthesis

Sequence feature	Effect on price	Reason(s)
Length	Longer → more expensive	PCR efficiency declines with length; mispriming increases with number of oligos; errors accumulate with length
Global GC or AT content	More extreme → more expensive	Extremes of base composition reduce information content of DNA, leading to mispriming
Homopolymers/copolymers	Longer → more expensive	Polymerase slippage leads to expansion/contraction of runs
Repeats in sequence	More/longer repeats → more expensive	Repeats lead to misassembly

over the work of joining oligos into gene-sized parts (at the time of writing, the length limit for commercial synthesis is typically less than 5 kb) using column- or microarray-synthesized oligonucleotides as building blocks. PCA-based methods often include an intrinsic error-correction step that employs nucleases that recognize and cut unpaired or improperly paired sequences; the action of these nucleases can weed out molecules that are mismatched due to errors in the underlying oligonucleotides. The most common type of error, a single-base deletion, greatly distorts the DNA helix, making it relatively easy to recognize and cut. A challenge is that these nucleases are mostly temperature sensitive so that including them in each cycle of the PCR is not practical, but adding them as often as possible should provide the highest degree of error correction. The oligo-to-gene synthesis procedures also involve subsequent validation by sequencing; if some parts fail to pass sequence verification, additional postsynthesis error-correction steps may be needed to produce final desired sequences, and the resulting price ranges from $0.10 to $0.15 per base pair in 2019 for finished cloned sequences. However, commercial prices for finished cloned sequences are typically subject to filtering rules that eliminate from consideration sequences that are difficult to synthesize, difficult to amplify, or even just difficult to sequence verify, such as molecules with extensive runs of homopolymers (e.g., A_n) or homocopolymers (e.g., $[AT]_n$), other direct repeats, or global or local extremes of AT or GC content. Special (i.e., higher) prices typically apply to such sequences and can raise the price of synthesis dramatically. Typical filtering rules and their effects on price are summarized in **Table 1**. Note that all of these sequence features are very commonly found in biological genomes, substantially increasing the price of de novo synthesis of native-like sequences.

Importantly, pooled oligonucleotide synthesis has the potential to radically decrease synthesis costs, but despite this, the challenges of converting these pools to cloned, (ideally) error-free DNAs have not yet resulted in precipitous price drops in the overall cost of gene-sized pieces that are as dramatic as the superexponential price drops that occurred in the development of DNA sequencing technology.

Alternatively, handing this labor-intensive gene synthesis work to undergraduate students—for example, the build-a-genome courses at Johns Hopkins University (49) and Tianjin University (50)—represent an approach to teach synthetic biology students practical gene and genome synthesis knowledge and to move synthetic genome projects forward in parallel.

Depending on the size and DNA content of the final assembly product, both in vitro and in vivo assembly methods have been used to assemble larger DNA from <5-kb building blocks (**Figure 2**). Among those in vitro DNA assembly methods, BioBricks and BglBricks are suitable for sequential assembly of several gene parts (51–53). However, the 8- or 6-bp scars generated at every junction are highly undesirable. Golden Gate/MoClo assembly was developed as a seamless way to ligate double-stranded DNA sequences in a defined order using type IIS restriction enzymes (54–56). This method has been demonstrated for highly efficient assembly of yeast transcription

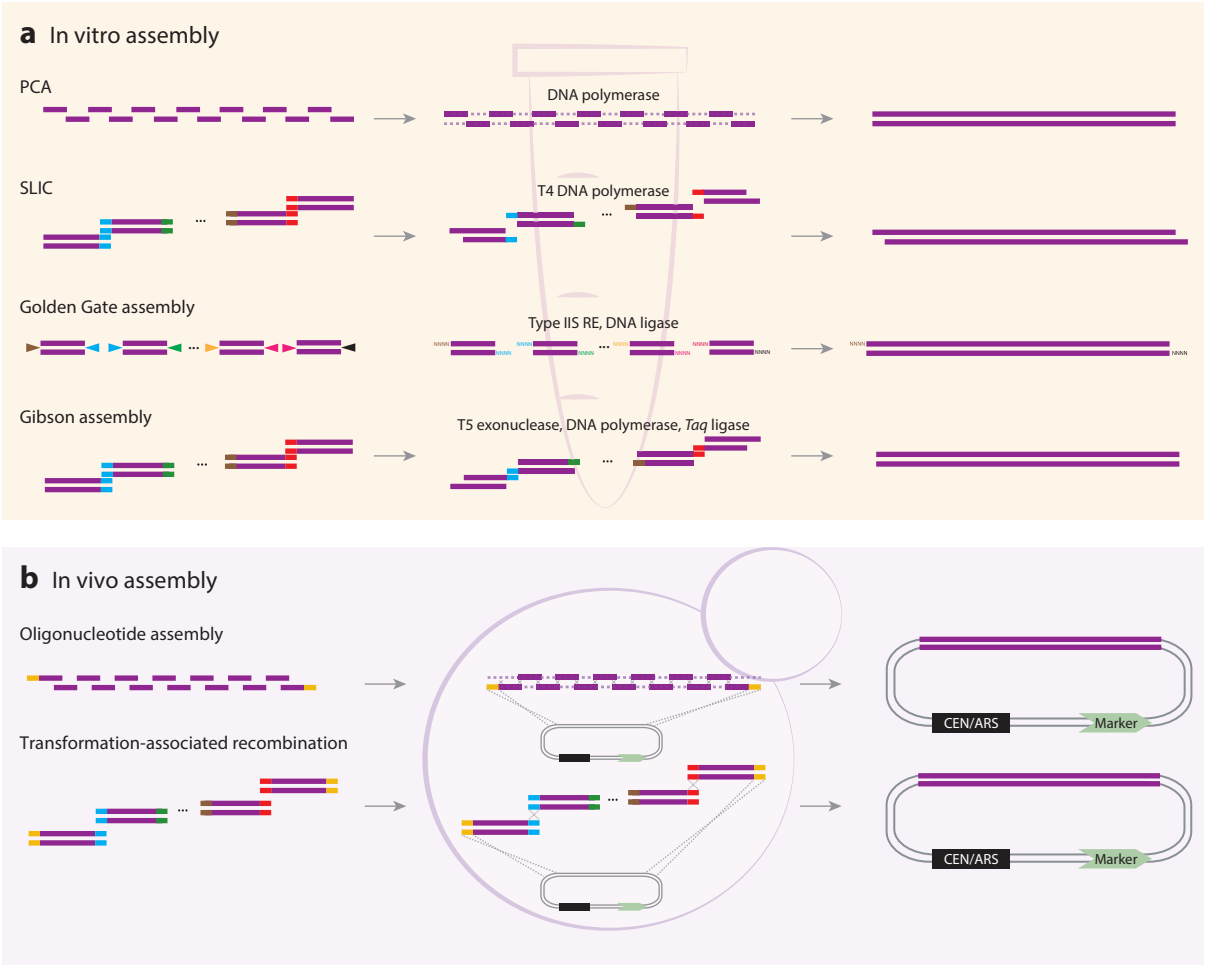

Figure 2

DNA assembly methods. (*a*) Four in vitro assembly methods. Polymerase cycling assembly (PCA) uses multiple oligonucleotides that share overlaps with adjacent ones as starting material. The oligonucleotides are annealed and joined during polymerase cycling, which is followed by regular PCR using two primers at each end to abundantly amplify the completely assembled product. Sequence- and ligation-independent cloning (SLIC) uses T4 DNA polymerase to digest pooled double-stranded linear DNA fragments that contain 20 base pairs or more of overlap to generate overhangs. DNA fragments with compatible overhangs are incubated with or without RecA to be linked together. The Golden Gate assembly method relies on two adapters that contain type IIS restriction enzyme (RE) sites on both sides of the double-stranded DNA fragments. The adapters are carefully designed to produce unique four-base-pair overhangs, which are compatible with the adjacent ones to make sure all digested fragments can be ligated in a defined order by T4 DNA ligase. The colored triangles represent double-stranded type IIS RE sites. Rightward triangles generate a top-strand overhang; leftward triangles generate a bottom-strand overhang. Overhangs are shown as NNNN. Gibson assembly also requires overlapping sequences between two adjacent double-stranded DNA fragments. The heat-labile T5 exonuclease is used to chew back one strand of the overlapping sequence in the early phase of the reaction. Adjacent pairs of DNA fragments are joined by annealing. DNA polymerase is used to fill in gaps, and *Taq* DNA ligase is used to seal the nicks. (*b*) Two in vivo assembly methods. Overlapping oligonucleotides with overlaps as short as 20 base pairs can be assembled into an acceptor vector in yeast. Overlapping double-stranded DNA fragments sharing overlaps by as short as 40 base pairs can be readily assembled in yeast via homologous recombination.

units (57) and for exogenous pathway assemblies that are up to six transcription units (58). The aforementioned restriction-enzyme-based assembly methods rely heavily on the sequence of the individual parts and more specifically on the presence and/or absence of the sites for efficient implementation. For example, the internal presence of a type IIS site used for Golden Gate assembly could cripple assembly efficiency if not appropriately accounted for. Another widely used in vitro assembly method relies on terminal sequence homology of adjacent parts, like a host of precursor methods such as sequence- and ligation-independent cloning. Gibson et al. (4) first developed this in vitro recombination method while cloning the synthetic *Mycoplasma genitalium* genome. This method requires three enzymes to sequentially chew back, anneal, and fill in to seamlessly join overlapping fragments. Further improvement of this method by using an isothermal (50°C) reaction consisting of heat-labile 5′ exonuclease, high-fidelity DNA polymerase, and *Taq* DNA ligase to increase the efficiency and the maximal size of assembly makes it a simple and robust cloning method (59). The Gibson assembly method has been commercialized and is widely used to make constructs in the tens of kilobases, with one publication providing evidence of a ∼900-kb assembly starting from very large pieces (59). The number of parts for Gibson assembly in a single step must be limited to maintain high efficiency (4, 60), meaning that, in practice, multiple intermediate steps may be required if the final DNA assembly product is large. Another limitation relative to restriction-enzyme-based parts strategies is that PCR introduces errors, and moreover, primers must be designed and ordered before DNA assembly can begin. Finally, not all primer pairs work in PCR, whereas restriction digests are much more reliable.

The budding yeast *S. cerevisiae* was first developed as a host for molecular cloning in the late 1970s and early 1980s (61–63). The ability of the organism to carry out general homologous recombination (HR) in mitotic cells at very high frequencies made it a natural and powerful crucible for DNA assembly reactions. Studies in the 1990s demonstrated that human genomic fragments hundreds of kilobases in size could be isolated by using transformation-associated recombination technology (64–66), an extrapolation of earlier classic work exploiting yeast HR (62, 63). This series of studies defined yeast as a preferred platform for large DNA cloning. Additionally, yeast can efficiently join multiple overlapping fragments via HR. For example, Gibson et al. (67) demonstrated that 25 pieces of ∼24-kb fragments can be assembled in yeast in one step by HR with overlaps as short as 80 bp. Gibson et al. (68) also assembled 10 pieces of 1-, 10-, or 100-kb fragments using yeast HR with high accuracy. By introducing orthogonal sequences as adapters, Mitchell et al. (69) designed a method called versatile genetic assembly system (VEGAS) to simultaneously assemble and test genetic pathways in yeast. Yeast is also capable of assembling Sc2.0 genome fragments, which are highly similar to its own genome (70–72). For example, sets of four ∼750-bp DNA building blocks were assembled into ∼2- to 4-kb minichunks, and four minichunks with one building block of overlap were subsequently assembled into larger chunks in yeast by using HR. In addition to yeast, *Bacillus subtilis* has been used as a DNA assembly host. Itaya et al. (73) integrated the whole genome of *Synechocystis* PCC6803 (3.5 Mb) into the *B. subtilis* genome via an HR-based approach, generating a 7.7-Mb composite genome. Using a modified *B. subtilis* genome (BGM) as a cloning vector, Itaya et al. (74) developed a more practical cloning protocol, called the domino method, that allows smaller contiguous DNAs from either commercial synthesis or PCR to integrate into BGM in a defined order. The cloned DNA in BGM can be transferred to another plasmid backbone in a practical way, broadening the scope of this method.

Until now, many synthetic genomes have been assembled by using combinatorial genome assembly tools. For example, restriction cloning and ligation methods were used for virus and bacteriophage synthetic genome assembly (2, 3). Gibson assembly and yeast HR-based assembly were used to assemble several bacterial genomes and subassemblies thereof, including *Mycoplasma*

genitalium (4), *Mycoplasma pneumoniae* (75), *Mycoplasma mycoides* (68), *Caulobacter ethensis-2.0* (*C. eth-2.0*) (76), and *E. coli* (77), as well as eukaryotic chromosomes of *S. cerevisiae* (24, 26, 50, 71, 78, 79) and *Phaeodactylum tricornutum* (80). A recent study showed the synthesis of a ~4-Mb *E. coli* genome with 61 codons, which set a new record for synthetic genome size in May 2019 (77). Together, the above-described genome assembly methods have proven powerful enough for us to build chromosomes and genomes designed in silico that are several megabases in size.

Synthetic Genome Delivery

Synthetic genomes have to be delivered and tested in living systems. Bacteriophage genomes can be directly introduced into a host as a function of their relatively small genome size and their ability to self-assemble (3, 5); this is not the case for bacterial and fungal genomes. Two major strategies have proven useful to deliver larger synthetic genomes: one-step delivery (**Figure 3*a***) and stepwise substitution (**Figure 3*b***).

One-step delivery. To transplant genomes between bacterial species, Lartigue et al. (81) developed technologies to isolate intact genomic DNA from *Mycoplasma mycoides* strain GM12 and move it into *Mycoplasma capricolum* recipient cells. Labroussaa et al. (82) used seven bacterial species as genome donors, with increasing phylogenetic distance from the *M. capricolum* recipient, and observed a negative correlation between genome transplantation efficiency and phylogenetic distance. However, to our knowledge, bacterial genome transplantation has been successfully demonstrated only in the Mollicutes (the bacterial class to which *Mycoplasma* spp. belong) and, with one exception, entirely within the genus *Mycoplasma* (a single instance of transplantation from the closely related genus *Mesoplasma* was also reported) (82, 83). Lartigue et al. (84) also developed a strategy to manipulate bacterial genomes in yeast by adding yeast propagation DNA elements into the wild-type genome of *M. mycoides* and delivered such genomic DNA into yeast via spheroplast transformation. The *M. mycoides* genome could then be engineered in yeast by using standard yeast genetic tools. However, transplanting such an engineered genome from yeast back into a recipient bacterium requires either inactivation of endogenous restriction enzyme systems in the recipient cells or in vitro methylation of the donor to protect the genome from restriction enzyme cleavage prior to transplantation (84). Genome delivery from bacterium A to yeast to bacterium B enables researchers to assemble designer synthetic bacterial genomes directly in yeast from smaller parts. An example is building the synthetic genome of *M. mycoides*, named JCVI-syn1.0, which was assembled in yeast and then transplanted into *M. capricolum* (68). Isolating genomes from bacteria or yeast may break the intact genomic DNA through shearing, especially for genomes larger than 1 Mb. Karas et al. (85) eliminated the genome isolation step by using cell fusion, in which bacterial genomes up to 1.8 Mb were delivered to yeast cells by membrane fusion between the bacteria and yeast spheroplasts. Cell fusion is also applicable for delivering DNA from yeast to mammalian cells. Brown et al. (86) reported that yeast centromeric plasmids could be delivered into cultured mammalian cell lines in a size-independent manner after optimizing the conditions of yeast spheroplast and mammalian cell fusion. Human and mouse artificial chromosomes have the ability to load more DNA than yeast artificial chromosomes, which in turn have higher capacity than bacterial artificial chromosomes (87–89). Human and mouse artificial chromosomes can also be transferred across different mammalian cell lines by using microcell-mediated chromosome transfer technology (90–92). Microcell-mediated chromosome transfer technology opens a new avenue for mammalian synthetic genomics research, and much technology development remains to be performed.

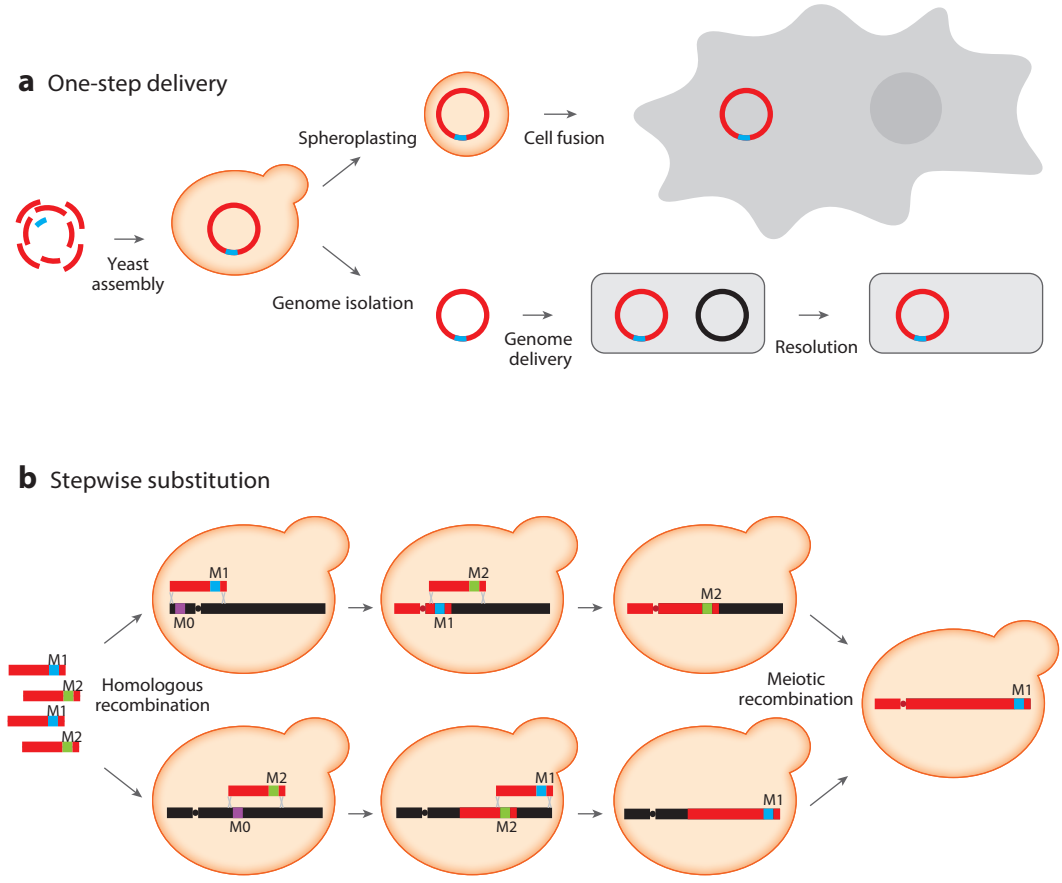

Figure 3

Genome delivery strategies. (*a*) Illustration of one-step genome delivery. Overlapping DNA fragments, including intermediate synthetic genome and yeast propagation core elements (*blue*), are assembled in yeast. The cell walls of yeast cells carrying a synthetic genome are digested to form spheroplasts, which can be fused with mammalian cells via polyethylene glycol treatment. Alternatively, the synthetic genome can be isolated from yeast and delivered into bacteria to replace the wild-type bacterial genome. (*b*) Illustration of stepwise substitution genome delivery. Intermediate synthetic genome fragments are preassembled using various assembly methods. Each intermediate fragment encodes one of two selection markers (M1 and M2) near the right end, and adjacent incoming fragments have different selection markers. Two entry strains are first engineered with a genetic marker (M0) located in different regions of the native chromosome. Following transformation, synthetic genome fragments are introduced into the host to replace the native counterpart by endogenous homologous recombination. The correct integrations are identified by selecting transformants with the $M1^{+}M0^{-}$ or $M2^{+}M0^{-}$ genotypes. Once one strain has the synthetic fragment integrated, eliminating the wild-type counterpart, it can be used as an entry strain for the next round of synthetic genome fragment integration. Finally, two intermediate synthetic strains containing different synthetic genome regions and sharing some overlap of synthetic sequences are first mated and then subjected to meiosis. By selecting the $M1^{+}M2^{-}$ genotype of the offspring, different synthetic genome regions are combined into one strain by meiotic recombination. Red lines represent double-stranded synthetic DNA fragments or genome, and black lines represent double-stranded wild-type chromosome.

Stepwise substitution delivery. The stepwise genome substitution method is suitable for large genome rewriting. This requires high HR efficiency of the host and watermarks for each step of substitution to enable rapid and inexpensive testing for replacement of wild-type with synthetic DNA. In the Sc2.0 project, each chromosome is divided into 30–60-kb megachunks, and adjacent megachunks are assigned one of two auxotrophic markers (*URA3* or *LEU2*) that are positioned

near the right end of each megachunk to facilitate switching auxotrophies progressively for integration (SwAP-In) (16). Taking advantage of endogenous HR in yeast, megachunks are iteratively introduced to overwrite the wild-type counterpart. PCRTags are introduced into most open reading frames during design of the synthetic chromosomes and are used to confirm the presence of synthetic DNA and the absence of wild-type DNA after each round of replacement. Similar partial genome replacement was performed in the bacterium *Salmonella typhimurium* to generate 1,557 synonymous leucine codon replacements in a 200-kb region (93). The replacement was termed stepwise integration of rolling circle amplified segments, as rolling circle amplification was used to generate the synthetic DNA fragments. Gene targeting in *E. coli* mainly relies on λ-Red-mediated recombination, which allows only short (2–3-kb) DNA integration. To increase the recombination efficiency, particularly for long (>100-kb) DNA replacement in *E. coli*, Wang et al. (94) developed a SwAP-In-like method useful in *E. coli* called replicon excision for enhanced genome engineering through programmed recombination (REXER). This system uses CRISPR–Cas9 to release double-stranded DNA of interest in vivo. Iterative REXER allows genome-wide substitution in a stepwise manner. The same group used REXER to systematically replace the *E. coli* genome in seven starting strains and eventually generated a ∼4-Mb recoded *E. coli* genome by combining the seven intermediate strains using conjugation-based assembly (77).

In a best-case scenario, one-step genome delivery can be faster than stepwise substitution. Unfortunately, this strategy is not robust and to date has failed. If the synthetic genome is unable to support cell viability due to design bugs, extensive debugging processes are necessary to locate the problematic design elements, figure out the reasons, and fix them. And obviously, if the bug leads to inviability, this can be very challenging. For example, Hutchison et al. (95) initially designed a hypothetical minimal genome (HMG) of *M. mycoides* that could not support cell viability. When eighths of the HMG were incorporated in the context of JCVI-syn1.0 to construct eight semisynthetic genomes, only one of them produced viable cells, consistent with multiple design bugs that needed to be corrected in the HMG design. Great efforts were taken to reclassify and test gene essentiality in *M. mycoides*. Eventually, three more rounds of the design-build-test cycle were performed to massively modify the sequence of HMG, which produced a viable strain supported by only 473 genes. In contrast, a stepwise substitution is more convenient for debugging. For the Sc2.0 project, if one megachunk incorporation causes an undesired phenotype, it is relatively easy to focus on a 30–60-kb synthetic sequence to map the problems. For the *E. coli* codon compression whole-genome synthesis project, Fredens et al. (77) replaced a ∼100-kb *E. coli* segment and then performed fitness test assays, and it was also relatively easy to pinpoint the problematic sequences and fix them.

TOP-DOWN GENOME EDITING

De novo genome synthesis can be used to construct designer genomes with as many useful features as the designer can envision. However, unless the density of changes is extremely high—as in Sc2.0, in which clusters of edits are positioned every ∼400 bp (16)—for larger genomes it is not necessarily cost effective to synthesize the entire genome from scratch. Alternatively, modifying an existing genome within its host cell offers an option to engineer the whole genome, for example, when only some changes are needed that may not cluster within one region of the genome. Studies have elucidated that DNA double-strand breaks boost the efficiency of DNA recombination across different organisms (96–99), leading researchers to discover DNA-targeting reagents to specifically break the genome and stimulate DNA repair via HR. Genome-editing tools including zinc finger nucleases, transcription-activator-like effector nucleases, and the widely adopted CRISPR–Cas9 system have been developed for genome modification in bacteria, fungi, invertebrates, vertebrates,

and plants. The detailed mechanisms and applications for each genome-targeting tool have been reviewed elsewhere (100, 101). Recently, CRISPR–Cas9 was used to facilitate chromosome fusion in yeast to massively change the three-dimensional structure of the genome (102, 103).

Determining how to edit genomes in a high-throughput but precise manner has always been a challenge. Zinc finger nucleases and transcription-activator-like effector nucleases rely on their DNA-binding domains to target breaks to specific DNA sequences, but it is difficult to routinely engineer tens to thousands of those proteins for high-throughput genome modification. The CRISPR–Cas9 system uses a single guide RNA (sgRNA) to specify the target sequence, and multiplexing sgRNAs to achieve genome-wide editing is relatively easy. Using a pooled library containing 73,151 sgRNAs to target 7,114 genes, Wang et al. (104) demonstrated the feasibility of using high-throughput CRISPR–Cas9 to perform genetic screens in human cell lines. Similarly, Joung et al. (105) generated an sgRNA library to target 10,504 intergenic long noncoding RNA transcriptional start sites to identify functional long noncoding RNA loci in human melanoma cells to explore the non-protein-coding genome. A high-throughput CRISPR interference experiment was performed in erythroleukemia cells to identify regulatory elements and dissect the network of noncoding variants to human disease (106). By harnessing CRISPR–Cas9 in a porcine kidney epithelial cell line, Yang et al. (107) disrupted 62 copies of the porcine endogenous retrovirus *pol* gene with just two sgRNAs, demonstrating that CRISPR–Cas9 is capable of targeting multiple near-identical loci without requiring multiple sgRNAs.

The efficiency of HR in *E. coli* was significantly increased by introducing the bacteriophage λ Red system, in which β-protein-mediated single-stranded DNA recombination has been widely used for whole-genome editing. A key study showed that synthetic single-stranded DNAs with homology arms as short as 30 bp can be incorporated into the lagging strand during DNA replication (108). To program *E. coli* cells in a large-scale fashion, Wang et al. (109) optimized a list of parameters for oligo-mediated allelic replacement and developed multiplex automated genome engineering (MAGE). By using a prototype device to automate the iterative MAGE procedure, they showed how MAGE could optimize metabolic flux. To select MAGE-modified cells more efficiently, Wang and colleagues (110–112) developed the coselection MAGE strategy by adding selectable markers near the modification region, and they used coselection MAGE for rare arginine codon replacement in essential genes of *E. coli* (113). Yeast oligo-mediated genome engineering is similar to MAGE but is applied in yeast (114, 115). In addition to metabolic flux optimization, MAGE was also used to replace all 314 TAG stop codons with TAA. Isaacs et al. (116) started the TAG-to-TAA replacement from 32 strains in parallel, in which each one had 10 or fewer TAG–TAA swaps. After ensuring that each strain had all designed alterations, they combined strains via conjugative assembly genome engineering to produce one *E. coli* strain with all 314 TAG–TAA changes. Although the TAG codon replacement led to poor fitness, adaptive evolution recovered nearly all the lost fitness after ∼1,100 generations (117). This reserving of the TAG codon for reassignment to a different function allowed for the subsequent introduction of orthogonal tRNA/synthetase systems that can incorporate unnatural amino acids at TAG codons strategically engineered into the genomically recoded *E. coli* genome, e.g., to provide control of growth of the organism to the investigator via provision of the unnatural amino acid in what is referred to generically as a genetically recoded organism (118). Genetic code expansion technologies and their challenges have been reviewed elsewhere (119, 120).

Bottom-up genome synthesis and top-down genome editing have distinct advantages and disadvantages. Bottom-up synthesis permits closely spaced designer changes positioned densely across a synthetic genome. Using a bottom-up synthesis approach, it is more feasible to incorporate all the designed elements at once. Also, with extensive DNA sequencing at every step of genome assembly, the final synthetic genome assembly to be delivered can be highly accurate.

However, at the time of writing, bottom-up synthesis remains prohibitively expensive for targets that are a megabase or larger. In addition, certain sequences, such as repeats required for telomeres and centromeres, can be exceedingly challenging for DNA synthesis and assembly. Some of these disadvantages are being addressed: As demand increases and oligonucleotide synthesis automation and miniaturization continue (121), the cost of DNA synthesis and assembly will continue to drop. Through the use of ligation-based cloning methods, more and more highly repetitive DNA sequences can be assembled accurately (122, 123).

Top-down genome editing uses site-specific targeting tools to introduce designed changes, which has the advantage that the genome size of a host is less relevant, provided that host-specific editing systems have been developed. Top-down genome editing is particularly suited to introducing a few changes quickly and economically. However, a conspicuous disadvantage of top-down editing is off-target editing caused by the imperfect specificity of targeting tools. The off-target probability is further increased by the multiple rounds of editing processes involved in top-down genome editing. To address the off-target problem, Zhang's group (124, 125) combined a Cas9 nickase mutant with a pair of guide RNAs to introduce DNA double-strand breaks. Several other groups fused cytidine or adenine deaminase to different DNA-binding modules including catalytically dead Cas9 (dCas9), Cas9 nickase, zinc finger, and transcription-activator-like effector to convert C to T or A to G without introducing DNA double-strand breaks (126–129). But this problem is far from being solved thoroughly right now, and more precise and accurate editing tools are sorely needed to achieve this goal.

INSIGHTS FROM SYNTHETIC GENOMES

Constructing synthetic genomes requires careful design, new technology development, and laborious bench work. In turn, synthetic genome projects pay back with opportunities for rich biological discovery. This section focuses on insights derived from synthetic genomes.

The initial design of the minimal *Mycoplasma* genome included only essential genes as determined by previous experiments (95). Fitness measurements led to discoveries of quasi-essential genes that are crucial for viability of a minimal genome. The Sc2.0 project has identified a series of bugs that affect cell growth, such as synonymous PCRTag recoding in *PRE4* (24), *FIP1* (26), and *MMM1* (79) that impaired their expression; a loxPsym site insertion that disrupted the promoter of the adjacent *ATP2* gene (26); and a putative intron deletion of *YLR202C* that abolished adjacent *COQ9* gene expression (79). Global elimination of TCG, TCA, and TAG codons in *E. coli* showed that recoding of the *yceQ* and *yccY* genes altered the 5′ untranslated region of *rne*, where its promoter elements are located (77). In the process of building a 57-codon *E. coli* genome, 55 of 87 synthetic chromosomal segments were tested individually by a complementation experiment. Codon changes in the *yceD* gene were found to disrupt the promoter of operon *rpmF–accC* and cause a fitness defect (130). Moreover, Venetz et al. (76) evaluated the functionality of the synthetic *C. eth-2.0* segments in merodiploid strains and found that 432 of 530 extensively recoded essential and semiessential genes in *C. eth-2.0* were functional. In other words, a high percentage of previously predicted genetic features within essential coding sequences, such as alternative open reading frames and predicted internal transcriptional start sites, actually turned out to be nonessential for function.

From an applications perspective, synthetic genomes bring many benefits that are otherwise difficult to obtain. In terms of viruses, by using over- or underrepresented synonymous codons in the poliovirus capsid coding sequence, the virus could be largely attenuated for vaccine development (131, 132). The J. Craig Venter Institute applied their genome assembly technology to synthesize influenza vaccine viruses, potentially greatly accelerating seasonal influenza vaccine

production (133). For bacteria, precise genome reduction in *E. coli* achieved higher electroporation efficiency and higher fidelity for plasmid propagation compared with wild type (13). By comparing T7 bacteriophage fitness and lysis time in genomically recoded *E. coli*, Lajoie et al. (118) concluded that the recoded *E. coli* strain has increased resistance to T7 bacteriophage after the deletion of release factor 1. More codon compression studies will further expand the genetic code and endow *E. coli* with new biological functions in the near future (77, 130), potentially leading to complete resistance to natural phages. In terms of yeast, the Sc2.0 design specifies loxPsym site placement downstream of nonessential genes across the entire genome. Insertion of loxPsym sites enables whole-genome rearrangement upon Cre activation (16), a technique termed synthetic chromosome rearrangement and modification by loxP-mediated evolution (SCRaMbLE). Using both linear and circular synthetic chromosome IXR, SCRaMbLE can generate abundant genome diversity after several hours of Cre induction (134, 135). SCRaMbLE has been developed for both haploid and heterozygous diploid synthetic yeast strains. SCRaMbLE regulation systems have been optimized and reconfigured, including light-controlled (136) and galactose-controlled (137) Cre recombinase combined with a SCRaMbLE indicator (138). These systems have shown that SCRaMbLE can be used not only to screen yeast strains according to a fitness phenotype, with high tolerance to temperature, ethanol, caffeine, acetic acid, and xylose (138–140), but also to select for genomic backgrounds that increase production of useful compounds such as β-carotene, violacein, and penicillin (140–142). The mechanisms underlying these characteristics have been dissected through whole-genome sequencing of SCRaMbLEd strains, directing the more purposeful design of genomes for biotech applications. In terms of mammalian cells, genome-wide disruption of all porcine endogenous retrovirus *pol* genes generates a genetically modified porcine kidney epithelial cell line (PK15) that has reduced porcine endogenous retrovirus transmission from PK15 cells to human HEK293 cells by up to 1,000-fold (107). This work sheds light on future clinical applications such as organ transplantation from pig to human.

ETHICAL CONSIDERATIONS

Genome synthesis and genome editing provoke substantial ethical considerations, particularly for edited organisms that may be designed to exist outside the laboratory. A natural progression exists from editing of agricultural products, to creation of gene drives to modify wild populations and restore extinct species, to, finally and most controversially, editing leading to humans with designed genomes.

For agricultural products, CRISPR–Cas9 and similar editing tools can be used to combine naturally occurring beneficial alleles that would otherwise require extensive breeding programs. Although the product genome could have been generated naturally, the process itself involves engineering. This has led to divergent regulation in Europe and the United States. European courts focus on the process and regulate these as genetically modified organisms (143). In contrast, the US Department of Agriculture does not consider these to be genetically modified organisms (144).

Gene drives, which harness CRISPR–Cas9 to spread an engineered change through a wild population, often with the goal of pathogen control, have generated substantial discussion about appropriate regulation (145). Several gene drive systems have been tested in large, contained populations, and work on mosquitos and rodents is nearing the point at which uncontained tests may be considered, possibly in isolated ecosystems such as islands (146).

Human genome editing is considered in depth in a joint report by the US National Academies of Science, Engineering, and Medicine (147). This report separates genome editing into somatic (nonheritable) and germ line (heritable) and separates types of changes into disease correction

and enhancement. For disease correction by somatic editing, frameworks for gene therapy already exist. New recommendations for disease correction by germ-line editing are to restrict envisioned therapies to reversion of strongly penetrant (Mendelian) disease-causing alleles back to the natural alleles observed in unaffected individuals. For genomic enhancement, the recommendation is to discuss governance of somatic editing for enhancement, but heritable editing for enhancement is not to be authorized by any regulatory agency. Challenges in distinguishing between disease correction and enhancement remain. Also challenging is proponents of genome editing making public statements that no experiments will be conducted without proper regulatory oversight, eliding the National Academies recommendation that no such experiments should be considered.

FUTURE DIRECTIONS—WHAT IS NEXT?

Genome synthesis enables us to understand biological systems and benefit from them as never before. The growing ability to write DNA motivates studies of biological questions across different organisms, including humans. Genome Project-write (GP-write) proposed to understand the genetic blueprints for genome construction, including in animals, plants, and humans (148). GP-write hopes to stimulate the rate of progress in technology development in all aspects of genome writing. In particular, dramatic decreases in the cost of building and testing large genomes are being sought. In the meantime, de novo genome synthesis relies on thoroughly understanding the sequence of a genome, and many genome sequences are still underexplored, especially centromeric and other heterochromatic, repeat-ridden regions of the genomes of multicellular organisms (149). New DNA sequencing technologies, particularly those exploiting long reads, need further development to uncover this dark matter of the genome sequence. In addition, to design more complex genomes, we may ultimately need more information—such as metabolic data, clinical data, genome-wide association study data, chromatin immunoprecipitation sequencing data, assay for transposase-accessible chromatin sequencing data, and chromatin three-dimensional organization information—to integrate into design algorithms. A surprising aspect of Sc2.0 has been that the design successfully ignored these considerations; they may, however, be important for mammalian chromosome design.

Recently, the GP-write consortium announced as its first target the engineering of a human cell line that is resistant to natural viruses, referred to as an ultrasafe cell line. By compressing the genetic code in these cell lines and deleting components such as the specific tRNAs that are required to decode those eliminated codons, it should be possible to produce cells that resist viruses due to a lack of the comprehensive translation machinery that is required by all natural viruses. To survive in such a cell, a virus would need to mutate many of its codons to the compressed code. Paradoxically, such a virus could be readily synthesized.

Genome synthesis, if done carefully, takes a huge amount of effort and resources. The experience of the Sc2.0 project has shown that a consortium can be highly effective in building genomes, with a master design and adoption of individual chromosomes by consortium members to permit rapid innovation and knowledge sharing.

As we consider the rapid pace of technology development, we also ponder future versions of synthetic genomes that will allow biologists to probe genome regulation, the limits of genome change, and how solid (or porous) species boundaries are. Also, we wonder, can the field ultimately deliver on the popular and overarching synthetic biology vision that genome engineers will someday be able to sit down at a computer, sketch out a vision for a genome in a truly novel design, and build the genome of an organism truly from scratch to perform a specific biological or technological function?

The practical applications of genome writing technology are manifold and have been reviewed elsewhere (101); they range from biomedicine, to agriculture, to the design of new living materials, to the transition from the petroleum economy to the bioeconomy. In addition to glamorous de-extinction projects like the passenger pigeon and the woolly mammoth, genome writing could be used to restore genetic variation to existing near-extinct populations suffering from population bottlenecks, such as the black-footed ferret. However, in this review we focus on how synthetic genomics will help us learn new biology.

Synthetic genomics will deliver much new information on how genomes are wired. The Dark Matter Project we have launched seeks to employ genome synthesis methods to tackle big, open questions surrounding the fundamentals of gene regulatory control, not by complete genome synthesis but by the construction of tens to hundreds of assemblons of 100 kb and more that feature either systematically designed and built variants or combinations of natural variants not readily available in nature. These can then be delivered specifically to predetermined sites in embryonic stem cells and other cells (150, 151), opening up the possibility of extensive functional analysis in multiple cell types or in living mice. Even though this scale permits the analysis of only one or a few genes at a time when deployed in mammalian cells, we believe it will be fertile ground for mechanistic discovery that could form the basis for modeling of genome-wide properties. Assemblon technology can be deployed to probe interactions among enhancers, insulators, and sites of binding of other proteins as well as the effects of all of these on expression. By systematic variation of introns, the importance of splicing can be systematically evaluated. Another important application is the dissection of genome-wide association study hits, which are mapped locations in the human genome that underlie genetic disease, sometimes identifying a causal variant but usually extending over a large haplotype with many candidate variants. These often fall in noncoding DNA, making it difficult to assign the target gene, since noncoding elements can operate over distances as large as a megabase in mammals. Finally, the limits of natural genome variation may become accessible in organismal oddities with unconventional genome structures and features, like sponges and octopuses (152, 153), if the appropriate gene delivery tools can be developed.

What are the limits to genome change? The Sc2.0 and *E. coli* genome resynthesis projects are good examples of this approach. In both cases, there is change to the genetic code, applied on a massive scale. Certainly, on a scale of deleting and reassigning one, three, or even seven codons, it seems that life can work. Impressively, a component of the decoding machinery can also be deleted in at least one case, namely the UAG-specific release factor RF1 in *E. coli* (154), and can even be re-engineered to reassign UAG to specify an unnatural amino acid (155, 156). What are the limits to such genetically recoded organisms? In theory, all 64 codons in the table could be assigned to as many as 63 amino acids and 1 stop codon, but this seems unlikely because of limits already encountered in deleting just 3 or 7 codons from the code. Also, an essential tRNA could be deleted from a synthetic *E. coli* genome; Fredens et al. (77) were able to delete *serT*, an essential tRNA gene decoding the TCA codon, as well as *serU*, a nonessential tRNA gene decoding the TCG codon, and the RF1 gene (*prfA*), because all three corresponding codons were removed in the synthetic *E. coli* with 61 codons. *E. coli*-57, a genome using just 57 sense codons, has been designed but only partially built, and segments were identified that led to inviability. Similarly, with the Sc2.0 genome, once completed, we will be in a position to ask whether we can remove all introns from this eukaryotic genome, and if so, whether we can delete part or all of the genes encoding the machinery that performs splicing. Another series of questions revolves around how much we can change genome structure. Can we circularize all the chromosomes by design, as has been done with some of the yeast chromosomes (24, 50, 157, 158)? Is it possible to increase chromosome number by chromosome splitting and still retain stability of a very large set of smaller chromosomes (159)?

Can we exceed current limits to ploidy through similar tricks of genome engineering (160)? And finally, what are the limits to expanded alphabets of nucleotides?

A final consideration relates to the real distinction between genome editing and genome writing. Will we ever be able to really design genomes from scratch—that is, can we ever break away from editing of existing code? For example, will a future genome designer be able to tap into a vast database of information on plant physiology, genetics, infectious disease, and metabolism and design a food crop that combines drought resistance, insect and fungus resistance, and high nutrition levels by selecting these traits from across 25 or 2,500 different species of plants and produce a living organism? How would such a designer go about producing the first plant cell that combined all of these traits? Stepwise acquisition would likely produce inviable intermediate species and would be intolerably slow. These are just a few of the formidable challenges that make most biologists intensely skeptical that such a future could ever exist. However, our colleagues in engineering are equally confident that if enough facts and specifications are known, anything can be built. This creative tension between biologists and engineers conspires to make synthetic genomics one of the most exciting scientific fields to be working in at this time.

DISCLOSURE STATEMENT

J.D.B. is a founder and director of Neochromosome Inc., the Center of Excellence for Engineering Biology, and CDI Laboratories Inc. and serves or has served on the scientific advisory boards of Modern Meadow Inc., Recombinetics Inc., Sample6 Inc., and Sangamo Therapeutics Inc. J.S.B. is a founder and director of Neochromosome Inc. and serves on the scientific advisory board of AI Therapeutics Inc. L.A.M. is a founder and president of Neochromosome Inc.

ACKNOWLEDGMENTS

J.D.B. and J.S.B. thank the National Science Foundation and National Institutes of Health for grant support.

LITERATURE CITED

1. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, et al. 2018. Earth BioGenome Project: sequencing life for the future of life. *PNAS* 115(17):4325–33
2. Agarwal KL, Büchi H, Caruthers MH, Gupta N, Khorana HG, et al. 1970. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* 227(5253):27–34
3. Cello J, Paul AV, Wimmer E. 2002. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* 297(5583):1016–18
4. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, et al. 2008. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 319(5867):1215–20
5. Smith HO, Hutchison CA, Pfannkoch C, Venter JC. 2003. Generating a synthetic genome by whole genome assembly: φX174 bacteriophage from synthetic oligonucleotides. *PNAS* 100(26):15440–45
6. Sanger F, Coulson AR, Friedmann T, Air GM, Barrell BG, et al. 1978. The nucleotide sequence of bacteriophage φX174. *J. Mol. Biol.* 125(2):225–46
7. Chan LY, Kosuri S, Endy D. 2005. Refactoring bacteriophage T7. *Mol. Syst. Biol.* 1:2005.0018
8. Temme K, Zhao D, Voigt CA. 2012. Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *PNAS* 109(18):7085–90
9. Ren H, Biswas S, Ho S, van der Donk WA, Zhao H. 2018. Rapid discovery of glycocins through pathway refactoring in *Escherichia coli*. *ACS Chem. Biol.* 13(10):2966–72
10. Chao L, Vargas C, Spear BB, Cox EC. 1983. Transposable elements as mutator genes in evolution. *Nature* 303(5918):633–35

11. Yu BJ, Sung BH, Koob MD, Lee CH, Lee JH, et al. 2002. Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/*loxP* excision system. *Nat. Biotechnol.* 20:1018–23
12. Kolisnychenko V, Plunkett G, Herring CD, Fehér T, Pósfai J, et al. 2002. Engineering a reduced *Escherichia coli* genome. *Genome Res.* 12(4):640–47
13. Pósfai G, Plunkett G, Fehér T, Frisch D, Keil GM, et al. 2006. Emergent properties of reduced-genome *Escherichia coli*. *Science* 312(5776):1044–46
14. Hashimoto M, Ichimura T, Mizoguchi H, Tanaka K, Fujimitsu K, et al. 2005. Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.* 55(1):137–49
15. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277(5331):1453–62
16. Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, et al. 2017. Design of a synthetic yeast genome. *Science* 355(6329):1040–44
17. Dymond J, Boeke J. 2012. The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioengineered* 3(3):168–71
18. Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, et al. 1999. Global transposon mutagenesis and a minimal mycoplasma genome. *Science* 286(5447):2165–69
19. Ross-Macdonald P, Coelho PSR, Roemer T, Agarwal S, Kumar A, et al. 1999. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* 402(6760):413–18
20. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, et al. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2:2006.0008
21. Giaever G, Chu AM, Ni L, Connelly C, Riles L, et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418(6896):387–91
22. Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, et al. 2003. Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *PNAS* 100(24):14339–44
23. Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, et al. 2011. The essential genome of a bacterium. *Mol. Syst. Biol.* 7:528
24. Mitchell LA, Wang A, Stracquadanio G, Kuang Z, Wang X, et al. 2017. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* 355(6329):eaaf4831
25. Lajoie MJ, Kosuri S, Mosberg JA, Gregg CJ, Zhang D, Church GM. 2013. Probing the limits of genetic recoding in essential genes. *Science* 342(6156):361–63
26. Wu Y, Li B-Z, Zhao M, Mitchell LA, Xie Z-X, et al. 2017. Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* 355(6329):eaaf4706
27. Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12(1):32–42
28. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, et al. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141(2):344–54
29. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324(5924):218–23
30. Hoshika S, Leal NA, Kim M-J, Kim M-S, Karalkar NB, et al. 2019. Hachimoji DNA and RNA: a genetic system with eight building blocks. *Science* 363(6429):884–87
31. Biondi E, Benner SA. 2018. Artificially expanded genetic information systems for new aptamer technologies. *Biomedicines* 6(2):53
32. Yaren O, Alto BW, Gangodkar PV, Ranade SR, Patil KN, et al. 2017. Point of sampling detection of Zika virus within a multiplexed kit capable of detecting dengue and chikungunya. *BMC Infect. Dis.* 17(1):293
33. Zhang Y, Ptacin JL, Fischer EC, Aerni HR, Caffaro CE, et al. 2017. A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* 551(7682):644–47
34. Michelson AM, Todd AR. 1955. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3′:5′-internucleotidic linkage. *J. Chem. Soc. Resumed* 1955:2632–38
35. Beaucage SL, Caruthers MH. 1981. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* 22(20):1859–62
36. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* 251(4995):767–73

37. LeProust EM, Peck BJ, Spirin K, McCuen HB, Moore B, et al. 2010. Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res.* 38(8):2522–40
38. Kosuri S, Church GM. 2014. Large-scale *de novo* DNA synthesis: technologies and applications. *Nat. Methods* 11(5):499–507
39. Hughes RA, Ellington AD. 2017. Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.* 9(1):a023812
40. Motea EA, Berdis AJ. 2010. Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase. *Biochim. Biophys. Acta Proteins Proteom.* 1804(5):1151–66
41. Bollum FJ. 1960. Calf thymus polymerase. *J. Biol. Chem.* 235(8):2399–403
42. Bollum FJ. 1962. Oligodeoxyribonucleotide-primed reactions catalyzed by calf thymus polymerase. *J. Biol. Chem.* 237:1945–49
43. Tjong V, Yu H, Hucknall A, Rangarajan S, Chilkoti A. 2011. Amplified on-chip fluorescence detection of DNA hybridization by surface-initiated enzymatic polymerization. *Anal. Chem.* 83(13):5153–59
44. Tauraitė D, Jakubovska J, Dabužinskaitė J, Bratchikov M, Meškys R. 2017. Modified nucleotides as substrates of terminal deoxynucleotidyl transferase. *Molecules* 22(4):672
45. Kuwahara M, Obika S, Takeshima H, Hagiwara Y, Nagashima J, et al. 2009. Smart conferring of nuclease resistance to DNA by 3′-end protection using 2′,4′-bridged nucleoside-5′-triphosphates. *Bioorg. Med. Chem. Lett.* 19(11):2941–43
46. Stemmer WPC, Crameri A, Ha KD, Brennan TM, Heyneker HL. 1995. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 164(1):49–53
47. Xiong A-S, Peng R-H, Zhuang J, Liu J-G, Gao F, et al. 2008. Non-polymerase-cycling-assembly-based chemical gene synthesis: strategies, methods, and progress. *Biotechnol. Adv.* 26(2):121–34
48. Gibson DG. 2009. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.* 37(20):6984–90
49. Dymond JS, Scheifele LZ, Richardson S, Lee P, Chandrasegaran S, et al. 2009. Teaching synthetic biology, bioinformatics and engineering to undergraduates: the interdisciplinary Build-a-Genome course. *Genetics* 181(1):13–21
50. Xie Z-X, Li B-Z, Mitchell LA, Wu Y, Qi X, et al. 2017. "Perfect" designer chromosome V and behavior of a ring derivative. *Science* 355(6329):eaaf4704
51. Knight T. 2003. *Idempotent vector design for standard assembly of biobricks*. Rep., MIT Synth. Biol. Work. Group, Cambridge, MA. **http://hdl.handle.net/1721.1/21168**
52. Shetty RP, Endy D, Knight TF. 2008. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* 2:5
53. Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA, et al. 2010. BglBricks: a flexible standard for biological part assembly. *J. Biol. Eng.* 4:1
54. Engler C, Kandzia R, Marillonnet S. 2008. A one pot, one step, precision cloning method with high throughput capability. *PLOS ONE* 3(11):e3647
55. Engler C, Gruetzner R, Kandzia R, Marillonnet S. 2009. Golden Gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLOS ONE* 4(5):e5553
56. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. 2011. A modular cloning system for standardized assembly of multigene constructs. *PLOS ONE* 6(2):e16765
57. Agmon N, Mitchell LA, Cai Y, Ikushima S, Chuang J, et al. 2015. Yeast Golden Gate (yGG) for the efficient assembly of *S. cerevisiae* transcription units. *ACS Synth. Biol.* 4(7):853–59
58. Guo Y, Dong J, Zhou T, Auxillos J, Li T, et al. 2015. YeastFab: the design and construction of standard biological parts for metabolic engineering in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 43(13):e88
59. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6(5):343–45
60. Mitchell LA, Cai Y, Taylor M, Noronha AM, Chuang J, et al. 2013. Multichange isothermal mutagenesis: a new strategy for multiple site-directed mutations in plasmid DNA. *ACS Synth. Biol.* 2(8):473–77
61. Hinnen A, Hicks JB, Fink GR. 1978. Transformation of yeast. *PNAS* 75(4):1929–33

62. Winston F, Chumley F, Fink GR. 1983. Eviction and transplacement of mutant genes in yeast. *Methods Enzymol.* 101:211–28
63. Orr-Weaver TL, Szostak JW, Rothstein RJ. 1981. Yeast transformation: a model system for the study of recombination. *PNAS* 78(10):6354–58
64. Larionov V, Kouprina N, Graves J, Chen XN, Korenberg JR, Resnick MA. 1996. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *PNAS* 93(1):491–96
65. Larionov V, Kouprina N, Solomon G, Barrett JC, Resnick MA. 1997. Direct isolation of human *BRCA2* gene by transformation-associated recombination in yeast. *PNAS* 94(14):7384–87
66. Kouprina N, Annab L, Graves J, Afshari C, Barrett JC, et al. 1998. Functional copies of a human gene can be directly isolated by transformation-associated recombination cloning with a small 3′ end target sequence. *PNAS* 95(8):4469–74
67. Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, et al. 2008. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *PNAS* 105(51):20404–9
68. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, et al. 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329(5987):52–56
69. Mitchell LA, Chuang J, Agmon N, Khunsriraksakul C, Phillips NA, et al. 2015. Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Res.* 43(13):6620–30
70. Lin Q, Jia B, Mitchell LA, Luo J, Yang K, et al. 2015. RADOM, an efficient *in vivo* method for assembling designed DNA fragments up to 10 kb long in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 4(3):213–20
71. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, et al. 2014. Total synthesis of a functional designer eukaryotic chromosome. *Science* 344(6179):55–58
72. Muller H, Annaluru N, Schwerzmann JW, Richardson SM, Dymond JS, et al. 2012. Assembling large DNA segments in yeast. *Methods Mol. Biol.* 852:133–50
73. 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. *PNAS* 102(44):15971–76
74. Itaya M, Fujita K, Kuroki A, Tsuge K. 2008. Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nat. Methods* 5(1):41–43
75. Benders GA, Noskov VN, Denisova EA, Lartigue C, Gibson DG, et al. 2010. Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.* 38(8):2558–69
76. Venetz JE, Medico LD, Wölfle A, Schächle P, Bucher Y, et al. 2019. Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *PNAS* 116(16):8070–79
77. Fredens J, Wang K, de la Torre D, Funke LFH, Robertson WE, et al. 2019. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* 569(7757):514–18
78. Shen Y, Wang Y, Chen T, Gao F, Gong J, et al. 2017. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* 355(6329):eaaf4791
79. Zhang W, Zhao G, Luo Z, Lin Y, Wang L, et al. 2017. Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science* 355(6329):eaaf3981
80. Karas BJ, Molparia B, Jablanovic J, Hermann WJ, Lin Y-C, et al. 2013. Assembly of eukaryotic algal chromosomes in yeast. *J. Biol. Eng.* 7:30
81. Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, et al. 2007. Genome transplantation in bacteria: changing one species to another. *Science* 317(5838):632–38
82. Labroussaa F, Lebaudy A, Baby V, Gourgues G, Matteau D, et al. 2016. Impact of donor-recipient phylogenetic distance on bacterial genome transplantation. *Nucleic Acids Res.* 44(17):8501–11
83. Baby V, Labroussaa F, Brodeur J, Matteau D, Gourgues G, et al. 2018. Cloning and transplantation of the *Mesoplasma florum* genome. *ACS Synth. Biol.* 7(1):209–17
84. Lartigue C, Vashee S, Algire MA, Chuang R-Y, Benders GA, et al. 2009. Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* 325(5948):1693–96
85. Karas BJ, Jablanovic J, Sun L, Ma L, Goldgof GM, et al. 2013. Direct transfer of whole genomes from bacteria to yeast. *Nat. Methods* 10(5):410–12

86. Brown DM, Chan YA, Desai PJ, Grzesik P, Oldfield LM, et al. 2017. Efficient size-independent chromosome delivery from yeast to cultured cell lines. *Nucleic Acids Res.* 45(7):e50
87. Kazuki Y, Oshimura M. 2011. Human artificial chromosomes for gene delivery and the development of animal models. *Mol. Ther.* 19(9):1591–601
88. Takiguchi M, Kazuki Y, Hiramatsu K, Abe S, Iida Y, et al. 2014. A novel and stable mouse artificial chromosome vector. *ACS Synth. Biol.* 3(12):903–14
89. Shizuya H, Kouros-Mehr H. 2001. The development and applications of the bacterial artificial chromosome cloning system. *Keio J. Med.* 50(1):26–30
90. Fournier RE, Ruddle FH. 1977. Microcell-mediated transfer of murine chromosomes into mouse, Chinese hamster, and human somatic cells. *PNAS* 74(1):319–23
91. Iida Y, Kim J-H, Kazuki Y, Hoshiya H, Takiguchi M, et al. 2010. Human artificial chromosome with a conditional centromere for gene delivery and gene expression. *DNA Res.* 17(5):293–301
92. Oshimura M, Uno N, Kazuki Y, Katoh M, Inoue T. 2015. A pathway from chromosome transfer to engineering resulting in human and mouse artificial chromosomes for a variety of applications to biomedical challenges. *Chromosome Res.* 23(1):111–33
93. Lau YH, Stirling F, Kuo J, Karrenbelt MAP, Chan YA, et al. 2017. Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA. *Nucleic Acids Res.* 45(11):6971–80
94. Wang K, Fredens J, Brunner SF, Kim SH, Chia T, Chin JW. 2016. Defining synonymous codon compression schemes by genome recoding. *Nature* 539(7627):59–64
95. Hutchison CA, Chuang R-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, et al. 2016. Design and synthesis of a minimal bacterial genome. *Science* 351(6280):aad6253
96. Rouet P, Smih F, Jasin M. 1994. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *PNAS* 91(13):6064–68
97. Smih F, Rouet P, Romanienko PJ, Jasin M. 1995. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res.* 23(24):5012–19
98. Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, et al. 2001. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.* 21(1):289–97
99. Bibikova M, Golic M, Golic KG, Carroll D. 2002. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161(3):1169–75
100. Carroll D. 2014. Genome engineering with targetable nucleases. *Annu. Rev. Biochem.* 83:409–39
101. Chari R, Church GM. 2017. Beyond editing to writing large genomes. *Nat. Rev. Genet.* 18(12):749–60
102. Shao Y, Lu N, Wu Z, Cai C, Wang S, et al. 2018. Creating a functional single-chromosome yeast. *Nature* 560(7718):331–35
103. Luo J, Sun X, Cormack BP, Boeke JD. 2018. Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* 560(7718):392–96
104. Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343(6166):80–84
105. Joung J, Engreitz JM, Konermann S, Abudayyeh OO, Verdine VK, et al. 2017. Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* 548(7667):343–46
106. Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, et al. 2016. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 354(6313):769–73
107. Yang L, Güell M, Niu D, George H, Lesha E, et al. 2015. Genome-wide inactivation of porcine endogenous retroviruses (PERVs). *Science* 350(6264):1101–4
108. Ellis HM, Yu D, DiTizio T, Court DL. 2001. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *PNAS* 98(12):6742–46
109. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, et al. 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460(7257):894–98
110. Wang HH, Kim H, Cong L, Jeong J, Bang D, Church GM. 2012. Genome-scale promoter engineering by co-selection MAGE. *Nat. Methods* 9(6):591–93
111. Carr PA, Wang HH, Sterling B, Isaacs FJ, Lajoie MJ, et al. 2012. Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res.* 40(17):e132

112. Gregg CJ, Lajoie MJ, Napolitano MG, Mosberg JA, Goodman DB, et al. 2014. Rational optimization of *tolC* as a powerful dual selectable marker for genome engineering. *Nucleic Acids Res.* 42(7):4779–90
113. Napolitano MG, Landon M, Gregg CJ, Lajoie MJ, Govindarajan L, et al. 2016. Emergent rules for codon choice elucidated by editing rare arginine codons in *Escherichia coli*. *PNAS* 113(38):E5588–97
114. DiCarlo JE, Conley AJ, Penttilä M, Jäntti J, Wang HH, Church GM. 2013. Yeast oligo-mediated genome engineering (YOGE). *ACS Synth. Biol.* 2(12):741–49
115. Barbieri EM, Muir P, Akhuetie-Oni BO, Yellman CM, Isaacs FJ. 2017. Precise editing at DNA replication forks enables multiplex genome engineering in eukaryotes. *Cell* 171(6):1453–67
116. Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, et al. 2011. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* 333(6040):348–53
117. Wannier TM, Kunjapur AM, Rice DP, McDonald MJ, Desai MM, Church GM. 2018. Adaptive evolution of genomically recoded *Escherichia coli*. *PNAS* 115(12):3090–95
118. Lajoie MJ, Rovner AJ, Goodman DB, Aerni H-R, Haimovich AD, et al. 2013. Genomically recoded organisms expand biological functions. *Science* 342(6156):357–60
119. Lajoie MJ, Söll D, Church GM. 2016. Overcoming challenges in engineering the genetic code. *J. Mol. Biol.* 428(5, Pt. B):1004–21
120. Mukai T, Lajoie MJ, Englert M, Söll D. 2017. Rewriting the genetic code. *Annu. Rev. Microbiol.* 71:557–77
121. Mitchell LA, Phillips NA, Lafont A, Martin JA, Cutting R, Boeke JD. 2015. qPCRTag analysis—a high throughput, real time PCR assay for Sc2.0 genotyping. *J. Vis. Exp.* 99:e52941
122. Scior A, Preissler S, Koch M, Deuerling E. 2011. Directed PCR-free engineering of highly repetitive DNA sequences. *BMC Biotechnol.* 11(1):87
123. Blacketer MJ, Gannon MK, Young IA, Shogren-Knaak MA. 2017. Solid-phase synthesis of highly repetitive chromatin assembly templates. *Anal. Biochem.* 531:12–15
124. Ran FA, Hsu PD, Lin C-Y, Gootenberg JS, Konermann S, et al. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154(6):1380–89
125. Trevino AE, Zhang F. 2014. Genome editing using Cas9 nickases. *Methods Enzymol.* 546:161–74
126. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533(7603):420–24
127. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, et al. 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551(7681):464–71
128. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, et al. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353(6305):aaf8729
129. Yang L, Briggs AW, Chew WL, Mali P, Guell M, et al. 2016. Engineering and optimising deaminase fusions for genome editing. *Nat. Commun.* 7:13330
130. Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, et al. 2016. Design, synthesis, and testing toward a 57-codon genome. *Science* 353(6301):819–22
131. Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S. 2008. Virus attenuation by genome-scale changes in codon pair bias. *Science* 320(5884):1784–87
132. Mueller S, Coleman JR, Papamichail D, Ward CB, Nimnual A, et al. 2010. Live attenuated influenza virus vaccines by computer-aided rational design. *Nat. Biotechnol.* 28(7):723–26
133. Dormitzer PR, Suphaphiphat P, Gibson DG, Wentworth DE, Stockwell TB, et al. 2013. Synthetic generation of influenza vaccine viruses for rapid response to pandemics. *Sci. Transl. Med.* 5(185):185ra68
134. Dymond JS, Richardson SM, Coombes CE, Babatz T, Muller H, et al. 2011. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* 477(7365):471–76
135. Shen Y, Stracquadanio G, Wang Y, Yang K, Mitchell LA, et al. 2016. SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Res.* 26(1):36–49
136. Hochrein L, Mitchell LA, Schulz K, Messerschmidt K, Mueller-Roeber B. 2018. L-SCRaMbLE as a tool for light-controlled Cre-mediated recombination in yeast. *Nat. Commun.* 9:1931
137. Jia B, Wu Y, Li B-Z, Mitchell LA, Liu H, et al. 2018. Precise control of SCRaMbLE in synthetic haploid and diploid yeast. *Nat. Commun.* 9:1933

138. Luo Z, Wang L, Wang Y, Zhang W, Guo Y, et al. 2018. Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES. *Nat. Commun.* 9:1930
139. Shen MJ, Wu Y, Yang K, Li Y, Xu H, et al. 2018. Heterozygous diploid and interspecies SCRaMbLEing. *Nat. Commun.* 9:1934
140. Blount BA, Gowers G-OF, Ho JCH, Ledesma-Amaro R, Jovicevic D, et al. 2018. Rapid host strain improvement by in vivo rearrangement of a synthetic yeast chromosome. *Nat. Commun.* 9:1932
141. Liu W, Luo Z, Wang Y, Pham NT, Tuck L, et al. 2018. Rapid pathway prototyping and engineering using in vitro and in vivo synthetic genome SCRaMbLE-in methods. *Nat. Commun.* 9:1936
142. Wu Y, Zhu R-Y, Mitchell LA, Ma L, Liu R, et al. 2018. In vitro DNA SCRaMbLE. *Nat. Commun.* 9:1935
143. Court Justice EU. 2018. *Organisms obtained by mutagenesis are GMOs and are, in principle, subject to the obligations laid down by the GME Directive*. Press Release 111/18, July 28. **https://curia.europa.eu/jcms/upload/docs/application/pdf/2018-07/cp180111en.pdf**
144. US Dep. Agric. 2018. *Secretary Perdue issues USDA statement on plant breeding innovation*. Bulletin, Mar. 28. **https://content.govdelivery.com/accounts/USDAAPHIS/bulletins/1e599ff**
145. Oye KA, Esvelt K, Appleton E, Catteruccia F, Church G, et al. 2014. Regulating gene drives. *Science* 345(6197):626–28
146. Scudellari M. 2019. Self-destructing mosquitoes and sterilized rodents: the promise of gene drives. *Nature* 571(7764):160–62
147. Natl. Acad. Sci. Eng. Med. 2017. *Human Genome Editing: Science, Ethics, and Governance*. Washington, DC: Natl. Acad. Press
148. Boeke JD, Church G, Hessel A, Kelley NJ, Arkin A, et al. 2016. The Genome Project-Write. *Science* 353(6295):126–27
149. Miga KH. 2019. Centromeric satellite DNAs: hidden sequence variation in the human population. *Genes* 10(5):352
150. Mitchell LA, McCulloch LH, Pinglay S, Berger H, Bosco N, et al. 2019. *De novo* assembly, delivery and expression of a 101 kb human gene in mouse cells. bioRxiv 423426. **https://doi.org/10.1101/423426**
151. Laurent JM, Fu X, German S, Maurano MT, Zhang K, Boeke JD. 2019. Big DNA as a tool to dissect an age-related macular degeneration-associated haplotype. *Precis. Clin. Med.* 2(1):1–7
152. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, et al. 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* 466(7307):720–26
153. Zarrella I, Herten K, Maes GE, Tai S, Yang M, et al. 2019. The survey and reference assisted assembly of the *Octopus vulgaris* genome. *Sci. Data* 6(1):13
154. Johnson DBF, Wang C, Xu J, Schultz MD, Schmitz RJ, et al. 2012. Release factor one is nonessential in *Escherichia coli*. *ACS Chem. Biol.* 7(8):1337–44
155. Rovner AJ, Haimovich AD, Katz SR, Li Z, Grome MW, et al. 2015. Recoded organisms engineered to depend on synthetic amino acids. *Nature* 518(7537):89–93
156. Mandell DJ, Lajoie MJ, Mee MT, Takeuchi R, Kuznetsov G, et al. 2015. Biocontainment of genetically modified organisms by synthetic protein design. *Nature* 518(7537):55–60
157. Naito T, Matsuura A, Ishikawa F. 1998. Circular chromosome formation in a fission yeast mutant defective in two ATM homologues. *Nat. Genet.* 20(2):203
158. Haber JE, Rogers DT, McCusker JH. 1980. Homothallic conversions of yeast mating-type genes occur by intrachromosomal recombination. *Cell* 22(1, Pt. 1):277–89
159. Ueda Y, Ikushima S, Sugiyama M, Matoba R, Kaneko Y, et al. 2012. Large-scale genome reorganization in *Saccharomyces cerevisiae* through combinatorial loss of mini-chromosomes. *J. Biosci. Bioeng.* 113(6):675–82
160. Selmecki AM, Maruvka YE, Richmond PA, Guillet M, Shoresh N, et al. 2015. Polyploidy can drive rapid adaptation in yeast. *Nature* 519(7543):349–52

Annual Review of Biochemistry

Volume 89, 2020

Contents

Errata

An online log of corrections to *Annual Review of Biochemistry* articles may be found at http://www.annualreviews.org/errata/biochem