



Building genomes to understand biology

Alessandro L. V. Coradini^{1,2}, Cara B. Hull ^{1,2} & Ian M. Ehrenreich ¹✉

Genetic manipulation is one of the central strategies that biologists use to investigate the molecular underpinnings of life and its diversity. Thus, advances in genetic manipulation usually lead to a deeper understanding of biological systems. During the last decade, the construction of chromosomes, known as synthetic genomics, has emerged as a novel approach to genetic manipulation. By facilitating complex modifications to chromosome content and structure, synthetic genomics opens new opportunities for studying biology through genetic manipulation. Here, we discuss different classes of genetic manipulation that are enabled by synthetic genomics, as well as biological problems they each can help solve.

Most biologists seek to understand life as it exists¹. For many, this entails characterizing how genomes and molecular systems give rise to cellular life and its diversity². The genetic manipulation of organisms has long played a critical role in this endeavor^{3,4}, and innovations in methods for genetic manipulation continually expand the biological research that is possible^{5–7}. In recent years, the synthesis of chromosomes, known as synthetic genomics, has emerged as a new form of genetic manipulation⁸. Megabase-sized chromosomes can now be generated from components synthesized de novo, obtained from naturally occurring genomes and other existing molecules, or a mixture of the two (Box 1, Fig. 1). During construction, the content and structure of these chromosomes can be modified relative to their natural templates to enable biological hypothesis testing^{9,10}, as well as to make organisms more amenable to research and bioengineering¹¹.

Because of their comparatively small size, viral chromosomes were the first to be synthesized. This began in the early 2000s¹² and a number of viral chromosomes have been constructed to date, including hepatitis C¹³, polio¹⁴, ΦX174¹⁵, a SARS-like coronavirus¹⁶, herpes type 1¹⁷, and the coronavirus that causes COVID-19¹⁸. However, by 2008, chromosome synthesis had progressed to cellular life forms, beginning with the human pathogenic bacterium *Mycoplasma genitalium*¹⁹. Subsequently, complete chromosomes were constructed for a number of other bacteria, including *Mycoplasma mycoides*^{9,19,20}, *Caulobacter crescentus*²¹, and *Escherichia coli*¹¹, as well as for the budding yeast *Saccharomyces cerevisiae*^{10,22–26}. In some cases, these chromosomes were re-engineered in remarkable ways, such as through the complete elimination of nonessential genes⁹ or particular codons¹¹, providing striking examples of the large-scale genetic manipulations that are enabled by synthetic genomics.

These foundational studies illustrate the tremendous advances that have occurred in synthetic genomics in roughly a single decade. They also suggest that moving forward, continued progress in synthetic genomics will facilitate increasingly complex genetic manipulations across a broader range of cellular organisms^{27–29}. Here, we assume such progress and discuss different classes of genome-wide genetic manipulation that either are presently possible or will likely become feasible in the future. We focus on six specific classes of genetic manipulation enabled by synthetic genomics: restructuring, recoding, minimization, chimerism, organelle reengineering, and

¹ Molecular and Computational Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089-2910, USA.

² These authors contributed equally: Alessandro L.V. Coradini, Cara B. Hull. ✉email: ian.ehrenreich@usc.edu

Box 1 | Methods for building and activating chromosomes

Assembly of large DNA molecules: Chromosomes are synthesized through the hierarchical assembly of smaller DNA molecules into progressively larger ones. This process begins with molecules that are hundreds of base pairs in length (“fragments”), which can be readily ordered from commercial vendors and are reviewed elsewhere^{134,135}. Fragments are typically combined together and cloned into bacterial plasmids by Gibson^{113,136,137} or Golden Gate^{138,139} assembly. These are highly specific and scarless in vitro techniques that rely on annealing or restriction digestion and ligation of partially overlapping fragments¹⁴⁰, respectively (Fig. 1a). Assemblies generated with these techniques are usually on a scale of 5–10 kb but can be as large as hundreds of kilobases^{113,136–139}. However, larger in vitro assemblies are susceptible to shearing during handling¹⁴¹ and may not be tolerated by *E. coli*, which is used to store and replicate these molecules¹⁹. Thus, often smaller in vitro assemblies are constructed and then combined together into larger molecules using yeast, which is highly recombinogenic¹⁹. In yeast, multiple smaller assemblies can be co-transformed and combined through homologous recombination into a larger assembly contained within a single centromeric plasmid^{19,44,45,142–144} (Fig. 1a). This approach has been used to produce molecules greater than a megabase in size^{9,19–21}. In the future, cell-free DNA replication systems may make it possible to bypass host organisms, such as *E. coli* and yeast, and to construct chromosome-sized molecules entirely in vitro^{145,146}.

Bringing synthetic chromosomes to life: Chromosomes constructed in yeast must be transplanted into living cells. This involves transforming these molecules into suitable recipient cells and selecting for their retention, resulting in the complete replacement of a cell’s original chromosome¹³². To date, such chromosome transplantation has only been demonstrated in *Mycoplasma*, a bacterial genus that possesses the smallest known genomes of all independently replicating organisms^{9,20,132,133}. Thus, for most species, synthetic chromosomes must instead be built through an iterative process in which existing genomic segments are successively replaced with synthetic assemblies using homologous recombination and selection. Several examples of this strategy have been reported, including switching auxotrophies progressively for integration²² in the yeast Sc2.0 project^{10,22–26} (Fig. 1b); stepwise integration of rolling circle amplified segments in *Salmonella typhimurium*⁷¹; conjugative assembly genome engineering (CAGE) in *E. coli*⁶⁹; and replicon excision for enhanced genome engineering through programmed recombination in *E. coli*^{11,70}. Over multiple cycles, these iterative, in vivo approaches can be used to produce entirely synthetic chromosomes that are biologically active.

genome resurrection. For each class of genome-wide genetic manipulation, we explore fundamental biological questions that they can be used to address. Our goal in this perspective is to inspire scientists to utilize synthetic genomics to advance understanding of biological systems.

Genetic manipulations enabled by synthetic genomics

Restructuring. Genome structure is thought to play an important role in evolution and phenotype^{30,31}. Supporting this possibility, genome structure varies significantly across species³², and changes within species are also prevalent, often impacting traits³³. While sequencing has led to the detection of many chromosome structural differences within and between species, rarely is it possible to know their biological significance through purely bioinformatic approaches³². Synthetic genomics enables the intentional restructuring of chromosomes, thereby facilitating research aimed at empirically determining the functional and phenotypic consequences of changes in genome structure. By restructuring chromosomes at different scales, scientists can directly probe the relationship between the layout of a genome and the features of an organism^{32,34–40}.

At the finest scale, the structures of genes and other functional elements can be altered. For example, many genes overlap another gene on the opposing DNA strand. Because of these overlaps, a genetic change in a gene on one strand might also affect the sequence of a gene on the other strand. Work to remove these overlaps from a large portion of the phage T7 genome showed that the elimination of such overlaps often has little to no phenotypic consequence⁴¹. This form of restructuring, which has been called refactoring⁴¹ (Fig. 2a), makes it possible to not only study the biological impacts of local genome structure but also to produce organisms that are more amenable to research. For example, refactoring may be necessary if one wants to globally replace certain codons with alternative codons¹¹. However, it also bears mention that generating new overlaps between genes may be desirable in some cases, such as to limit the evolutionary potential of synthetic organisms⁴².

At an intermediate scale, restructuring can be used to rearrange genetic material within a genome. For example, in one-eighth of a synthetic *M. mycoides* genome, genes were reordered into modules based on common function⁹. This restructured genome

produced a viable organism, raising questions about whether gene arrangement along a chromosome is functionally important and suggesting that someday it may be possible to generate organisms with completely modular genomes (Fig. 2a). However, similar work in other organisms, including multicellular species, is needed to more generally assess the functional and phenotypic consequences of gene arrangement. In these organisms, gene arrangement might have greater biological significance, for example by affecting the three-dimensional conformation of the genome in the nucleus.

Lastly, at the macro scale, restructuring can be used to modify how entire genomes are packaged into chromosomes (Fig. 2b). Strategies for chromosome restructuring inside living cells have been reported using CRISPR/Cas9 or restriction enzymes^{32,34–39,43}. Most applications of these approaches sought to introduce one or a small number of targeted or random changes in genome structure. However, in budding yeast, these approaches were used to more extensively modify the organism’s genome structure by concatenating the 16 nuclear chromosomes into as few as one or two molecules^{32,38}. Subsequently, it was shown that yeast, which naturally possesses linear chromosomes, will even tolerate a genome restructured into a single, circular chromosome³⁹. These changes in genome structure had little impact on gene expression, but affected growth across environments, reproductive ability, and other cellular features. This suggests that the phenotypic impacts of restructuring may be mediated through non-transcriptional mechanisms.

Synthetic genomics enables even more extensive restructuring than what was described in the preceding paragraphs. For example, a genome could be simultaneously restructured at multiple of the aforementioned scales. This is a key part of the Sc2.0 project, in which loxP sites have been added between each pair of genes on a given synthetic chromosome to allow for random Cre-mediated site-specific recombination^{10,22–26}. Because of these loxP sites, synthetic chromosomes constructed as a part of Sc2.0 facilitate the generation of large libraries of cells with diverse chromosome or genome structures through the induction of synthetic chromosome recombination and modification by LoxP-mediated evolution^{44–51}. Analysis of cells produced by SCRaMble can reveal how changes in genome structure influence particular phenotypes or can be used to screen for

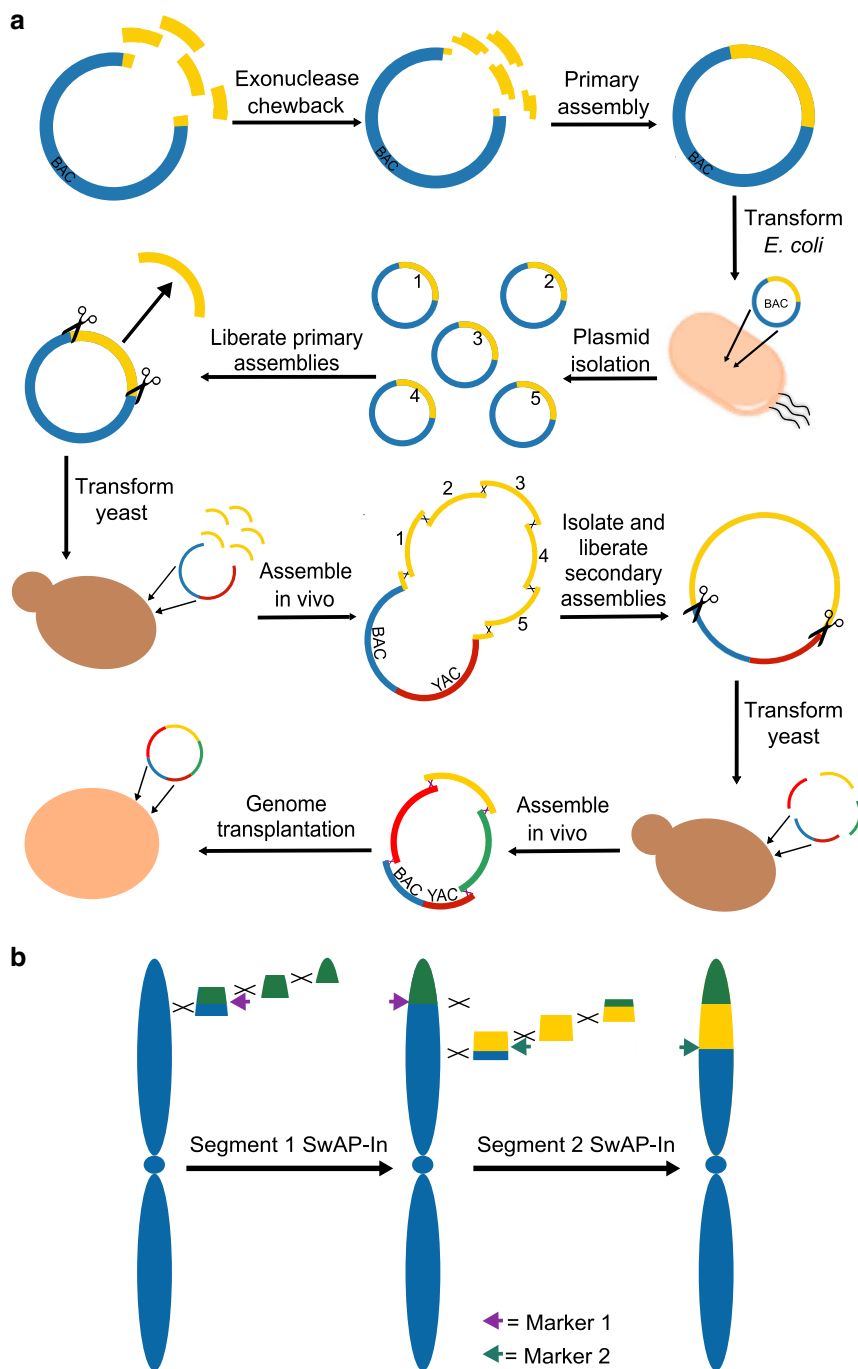


Fig. 1 Methods for building chromosomes. **a** Synthetic chromosomes are constructed using a hierarchical assembly process¹⁹. This involves building primary assemblies from DNA fragments in vitro and then transforming these assemblies into *E. coli*. These primary assemblies can then be liberated from their plasmid backbones and co-transformed into yeast. Homologous recombination in yeast can then combine these primary assemblies to build a larger secondary assembly. If necessary, such secondary assemblies can be combined together in yeast to produce even larger molecules. In *Mycoplasma* (and potentially other organisms in the future), synthetic chromosomes can be activated through transplantation^{9,20,132,133}. **b** In organisms that do not permit chromosome transplantation, an alternative strategy is to build chromosomes in vivo by iteratively replacing segments of the host cell's genome. This figure is based on the SwAP-In approach described in yeast^{10,22}.

genome structures that confer traits of interest. While SCRaMBLE is powerful and highly scalable, it also has some constraints, such as a need for de novo sequencing to determine the exact structures of output genomes and a likelihood that output genomes will vary in their gene content.

Moving forward, it will be desirable to generate similar genome restructuring libraries in other species. Such work may help

clarify the role of genome restructuring in evolutionary change and other biological phenomena, such as the emergence of cancer. On the latter topic, genome restructuring events are observed in many cancers⁵². In some cases, it is even thought that massive, one-step genome restructuring events, known as chromothripsis, can drive oncogenic transformations⁵³. Yet, our understanding of the relationship between cancer and genome

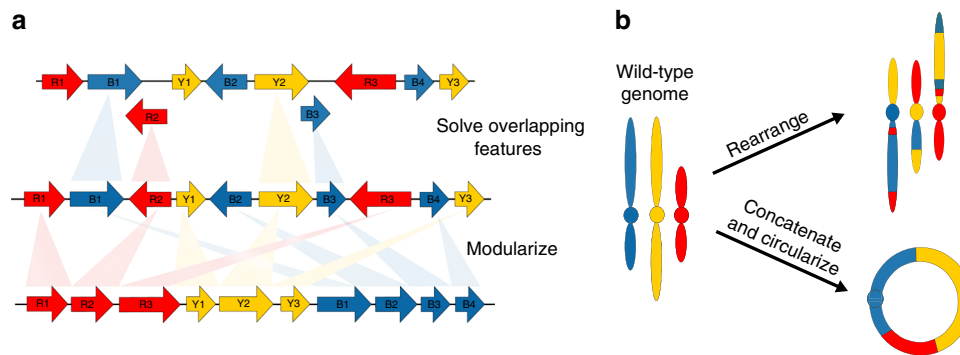


Fig. 2 Restructuring. **a** Synthetic genomics makes it possible to engineer diverse chromosome structures. At a micro-scale, the arrangement of genetic material within genomic regions can be altered. For example, structural overlaps between adjacent genes can be eliminated (i.e., refactoring) or genes with related functions can be positioned next to each other (i.e., modularization). Such modifications can be implemented chromosome-wide. **b** On a macro scale, this includes the introduction of gross chromosomal rearrangements, such as deletions, insertion, inversions, and translocations, as well as the modification of chromosome number and conformation.

restructuring is limited to what has been observed through genome sequencing. Synthetic genomic approaches may enable the controlled experimental characterization of the role of genome restructuring in cancer. Genome restructuring experiments also provide numerous other opportunities for biological discovery. For example, they may provide insights into how genome structure relates to three-dimensional genome conformation, the regulation of transcription, and the expression of phenotype in multicellular eukaryotes, similar to recent efforts to explore these questions in yeast⁴⁰. Research along these lines could improve our understanding of development, thereby enhancing efforts to reprogram and differentiate cells in a controlled manner⁵⁴.

Recoding. Codon usage is another feature of genomes that varies among species^{55–57}. Differences in codon usage between genes in the same genome or between different genomes can affect transcription, translation, and other molecular processes within an organism^{55,58,59}, providing a potential substrate for natural selection^{60–62}. The importance of codon usage can be explored experimentally using recoding—the genome-wide replacement of particular codons with their synonyms (Fig. 3a, b). Recoding can be used to address questions about the molecular and evolutionary impacts of changes in codon usage^{63,64}, and to produce organisms poised for biotechnology applications, such as the generation of organisms that utilize non-natural amino acids^{65–67} (Fig. 3c).

Both genome editing and synthetic genomics approaches enable recoding on a genome-wide scale^{11,64,65,68–71}. Genome-wide manipulation of codon usage was first performed in *E. coli* with the objective of generating a strain in which all 314 TAG stop codons were converted to TAA⁷². Doing this required combining a genome editing strategy—multiplexed automated genome editing (MAGE)⁶⁸—with a conjugation-based assembly method—CAGE⁶⁹. MAGE produces site-specific nucleotide changes by utilizing short oligonucleotides and the λ phage β protein⁶⁸ to catalyze homologous recombination between single-stranded DNA and a chromosome⁷³. Although highly effective, MAGE is only capable of introducing a limited number of edits at a time. Thus, different regions of the genome were edited in parallel in distinct strains and then combined together using CAGE to produce strains containing as many as 80 synonymous codon changes⁷². This work highlights limitations to genome recoding through genome editing, including the number of codons that can be simultaneously modified, challenges in completely eliminating particular codons, and the need for

sequence complementarity between editing reagents and a genome.

In contrast to genome editing-based recoding strategies, synthetic genomics is constrained only by what an organism will tolerate^{11,70,71,74}. The most extensive, synthetic genomic-based recoding efforts thus far also occurred in *E. coli*. In one study, a 57 codon genome lacking the TAG stop codon, two arginine codons, two leucine codons, and two serine codons was designed and constructed as a series of ~50 kb assemblies⁷⁴. These segments were assayed for viability using complementation tests but have yet to be assembled into a recoded genome. In a second study, a strategy for genome-wide recoding using direct replacement of *E. coli* genomic segments with DNA assemblies was developed⁷⁰. This approach was recently applied genome-wide to recode certain serine codons from TCG to AGC and TCA to AGT, as well as to replace TAG stop codons with TAA¹¹. The *E. coli* Syn61 strain produced by this work contains 18,214 codon modifications in total but shows little phenotypic change relative to its progenitor, illustrating the malleability of the genetic landscape. This work also demonstrated the potential to obtain new insight into functional differences between synonymous codons through the identification of idiosyncratic codons that do not tolerate recoding^{70,74}.

The above projects show that genome-wide recoding is now feasible, at least in certain organisms. Based on these advances, a number of fundamental questions can now be explored using recoding. For example, to what extent can the genetic code be reduced? At the extreme, one could imagine trying to produce an organism that utilizes only 21 codons, one for each of the 20 amino acids and one for the stop. The generation of organisms with simplified genetic codes may make it possible to explore the consequences of tRNA utilization for fitness and evolvability across environments. Recoding may also make it possible to generate particular nonsynonymous changes globally. Such work could clarify the exchangeability (or lack thereof) of different amino acids and can reveal mechanisms by which the translation apparatus recovers from problems in translation⁷⁵. Recoding may even shed light on ecological interactions between organisms or between living species and viruses²⁹. Indeed, evidence suggests that such interactions can shape patterns of codon usage, likely due to shared environmental resources and selection upon translation⁷⁶.

Minimization. A fundamental question in biology regards the minimal set of genes and physiological functions needed to support cellular life^{77,78}. Synthetic genomics has made it possible

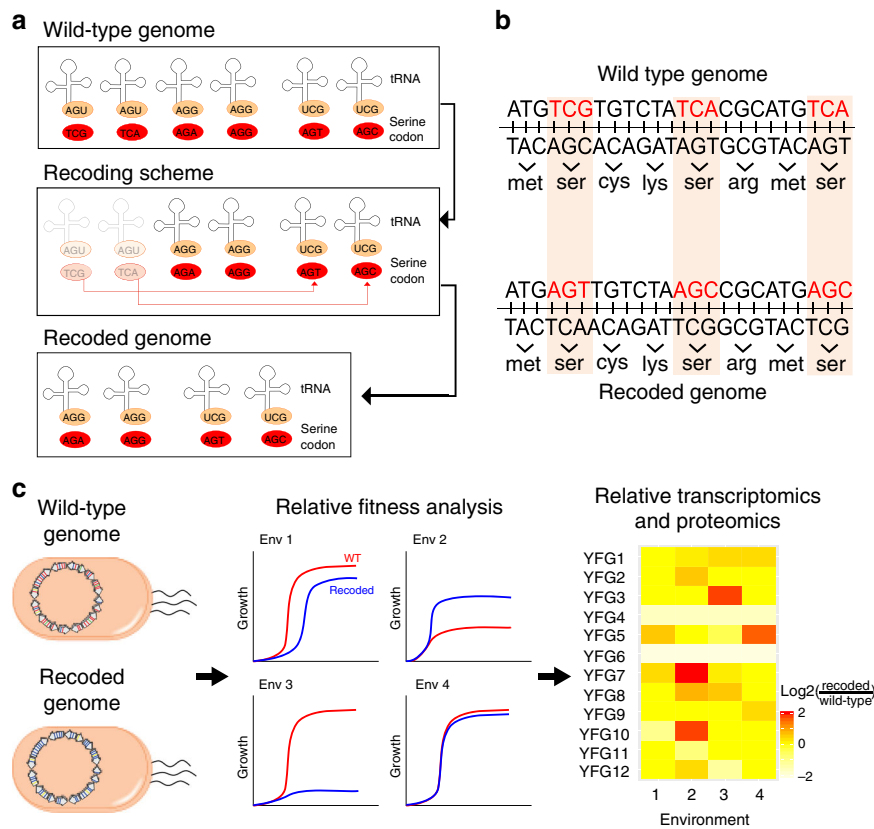


Fig. 3 Recoding. **a, b** Certain codons can be replaced with alternative codons on a genome-wide scale^{11,66,70}. Typically, such recoding involves substituting a codon with a synonymous codon, followed by the deletion of the corresponding tRNA for the original codon. Here, the number of synonymous codons for the serine amino acid is reduced from six to four by the genome-wide substitution of TCG and TCA codons by the synonyms AGT and AGC, respectively. This makes it possible for the codon and tRNA to be repurposed, for example, to incorporate nonnatural amino acids into proteins. However, in the future, recoding efforts may also involve non-synonymous codon substitutions. **c** Genome-wide recoding can be used to explore how codon usage affects the transcriptome, proteome, and fitness.

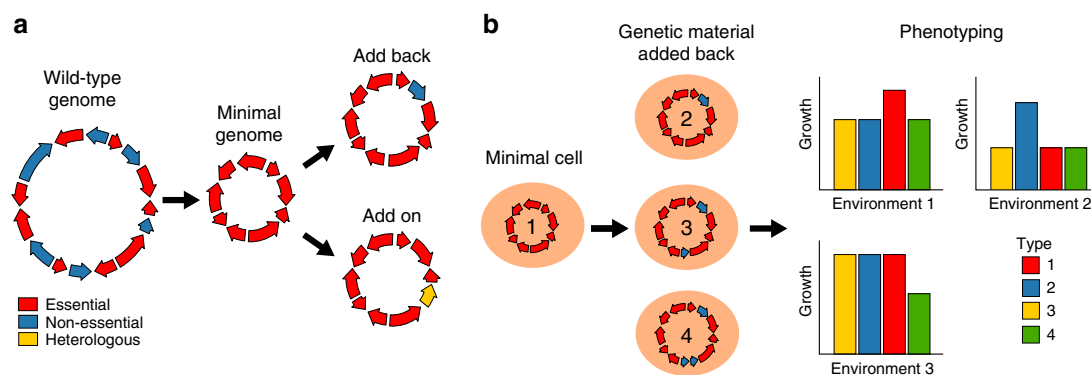


Fig. 4 Minimization. **a** Minimal genomes lack non-essential genes⁹. After a minimal genome has been generated, non-essential genes can be added back to explore how they impact function and phenotypic outcomes. In addition, heterologous genes (add on) can be inserted to create new cell phenotypes. **b** Genetic material added back to minimal genomes may help elucidate how certain genes contribute to phenotype across environments.

to experimentally answer this question through genome minimization, the elimination of all non-essential genes from an organism’s genome⁹ (Fig. 4a). The output of genome minimization is an organism that only carries a core set of genes needed to sustain life and reproduction in a given environment. In the first and only genome minimization to date, a minimal *M. mycoides* genome was produced. *Mycoplasma* species have small genomes, with the smallest known genome of a cellular life form being that of *M. genitalium* (580 kb containing around 525 genes). Naturally

occurring *M. mycoides* has a roughly one Mb genome containing nearly 1,000 genes. However, using multiple rounds of global transposon mutagenesis and a genome design-build-test cycle to eliminate non-essential genes, researchers produced a 531 kb synthetic *M. mycoides* genome with only 473 genes⁹. This minimal organism required genes involved in the expression of genomic information (i.e., transcription and translation), the cell membrane, cytoplasmic metabolism, and preservation of genome information, though 149 genes had unclear functions.

Mycoplasmas are extracellular parasites or commensals that have streamlined genomes adapted to life within a host⁷⁹ and may have a single minimal genome state. In contrast, most unicellular organisms used in biological research, including *E. coli* and *S. cerevisiae*, have larger numbers of genes, more complicated genomes, and a greater tolerance for diverse environments^{80,81}. For organisms like these, whether one or multiple potential minimal genomes exists, as well as how these minimal gene sets might depend on the environment, is not entirely clear. Because these organisms possess diverse metabolic modes and an abundance of functional redundancies at the genetic level⁸², there may be a potential for multiple minimal genomic states in a given condition. This question should be directly addressable in the future through synthetic genomic-based genome minimization efforts.

Minimization also has the potential to provide insights into multicellular organisms by improving understanding of how a single genome encodes diverse cell types. Minimization could be used to identify genes that are essential throughout all cell types, as well as genes that are essential for only particular cell types. By minimizing the genomes of organoids instead of cells, the essential gene set for cell–cell interactions that produce a given tissue might even be identified. However, a challenge with multicellular species, and potentially some unicellular species as well, will be alternative splicing, which can significantly impact the enzymatic activities, protein–protein interactions, and other functional attributes of proteins⁸³. One could imagine performing minimization in a way that eliminates certain isoforms through the removal of only some exons or introns of a gene; however, doing this in a scalable manner will likely require technical innovation.

Across species, minimization also has the potential to help characterize the essential non-coding portion of genomes. In multicellular organisms, most transcripts are noncoding and many of these molecules have unknown biological significance^{84,85}. Minimization of the noncoding portion of the genome could elucidate which of these transcripts are essential for life, which can serve as a starting point for determining their biological functions and mechanisms of actions. Similar logic can be employed for mobile genetic elements, which comprise most of the human genome and are abundant throughout the genomes of many other organisms as well^{86,87}. Production of genomes in which such mobile elements have been removed could also have applied benefits, such as the production of crops that grow faster due to their reduced genome sizes or cell lines with enhanced genomic stability.

Minimization can also provide a foundation for studying the non-essential portions of genomes. Minimal genomes can serve as chassis to which nonessential genetic material can be added back or added on. We use these terms to refer to the analysis of nonessential genes or other DNA elements from the same or different organisms, respectively, through their addiction to a minimal genome (Fig. 4a, b). Such work may address how particular nonessential genes and pathways influence features of an organism, including fitness in a focal environment, tolerance to different environments, mutational robustness, and evolvability^{88–90}. In multicellular organisms, add back of cell type-specific essential genes be useful for studying the roles of nonessential genes in enabling a core genome to produce a diversity of cell types. Add on genetics could also possibly be used to probe questions about how the addition of genetic material facilitates the evolution of new traits and species.

Chimerism. Chimerism is the combining of genetic material from multiple strains or species into a single organism (Fig. 5). This is standard practice in biology and bioengineering; for

example, heterologous expression of pathways is often used to produce drugs and other valuable compounds^{91–94}. However, synthetic genomics makes it possible to consider new forms of chimerism either through the programmed combination of specific genomic segments from different organisms, as has been shown in *E. coli*⁹⁵, or random approaches, as has been shown with the application of SCRaMBLE to yeast interspecies hybrids⁴⁸. However, present limitations in the generation of chimeras are the inability to easily hybridize more than a small number of loci⁹⁵ and to conduct SCRaMBLE between chromosomes from different species⁴⁸.

In the future, building and phenotyping libraries of genomes that are mosaics between two or more strains or species may enable the identification of functional genetic differences in situations that do not permit standard genetic mapping approaches, such as linkage and association mapping^{96,97} (Fig. 5). This might allow for mapping of loci that contribute to genetic incompatibilities or traits of interest between individuals or species that are not amenable to controlled crossing because hybrids are inviable or produce defective offspring. The generation of chimeras might also allow geneticists to map traits across larger evolutionary distances than can typically be experimentally explored by crosses, such as at the genus level or higher. Using chimerism for genetic mapping will require approaches for making recombinant chromosome libraries with the same genome structure but a high degree of genetic shuffling, as would typically be obtained from meiosis.

The production of chimeric genomes may also be of value in understanding the functional and phenotypic significance of species' pan-genomes. Work in prokaryotes^{98–100}, as well as both unicellular and multicellular eukaryotes^{101,102}, has shown that individuals within species commonly harbor extensive differences in gene content. This gene content variation clearly plays a major role in many heritable phenotypes, but often it can be difficult to connect specific gene content differences to particular traits. For example, any two *E. coli* isolates may only share the minority of their genes^{103–105}. Producing libraries of chimeric *E. coli* genomes (Fig. 5) could make it possible to examine how this substantial variation in gene content influences ecologically and clinically relevant traits, including antibiotic resistance, tolerance for different environmental conditions, and pathogenicity. It may also enable the identification of genetic interactions involving genes in the pan-genome, thereby shedding light on their functions⁸².

A probable limitation of chimerism will be transcriptional dysregulation resulting from the combining of chromosome segments from genetically diverged organisms. To overcome such issues, it may help to generate chimeras in which the protein-coding portions of the genomes vary while the noncoding portions remain the same. This could enable examination of how the proteome alone differs in function between contributing genomes. Of course, understanding how transcriptional regulation evolves within and between species is an important question as well. To address this, one could envision keeping the protein-coding portion of the genome the same but varying the noncoding portion. These ideas speak to how chimerism can be employed in sophisticated ways that go beyond crudely combining together DNA segments from different sources.

Organelle reengineering. Nearly all eukaryotes possess mitochondria, with photosynthetic eukaryotes also harboring chloroplasts. These organelles arose through the internalization and co-option of bacteria^{106,107}. Since their origins, organelles have lost the majority of their genes through mutational degeneration

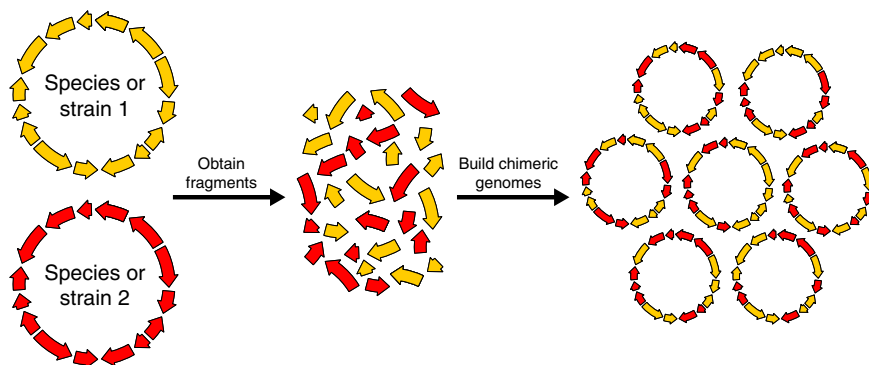


Fig. 5 Chimerism. Synthetic genomic approaches can enable the generation of chimeric chromosomes that contain mosaics of genetic material from two or more strains or species. By generating and screening libraries of chimeric chromosomes, it may be possible to determine the genetic basis of traits that differentiate reproductively isolated organisms.

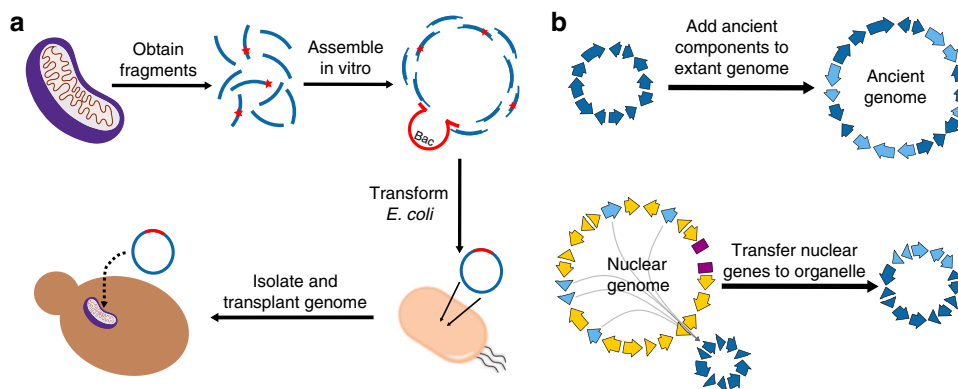


Fig. 6 Organellar reengineering. **a** Organellar genomes can be constructed by harvesting and modifying natural genetic material or by performing de novo synthesis¹¹³. **b** Synthesis of organellar genomes can be used to study a number of problems, including, but not limited to, the early evolutionary stages following endosymbiosis and cytonuclear genetic interactions. Note: the purple rectangles represent telomeres.

or translocation to the nuclear genome¹⁰⁸. This means that only a small fraction of the essential genes for a given organelle are encoded within the organelle itself. Additionally, ongoing molecular evolution has resulted in substantial variability in organellar genome size and content across species. For example, mitochondrial genomes range from 6 kb in the malaria parasite *Plasmodium falciparum* to 11 Mb in the plant *Silene conica*¹⁰⁹. Although the origins and ongoing evolution of these organelles can be studied using sequencing and bioinformatics, such approaches are limited in their abilities to explore the early stages of organelle evolution and do not permit mechanistic hypothesis testing by experimentation^{109,110}. In addition, techniques for genetically manipulating organellar genomes are less developed than methods for modifying the nuclear genome, though methods for directing specific metabolic enzymes¹¹¹ or base editors to the mitochondria have been described¹¹².

To help address gaps in our understanding of evolution and genetics, organellar genomes can be synthesized¹¹³ (Fig. 6a). Relative to nuclear chromosomes, organellar genomes are easier to synthesize because they are typically smaller and contain fewer genes. Once constructed, these synthetic organellar genomes can then be transformed into organelles within cells using biolistic approaches¹¹⁴, though higher throughput transplantation methods are likely to emerge in the future²⁸. By constructing likely evolutionary intermediates and activating them in organisms, it may be possible to reproduce key steps in organellar evolution (Fig. 6b). Furthermore, transferring genes back from the nuclear genome to the mitochondrial genome may further clarify the

benefits of encoding these genes in the nucleus^{115,116} (Fig. 6b). Because genetic variants in the organellar genomes frequently contribute to incompatibilities between strains and species, such work can also provide insight into the role of organellar evolution on reproductive isolation¹¹⁵. Beyond these evolutionary questions, synthetic genomics might simply aid in mapping organellar mutations and genetic variants with phenotypic effects. For example, to date, more than 250 point mutations in the human mitochondrial genome have been implicated with roles in disease¹¹⁷. The human mitochondrial genome is only ~17 kb, suggesting that it should be possible to build and transplant comprehensive variant libraries. Such work could further clarify the contribution of the mitochondria to disease and other traits.

Genome resurrection. Further in the future, another biological application of synthetic genomics may be the synthesis and activation of extinct organisms' genomes. We refer to this as a genome resurrection due to its similarity to the resurrection and study of ancient biomolecules^{118–120}. Extinct genomes sequences may be determined by sequencing ancient DNA, bioinformatic inference from extant genomes, or a combination of the two¹²¹. Synthesis of all or part of extinct genomes can provide substrates for the broader analysis of the molecular evolution of gene and pathway functions^{118–120}. This is true regardless of whether these genomes are activated to produce living organisms. In cases in which it is appropriate, activation of these genomes can be used in de-extinction^{122–126}. Presently, using genome editing to make extant species similar to their extinct relatives is being discussed^{122,123}. Yet,

the number of edits needed to convert one species into another could prove prohibitive, requiring the use of synthetic genomics in the future. There may be benefits to de-extinction in some cases, though careful ethical consideration will be imperative^{124–126}.

Genome resurrection could also potentially enable research aimed at understanding key steps in molecular and phenotypic evolution. For example, it could help elucidate the origins and diversification of eukaryotes. The favored theory for the origin of eukaryotes is that the ancestral eukaryotic cell was produced through the fusion of an archaeon and a bacterium¹²⁷. The recent isolation of a species of Asgard archaea, which are the closest extant relatives of the archaeon that gave rise to eukaryotes, specifically suggests eukaryotes evolved due to their archaeal ancestor engulfing and endogenizing a syntrophic bacterium¹²⁸. Utilizing this information, it might be possible to construct and activate a genome resembling the earliest eukaryote. This would provide a model system for studying the initial evolution of eukaryotes, as well as for examining how additional defining eukaryotic features subsequently evolved. For example, what were the origins of the nucleus, cytoskeleton, and specialized organelles¹²⁹? The addition of genetic material to the ancestral eukaryote model could be used to study the mechanisms underlying these intermediate steps in eukaryotic evolution. Similar logic could be applied to later eukaryotic innovations, such as multicellularity and potentially even the nervous system.

Conclusion

During the last decade, synthetic genomics has gone from science fiction to reality, giving rise to fundamentally new ways to genetically manipulate biological systems. Relative to more conventional approaches for genetic manipulation, synthetic genomics is distinguished by both the large numbers and diversity of genetic changes that can be introduced. Indeed, as discussed throughout this paper, for certain single-celled organisms, it is now possible to entirely modify genome structure and content. In the future, technological advances should enable similar manipulations in a much broader array of species, including multicellular organisms. As we have attempted to emphasize, the ability to perturb biological systems in such a manner has the potential to provide deep insights into a wide range of fundamental topics in biology that remain only partially understood. Arguably, the overarching theme connecting all of the biological disciplines, classes of genome-wide genetic manipulation, and specific questions that we have discussed are that synthetic genomics will likely play a key role in achieving a deep mechanistic understanding of how biological systems work and came to be.

Of course, as synthetic genomics progresses, ethical considerations will be a major concern. Already with efforts to date, questions have been raised about the acceptable bounds of genetic modification in existing life forms^{130,131}. The degree of manipulation that is now possible with synthetic genomics challenges our notions of species, as many of these genetic manipulations that we have described will produce organisms that significantly differ at the sequence level and may be reproductively isolated from their progenitors. Is it acceptable for people to generate such organisms? We believe that such manipulations are necessary to advance biology as a field, though valid questions exist about such manipulations. For example, would such manipulations be allowed in vertebrates or human pathogens? A more practical ethical concern is the escape of organisms whose genomes have been heavily manipulated into the environment. However, as others have argued, ethical, governance, and containment policies, such as proposed for Sc2.0¹³¹, as well as modification of

synthetic genomes to make them unfit outside the lab, can help alleviate many ethical concerns.

In conclusion, synthetic genomics represents the genesis of a new era of biology in which scientists will increasingly transition from simply reading genomes to writing them^{28,29}. Many fundamental questions about biological systems that historically could only be addressed through inferential, bioinformatic analyses may now or will soon be amenable to experimentation using synthetic genomics strategies. It is hard to imagine that modifying biological systems through synthetic genomic approaches will not produce new, previously inaccessible insights. This speaks to how building genomes can be used to determine the design rules that underpin life. We expect that work along these lines will not only enhance our understanding of biology, but it will help maximize the potential for genetically programming cellular organisms for human benefit in the future.

Received: 27 July 2020; Accepted: 29 October 2020;

Published online: 02 December 2020

References

1. Elowitz, M. & Lim, W. A. Build life to understand it. *Nature* **468**, 889–890 (2010).
2. Kirschner, M. W. The meaning of systems biology. *Cell* **121**, 503–504 (2005).
3. Flavell, R. A., Sabo, D. L. O., Bandle, E. F. & Weissmann, C. Site-directed mutagenesis: effect of an extracistronic mutation on the in vitro propagation of bacteriophage Qbeta RNA. *Proc. Natl Acad. Sci. USA* **72**, 367–371 (1975).
4. Hutchison, C. A. et al. Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253**, 6551–6560 (1978).
5. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
6. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* (2013).
7. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* (2013).
8. Mitchell, L. A. & Ellis, T. Synthetic genome engineering gets infectious. *Proc. Natl Acad. Sci. USA* **114**, 11006–11008 (2017).
9. Hutchison, C. A. et al. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253–aad6253 (2016). The authors construct a minimal bacterial genome that supports cellular life.
10. Dymond, J. S. et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011). This paper lays the technical and conceptual groundwork for the Sc2.0 project, including the SCRaMble technique.
11. Fredens, J. et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019). The authors produce an *E. coli* strain with a completely synthetic, recoded genome.
12. Wimmer, E., Mueller, S., Tumpey, T. M. & Taubenberger, J. K. Synthetic viruses: a new opportunity to understand and prevent viral disease. *Nat. Biotechnol.* **27**, 1163–1172 (2009).
13. Blight, K. J. Efficient initiation of HCV RNA replication in cell culture. *Science* **290**, 1972–1974 (2000).
14. Cello, J., Paul, A. V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297**, 1016–1018 (2002).
15. Smith, H. O., Hutchison, C. A., Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: X174 bacteriophage from synthetic oligonucleotides. *Proc. Natl Acad. Sci. USA* **100**, 15440–15445 (2003).
16. Becker, M. M. et al. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc. Natl Acad. Sci. USA* **105**, 19944–19949 (2008).
17. Oldfield, L. M. et al. Genome-wide engineering of an infectious clone of herpes simplex virus type 1 using synthetic genomics assembly methods. *Proc. Natl Acad. Sci. USA* **114**, E8885–E8894 (2017).
18. Thi Nhu Thao, T. et al. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561–565 (2020).
19. Gibson, D. G. et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008). The authors demonstrate that it is possible to build synthetic DNA molecules on the scale of entire chromosomes.

20. Gibson, D. G. G. et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010). This paper reports the first transplantation of a completely synthetic genome into a living cell.
21. Venetz, J. E. et al. Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *Proc. Natl Acad. Sci. USA* **116**, 8070–8079 (2019).
22. Richardson, S. M. et al. Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).
23. Xie, Z. X. et al. ‘Perfect’ designer chromosome v and behavior of a ring derivative. *Science* **355**, eaaf4704 (2017).
24. Wu, Y. et al. Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* **355**, eaaf4706 (2017).
25. Shen, Y. et al. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* **355**, eaaf4791 (2017).
26. Zhang, W. et al. Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science* **355**, eaaf3981 (2017).
27. Annaluru, N., Ramalingam, S. & Chandrasegaran, S. Rewriting the blueprint of life by synthetic genomics and genome engineering. *Genome Biol.* **16**, 125 (2015).
28. Ostrov, N. et al. Technological challenges and milestones for writing genomes. *Science* **366**, 310–312 (2019).
29. Boeke, J. D. et al. The Genome Project-Write. *Science* **353**, 126–127 (2016).
30. Peichel, C. L. Chromosome evolution: molecular mechanisms and evolutionary consequences. *J. Hered.* **108**, 1–2 (2017).
31. Hufton, A. L. & Panopoulou, G. Polyploidy and genome restructuring: a variety of outcomes. *Curr. Opin. Genet. Dev.* **19**, 600–606 (2009).
32. Shao, Y. et al. Creating a functional single-chromosome yeast. *Nature* **560**, 331–335 (2018).
33. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85–97 (2006).
34. Muramoto, N. et al. Phenotypic diversification by enhanced genome restructuring after induction of multiple DNA double-strand breaks. *Nat. Commun.* **9**, 1995 (2018).
35. Fleiss, A. et al. Reshuffling yeast chromosomes with CRISPR/Cas9. *PLOS Genet.* **15**, e1008332 (2019).
36. Schmidt, C., Schindele, P. & Puchta, H. From gene editing to genome engineering: restructuring plant chromosomes via CRISPR/Cas. *aBIOTECH* **1**, 21–31 (2020).
37. Beying, N., Schmidt, C., Pacher, M., Houben, A. & Puchta, H. CRISPR–Cas9-mediated induction of heritable chromosomal translocations in Arabidopsis. *Nat. Plants* **6**, 638–645 (2020).
38. Luo, J., Sun, X., Cormack, B. P. & Boeke, J. D. Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* **560**, 392–396 (2018).
39. Shao, Y. et al. A single circular chromosome yeast. *Cell Res.* **29**, 87–89 (2019).
40. Mercy, G. et al. 3D organization of synthetic and scrambled chromosomes. *Science* **355**, eaaf4597 (2017).
41. Chan, L. Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol. Syst. Biol.* **1**, 2005.0018 (2005).
42. Blazejewski, T., Ho, H.-I. & Wang, H. H. Synthetic sequence entanglement augments stability and containment of genetic information in cells. *Science* **365**, 595–598 (2019).
43. Brunet, E. & Jasin, M. Induction of chromosomal translocations with CRISPR-Cas9 and other nucleases: understanding the repair mechanisms that give rise to translocations. *Adv. Exp. Med. Biol.* **1044**, 15–25 (2018).
44. Gowers, G.-O. F. et al. Improved betulinic acid biosynthesis using synthetic yeast chromosome recombination and semi-automated rapid LC-MS screening. *Nat. Commun.* **11**, 868 (2020).
45. Wang, J. et al. Ring synthetic chromosome V SCRaMbLE. *Nat. Commun.* (2018).
46. Jia, B. et al. Precise control of SCRaMbLE in synthetic haploid and diploid yeast. *Nat. Commun.* **9**, 1933 (2018).
47. Blount, B. A. et al. Rapid host strain improvement by in vivo rearrangement of a synthetic yeast chromosome. *Nat. Commun.* **9**, 1932 (2018).
48. Shen, M. J. et al. Heterozygous diploid and interspecies SCRaMbLEing. *Nat. Commun.* **9**, 1934 (2018).
49. Liu, W. et al. Rapid pathway prototyping and engineering using in vitro and in vivo synthetic genome SCRaMbLE-in methods. *Nat. Commun.* **9**, 1936 (2018).
50. Luo, Z. et al. Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES. *Nat. Commun.* **9**, 1930 (2018).
51. Hochrein, L., Mitchell, L. A., Schulz, K., Messerschmidt, K. & Mueller-Roeber, B. L-SCRaMbLE as a tool for light-controlled Cre-mediated recombination in yeast. *Nat. Commun.* **9**, 1931 (2018).
52. Willis, N. A., Rass, E. & Scully, R. Deciphering the code of the cancer genome: mechanisms of chromosome rearrangement genomic instability and the evolution of a cancer. *Trends Cancer* **1**, 217–230 (2015).
53. Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
54. Takahashi, K. & Yamanaka, S. A developmental framework for induced pluripotency. *Development* **142**, 3274–3285 (2015).
55. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42 (2011).
56. LaBella, A. L., Opulente, D. A., Steenwyk, J. L., Hittinger, C. T. & Rokas, A. Variation and selection on codon usage bias across an entire subphylum. *PLOS Genet.* **15**, e1008304 (2019).
57. Sharp, P. M., Emery, L. R. & Zeng, K. Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 1203–1212 (2010).
58. Zhoua, Z. et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl Acad. Sci. USA* **113**, E6117–E6125 (2016).
59. Frumkin, I. et al. Codon usage of highly expressed genes affects proteome-wide translation efficiency. *Proc. Natl Acad. Sci. USA* **115**, E4940–E4949 (2018).
60. Sharp, P. M., Emery, L. R. & Zeng, K. Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 1203–1212 (2010).
61. Sharp, P. M. & Matassi, G. Codon usage and genome evolution. *Curr. Opin. Genet. Dev.* **4**, 851–860 (1994).
62. Powell, J. R. & Moriyama, E. N. Evolution of codon usage bias in Drosophila. *Proc. Natl Acad. Sci. USA* **94**, 7784–7790 (1997).
63. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of expression in *Escherichia coli*. *Science* **324**, 255–258 (2009).
64. Martínez, M. A., Jordan-Paiz, A., Franco, S. & Nevot, M. Synonymous genome recoding: a tool to explore microbial biology and new therapeutic strategies. *Nucleic Acids Res.* **47**, 10506–10519 (2019).
65. Lajoie, M. J. et al. Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
66. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
67. Kuo, J. et al. Synthetic genome recoding: new genetic codes for new features. *Curr. Genet.* **64**, 327–333 (2018).
68. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
69. Ma, N. J., Moonan, D. W. & Isaacs, F. J. Precise manipulation of bacterial chromosomes by conjugative assembly genome engineering. *Nat. Protoc.* **9**, 2285–2300 (2014).
70. Wang, K. et al. Defining synonymous codon compression schemes by genome recoding. *Nature* **539**, 59–64 (2016).
71. Lau, Y. H. et al. Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA. *Nucleic Acids Res.* **45**, 6971–6980 (2017).
72. Isaacs, F. J. et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348–353 (2011).
73. Ellis, H. M., Yu, D., DiTizio, T. & Court, D. L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl Acad. Sci. USA* **98**, 6742–6746 (2001).
74. Ostrov, N. et al. Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
75. Ma, N. J., Hemez, C. F., Barber, K. W., Rinehart, J. & Isaacs, F. J. Organisms with alternative genetic codes resolve unassigned codons via mistranslation and ribosomal rescue. *Elife* **7**, 1–23 (2018).
76. Roller, M., Lucić, V., Nagy, I., Perica, T. & Vlahoviček, K. Environmental shaping of codon usage and functional adaptation across microbial communities. *Nucleic Acids Res.* **41**, 8842–8852 (2013).
77. Mushegian, A. R. & Koonin, E. V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl Acad. Sci. USA* **93**, 10268–10273 (1996).
78. Glass, J. I. et al. Essential genes of a minimal bacterium. *Proc. Natl Acad. Sci. USA* **103**, 425–430 (2006).
79. Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME J.* **8**, 1553–1565 (2014).
80. Goffeau, A. et al. Life with 6000 genes. *Science* **274**, 563–567 (1996).
81. Blattner, F. R. et al. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462 (1997).
82. Costanzo, M. et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420–aaf1420 (2016).
83. Yang, X. et al. Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* **164**, 805–817 (2016).
84. Djebali, S. et al. Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
85. Gao, F., Cai, Y., Kapranov, P. & Xu, D. Reverse-genetics studies of lncRNAs—what we have learnt and paths forward. *Genome Biol.* **21**, 1–23 (2020).
86. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

87. Tang, W., Mun, S., Joshi, A., Han, K. & Liang, P. Mobile elements contribute to the uniqueness of human genome with 15,000 human-specific insertions and 14 Mbp sequence increase. *DNA Res.* **25**, 521–533 (2018).
88. Kitano, H. Biological robustness. *Nat. Rev. Genet.* **5**, 826–837 (2004).
89. Félix, M. A. & Barkoulas, M. Pervasive robustness in biological systems. *Nat. Rev. Genet.* **16**, 483–496 (2015).
90. Payne, J. L. & Wagner, A. The causes of evolvability and their evolution. *Nat. Rev. Genet.* **20**, 24–38 (2019).
91. Ro, D. K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
92. Paddon, C. J. et al. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–532 (2013).
93. Luo, X. et al. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **567**, 123–126 (2019).
94. Galanie, S., Thodey, K., Trenchard, I. J., Interrante, M. F. & Smolke, C. D. Complete biosynthesis of opioids in yeast. *Science* **349**, 1095–1100 (2015).
95. Wang, K., de la Torre, D., Robertson, W. E. & Chin, J. W. Programmed chromosome fission and fusion enable precise large-scale genome rearrangement and assembly. *Science* **365**, 922–926 (2019).
96. Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234–237 (2013).
97. Atwell, S. et al. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627–631 (2010).
98. Medini, D., Donati, C., Tettelin, H., Massignani, V. & Rappuoli, R. The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15**, 589–594 (2005).
99. Rouli, L., Merhej, V., Fournier, P.-E. & Raoult, D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *N. Microbes N. Infect.* **7**, 72–85 (2015).
100. Tettelin, H., Riley, D., Cattuto, C. & Medini, D. Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* **11**, 472–477 (2008).
101. Peter, J. et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates species-wide genetic and phenotypic diversity. *Nature* **556**, 339–344 (2018).
102. Hübner, S. et al. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat. Plants* **5**, 54–62 (2019).
103. Rasko, D. A. et al. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* **190**, 6881–6893 (2008).
104. Gordienko, E. N., Kazanov, M. D. & Gelfand, M. S. Evolution of Pan-Genomes of *Escherichia coli*, *Shigella* spp., and *Salmonella enterica*. *J. Bacteriol.* **195**, 2786–2792 (2013).
105. Lukjancenko, O., Wassenaar, T. M. & Ussery, D. W. Comparison of 61 sequenced *Escherichia coli* genomes. *Microb. Ecol.* **60**, 708–720 (2010).
106. Sagan, L. On the origin of mitosing cells. *J. Theor. Biol.* **14**, 255–274 (1967).
107. Schwartz, R. M. & Dayhoff, M. O. Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science* **199**, 395–403 (1978).
108. Berg, O. G. & Kurland, C. G. Why mitochondrial genes are most often found in nuclei. *Mol. Biol. Evol.* **17**, 951–961 (2000).
109. Gray, M. W. Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* **4**, a011403 (2012).
110. Gray, M. W. Lynn Margulis and the endosymbiont hypothesis: 50 years later. *Mol. Biol. Cell* **28**, 1285–1287 (2017).
111. Avalos, J. L., Fink, G. R. & Stephanopoulos, G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat. Biotechnol.* **31**, 335–341 (2013).
112. Mok, B. Y. et al. A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **583**, 631–637 (2020).
113. Gibson, D. G., Smith, H. O., Hutchison, C. A., Venter, J. C. & Merryman, C. Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* **7**, 901–903 (2010).
114. Johnston, S. A., Anziano, P. Q., Shark, K., Sanford, J. C. & Butow, R. A. Mitochondrial transformation in yeast by bombardment with microprojectiles. *Science* **240**, 1538–1541 (1988).
115. Rand, D. M., Haney, R. A. & Fry, A. J. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol. Evol.* **19**, 645–653 (2004).
116. Kelly, S. The economics of endosymbiotic gene transfer and the evolution of organellar genomes. *bioRxiv* <https://doi.org/10.1101/2020.10.01.322487> (2020).
117. Greaves, L. & Taylor, R. Mitochondrial DNA mutations in human disease. *IUBMB Life* **58**, 143–151 (2006).
118. Thornton, J. W., Need, E. & Crews, D. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* **301**, 1714–1717 (2003).
119. Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375 (2004).
120. Jermann, T. M., Opitz, J. G., Stackhouse, J. & Benner, S. A. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* **374**, 57–59 (1995).
121. Zausa, J. & Heddle, J. G. Resurrecting the dead (molecules). *Comput. Struct. Biotechnol. J.* **15**, 351–358 (2017).
122. Novak, B. J. De-extinction. *Genes* **9**, 548 (2018).
123. Seddon, P. J., Moehrenschrager, A. & Ewen, J. Reintroducing resurrected species: selecting DeExtinction candidates. *Trends Ecol. Evol.* **29**, 140–147 (2014).
124. Cohen, S. The ethics of de-extinction. *Nanoethics* **8**, 165–178 (2014).
125. McCauley, D. J., Hardesty-Moore, M., Halpern, B. S. & Young, H. S. A mammoth undertaking: harnessing insight from functional ecology to shape de-extinction priority setting. *Funct. Ecol.* **31**, 1003–1011 (2017).
126. Shultz, D. Should we bring extinct species back from the dead? *Science* <https://doi.org/10.1126/science.aah7343> (2016).
127. Margulis, L., Dolan, M. F. & Guerrero, R. The chimeric eukaryote: origin of the nucleus from the karyomastigote in amitochondriate protists. *Proc. Natl Acad. Sci. USA* **97**, 6954–6959 (2000).
128. Imachi, H. et al. Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**, 519–525 (2020).
129. Martin, W. F., Garg, S. & Zimorski, V. Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140330 (2015).
130. Douglas, T. & Savulescu, J. Synthetic biology and the ethics of knowledge. *J. Med. Ethics* **36**, 687–693 (2010).
131. Sliva, A., Yang, H., Boeke, J. D. & Mathews, D. J. H. Freedom and responsibility in synthetic genomics: the synthetic yeast project. *Genetics* **200**, 1021–1028 (2015).
132. Lartigue, C. et al. Genome transplantation in bacteria: changing one species to another. *Science* **317**, 632–638 (2007). The authors report the first example of genome transplantation.
133. Labrousseau, F. et al. Impact of donor-recipient phylogenetic distance on bacterial genome transplantation. *Nucleic Acids Res.* **44**, 8501–8511 (2016).
134. Tang, N., Ma, S. & Tian, J. New tools for cost-effective DNA synthesis. *Synth. Biol.* **3**–21 (2013).
135. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507 (2014).
136. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
137. Gibson, D. G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol.* **498**, 349–361 (2011).
138. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**, e3647 (2008).
139. Engler, C., Gruetznher, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot dna shuffling method based on type iis restriction enzymes. *PLoS ONE* **4**, e5553 (2009).
140. Ellis, T., Adie, T. & Baldwin, G. S. DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol.* **3**, 109–118 (2011).
141. Herschleb, J., Ananiev, G. & Schwartz, D. C. Pulsed-field gel electrophoresis. *Nat. Protoc.* **2**, 677–684 (2007).
142. Gibson, D. G. et al. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl Acad. Sci. USA* **105**, 20404–20409 (2008).
143. Benders, G. A. et al. Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.* **38**, 2558–2569 (2010).
144. Kouprina, N. & Larionov, V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* **125**, 621–632 (2016).
145. Hasebe, T., Narita, K., Hidaka, S. & Su’etsugu, M. Efficient arrangement of the replication fork trap for in vitro propagation of monomeric circular DNA in the chromosome-replication cycle reaction. *Life* **8**, 43 (2018).
146. Su’etsugu, M., Takada, H., Katayama, T. & Tsujimoto, H. Exponential propagation of large circular DNA by reconstitution of a chromosome-replication cycle. *Nucleic Acids Res.* **45**, 11525–11534 (2017). The authors reconstitute the *E. coli* replisome in vitro, showing that it can be used to assemble and replicate large DNA molecules.

Acknowledgements

We thank Sudarshan Pinglay, and members of the Ehrenreich lab for helpful input regarding this manuscript. I.M.E. also thanks to the JCVI Synthetic Biology and Bioenergy Minimal Cell team, in particular Dan Gibson, Clyde Hutchison III, John Glass, Chuck Merryman, Hamilton Smith, Yo Suzuki, and Kim Wise, for a sabbatical that motivated this paper. This paper was supported by grant R35GM130381 from the National Institutes of Health (I.M.E.), funds from the University of Southern California (I.M.E.), and National Institutes of Health Chemistry–Biology Interface training grant fellowship (T32GM118289; C.B.H.).

Author contributions

A.L.V.C., C.B.H., and I.M.E. conceived of, wrote, and revised this paper.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to I.M.E.

Peer review information *Nature Communications* thanks Stanley Fields and Liam Holt for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020