Genome Design and Engineering Postprint

Authors: Xue Xiaoli, Qin Zhongjun

Date: 2023-03-19T00:00:00+00:00

Abstract

Synthetic biology, which employs engineering principles to design and construct novel biological genomes, has become a prominent research focus in recent years. Notable successes have been achieved in the de novo synthesis of natural minimal genomes from prokaryotic mycoplasma cells and their subsequent design and construction into minimal genomes, the continuous reduction of genomes in the prokaryotic model organism Escherichia coli along with whole-genome codon deglomerization design and artificial synthesis testing, and the artificial design and synthesis testing of the genome of the eukaryotic model organism Saccharomyces cerevisiae. These accomplishments have significantly advanced our understanding of life. Although research on genome design and engineering construction in China started relatively late, it has achieved internationally remarkable results in recent years. The design and construction of genomes have laid the foundation for deeper understanding of the origin and evolution of life, and for further construction of novel life forms with powerful application functions.

Full Text

Genome Design and Engineering Construction

Synthetic biology, which applies engineering principles to design and construct novel biological genomes, has emerged as a major research frontier in recent years. Remarkable successes have been achieved in the de novo synthesis of natural minimal genomes of prokaryotic Mycoplasma and the subsequent design and construction of even more streamlined genomes, the continuous deletion and genome-wide codon simplification of the prokaryotic model organism *Escherichia coli*, and the artificial design and synthesis of the eukaryotic model organism *Saccharomyces cerevisiae* genome. These advances have profoundly enhanced our understanding of life. Although genome design and engineering construction research in China started relatively late, it has recently achieved internationally recognized breakthroughs. Genome design and construction not

only provides deep insights into the origin and evolution of life but also lays the foundation for creating novel life forms with powerful functional applications.

Artificial Design and Synthesis of Minimal Prokaryotic Genomes

In 2010, American scientist J. Craig Venter and his team published the world's first prokaryotic organism controlled by a chemically synthesized genome [1]. On this basis, through extensive design and testing, they constructed the currently known minimal genome of Mycoplasma [2]. This achievement marked a major milestone in synthetic biology. In comparison with the rapid international development of synthetic biology, China's research in this field began later. In 2011, China launched ten synthetic biology-related "973" projects, which greatly promoted research in artificial biological system design and synthesis, modularization and controllable operation of cellular functional groups, development of streamlined genome reconstruction and editing technologies, and "plug-and-play" assembly of biological modules in chassis cells.

Building a minimal genome requires enormous design, synthesis, and testing efforts. Initially, the researchers found that designing a minimal genome based on existing knowledge was not feasible. By integrating existing transposon mutagenesis and knockout data with accumulated molecular biology knowledge, they identified 440 non-essential genes that could be deleted from the 1.08 Mbp *Mycoplasma mycoides* genome JCVI-syn1.0, reducing it to 483 kbp. This presumed minimal genome was divided into eight large segments for synthesis and functional testing. However, only one of the eight segments was functional, and even that design resulted in poor bacterial growth. A major reason for this design failure was the existence of quasi-essential genes that significantly affect growth, lying between essential genes (e) and non-essential genes (n). Based on their impact on growth, these were classified as intensely affecting (ie), affecting (i), or mildly affecting (in) growth. The researchers concluded that essential genes (e) and growth-affecting genes (ie and i) should be retained, while only mildly affecting (in) and non-essential (n) genes should be candidates for deletion. Beyond quasi-essential genes, there are also redundant genes performing essential functions. For instance, if genes A and B both perform essential function E, single knockout of either A or B does not affect function E, leading to the misconception that both are non-essential. However, simultaneous deletion of both A and B would eliminate essential function E and kill the bacterium. Therefore, extensive testing is required to identify these quasi-essential genes and functionally redundant essential genes.

The successful synthesis of the 531 kbp minimal genome JCVI-syn3.0 [2] involved dividing the genome into 87 fragments of approximately 50 kbp each, with 55 fragments synthesized and tested individually. The researchers found that 91% of essential genes remained functional and did not affect growth after these modifications, demonstrating substantial genomic plasticity. Notably, there is a trade-off between genome size and growth rate during genome reduction. As the genome shrinks, growth rate decreases dramatically. The engineered JCVI-

syn3.0 has a generation time of 180 minutes, significantly slower than the 60 minutes of the parental 1.08 Mbp JCVI-syn1.0 strain. Moreover, unlike the parental strain that forms uniform suspensions in liquid media, JCVI-syn3.0 tends to precipitate and forms long filamentous networks and large vesicles under microscopic examination. Thus, this near-minimal genome cell still exhibits growth defects and is not perfect. Nevertheless, it represents a successful paradigm for designing and synthesizing a minimal genome.

Artificial Design and Synthesis of the *Escherichia coli* Genome

E. coli is the most extensively studied prokaryotic model organism. It grows rapidly and is relatively simple to manipulate genetically—many genetic engineering techniques were first established in *E. coli*. Comprehensive omics-level research, bioinformatics analysis, and modeling have enhanced our understanding of *E. coli* systems, enabling more rational modifications to create desired super-strains for industrial production. Genome minimization reduces complexity and eliminates non-essential metabolic pathways, facilitating the integration of heterologous high-efficiency production elements to obtain optimized high-yield strains. The *E. coli* genome is 4–5 Mbp, substantially larger than the 1 Mbp Mycoplasma genome, making a "bottom-up" de novo synthesis approach for constructing a streamlined genome extremely challenging. Consequently, most reported research on *E. coli* genome reduction has employed a "top-down" deletion strategy.

Three major classes of genome-reduced *E. coli* strains have been constructed [9]: (1) The Delta series rapidly reduces the genome through large-scale deletion of non-essential regions, but the resulting strains exhibit growth defects due to the lack of rational design. (2) The MGF series accumulates only small fragment deletions that do not affect growth, yielding strains with good growth performance. (3) The MDS series focuses on deleting insertion sequences, prophages, and other elements that do not affect growth but contribute to genome stability, producing strains considered to have "clean genomes" with robust growth.

In 2016, the Church research group reported the fully synthesized codon-compressed *E. coli* genome [9]. The genetic code consists of 64 codons (three-base combinations), with 61 codons corresponding to 20 amino acids and 3 stop codons. Multiple codons can encode the same amino acid. Genome-wide synonymous codon replacement can create genetically isolated organisms with novel biological functions. Genetic isolation means that foreign DNA, including from viruses, plasmids, or other cells, cannot be properly expressed in the engineered organism, making it resistant to infection and horizontal gene transfer. Additionally, the freed-up codons can be repurposed to encode non-standard amino acids beyond the 20 standard ones, enabling synthesis of proteins with novel chemical activities. The Church group reprogrammed the *E. coli* genome by replacing 7 codons (two serine codons, two leucine codons, two arginine codons, and one stop codon) with synonymous alternatives, reducing the genome's codon usage from 64 to 57 [9]. This was a bold undertaking, involving

over 60,000 modifications. The *E. coli* genome was divided into 87 fragments of approximately 50 kbp, with 55 fragments synthesized and tested individually. The results showed that 91% of essential genes remained functional and did not affect growth after these extensive modifications, demonstrating remarkable genomic plasticity.

Artificial Design and Synthesis of the *Saccharomyces cerevisiae* Genome

Saccharomyces cerevisiae is the simplest single-celled eukaryotic model organism. Redesigning and synthesizing its genome for functional studies will greatly advance human understanding and ability to engineer life. The "Synthetic Yeast Genome Project" (Sc2.0) represents humanity's first attempt to redesign and synthesize a eukaryotic organism from scratch. Initiated by National Academy of Sciences member Jef Boeke and involving multiple international research institutions, Sc2.0 aims to artificially redesign and chemically reconstruct all 16 chromosomes of the eukaryotic model organism *S. cerevisiae*, representing another major landmark international collaboration in synthetic biology following the prokaryotic Mycoplasma genome synthesis project. The first completed milestone was the total synthesis and replacement of chromosome 3, reported in *Science* in 2014 [10].

Chinese scientists (primarily from Tianjin University, Tsinghua University, and BGI-Shenzhen) collaborated with international partners including the Boeke laboratory to complete the de novo design and synthesis of 4 chromosomes (chromosomes 2, 5, 6, 10, and 12) [3-6], demonstrating the feasibility of artificial eukaryotic life systems. The design features for synthetic yeast chromosomes include: replacement of stop codon TAG with TAA, introduction of loxP recombination sites to enable rapid genome modifications, incorporation of PCR detection tags and restriction enzyme recognition sites, and deletion of tRNA genes and repetitive sequences. These modifications do not significantly affect yeast growth. The loxP sites introduced into early synthetic yeast chromosomes enable the SCRaMbLE (Synthetic Chromosome Recombination and Modification by LoxP-mediated Evolution) technology, where Cre recombinase cuts loxP sites, allowing sequences between any two loxP sites to invert, delete, or duplicate. This enables rapid chromosome rearrangement, and through directed screening, can quickly identify strains useful for improving microbial cell factories, accelerating evolution, and studying human chromosomal diseases.

In 2018, a research team led by Qin Chongjun at the Chinese Academy of Sciences created a yeast strain with a single linear chromosome (SY14) through 15 rounds of chromosome fusion, merging the natural 16 chromosomes of *S. cerevisiae* into one [7]. During this stepwise fusion process, a series of intermediate strains with progressively fewer chromosomes (from SY0 to SY13) were constructed. The final SY14 strain deleted 15 centromeres, 30 telomeres, and 19 long repeat sequences [7] [Figure 1: see original paper]. Although the single-chromosome yeast exhibits dramatic changes in three-dimensional chromosome

structure, its transcriptome and phenome remain similar to the wild-type strain. Moreover, the single-chromosome yeast retains the ability to undergo meiosis, albeit with slightly reduced sporulation efficiency and spore viability. These results demonstrate that single-chromosome *S. cerevisiae* can maintain normal cellular functions.

Discussion and Outlook

The complexity of life systems makes them difficult to analyze and manipulate. Simplifying and engineering genomes not only facilitates research on the origin, evolution, and metabolic regulation of life but also provides suitable expression hosts for biotechnological applications. From the perspective of genome minimization, deleting non-essential genes may be more complex than anticipated. Although 80–90% of genes in most bacterial and fungal genomes are non-essential for survival, these genes—including those with redundant essential functions—may become essential under different genome simplification backgrounds. The "bottom-up" artificial reconstruction and minimization of the naturally compact Mycoplasma genome demonstrate that even highly compact genomes can be further reduced. However, current strategies for rationally designing minimal genomes remain limited, relying heavily on extensive "design-synthesis-test" cycles. "Top-down" genome deletion in prokaryotic model organisms has also progressed slowly; despite nearly two decades of research on *E. coli* genome minimization, strains constructed through continuous deletion from various starting strains (with genomes of only ~3 Mbp) remain far from the theoretical minimal genome (~300 genes, ~0.3 Mbp). This indicates that rational design of minimal-genome artificial life remains in its early stages.

Nevertheless, successful attempts at genome simplification from alternative perspectives—such as deleting telomeres and centromeres from multiple yeast chromosomes and fusing the natural 16 chromosomes into one—have substantially simplified both inter-chromosomal and intra-chromosomal interactions. This demonstrates that naturally complex life systems can be simplified through artificial design and engineering to manifest as entirely new life forms.

References

1. Hutchison 3rd C A, Chuang R Y, Noskow V N, et al. Design and synthesis of a minimal bacterial genome. Science, 2016, 351: aad6253.
2. Zhang W, Zhao G, Luo Z, et al. Engineering the ribosomal DNA in a megabase synthetic chromosome. Science, 2017, 355: aaf3981.
3. Xie Z X, Li B Z, Mitchell L A, et al. "Perfect" designer chromosome V and behavior of a ring derivative. Science, 2017, 355: aaf4704.
4. Wu Y, Li B Z, Zhao M, et al. Bug mapping and fitness testing of chemically synthesized chromosome X. Science, 2017, 355: aaf4706.

5. Shen Y, Wang Y, Chen T, et al. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. Science, 2017, 355: aaf4791.

6. Shao Y, Lu N, Wu Z, et al. Creating a functional single-chromosome yeast. Nature, 2018, 560: 331-335.

7. Ostrov N, Landon M, Guell M, et al. Design, synthesis, and testing toward a 57-codon genome. Science, 353(6301): 819-822.

8. Annaluru N, Muller H, Mitchell L A, et al. Total synthesis of a functional designer eukaryotic chromosome. Science, 344(6179): 55-58.

9. 薛小莉, 覃重军. 大肠杆菌最小基因组分析和删减进展. 生命科学, 2013,10: 978-982.

Author Information

XUE Xiaoli is an Associate Researcher at the CAS Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. She received her B.Sc. in 2005 and M.Sc. in 2007 from Zhongshan University in Guangzhou, and her Ph.D. in 2011 from the Helmholtz Centre for Infection Research/Technical University of Braunschweig, Germany. Her primary research interest is synthetic biology, focusing on the artificial design and reconstruction of genomes in *Saccharomyces cerevisiae* and *Escherichia coli*.

Corresponding author. E-mail: xlxue@sibs.ac.cn

Note: Figure translations are in progress. See original paper for figures.