Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Synthetic genomics: a new venture to dissect genome fundamentals and engineer new functions

Daniel Schindler[1], Junbiao Dai[2] and Yizhi Cai[1,2]

Since the first synthetic gene was synthesized in 1970s, the efficiency and the capacity of made-to-order DNA sequence synthesis has increased by several orders of magnitude. Advances in DNA synthesis and assembly over the past years has resulted in a steep drop in price for custom made DNA. Similar effects were observed in DNA sequencing technologies which underpin DNA-reading projects. Today, synthetic DNA sequences with more than 10 000 bps and turn-around times of a few weeks are commercially available. This enables researchers to perform large-scale projects to write synthetic chromosomes and characterize their functionalities *in vivo*. Synthetic genomics opens up new paradigms to study the genome fundamentals and engineer novel biological functions.

Addresses
[1] Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, M1 7DN Manchester, UK
[2] Centre for Synthetic Genomics, Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

Corresponding author: Cai, Yizhi (yizhi.cai@manchester.ac.uk)

Current Opinion in Chemical Biology 2018, **46**:56–62

This review comes from a themed issue on **Synthetic biomolecules**

Edited by **Richard J Payne** and **Nicolas Winssinger**

https://doi.org/10.1016/j.cbpa.2018.04.002

Introduction

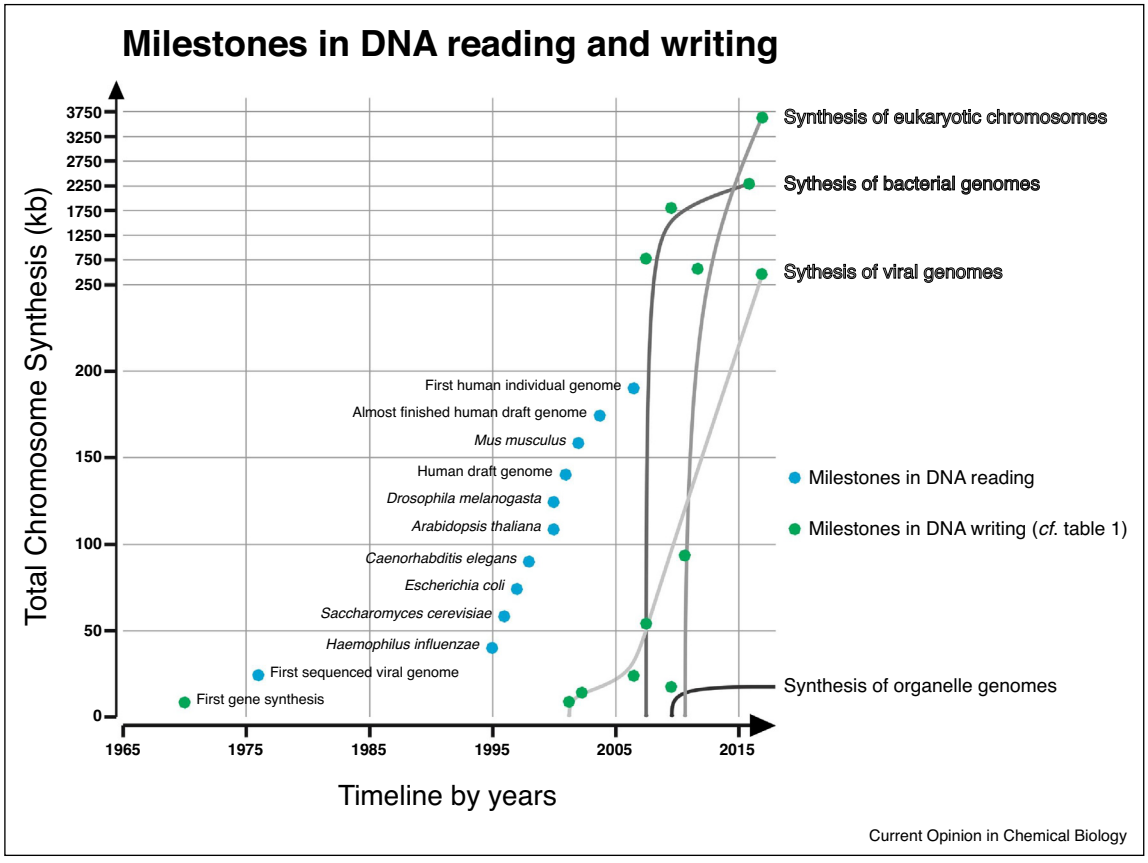
One of the major challenges in biological sciences was the determination of DNA sequences. In the beginning, only single DNA fragments were sequenced using the chain termination sequencing technique [1]. However, the Human Genome Project (GP-Read) accelerated the evolution of new sequencing techniques by having the ambitious goal to sequence the human genome within 15 years. The development of Next Generation Sequencing techniques today allows sequencing of a human genome within days. However, most eukaryotic genomes are not fully sequenced and new sequencing techniques are still being developed. As exemplary achievement of this development, in 2017 sequencing of one of the highly repetitive human centromeres was achieved [2•]. Scientists are now performing well in reading genomes, a measurable output being the growing number of genome sequences in public databases. However, reading a book alone does not make a good writer, instead it requires one to start writing extensively and creatively to master the art and ultimately it leads to a better understanding of grammar and expression. In this case, one needs to write synthetic DNA sequences in order to better understand the grammar of life.

Writing DNA starts with short single-stranded fragments: the oligonucleotides. Since the development of the Polymerase Chain Reaction and the first complete synthesis of a gene, writing DNA *in vitro* has progressed impressively (Figure 1) [3,4]. Recent drops in DNA synthesis costs and the improved capability of synthesizing longer stretches of DNA allow the design and construction of whole synthetic chromosomes in the mega-base range. Recent publications report the construction of viral and microbial synthetic genomes, and the Sc2.0 project aims to generate the first synthetic eukaryotic genome. It is an open discussion how to define whether a chromosome or genome is synthetic. In this review, chromosomes and genomes are defined as synthetic when all building blocks of the final DNA molecule are generated by chemical synthesis. Chromosomes and genomes which are not completely synthesized are considered 'engineered' or 'modified' and are outside the scope of this review. We define synthetic genomics to be a new field where biology is being engineered at the genome level, and it is an intersection of synthetic biology and systems biology. This review neither aims to discuss assembly methods nor the dual-use character of synthetic genomics. The authors are fully aware of the potential dual use character, especially for the synthesis of viral genomes. However, these issues are discussed and reviewed extensively elsewhere [5–7].

Design concepts and assembly strategies for synthetic chromosomes

Computer-assisted design software (CADs) have been developed to ensure efficient and consistent design of synthetic DNA sequences at the genome scale [8••,9•]. The design space of synthetic DNA is enormous and many (if not infinite) design blueprints are possible, as long as they can result in the viability of the cell, to achieve the design intention [10]. Initial projects aiming to synthesize a whole genome were conservative in

Figure 1



Milestones in DNA reading and writing. Reading and writing technologies depend on technological breakthroughs. Large scale DNA-reading projects (examples in blue) were accomplished after development of Sanger Sequencing, PCR and Next Generation Sequencing. The number of studies utilizing new technologies grows quickly after a developmental lag phase. The number of genome sequences uploaded to databases is exploding and it is impossible to give a number which would be accurate and valid for some time. The knowledge gained by genome sequencing and advantages in gene synthesis is the dawn of writing chromosomes. By now the number of bases incorporated into completely synthetic chromosomes is: 6.1 mb. The cost by today would be roughly $425 000 assuming the current competitive price rate of 7 cents per base for non-clonal 1.8 kb DNA-fragments. The synthesis cost for a haploid human genome would by today be roughly $45 000 000. However, lowering DNA synthesis costs is one of the major goals of GP-write. In the future, the DNA synthesis cost of a human genome in will be less than the price of the *Mycoplasma mycoides* JCVI-syn1.0 project (estimated $40 000 000 [33]).

changes to the genetic content, but nonetheless resulted in the breakthrough in synthesizing, assembling and ultimately transplanting chromosome-scale synthetic DNA [11,12•]. With increasing knowledge and progress in chromosome-scale DNA synthesis, the designs of synthetic sequences are becoming more complex and ambitious [8••,13••]. Many genome synthesis projects utilize a hierarchical genome assembly strategy starting with small building blocks which are assembled, by the technique of choice, to larger building blocks of around 50–100 kb. These fragments are used to further assemble the synthetic chromosome in a heterologous host or to replace the corresponding wild-type sequence in a stepwise manner. Each of the techniques have advantages and disadvantages (Box 1), and should be chosen carefully based on the use cases.

Synthesizing DNA goes viral

Although viruses and phages are not considered to be 'alive' they have a genome. They can reproduce themselves by leveraging the resources from a host. Viral genomes are rather small, with sizes between 1759 bps (*Porcine circovirus* [14]) and 1259 kb (*Megavirus chilensis* [15]) and can consist of DNA or RNA. The first complete synthesis of a viral RNA genome, the polio virus, was accomplished in 2002 [16••]. The 7.5 kb synthesized cDNA genome was *in vitro* transcribed by RNA polymerase and can generate infectious virus particles after transfer into a cell free extract. Further viral RNA and DNA genomes were synthesized up to a size of 212 kb in recent years (Table 1). Synthesizing, as well as engineering variations of viral genomes to produce genome libraries, has an enormous potential for therapeutic applications. Vaccines and drugs could be quickly generated in