What is synthetic genomics anyway?

Tom Ellis
(Imperial College London, UK)

You may have heard of synthetic genomics. This headline-grabbing, high-profile, big science topic is starting to emerge catalysed by the pioneering work of famous names in synthetic biology and biotechnology like George Church and Craig Venter. But what is synthetic genomics and what is it being used for? As a prominent researcher at a recent UK meeting said: "Is it just synthetic biology with bigger bits of DNA?" Well no, not quite...

Synthetic genomics and synthetic biology

Synthetic genomics certainly owes a lot to synthetic biology, making use of many of the methods, resources and jargon involved, yet it differs in crucial ways too. Synthetic biology—a major new interdisciplinary subject that has emerged this century—is focused on rewriting and reprogramming the DNA in cells using cycles of design and engineering to obtain improvements. It aspires to do this using engineering principles such as modularity and standardization that enable researchers to more quickly get to the ultimate aim of tailoring cells as technologies for specific tasks. Right now, synthetic genomics lacks these formalities, as the goal of the work is not to optimize one cell behaviour over the rest but to produce a new understanding of DNA and biology, either directly or by enabling new experiments that can't be done any other way.

New understanding of biology has already been one of the main outcomes of two decades of synthetic biology research. By building up synthetic gene systems from first principles, scientists are better able to understand and mathematically model the key factors that define important networks and pathways where genes interact together in cells. Proponents of this aspect of synthetic biology often use a famous quote from physicist Richard Feynman: "What I cannot create, I do not understand", which concisely postulates that the best way to learn about how something works is by trying to build it. Indeed, wanting to determine the minimal requirements for cells to genetically encode memory and rhythms led to the first significant achievements in synthetic biology; synthetic gene circuits that act as switches and oscillators.

Twenty years since these first steps in synthetic biology, academic labs and biotech companies around

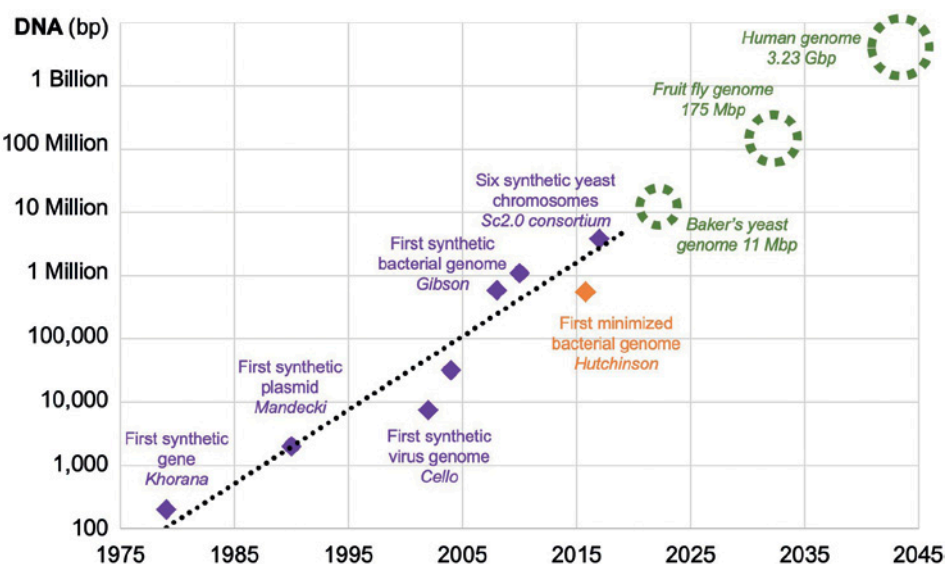

Figure 1. Progress in the scale of DNA synthesis and assembly. Landmark publications constructing with synthetic DNA are shown going from Khorana's 1979 work to chemically synthesize a tRNA gene, to the completion of six synthetic yeast chromosomes in 2017. Assuming continued exponential progress, estimate dates for completion of yeast, drosophila and human genomes are shown.

the world now use synthetic DNA to build a lot bigger than just systems of two or three genes (Figure 1). It is becoming routine to see dozens of genes used in synthetic DNA constructs for various tasks, and so naturally the cell's own operating system, its genome, is increasingly within our sights. However, the true synthetic biology version of a synthetic genome, a genome designed and built using first principles from a kit of modular parts, is still a long way off, looming as a grand challenge that could even take another couple of decades to achieve. Right now, we simply don't know enough about all the genes and genetic regulation that is required to direct a cell to grow and perform a cell cycle, and so we cannot yet write a genome from scratch. The task also gets more complex by the day as researchers in cell and genome science continue to uncover new unexpected ways that DNA encodes regulation and function that will need to be taken into account.

So, for now and the near future, synthetic genomics is best placed to help us understand what we do and don't know about cell biology and especially how the genome encodes an organism. Constructing and testing synthesized genomes and chromosomes that are increasingly different compared with natural genomes enables us to test our current understandings of genome biology, whilst also developing the methods and tools to one day build custom genomes to design. Therefore, most synthetic genomics projects right now aim to deliver new knowledge of genome coding, content and organization—aspects that are hard to determine by other approaches. By tackling these interesting questions using a new synthetic approach to genome manipulation, these projects both push and pull the development of new technologies that one day will enable broader use of synthetic genomics within research or applied synthetic biology.

A decade of synthetic genome progress

Impressively in just over 10 years, synthetic genomics efforts in bacteria have already advanced what is possible by several steps (Figure 2). In 2008, a full copy of a 580,000 bp *Mycoplasma* genome was constructed from chemically synthesized DNA, and then in 2010 the same team showed that a synthetic copy of a 1 million bp *Mycoplasma* genome could replace a natural genome and support the growth and division of a cell. This landmark work by the J. Craig Venter Institute gave us the first cell with a synthesized genome, albeit one with no major changes to its DNA sequence—it simply showed that synthesis and construction was possible.

In 2013, a team from Yale and Harvard then showed that a bacterial genome could be 'recoded' by using site-specific mutation (not genome synthesis) to remove all 321 occurrences of the rarest codon used in protein synthesis in *Escherichia coli*. This Genomically Recoded Organism (GRO) now differed from almost all of the rest of natural biology in not using the same 64 codons in its genes to direct which amino acids are used to make its proteins. It now only used 63, and so the spare codon in this cell could be reassigned to make *E. coli* add non-standard amino acids into proteins; a feature useful for both research and biotechnology applications.

While altering only 321 bases in a 4.6 Mbp genome may seem like a minor change, this work showed that genomes could be made with recoding throughout their genes, changing the DNA that encodes the proteins without altering the protein itself. UK and US teams are now pushing to produce *E. coli* and *Salmonella* bacteria with substantially more DNA recoding in their genes and more codon reassignment, in all cases now doing it by constructing the recoded genomes from synthesized DNA, rather than by mutation.

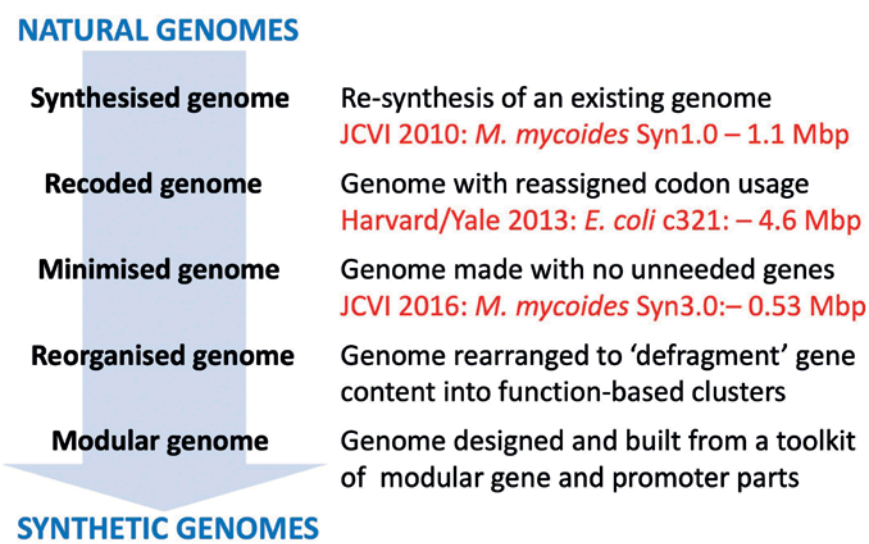

Figure 2. Five steps from natural genomes to fully synthetic genomes. Overview of the steps from being able to build a synthetic copy of an existing genome to being able to build custom genomes from modular parts. The first synthetic genomics project to achieve each step is shown in red text.

The next step that bacterial synthetic genomics has taken beyond recoding is in genome minimization. In 2016, the J. Craig Venter Institute constructed a synthetic, redesigned version of their 2010 *Mycoplasma* genome, leaving out the genes and DNA that they deemed not to be essential for growing this cell in the lab, which amounted to roughly half of the genome. No recoding of genes was done in this work, and where DNA remained it was the equivalent to its natural sequence. However, this achievement represents our 'most synthetic' genome to date as it has such huge differences in its gene content and layout compared with its natural equivalent.

Interestingly, in this minimized genome project the team tried pushing their work to an even further step towards the long-term goal of a fully modular synthetic genome. As they synthesized and constructed their minimized genome, they also made a version where the order and layout of the remaining genes on the bacterial genome was totally changed, with the genes now arranged along the chromosome according to function. The team called this version 'defragmented', making an analogy to the process where computer files in a hard drive are relocated to common clusters to improve storage efficiency. For a 1/8th segment of the genome this defragmented design could replace its natural equivalent, but for the rest of the genome it could not. This tells us that the layout and order of the genes in the genome play a crucial role in whether they work correctly—revealing important new information on 'genome design rules' that will need to be considered in future efforts to construct custom genomes from modular DNA parts.

Synthetic genomes beyond bacteria

A synthetic genome for a eukaryote has yet to be realized, but the international synthetic yeast genome project (Sc2.0) is rapidly approaching that goal by having a community of research groups around the world build synthetic chromosomes to a common new design. The baker's yeast *Saccharomyces cerevisiae* has an 11 Mbp genome naturally split into 16 different chromosomes, and synthetic versions of seven of these have now been completed. The design of the synthetic genome includes gene recoding and some minimization too, via the removal of unneeded non-coding elements such as transposons and introns. It also has an element of defragmenting as all transfer RNA (tRNA) genes are being removed from their normal locations in the main chromosomes to be now placed on a new synthetic tRNA chromosome. The Sc2.0 genome also has an inbuilt design feature that means further minimization and gene rearrangement can be done when desired. This is achieved by an inbuilt system called 'SCRaMbLE' where genes within the synthetic chromosomes can be randomly removed and rearranged inside the living yeast cells when they are given a specific chemical stimulus (Figure 3). Theoretically, continued SCRaMbLE of the complete Sc2.0 genome inside yeast growing under lab conditions would eventually lead to a genome only containing the required genes for lab-based growth, and with these genes in a new layout that enabled this genome to function well.

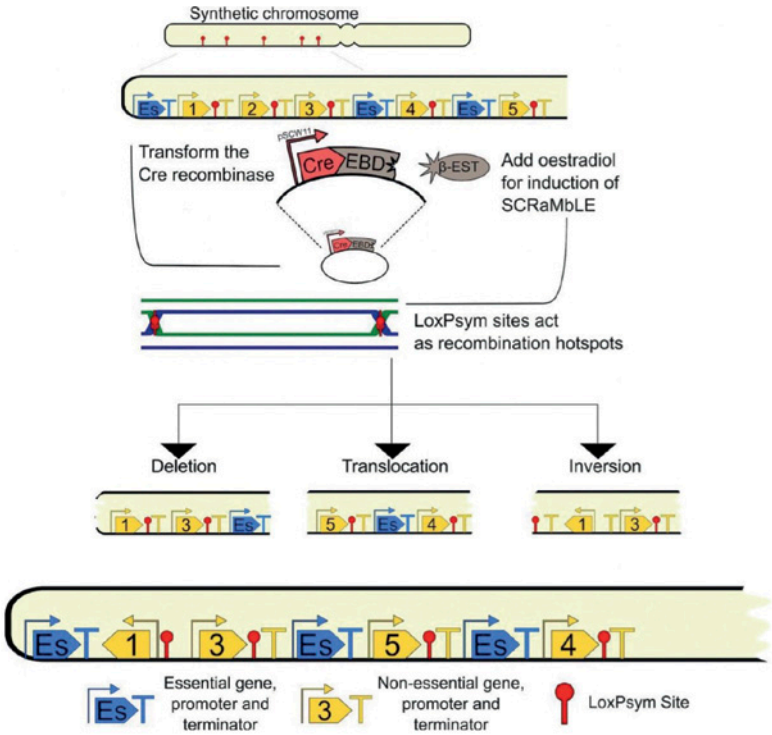

Figure 3. The SCRaMbLE system for rearranging synthetic yeast chromosomes. Diagram shows the rearrangements typical when the SCRaMbLE system is used on a region of a synthetic yeast chromosome. SCRaMbLE requires yeast cells to contain a plasmid that expresses Cre recombinase that binds and recombines pairs of loxPsym DNA sites when an oestradiol inducer is given.

While SCRaMbLE is not a direct way to remove or relocate large portions of the genome as desired, it still provides a powerful method to explore what genes are essential for a cell in various conditions and what gene order and genome arrangements are tolerated (and which ones aren't). Work with SCRaMbLE on the completed synthetic yeast chromosomes has already shown that the yeast genome can handle some serious rearrangement of its genes without many problems. Two teams have also shown that the 16 chromosomes of yeast can also be fused together so that the genome of yeast can be put on only two chromosomes with the cell functioning just fine. The whole genome can even be completely placed on just a single chromosome and still power a growing cell, albeit one that grows slower than usual. Clearly, there is significant plasticity in the chromosomal structure and gene layout in the yeast genome, which is a clue that eukaryotic genomes may in the end be more amenable than those of bacteria for the next steps for synthetic genomics, such as full genome reorganization and ultimately modular design and construction.

So it seems after only 10 years of synthetic genomics that genome recoding, genome minimization and large-scale synthetic chromosomal reorganization are all possible

both in prokaryote and eukaryote microbes. These efforts are redefining how we think about genomes and the relative (lack of) importance of naturally evolved sequences, gene content and layout. We now know that cells can happily exist outside of nature's standard genetic code where the same 64 codons encode the 20 amino acids of all proteins, and we've proven that genomes have no need to host transposable elements despite their ubiquity. Chromosome layout and content can be altered far beyond what we see in natural variation within species, but so long as key genes remain and are appropriately regulated, then cells are still viable and can even grow just fine.

These recent and ongoing advances all help towards the next major goals for synthetic genomics, which are to make viable, fully refactored genomes and eventually realize completely modular genomes that are built-to-design from standard parts. At that point, synthetic genomics would indeed return to being an engineering discipline like that of synthetic biology, where engineering tools (design and construction automation) and engineering principles (modularity, standardization) can be used to accelerate and industrialize the work of making cells as technologies.

While teams work towards achieving these goals in model microbes, the technologies for performing synthetic genomics can also benefit research elsewhere. For example, in more complex organisms like humans and mice, biomedical research is continually seeking to better understand how the DNA sequence and the organization of regulatory regions is important in determining gene function and how mutations lead to pathogenicity. The same tools and methods used to design and make Mb chromosomes for microbes can be used to recode and reorganize similar-sized regions within mammalian genomes, providing a new way to ask and answer questions on genome biology. Being able to synthesize, rearrange and relocate big DNA into mammalian genomes is now just beginning as a new approach to explore how the content and organization of the large stretches of non-coding sequence ('the dark matter of the genome') are involved in the regulation and correct functioning of genes and cells. Via big DNA design and synthesis, researchers can make and test synthetic variants of important genomic loci, like regions containing key genes associated with cancer or development. They can then learn how the sequences, features and arrangements in these loci define how they work, helping to better understand how our own genomes function and how mutation in them can lead to diseases.

It's early days, but if this 'learn-by-building' approach with big DNA pays off, then synthetic genomics in complex organisms may well become mainstream more quickly than we think. And so, while making synthetic human genomes seems decades away right now when synthesizing genomes a thousand times smaller is still an expensive and lengthy challenge, we need to be wary that technology in this area may well accelerate much faster than we think. Are we prepared for synthetic human genomes anytime soon? Custom-built synthetic microbial genomes is one thing, but the notion of synthetic human genomes raises many more pressing questions. They are not reality now, but very well may be real issues within in our lifetimes. It is therefore important that concerted efforts are made to engage widely, discuss and coordinate globally how synthetic genomics will advance over the next decades. These efforts have already begun, spearheaded by an international community of interested researchers, social scientists, engineers, lawyers and citizen science advocates who have formed the GP Write consortium. With community oversight, it is hoped that synthetic genomics can not only advance quickly to the benefit of science, but also advance safely to the benefit of society. ■

Tom Ellis is a Reader in Synthetic Genome Engineering in the Department of Bioengineering at Imperial College London, and Associate Faculty in Synthetic Genomics at the Wellcome Trust Sanger Institute, Hinxton. His group is part of the Imperial College Centre for Synthetic Biology and focuses on science and engineering research to advance synthetic genomics and modular synthetic biology. Dr Ellis is part of the international Sc2.0 consortium, leading work towards a synthetic version of yeast chromosome 11. He is also a member of the GP Write technical working group and the EU BioRoboost project to define standards for synthetic biology. Email: t.ellis@imperial.ac.uk

Further reading

- Hutchison, C.A. 3rd, Chuang, R.Y., Noskov, V.N., et al. (2016) Design and synthesis of a minimal bacterial genome. Science **351**. aad6253. doi:10.1126/science.aad6253
- Richardson, S.M., Mitchell, L.A., Stracquadanio, G., et al. (2017) Design of a synthetic yeast genome. Science **355**, 1040–1044. doi: 10.1126/science.aaf4557
- Wang, L., Jiang, S., Chen C., et al. (2018) Synthetic genomics: from DNA synthesis to genome design. Angew Chem. Int. Ed. **57**, 1748–1756. doi:10.1002/anie.201708741
- GP Write website containing information about the GP Write consortium, pilot and planned projects and roadmapping documents from the consortium working groups, https://engineeringbiologycenter.org/
- Meeting report from the UK meeting 'Synthesising a Human Genome: What could go right?' held September 2018. www.synbicite.com/media/uploads/files/Synthesising_a_Human_Genome_Meeting_Report.pdf