

STEVE SCHMEISSNER/SPL

The yeast *Saccharomyces cerevisiae* is the focus of a project to synthesize one of the first non-bacterial artificial genomes.

BUILDING BLOCKS FOR BESPOKE BIOLOGY

A powerful set of molecular tools helps synthetic biologists to assemble DNA of different sizes, from the gene to the chromosome scale. **By Michael Eisenstein**

Leslie Mitchell had no intention of doing a postdoc. After completing her PhD at the University of Ottawa, she had planned to move to industry. But then a member of her thesis committee, geneticist Jef Boeke, invited her to join his team at Johns Hopkins University in Baltimore, Maryland. Boeke was spearheading an ambitious effort to design and build an entire yeast genome from scratch, known as the Sc2.0 project. It was a once-in-a-lifetime opportunity, and one she just couldn't refuse. "I just thought that was the coolest way to study biology and really understand it," she says. "To build it from the ground up."

Eight years later, the Boeke lab (now at New York University's Langone Medical Center in New York City) and its collaborators in Europe, Asia and Australia are close to producing

recorded versions of all 16 *Saccharomyces cerevisiae* chromosomes, as well as a 17th, artificial, 'neochromosome'.

Only a handful of genomes have been synthesized so far, mostly for bacteria. Synthetic biologist Jason Chin and his colleagues at the MRC Laboratory of Molecular Biology in Cambridge, UK, have rewritten the genome of *Escherichia coli*¹, and researchers at the J. Craig Venter Institute (JCVI) in La Jolla, California, have constructed a 'minimal'

"The complexity of genome design remains much higher than our current tools can support."

genome for *Mycoplasma mycoides*, which has all non-essential genes deleted². Sc2.0 will synthesize the first genome of a eukaryote (a cell that has a nucleus enclosed within a membrane), and marks a huge advance in the engineering and assembly of DNA sequences.

"Twenty years ago, people struggled just to put a few genes together," says Patrick Cai, a synthetic biologist at the University of Manchester, UK, who is the international coordinator of Sc2.0. "Today, people are looking at chromosomes with thousands of components."

The tools and techniques used to synthesize genomes are proving powerful at smaller scales, too. They are, for example, allowing researchers to string together custom-built metabolic pathways so that cells can manufacture drugs such as opioids and antibiotics. But cells are not as easy to rewire as circuit boards, and the

field is still unable to achieve its ultimate goal: designing complex biological systems that give predictable results. “The complexity of genome design remains much higher than our current tools can support,” says Cai.

Budget base pairs

In the early years of synthetic biology, researchers faced two roadblocks: they had no easy way to assemble large sections of DNA, and could not afford to buy the components from commercial manufacturers. At the turn of the millennium, says JCVI synthetic biologist John Glass, researchers might have paid as much as US\$16 per nucleotide for a custom-made DNA sequence. A construct spanning a few thousand bases – roughly the length of a typical yeast gene – could carry a five-figure price tag.

Today, chromosome-scale construction is affordable, although still not cheap. Five years ago, when Cai’s team first started rebuilding a *S. cerevisiae* chromosome, it was paying DNA-synthesis companies about 30 cents per base. “To synthesize about 700 kilobases, that was roughly \$200,000 in raw material,” he says. A similar effort today would cost less than half that, says Tom Ellis, a synthetic biologist at Imperial College London who is working with Cai on the Sc2.0 project.

But it is unclear how much further prices can fall without a reinvention of synthesis technology. The standard technique, called phosphoramidite synthesis, is decades old and struggles to produce sequences longer than about 200 bases; anything bigger must be created by linking the fragments together.

Enzymatic synthesis methods are a promising alternative. In 2018, for instance, researchers led by synthetic biologist Jay Keasling and his then-PhD student Daniel Arlow at the University of California, Berkeley, demonstrated a process using enzymes that were cross-linked to nucleotides³, although the resulting sequences were just ten bases long. Last year, the Paris-based company DNA Script announced the synthesis of a 200-nucleotide sequence, the longest such construct reported so far. Several other companies are now moving in this direction, including Ansa Biotechnologies, co-founded by Arlow in Berkeley in 2018, and Molecular Assemblies in San Diego, California. “In five years, I think we’ll be looking at enzymatic-synthesis companies that are competitive with phosphoramidite synthesis,” says Ellis.

Bigger and better

Researchers now routinely outsource the production of fragments spanning a few thousand bases to companies such as Twist Bioscience in San Francisco, California, and Integrated DNA Technologies in Coralville, Iowa. Larger segments are available, but as the length increases so, too, does the cost per base. “It just depends how much you have in

your bank account versus how much time you want to spend putting DNA together,” Mitchell says. Junbiao Dai, director of the Shenzhen Key Laboratory of Synthetic Genomics in China, typically outsources the synthesis of pieces that are around 2,000–3,000 bases long, but estimates that the cost per base would double for a 10-kilobase fragment. “I would just do the assembly in my own lab, because we are experienced and I think we can do it much faster,” says Dai.

Fortunately, researchers seeking to construct assemblies measuring between 5,000 and 50,000 bases have several choices. One of these was used in the assembly of the minimal *M. mycoides* genome². Developed by Daniel Gibson and his colleagues at the JCVI, ‘Gibson Assembly’ makes use of DNA fragments that have matching overlapping sequences at their ends. An exonuclease enzyme is used to digest the ends of the DNA and leave complementary single-stranded sequences that readily pair up. Other enzymes then fill in any gaps and produce the finished molecule (see ‘Gene assembly’).

“The neochromosome is still super unstable: we’re putting all the bad eggs in one basket.”

Gibson Assembly can efficiently combine up to a dozen chunks of DNA in a single reaction, producing constructs longer than 50 kilobases. But it can stumble over repetitive sequences, and is less well suited to constructs that bring together multiple small chunks of DNA. “It’s really bad at assembling a really long DNA and a really short DNA,” says Nicola Patron, a molecular and synthetic biologist at the Earlham Institute in Norwich, UK. Much of her team’s work revolves around combining multiple genes and regulatory elements to alter the function of plant cells. Patron has found that a method known as Golden Gate assembly offers a better fit.

Developed by synthetic biologist Sylvestre Marillonnet and his colleagues at Icon Genetics in Halle, Germany, Golden Gate uses specialized proteins known as type IIS restriction enzymes to make targeted cuts in DNA strands⁴. The enzymes are guided by a ‘recognition sequence’, but make the cuts at a defined distance from the recognition site. Researchers can customize the resulting ‘overhangs’ so that different pieces can be assembled in a defined order. Users can typically combine 5–10 fragments in a single reaction, building up pieces that span tens of thousands of bases. However, the reliance on DNA-cutting enzymes means that researchers must ensure that none of their fragments contains an unwanted recognition site.

The synthetic-biology community has

extended Golden Gate by creating libraries of standardized parts, including genes, promoter sequences that guide where gene transcription starts, and other regulatory elements. “You can pick and choose pieces like Lego,” says Ellis. His group routinely engineers yeast gene circuits with this system, and Golden Gate kits for various species have been shared across labs or commercialized, with many available through AddGene, a non-profit reagent repository in Cambridge, Massachusetts. Patron’s group has developed a plant-specific Golden Gate library of roughly 350 parts, which other plant-biology groups have embraced. “Our toolkits have been distributed to over 200 labs, and I’d guess that every lab that gets it makes at least a couple of new parts,” she says.

Creating chromosomes

Both Gibson Assembly and Golden Gate are cost-effective – Ellis estimates that a typical reaction costs less than \$5. And the methods are sufficiently well established that new users do not take long to get up to speed. “We’ve got an undergraduate student in our lab and within the first three weeks they could assemble multi-gene constructs with Golden Gate,” says Patron.

But for assemblies that span hundreds of thousands or even millions of bases, the challenges intensify. At present, the only solution is to let living cells do the hard work. *Saccharomyces cerevisiae* has highly efficient DNA recombination mechanisms, and biologists can hijack these by feeding the cell with large fragments that have overlapping ends, similar to those used for Gibson Assembly. This means researchers can use a yeast cell to string the sequences together into constructs of 100 kilobases or more while they wait. “*In vivo* yeast assembly is the method being used for all of the large synthetic chromosome projects that I know of,” says Nili Ostrov, a geneticist in the lab of genomics researcher George Church at Harvard University in Cambridge, Massachusetts.

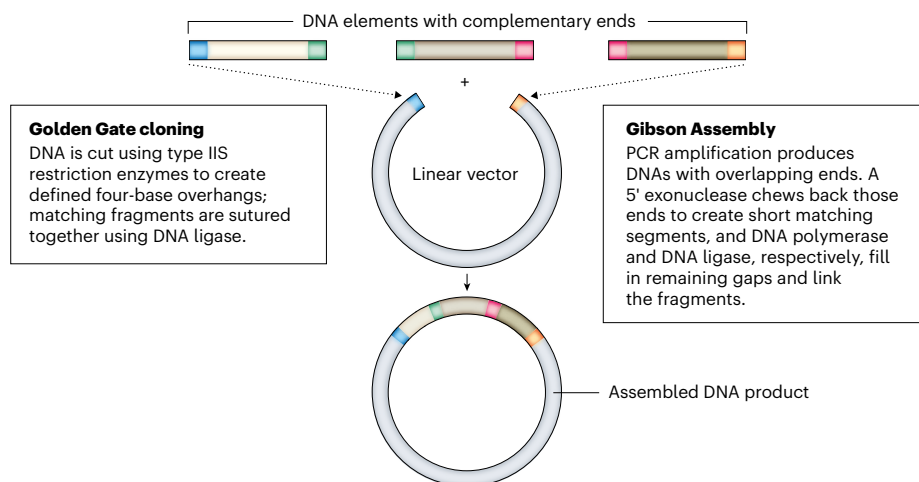
Assembly is typically achieved in a stepwise fashion, which allows for careful quality control and troubleshooting. For example, Dai’s Sc2.0 group used Golden Gate to build moderately large fragments which were then sequentially recombined into a yeast chromosome. “We would replace the native genome with three 10-kilobase fragments at a time, covering a chunk of 30 kilobases,” he says.

But long sequences are hard to handle. “As you get to 50 or 100 or 500 kilobases, it becomes exponentially more difficult,” says Glass. For example, routine laboratory procedures such as pipetting have minimal effect on sequences that are a few thousand bases long, but produce destructive shear forces on much larger fragments that can render the sequences unusable.

Nevertheless, stepwise assembly can yield extraordinary results. The minimal

GENE ASSEMBLY

To ensure DNA fragments are put together in the required order — for instance, to create a genetic circuit — researchers have several options, including Gibson Assembly and Golden Gate cloning.



M. mycoides genome contained more than one million bases². The longest yeast chromosome under construction for Sc2.0 is 50% larger than this — around 1.5 megabases. And researchers at the Chinese Academy of Sciences managed to pack the entire *S. cerevisiae* genome into a single chromosome spanning nearly 12 million base pairs⁵.

The next frontier

Most of the genome synthesis efforts so far have focused on rewriting existing material, rather than starting from scratch, but these early forays already hint at remarkable genomic flexibility. For example, Ostrov and her colleagues have been developing an *E. coli* derivative with a genetic code that uses only 57 of the usual 64 3-letter 'codons' found in nature⁶, freeing up the other 7 for future repurposing. "We tend to take the wild type as baseline and move a little bit to the left or right, but maybe we can try very radical changes," she says.

This mirrors Cai's experiences in constructing the Sc2.0 neochromosome. The neochromosome carries all the genes that encode yeast transfer RNA (tRNA) molecules, which were then deleted from their native locations in the other recoded chromosomes. "These are the troublemakers in the genome," Cai explains: tRNA genes tend to be sites of genomic damage and rearrangement. "We built the neochromosome, but it's still super unstable and you can imagine why: we're putting all the bad eggs in one basket."

Now, the field's pioneers are moving beyond the well-established unicellular laboratory models of yeast and *E. coli*, towards complex bacteria — and even into plant and mammalian cells. Many of these researchers have joined forces under the aegis of the GP-write Consortium, an international effort spearheaded by Church, Boeke and others to streamline the cost and labour associated with genome design and synthesis. Dai is coordinating a

parallel effort in China, which will initially focus on single-celled organisms and viruses but is also exploring plant species.

In policy paper published last October, GP-write members highlighted technological challenges confronting the field⁷. Most notably, these include the need for alternatives to yeast as a way to assemble large DNA fragments. This is because chromosome-scale DNA molecules assembled in yeast are exceedingly difficult to transfer to cells of other organisms, and some genomes will inevitably be incompatible with the mechanisms yeast uses. No alternative has yet been identified, but Glass suspects other hosts are out there waiting to be found. "Odds are, *E. coli* and yeast are not the best platforms in nature for doing what we do," says Glass. "They're just the best ones that we know about for now."

The computational design of synthetic genomes also remains a problem. Electrical engineer Douglas Densmore at Boston University in Massachusetts has spent much of his career designing DNA-based circuits that involve interconnected genes and regulatory elements in much the same way that electronic circuits use wires, resistors and capacitors. But unlike electronic circuits, he says, "DNA has a much richer functional language in terms of protein production and binding sites and things like that."

To Densmore, the modular assembly enabled by systems such as Golden Gate reflects how capacitors and resistors are soldered to a circuit board. But the relative lack of standardized parts can complicate synthesis. "Design tools exist, but they stumble because the libraries aren't well defined or characterized or modular," he says. It also remains unclear how well these platforms might translate across species, because the rules of gene regulation might differ in subtle but important ways.

Groups can use tools such as the cloud-based

software Benchling to plan and simulate Golden Gate and Gibson Assembly reactions, Ellis says. But these tools struggle with the complexity of large-scale projects, so most of the planning for the chromosome and genome assembly efforts so far have relied on homemade software. In an effort to make such projects more systematic and predictable, Boeke announced the launch of a challenge to create software for computational genome design at the GP-write meeting in New York City in November, which will offer a 'significant cash prize' for the winner.

Services rendered

Densmore also sees automation as a valuable opportunity to make synthetic-genomics efforts more consistent and reproducible. "The best way to get people to use and adhere to design rules is to get manual experiments out of the manufacturing loop," he says. Much of the labour underlying DNA assembly is tedious and repetitive, and many groups find it worthwhile to assign these tasks to robots.

But building up in-house automation capacity has potential pitfalls — not least that the knowledge gained about the equipment is often not shared widely. "Lots of people have the experience where you buy it, your amazing postdoc uses it, your amazing postdoc leaves, and then no one knows how to turn it on," says Patron. A growing number of institutions have therefore established biofoundries — core facilities that academics can use for DNA synthesis and assembly. One example is the Earlham BIO Foundry, which Patron helped to launch in 2016.

There are now more than 20 such facilities worldwide, and some researchers expect that much of the field will come to rely on outsourcing. "In my imagination, in the next few years we will be able to design a piece of DNA with your computer, and a foundry will manufacture the DNA and deliver it to you the next day," says Cai, who co-founded the Edinburgh Genome Foundry at the University of Edinburgh, UK, in 2016.

With more hands on deck, large-scale genome manipulation could become a mainstream tool for the next generation of researchers — an engine for executing bold ideas rather than an end in itself. "I'm of the opinion that academics should do more risk-taking," says Cai. "Getting into the unknowns, instead of duplicating things companies can do — and do better."

Michael Eisenstein is a science writer in Philadelphia, Pennsylvania.

1. Fredens, J. et al. *Nature* **569**, 514–518 (2019).
2. Hutchison, C. A. et al. *Science* **351**, aad6253 (2016).
3. Palluk, S. et al. *Nature Biotechnol.* **36**, 645–650 (2018).
4. Engler, C., Kandzia, R. & Marillonnet, S. *PLoS ONE* **3**, e3647 (2008).
5. Shao, Y. et al. *Nature* **560**, 331–335 (2018).
6. Ostrov, N. et al. *Science* **353**, 819–822 (2016).
7. Ostrov, N. et al. *Science* **366**, 310–312 (2019).