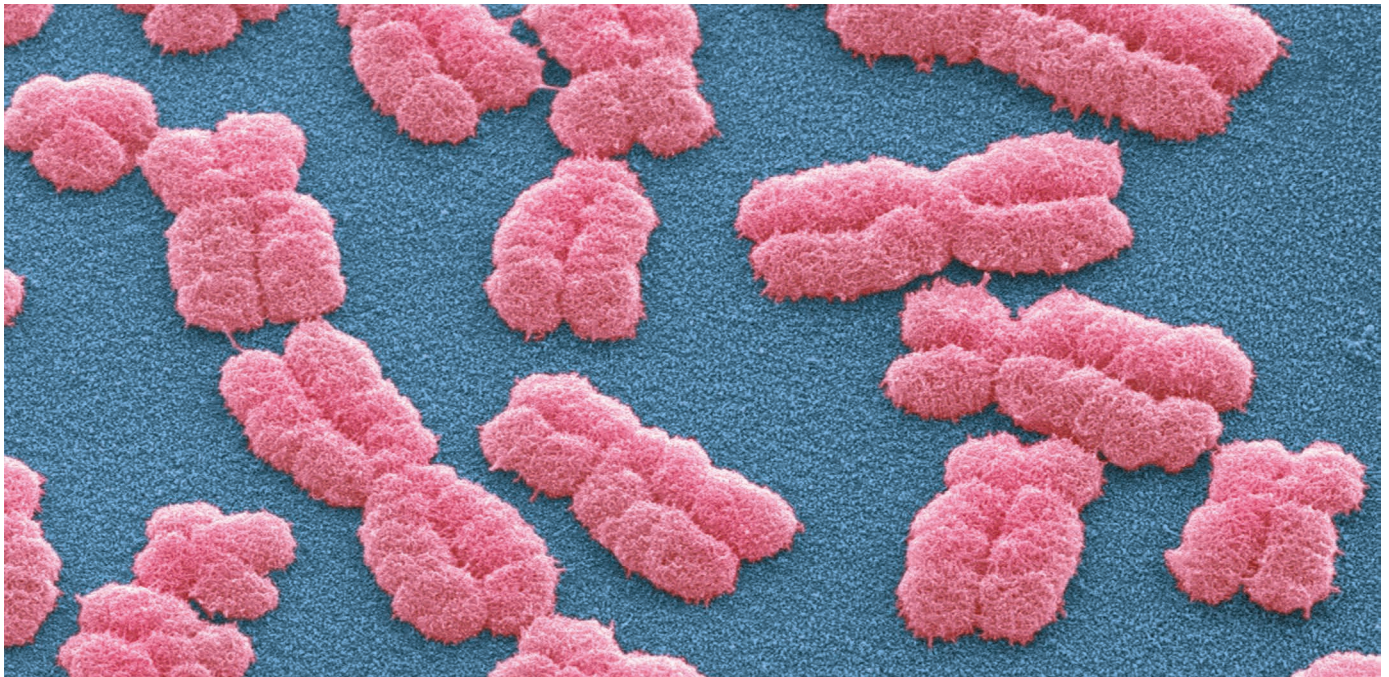


# HOW TO HACK THE GENOME

*After tackling the genomes of bacteria and yeast, synthetic biologists are setting their sights on rewriting those of more complex organisms, including humans.*

POWER AND SYRED/SCIENCE PHOTO LIBRARY



Human chromosomes under a scanning electron microscope.

BY JEFFREY M. PERKEL

Under typical laboratory conditions, strain JF1 of the bacterium *Escherichia coli* looks like any other — a spatter of yellow-tinged colonies on an amber agar plate. But bathe the colonies in wavelengths of red, green or blue light and their cells convert chemicals in the growth medium into pigments in a pattern that matches that of the coloured light to which they were exposed, yielding a muted and blurred image that is reminiscent of a 1970s Polaroid.

Christopher Voigt, whose lab at the Massachusetts Institute of Technology in Cambridge created the intricate genetic circuit that drives this transformation, reported in May 2017 that his team had used the system to recreate a multicoloured geometric illustration of lizards by Dutch artist M. C. Escher<sup>1</sup>. That exercise was just for fun, he says, and a way to demonstrate the state-of-the-art in synthetic biology. But it was not easy: the circuit contained 18 genes and 32 regulatory elements, spread over 4 small circular molecules of DNA known as plasmids, and 46,198 base pairs of DNA. It responds separately to red, green and

blue light. “When you add it all up, it’s quite a sophisticated project,” Voigt says.

And it’s not the only one. Synthetic biology is awash with projects of similar or even greater complexity. Improvements in techniques for synthesizing and editing DNA have brought reduced costs and enormous precision, helping biologists to build from scratch or re-engineer the genomes of microorganisms such as *E. coli* and brewer’s yeast (*Saccharomyces cerevisiae*). Synthetic-biology researchers are now having serious discussions about re-engineering the genomes of more-complex organisms, including humans, although substantial hurdles stand in the way. For instance, the manipulation of large pieces of DNA presents technical challenges, and despite the falling cost of DNA synthesis, the cost is still prohibitive when billions of bases must be rewritten.

“Results over the past two years have certainly increased my optimism that we may be able to do some really profound engineering in animals,” says Peter Carr, a synthetic biologist at the MIT Lincoln Laboratory in Lexington, Massachusetts.

In a 2015 letter<sup>2</sup> to the journal *Trends in Biotechnology*, Carr asked: “Is there a synthetic biology equivalent of the sound barrier, or of the speed of light?” The question was rhetorical because immutable limits clearly exist — the growth rate, for example, cannot be infinitely fast. But what constitutes a sound barrier in synthetic biology is evolving, he says. Designs that were on the cusp of feasibility a few years ago are now practical.

Researchers who once struggled to produce a few kilobases of synthetic DNA are now building whole genomes on the scale of megabases. In March 2016, sequencing and synthetic biology pioneer Craig Venter and his colleagues announced that they had pruned and rewritten the genome of the bacterium *Mycoplasma mycoides* from about 1 megabase to 531 kilobases to create a ‘minimal’ genome<sup>3</sup> — the smallest set of genes that is required for life.

In August 2016, researchers led by George Church, a geneticist at Harvard Medical School in Boston, Massachusetts, and Nili Ostrov, a postdoctoral researcher in his lab, reported that they had produced a bacterium, dubbed ‘rE.coli-57’, in which seven codons — the triplets of nucleotides that encode particular



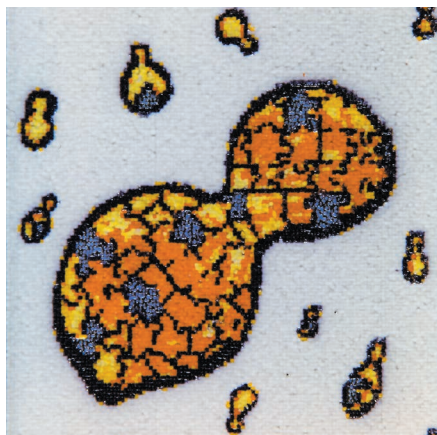
► amino acids — had been stripped out and replaced with synonymous alternatives<sup>4</sup> in a process known as genetic recoding. And in March 2017, a team led by Pamela Silver, a biochemist at the Wyss Institute for Biologically Inspired Engineering at Harvard University in Boston, Massachusetts, described its initial attempts to recode the genome of strain LT2 of the bacterium *Salmonella typhimurium*<sup>5</sup>, replacing around 200 kilobases of genomic DNA and eliminating a specific leucine codon in the hope of preventing the transfer of genes between pathogenic microbes.

Most dramatically, in March 2017, an international consortium led by Jef Boeke at the New York University Langone Medical Center and Joel Bader of the Johns Hopkins University in Baltimore, Maryland, reported the end-to-end rewrite of 5 of the 16 chromosomes of *S. cerevisiae*<sup>6</sup> — a milestone in an international project called the Synthetic Yeast Genome Project (Sc2.0). Sc2.0 aims to optimize and synthesize the complete genome of *S. cerevisiae* for both industrial and pure research applications. For instance, says Boeke, by removing all DNA sequences that do not encode proteins (introns), the team can assess the biological roles of the cellular machinery required to handle those genetic elements.

The artificial yeast chromosomes designed for Sc2.0 have been streamlined and stabilized by deleting repeated sequences and introns, by moving sequences that encode crucial pieces of the protein translational machinery to a dedicated chromosome, and by eliminating the codon TAG, which signals a stop in translation, and replacing it with an alternative stop codon, TAA, to facilitate protein engineering. A customized software package called BioStudio enabled the team to manage the genetic bookkeeping required to complete such a massive task.

The logistics of Sc2.0 were substantial, says Patrick Yizhi Cai at the University of Edinburgh, UK, who is the project's international coordinator. Yet the actual process of editing the yeast chromosomes was fairly routine, requiring just a few plates of yeast in the incubator, says Leslie Mitchell, a postdoc in Boeke's lab who led the subgroup that synthesized chromosome VI.

The Sc2.0 team uses a strategy called SwAP-In to rewrite chromosomes piece by piece (see 'Rebuilding a chromosome'). Researchers first assemble short single-stranded molecules of DNA known as oligonucleotides into building blocks of about 750 bases, and then into 'chunks' of 10 kilobases or fewer that, in turn, are combined into 'megachunks' of 30–60 kilobases. Each megachunk contains one of two marker genes that enable the selection of yeast carrying these genes — in this case, *URA3*, which enables yeast to grow in the absence of uracil, and



Yeast genetically engineered to produce pigments from bacteria, other fungi and plants.

*LEU2*, which enables growth when leucine is missing. The megachunks are then slotted into an existing chromosome through homologous recombination, a natural process in which one stretch of DNA is replaced with another, rewriting the DNA from one end to the other. As each subsequent segment is integrated, it replaces the marker gene of the previous segment, swapping the yeast cell's nutritional requirements between uracil and leucine. Following quantitative PCR analysis to ensure that each megachunk has been fully incorporated, the resulting yeast strain is tested for its ability to form colonies under relatively stringent conditions; its slowed growth or death in comparison to the wild type indicates a problem in need of repair. "The idea is: integrate the megachunk, test the fitness", and repeat, Mitchell explains.

#### WRITER'S BLOCK

A single megachunk can be integrated and tested in about two weeks, Mitchell says, assuming that there are no problems. Fitness testing and 'debugging' (error correction) "take longer than the actual build at this point", says Boeke.

Each chromosome completed so far presented only a handful of notable 'bugs', he says. Some stemmed from errors in genome annotation, whereas others were caused by codon replacements that, for instance, alter the secondary structure of RNA.

For the most part, the yeast rolled with the punches. Yet the glitches that did crop up hint at the challenges faced by a larger-scale engineering project called Genome Project-write (GP-write), which aims to rewrite the genomes of more-complex eukaryotes. Besides the obvious problem of scale — at 3 billion base pairs, the human genome is two orders of magnitude larger than that of *S. cerevisiae* — the genomes of more-complex organisms tend to be less well annotated. When Venter's team first tried to build a minimal genome in *M. mycoides*, it applied a rational design, using published genetic data to compile a list of essential genes — an approach that didn't work. "Our lack of basic biological knowledge, even with

the simplest bacterial genomes, is huge," Venter says. Success came instead from a top-down approach that whittled away the genome to arrive at a core set of 473 genes. But about one-third of those have no known function. "I found that to be just kind of a mind-blowing result," Silver says.

And there are further challenges. Sc2.0 and other genome-rewriting projects have tended to steer clear of the regulatory regions of genes, but in more-complex organisms such as eukaryotes, these are often located far from the genes that they influence, and might not yet be fully mapped. Researchers may therefore not know which segments to rewrite, and which to leave alone. It is also unclear how such large-scale genomic changes might affect chromatin architecture and therefore gene expression.

On a practical level, chromosome-sized molecules of DNA cannot be easily manipulated without being broken, and there is no efficient way to deliver them into most eukaryotic cells. Even if scientists can deliver the DNA, they might not be able to integrate it into the genome because most such cells are unable to perform homologous recombination as readily as yeast, and their slower growth drags out each experimental step.

There also is the cost of synthetic DNA to consider. Silver's team received funding from the US Defense Advanced Research Projects Agency for her work in *S. typhimurium*, which allowed it to negotiate a favourable price for DNA synthesis. But at a per-base price of US\$0.10, she says, it will cost more than \$1 million to complete her project; the human genome, by comparison, would cost hundreds of times more.

Yet Church says it is just a matter of time before technology catches up with ambition. "My guess is, it's going to get easier and easier with time to build large genomes."

#### PRECISION REWRITE

Genome rewrites so far have largely stuck to nature's recipe. But ultimately, biologists hope to impart new functions.

Several projects, including the rE.coli and *S. typhimurium* studies, are focusing on genetic recoding, in which codons removed from the genome are freed for other uses. Jason Chin, a synthetic biologist at the MRC Laboratory of Molecular Biology in Cambridge, UK, has done extensive work to manipulate the genetic code. He says that such recoding can advance protein engineering, not to mention the design, testing and synthesis of new chemical polymers built from monomers other than standard amino acids. Other possible applications include biocontainment (preventing release of an organism outside the lab) and genetic isolation (protecting organisms from viral infection).

In 2005, while working as a postdoc with Church, Farren Isaacs, now a bioengineer at Yale University in New Haven, Connecticut, began to pursue the idea of recoding the *E. coli*

genome, focusing his energy on replacing the stop codon TAG throughout.

Because *E. coli* contains just 321 TAG codons, Isaacs was able to accomplish this task by modifying an existing genome rather than synthesizing one from scratch<sup>7</sup>. Using a strategy called MAGE (multiplexed automated genome engineering) that enables multiple DNA sequences to be edited at once, Isaacs and his colleagues first divided the *E. coli* genome into 32 segments and altered the TAG codons in each to the synonymous codon TAA. Next, they joined the 32 modified segments into a single molecule by exploiting a natural process of genetic exchange between bacteria. To complete the recoding process, the team deleted a gene that encodes a protein known as RF1, which recognizes the codon TAG. (A related protein, RF2, recognizes the codon TAA.) Survival of the modified *E. coli* following removal of this otherwise essential gene showed that their recoding process had worked.

For many researchers, existing technologies provide all the power they need to hack the genome. eGenesis, located in Cambridge, Massachusetts, is using the gene-editing tool CRISPR to turn pigs into sources of transplantable organs. Luhan Yang, cofounder of the company, explains that the idea is to pare back the pig genome using CRISPR to remove sequences encoding proteins that might evoke an immune response in people. New genes encoding proteins that help to make the pig tissues compatible with humans can also be introduced. “We think dozens of modifications probably would suffice,” she says.

Yet the approach is different for more-expansive projects. Ostrov and Church’s rE.coli project, for instance, removed the TAG stop codon and two codons each for serine, arginine and leucine from *E. coli* to create a 57-codon strain<sup>4</sup>. That work required 62,214 changes, which the team made using bottom-up DNA synthesis rather than top-down editing. With so many necessary genetic modifications, Ostrov explains, “we might as well make the genome from scratch.”

None of these genome-hacking studies has actually built a chromosome-sized molecule of DNA in one continuous stretch. Most commercial suppliers of synthetic DNA rely on a decades-old method of synthesis that is unsuitable for producing molecules longer than about 200 nucleotides. Church’s team, as do most groups pursuing genome synthesis, assembled the DNA it required hierarchically. It purchased pre-made segments of genes that were 2–4 kilobases long, assembled them into 50-kilobase-long blocks in yeast using homologous recombination, and transferred these completed segments into *E. coli*. The team then deleted the corresponding region of the *E. coli* genome, and tested the resulting strain of bacteria for fitness.

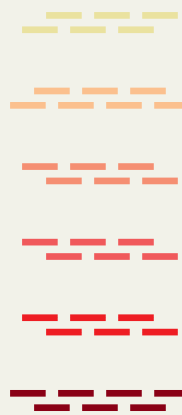
According to Ostrov, the recoding process went smoothly, despite a few bugs. For example,

## REBUILDING A CHROMOSOME

To rewrite yeast chromosome XII, the Synthetic Yeast Genome Project (Sc2.0) used a two-step process.

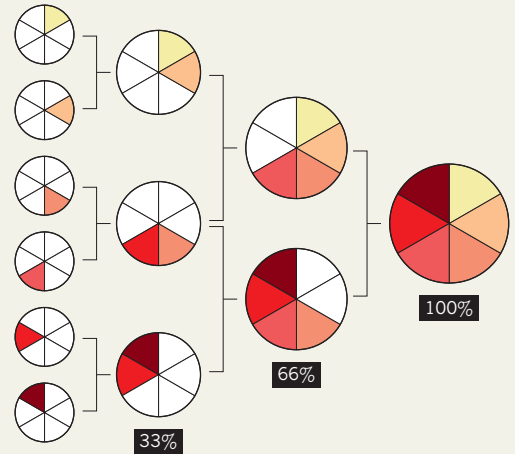
### Synthetic megachunks

Segments of synthetic DNA (coloured blocks) were inserted into the yeast genome to create six separate strains, each containing one-sixth of rewritten chromosome XII.



### Partial synthetic chromosome XII strains

Through a series of hierarchical mating steps, the segments in each partial strain were brought together to produce a single, fully synthetic chromosome XII.



altering the coding sequence of a particular gene inadvertently weakened the promoter of an overlapping gene, reducing the fitness of the strain.

“There’s idiosyncratic information in the genome,” explains Chin, and it can only be deciphered experimentally.

## CIRCUIT CITY

Other researchers are developing genetic circuitry to imbue genomes with new functionality.

Generally, these circuits — such as Voigt’s image-capturing strain of bacteria — are built up from simpler designs that use proteins called transcription factors as positive or negative input and output signals. Wilson Wong, a biomedical engineer at Boston University in Massachusetts, builds his designs instead using enzymes known as recombinases that invert or delete segments of DNA — a design strategy called BLADE (Boolean logic and arithmetic through DNA excision).

Wong says that BLADE frees researchers from the difficulty of linking circuits to one another, which requires the output strength of one circuit to be matched with the anticipated input of the next.

In one demonstration, Wong and his team created a Boolean logic look-up table<sup>8</sup> — a genetic circuit, about 10 kilobases long, that is capable of turning into any of 16 possible logic gates depending on whether 6 recombinases are present.

Wong’s team essentially designed that circuit using a pencil and paper. But ultimately, synthetic biologists hope to build their designs *in silico*. Voigt, working with Douglas Densmore, an electrical engineer at Boston University has developed a tool called Cello

(cellocad.org) to make that possible. Researchers specify genetic-circuit designs in a programming language called Verilog, and Cello produces the DNA sequences that are required to make them work<sup>9</sup>.

Despite their seeming simplicity, the genomes of microorganisms demonstrate an incredible capacity for subtle genetic control, Voigt says. “We’re almost taunted by what exists in nature.” But through a combination of genome editing, genome synthesis and cleverly designed tools, researchers are slowly rebalancing the scales. ■

**Jeffrey M. Perkel** is technology editor for Nature.

1. Fernandez-Rodriguez, J. *et al. Nature Chem. Biol.* **13**, 706–708 (2017).
2. Zakeri, B. & Carr, P. A. *Trends Biotechnol.* **33**, 57–58 (2015).
3. Hutchison, C. A. III *et al. Science* **351**, aad6253 (2016).
4. Ostrov, N. *et al. Science* **353**, 819–822 (2016).
5. Lau, Y. H. *et al. Nucleic Acids Res.* **45**, 6971–6980 (2017).
6. Richardson, S. M. *et al. Science* **355**, 1040–1044 (2017).
7. Isaacs, F. J. *et al. Science* **333**, 348–353 (2011).
8. Weinber, B. H. *et al. Nature Biotechnol.* **35**, 453–462 (2017).
9. Nielsen, A. A. K. *et al. Science* **352**, aac7341 (2016).

## CORRECTION

The Technology Feature ‘The architecture of structured DNA’ (*Nature* **546**, 687–689; 2017) erroneously credited Paul Rothemund for designing the experiments. In fact, the work was done by his postdoc, Ashwin Gopinath. Furthermore, the examples of shapes created by Shawn Douglas should have been protractors, twisted ribbons and an octahedral globe.