

TECHNOLOGY FEATURE

THE NEXT STEP FOR THE SYNTHETIC GENOME

Biologists have copied an existing genetic code, but haven't yet commercialized it or written their own. What will it take for a tour de force to reach industrial force?

H. WANG/CHURCH LAB/HARVARD UNIV.



Strains of *Escherichia coli* have been developed to produce lycopene, an antioxidant found in tomatoes.

BY MONYA BAKER

A year ago this week, headlines trumpeted that humans had created artificial life. Scientists at the J. Craig Venter Institute in Rockville, Maryland, had chemically synthesized DNA and placed it inside a bacterial cell emptied of its own genetic material. Tests a few days after the insertion showed that the 1-million-base-pair-long synthetic genome was able to run the cellular machinery¹.

Whole-genome engineering could one day create cells unbound by biochemistry as we know it, says George Church, a geneticist at Harvard Medical School in Boston, Massachusetts. Researchers might even be able to design a new genetic code, one that could incorporate more than the 20 or so amino acids used by natural living systems. That achievement is “going to be more than an increment”, says Church, “that’s going to be a game-changer”. But current reality is more prosaic. As Venter Institute staff celebrated their cell’s first birthday with a chocolate-and-spice layer cake topped by a miniature microscope made of sugar, they were well aware that the era of synthetic genomes still faces plenty of growing pains.

Breathless headlines notwithstanding, the Venter Institute team did not create life so much as copy an existing plan. In this case, they acted more like scribes than authors. Synthetic biologists are also working on changing DNA sequences — trying to engineer microbes for practical applications such as decontaminating toxic waste, tracking down tumours or secreting biofuels — but few work with more than ten genes at a time. The story of the field so far is, “can write DNA, nothing to say”, says Drew Endy, a synthetic biologist at Stanford University in California. “We can compile megabases of DNA, but no one is designing beyond the kilobase scale.”

“Most of us are still working on a small scale because there are interesting questions there and because that’s what we have the technology to build,” says James Collins, a biomedical engineer at Boston University in Massachusetts. ►

► “We frankly don’t understand biology well enough to start designing genomes *de novo*.”

Many technologies must fall into place before researchers will be able to routinely work with even tens of genes at a time. Putting together huge DNA molecules is time-consuming and expensive, and designing biological components to perform a particular task is a challenge for parts of genes, let alone whole genomes. Transplanting DNA molecules into cells is not easy, nor is getting the DNA to ‘boot up’ once it is in place. And because the genomes will be far from perfect, researchers will need ways to tweak and test many variants.

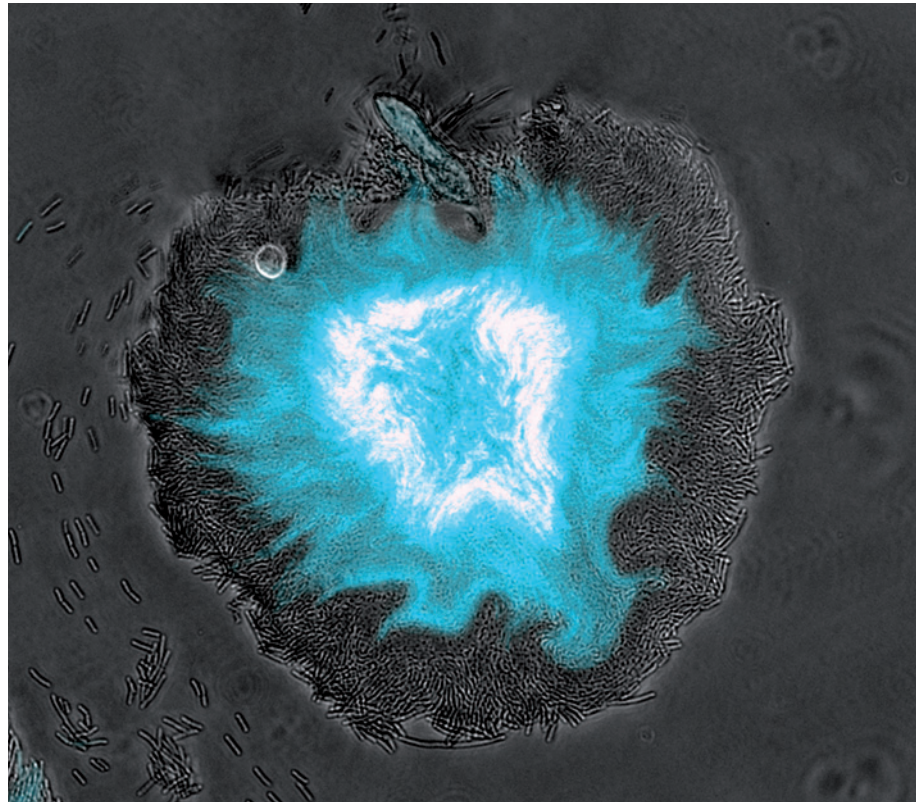
No one expects fast results, and much of the work will be tedious. The Venter Institute spent 15 years and US\$40 million creating the technology to build and transplant a genome. The 2010 paper lists two dozen authors. “This was a debugging process from the beginning,” says Craig Venter, founder of the institute. “99% of our experiments failed.” And failed experiments were costly: a single error in a million base pairs set the project back months. Not counting scientists and their equipment, four species were involved in the genome transplant: *Mycoplasma mycoides* to provide the source code, *Escherichia coli* to copy DNA pieces, baker’s yeast (*Saccharomyces cerevisiae*) to assemble them into a million-base-pair circle and *Mycoplasma capricolum* to provide the recipient shell. No wonder more synthetic biologists are thinking about parts of genes than are dreaming of constructing whole genomes.

LEARNING TO WRITE GENOMES

Synthetic biology often adopts the language of engineers: rather than talking about genes, networks and biosynthetic pathways, practitioners prefer to talk about parts, devices and modules. ‘Parts’ refer to the protein-coding section of a gene and sundry regulatory sequences that tune gene expression. A ‘device’ is an assembly of parts that together perform a particular function, often turning a protein’s production on or off. And a ‘module’ or pathway is a collection of devices that carry out more-complex functions, such as coordinating a chemical synthesis or shunting cells between ‘growth’ and ‘production’ modes.

Jeff Hasty, a bioengineer at the University of California, San Diego, used three genes to make bacteria light up in sync². Each gene is indirectly activated by the same small molecule: one controls the production of the molecule, another directs its degradation and the third makes the fluorescent protein that causes the cell to flash. The molecule diffuses between cells and coordinates bursts of protein production.

Another example of bioengineering involves a dozen or so genes from multiple species. Jay Keasling, a biologist at the University of California, Berkeley, engineered *E. coli* and yeast cells to make a precursor of the malaria drug artemisinin at one-tenth of the cost incurred



Engineered bacteria produce flashes of fluorescence, controlled by three genes in a synchronized circuit.

by the conventional method of production: extracting the natural product from sweet wormwood³. (More importantly, microbes grow faster than the plants, which are in limited supply.) Sanofi-aventis in Paris and the Institute of One World Health in South San Francisco, California, plan to start distributing the synthetic form of artemisinin next year.

But Keasling’s achievement is an object lesson in the time and expense involved. He and his colleagues began work a decade ago and had a \$43-million grant from the Bill & Melinda Gates Foundation in Seattle, Washington, to Keasling’s lab and to Amyris Biotechnologies in Emeryville, California, a company that Keasling co-founded in 2003. Researchers had to track down a previously unidentified enzyme, and engineered a dozen further yeast enzymes not just to work in *E. coli*, but also to operate at the right levels to move chemical intermediates towards a desired product without poisoning the cell or wasting resources.

To speed up such projects, the Massachusetts Institute of Technology (MIT) in Cambridge maintains a Registry of Standard Biological Parts (<http://partsregistry.org>) that lists thousands of components. However, descriptions of these parts are often incomplete, and they don’t all work as described.

To address this issue, Endy and bioengineers from the University of California, Berkeley, launched the International Open Facility Advancing Biotechnology (BioFab) in Emeryville in 2009, with a grant from the National Science Foundation. The BioFab aims to boost the

supply of working parts, both by optimizing the parts themselves and by developing systems to swiftly design genetic constructs. The goal, says Endy, is to create a set of genetic regulatory elements to precisely control the rates and levels of protein production. The BioFab currently provides 350 promoters, grouped into ten levels of protein production. Having a range of options is important, says Endy, because using the same sequences multiple times makes genetic constructs unstable. The team is assessing how these elements behave in systems and under different *E. coli* growth conditions.

Eventually, the researchers hope to create vast libraries combining variants of different parts. This will let them compare the parts’ performances, and pick the best ones. Computer analysis will then be used to model how different sequences affect gene expression, which can in turn predict how new combinations of parts will function. But the *in silico* design process can go only so far. “Models are not yet as predictive as they could be,” says Adam Arkin, a bioengineer at Berkeley and co-director of the BioFab. “In almost all cases of real application we are faced with some tinkering,” he adds. And the more parts are combined, the more unpredictable are the results.

SOME ASSEMBLY REQUIRED

The biological parts are generally easy to come by — short stretches of DNA can be ordered from a variety of companies (see ‘Making DNA on the cheap’) — but physically assembling multiple parts can be cumbersome and

Making DNA on the cheap

The description of the first chemical synthesis of a gene took up an entire issue of the *Journal of Molecular Biology* in December 1972. Then, copying an oligonucleotide of just 20 base pairs took a chemist two years¹¹. Now, researchers can order oligonucleotides 50–200 base pairs long from several vendors, for shipment within 24 hours.

“We ship tens of thousands of oligonucleotides every single day. That didn’t happen 20 years ago,” says John Havens, vice-president of business development at Integrated DNA Technologies in Coralville, Iowa, which supplied the oligonucleotides used by the J. Craig Venter Institute in Rockville, Maryland, to assemble the mouse mitochondrial genome⁵.

For an extra cost, oligonucleotides can be assembled into synthetic genes up to 1,500 base pairs long, which are stitched into plasmids — circles of DNA that can be propagated in bacteria. Companies including Integrated DNA Technologies, GeneArt in Regensburg, Germany, OriGene in Rockville, Maryland, and DNA2.0 in Menlo Park, California, can make genes, check for errors and ship them to clients in two weeks or less.

Prices have fallen, but much more slowly than those of multiplex gene sequencing (see ‘Reading and writing DNA’). Ten years ago, synthesis cost US\$25 per base and sequencing cost \$0.25, according to data compiled by Rob Carlson, a principal at Biodesic, a consulting firm in Seattle, Washington. Last August, the figures were \$0.35 per base for synthesis and \$0.00000317 for sequencing. Synthesis at prices comparable to sequencing would greatly accelerate synthetic biology, says Christopher Anderson, a genetic engineer at the University of California, Berkeley.

Cheaper DNA should be coming.

Oligonucleotides have conventionally been synthesized individually on glass-column supports, but now microarrays allow thousands of molecules to be assembled side-by-side; microfluidics, microelectrodes or even tiny beams of light can be programmed to direct the desired chemical reaction to each spot on the microarray. LC Sciences in Houston, Texas, MYcroarray in Ann Arbor, Michigan, and CustomArray in Mukilteo, Washington, all offer oligonucleotide libraries synthesized on microarrays; so does Agilent Technologies in Santa Clara, California, but only to a limited number of collaborators.

So far, oligonucleotides produced on microarrays have been difficult to use for gene synthesis: only very small amounts of each DNA sequence are made; thousands of sequences are mixed together and hard to separate; and the error rate is usually far higher than that of conventional techniques.

Work by George Church, a geneticist at Harvard Medical School in Boston, Massachusetts, and Jingdong Tian, a synthetic-systems biologist at Duke University in Durham, North Carolina, may help to change that. Tian designed a microarray in which groups of 60-mer oligonucleotides are synthesized in small wells, facilitating their assembly into genes. These microarrays could make many variants of the same gene, which could be quickly assessed for desirable properties, such as high levels of protein expression¹². Currently, the costs of gene synthesis can become prohibitive for such experiments.

Separately, Church and colleagues at Agilent, the Wyss Institute for Biological Engineering in Boston and Stanford University in California took another approach to assembling genes from a microarray. By carefully designing the oligonucleotides and PCR primers that target them, they could selectively amplify the oligonucleotides necessary for genes¹³. This method successfully made 40 out of an attempted 42 genes for single-chain antibodies, a particularly challenging target.

Achieving that kind of performance on a commercial scale will be difficult, but the Wyss Institute is testing the waters. Later this year, it plans to use the technique to launch a service that will construct genes for as little as \$10 for a 500-base-pair piece of DNA. “We are interested in what the scientific community can do with very cheap, though imperfect, sources of DNA,” says Church. **M.B.**

expensive. DNA molecules are either designed using complementary DNA sequences or mixed in with DNA complementary to the opposing ends of the molecules that are to be joined. These are combined with enzymes that cut and join DNA. Researchers can link elements using a system called BioBricks, in which sequences are cut out of circular genetic elements called plasmids by restriction enzymes specific to a particular series of nucleotides at the start and end of the sequence. The desirable parts are then stitched together into larger plasmids by other enzymes. (New England BioLabs in Ipswich, Massachusetts, sells a kit of the necessary enzymes and buffers.) Assembled sequences can then be replicated in bacteria.

Each assembled DNA piece starts and ends with the same sequences as the component parts, theoretically allowing larger and larger components to be assembled sequentially. But only three elements can be put together in a single reaction, which generally takes a couple of days. Reactions are also less successful with longer molecules, discouraging long assemblies.

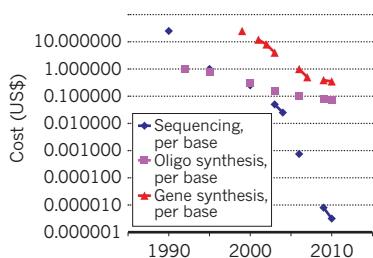
Tom Knight, a computer scientist and co-founder of start-up company Gingko BioWorks in Boston, invented BioBricks while working at MIT and has redesigned the system for industrial applications. The proprietary version can assemble up to ten parts in a single reaction, says company co-founder Barry Canton. This allows researchers to work on DNA molecules with as many as 100,000 base pairs, although most of the pathways that Gingko is working on are half that size. Just as importantly, most assembly steps can be performed by liquid-handling robots. For example, rather than bands of DNA being isolated from a gel, as in most methods, DNA molecules are collected onto and separated by suspended magnetic beads. Such automation speeds assembly and frees up lab technicians for more complicated tasks.

But BioBricks-type methods are limited by their use of restriction enzymes. Because the enzymes cut DNA whenever they encounter a particular series of nucleotides, there are ‘forbidden sequences’ that must be excluded from the genetic construct to avoid errant cutting. The larger a construct becomes, the harder it is to avoid such sequences. To circumvent this problem, researchers have developed assembly ‘overlap’ methods, in which opposite ends of molecules are joined as DNA is copied. Dozens of separate pieces of DNA can be assembled in the same reaction, often totalling a few thousand nucleotides. These methods have their own drawbacks, however. Most copy DNA using the polymerase chain reaction (PCR), which can introduce errors.

There is a bewildering array of overlap assembly techniques. ‘Gibson assembly’, invented by Daniel Gibson and his colleagues at the Venter Institute, allows many sequences to

READING AND WRITING DNA

The price of sequencing has fallen faster than that of making genes and oligonucleotides.



be assembled in parallel, and can even stitch together entire genomes⁴. In one demonstration, the team started with six hundred '60-mers' (oligonucleotides 60 base pairs long), and went on to assemble the 16.3-kilobase mouse mitochondrial genome⁵.

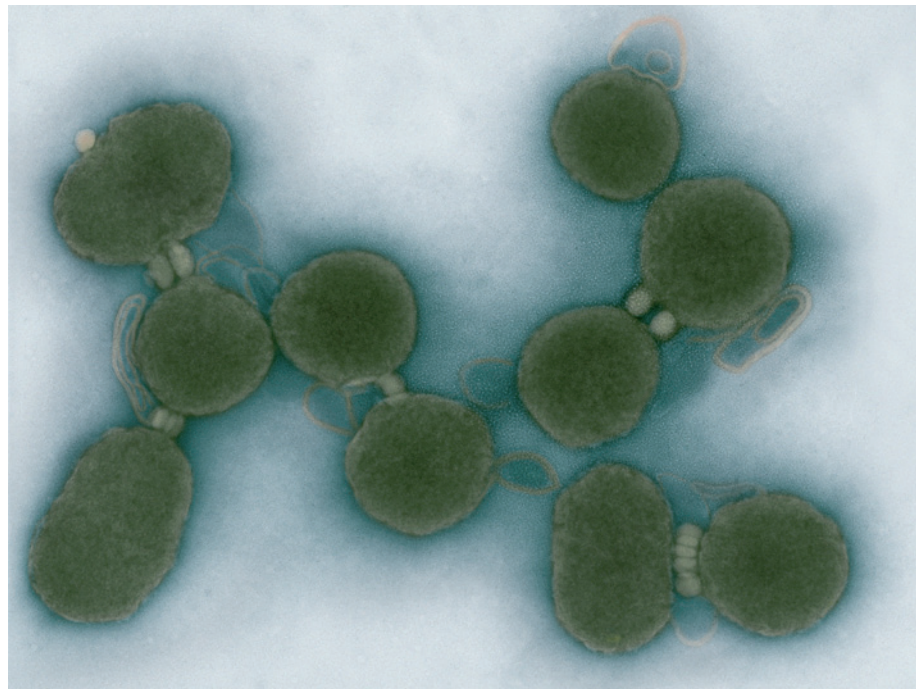
Other methods include Golden Gate Shuffling, sequence- and ligation-independent cloning (SLIC), splicing by overlapping extension (SOEing), enzymatic inverse PCR (EIPCR), overlap extension and more⁶. Some commercial kits are available: In-Fusion, from Clontech in Mountain View, California, has a mix of enzymes that can assemble 15-base-pair overlaps of any desired sequence. Life Technologies in Carlsbad, California, sells a plasmid-construction kit, MultiSite Gateway, that can join molecules with specific overlap sequences; it also markets the GeneArt High-Order Genetic Assembly System, which can assemble 10 DNA molecules, totalling up to 110 kilobases.

Researchers also design their own assembly reactions. To help this, the Joint BioEnergy Institute in Berkeley has invented a design tool, dubbed j5, that let researchers work with several DNA assembly protocols. It determines which overlap sequences to use, recommends the sequences to order from vendors and can instruct liquid-handling robots. Synthetic Genomics in La Jolla, California, which was co-founded by Venter, plans to start offering fee-for-assembly services later this year.

Assemblies larger than about 100 kilobases may be best put together inside cells, because big DNA molecules are fragile and difficult to manipulate. *In vitro* replication is also less accurate than cells' machinery. The Venter Institute team managed to assemble a 583-kilobase genome *in vitro*⁷, but it ultimately developed an *in vivo* assembly system for its synthetic genome.

Larger genomes than that of *M. mycoides* have been assembled inside cells, albeit not from synthetic starting points. In 2005, Mitsuhiro Itaya, a biochemist now at Keio University in Tsuroka, Japan, and his colleagues constructed a 3,500-kilobase cyanobacterium genome⁸. They cut the genome of the bacterium *Synechocystis* PCC6803 into large chunks and propagated them in specially prepared plasmids in *E. coli*. The plasmids were then transferred into a third species, *Bacillus subtilis*, where the DNA was stitched together.

Assembly methods aren't interchangeable. Overlap sequences that work for one method



Bacterial cells carrying a synthetic genome can grow and divide like normal cells.

often don't work for others, so researchers who run into problems with one technique have to start from scratch, says Tom Ellis, a synthetic biologist at Imperial College London. Ellis is working with Geoff Baldwin, a biochemist also at Imperial, and other colleagues to develop rules to find out which sequences will work with multiple overlap techniques, including recombination in yeast and *Bacillus*. That way, if one technique doesn't work, researchers can try others quickly.

These standards will also allow researchers to assemble DNA pieces in any order, says Ellis. Although a dictated order of assembly is fine for copying an existing genome, it does not let synthetic biologists test multiple possibilities. That issue is going to become more important as researchers move from working with thousands of base pairs to tens of thousands (see 'Sizing up synthetic DNA'). If researchers start building genomes or even large parts of genomes, they will have to think about how the DNA will wrap up on itself, and how they can place genes in chromosomes so that they end up in the right places, says Ellis. "It's a whole other aspect we'll have to uncover if we're going to do genome engineering."

EDITING IS ESSENTIAL

Jef Boeke, a molecular biologist at Johns Hopkins Medical Institute in Baltimore, Maryland, believes that genome-scale engineering is coming more quickly than many think. He is building artificial yeast chromosomes, each about the same size as the *M. mycoides* genome. Although he hasn't yet been able to design an entire new genome, he has developed techniques to make systematic alterations in existing genetic codes. "It opens the door to a lot

of imaginative change at the genome scale that wasn't possible before," he says. For example, one systematic study in 2008 deleted introns (regions within genes that don't code for protein) from many yeast genes individually, and found that the procedure had surprisingly little effect on the growth and fitness of cells⁹. Boeke wants to use his techniques to find out what will happen if all introns are removed from the genome at once.

But new possibilities introduce new problems. For the next few years, large genome assemblies are going to take months to build. With every assembly, researchers will detect unanticipated errors or realize after the fact that another sequence should work better, predicts Ellis. Then they will need to decide whether to assemble the whole genome again, or just edit it. "There has not been widespread acknowledgement in the synthetic-biology community that this is going to be an issue as we go into bigger assemblies," he says. The problem has already made itself felt: a quotation that the Venter Institute had incorporated into its synthetic genome turned out to contain a mistake, and is going to be altered.

Another use of editing is to produce and compare many gene variants. In a colourful demonstration in 2009, Church and his colleagues described a high-throughput editing system. Multiplex-automated genome engineering (MAGE) mixes bacteria with synthesized stretches of DNA that are designed to target many areas in the genome; carefully timed jolts of electricity cause the bacteria to take up the DNA as they grow in culture. Church used MAGE to alter 24 genes in *E. coli* at once, focusing on those involved in making lycopene, an antioxidant and pigment found



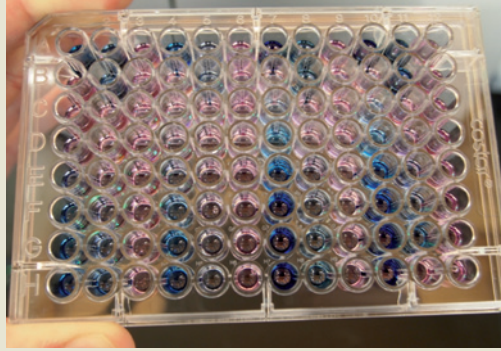
"It's quite likely that transplantation will be the unique step for each species."

Craig Venter

The useful genome

In the short term, advances in genome assembly are not likely to matter much commercially, says Rob Carlson, a principal at the engineering, consulting and design company Biodesic in Seattle, Washington. Getting the genes right is only part of the challenge; finding the best conditions for growing microbes and extracting their products is just as important. "People will build things that are as complex as can be understood and commercialized," says Carlson. Amyris Biotechnologies of Emeryville, California, is making biofuels by working with about a dozen genes, not a million-base-pair genome. "It's true that you can build something that size; it's not true that you would do that in a circumstance where you have to worry about the bottom line," says Carlson. "The problem is in getting stuff to work with the design and tinkering, rather than the assembly."

Amyris, Easel Biotechnologies in Costa Mesa, California, Gevo in Englewood,



Genetic assemblies are screened for desired effects.

Colorado, Joule Unlimited in Cambridge, Massachusetts, and LS9 in South San Francisco, California, are making biofuels and petroleum extracts from yeast, blue-green algae, *Escherichia coli* and other microbes. These efforts are backed by high-profile scientists and venture capitalists. Synthetic Genomics in La Jolla, California, has a US\$300-million partnership with

ExxonMobil and plans to use engineered microbes to make clean water, fuel and vaccines. Amyris, which went public last year, is now valued at more than \$1 billion. One company that folded, Codon Devices of Cambridge, Massachusetts, had focused not on making a product but on supplying services: providing synthetic genes to order and helping companies to develop synthetic-biology applications.

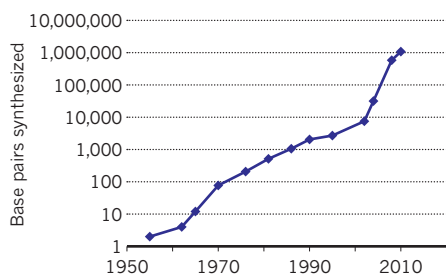
Gingko BioWorks in Boston, Massachusetts, originally planned to offer services for large DNA assemblies, then decided to focus on engineering the microbes themselves. Business partners and investors are less interested in new ways to assemble DNA than in better ways to manufacture products, says Barry Canton, the company's co-founder. "Companies may be shifting from a chemical synthesis platform to a biosynthesis platform," he adds. Those who control the microbes control the means of production. **M.B.**

in tomatoes. Within three days, some bacterial cells were making five times more of the red stuff than cells in the starting population¹⁰. The need for custom equipment and the difficulty of purifying transformed cells has kept researchers from widely adopting the technique, but the sheer number of genetic possibilities that can be tested using MAGE is a huge advantage, says Church. As many as 4 billion *E. coli* genomes were produced in the course of one experiment. "You're not resting on the outcome of one construct," he adds.

Mutagenesis and directed evolution of existing genomes could also help synthetic biologists to make up for current gaps in knowledge, says Collins. As more genes are brought into the system, he says, "uncertainty goes up exponentially, and you run up against the limits of what you can do modelling-wise". And although computational approaches are not yet sophisticated enough to design new genomes, they are

SIZING UP SYNTHETIC DNA

Artificial DNA has grown from two-nucleotide molecules in 1955 to more than one million in 2010.



good at modelling existing ones, he says. This understanding could help researchers to co-opt existing cellular networks to perform desirable tasks. "We are starting to see labs recognize that there is a lot to be exploited inside the cell," says Collins (see 'The useful genome').

BIOLOGY MATTERS

The most difficult problem may well be one of the least discussed: putting the genome to work. Although Itaya has synthesized large genomes inside cells, the introduced genomes do not go on to produce proteins. Venter's group had originally chosen *Mycoplasma genitalium* for the synthesis project because its genome was, at the time, the smallest known: only 583 kilobases. But *M. genitalium* grows so slowly that the team switched to its faster-growing cousin, even though its genome is twice the size. Making the DNA is not the rate-limiting step, says Venter. "It's much more dealing with the complexity of biology versus the chemical synthesis," he says.

In fact, Venter thinks that adapting genomes to work in different cell types may be one of the most difficult tasks. The creation of the first synthetic cell is illustrative: the team had to remove certain enzymes from recipient cells to keep them from cutting up the foreign DNA. And moving to other species is going to be even more difficult. Unlike *Mycoplasma*, many microbes contain tough cell walls that resist the introduction of DNA. "It's quite likely that transplantation will be the unique step for each species," says Venter.

Like a child learning to write, researchers must be able to copy natural genomes before they can create new ones. One day, geneticists will be able to design code on large scales, fueling as-yet-undreamed-of applications, says Venter. "After we sequenced the genome, analysts were arguing that there was no more need for sequencing, and I argued that this was the starting point." The question of whether whole-genome synthesis will be useful will prove foolish in time, Venter believes. "It's like asking, 'why would you want to invent an airplane when people already have horses?'" ■

Monya Baker is technology editor for *Nature* and *Nature Methods*.

- Gibson, D. G. *et al. Science* **329**, 52–56 (2010).
- Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. *Nature* **463**, 326–330 (2010).
- Ro, D.-K. *et al. Nature* **440**, 940–943 (2006).
- Gibson, D. G. *et al. Nature Meth.* **6**, 343–345 (2009).
- Gibson, D. G., Smith, H. O., Hutchison, C. A., Venter, J. C. & Merryman, C. *Nature Meth.* **7**, 901–903 (2010).
- Ellis, T., Adie, T. & Baldwin, G. S. *Integr. Biol.* **3**, 109–118 (2011).
- Gibson, D. G. *et al. Proc. Natl Acad. Sci. USA* **105**, 20404–20409 (2008).
- Itaya, M., Tsuge, K., Koizumi, M. & Fujita, K. *Proc. Natl Acad. Sci. USA* **102**, 15971–15976 (2005).
- Parenteau, J. *et al. Mol. Biol. Cell* **19**, 1932–1941 (2008).
- Wang, H. H. *et al. Nature* **460**, 894–898 (2009).
- Caruthers, M. H. *Biochem. Soc. Trans.* **39**, 575–580 (2011).
- Quan, J. *et al. Nature Biotechnol.* doi:10.1038/nbt.1847 (2011).
- Kosuri, S. *et al. Nature Biotechnol.* **28**, 1295–1299 (2010).