

# Pursuing the simple life

Michael Eisenstein

Efforts to pare away cellular genomes are yielding streamlined biosynthetic factories and deeper insights into the core processes of biology.

There are countless tales of hermits who have discarded the distractions of the material world, shedding all but the barest necessities of survival—down to just the shirt on their back. Through this austerity, they eventually gain deep wisdom and enlightenment. Scientific wisdom is more expensive than its philosophical counterpart, but researchers around the world are nevertheless embracing similar principles of enlightenment through simplicity as they delve into the core machinery of cellular life, in search of functions that are truly ‘essential’.

Evolution does not follow a straight path, and cellular genomes are cluttered with redundancies, hyperspecialized functions, and sometimes even outright junk. Researchers have been exploring the concept of bare-essentials ‘minimal genomes’ for decades. For example, in 1996, NIH researchers Arcady Mushegian and Eugene Koonin predicted that a set of just 256 conserved genes may be sufficient to sustain the microbe *Mycoplasma genitalium*<sup>1</sup>. Such minimal genomes could offer critical insights into the core processes that define life and the early stages of cellular evolution. In parallel, synthetic biologists are intrigued by the prospect of building predictable, streamlined cellular ‘chassis’ for biomanufacturing, where all the components are well known and understood.

Confidently determining which genes are essential in a given organism has proven a considerable challenge, but scientists are making headway. “This question has been around for a long while, but now there are the technologies for either reducing or synthesizing genomes chemically from DNA sequences that allow you to tackle it,” says Victor de Lorenzo of the Centro Nacional de Biotecnología in Madrid, Spain.

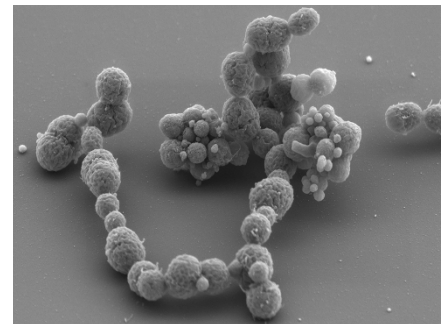
## Trimming the fat

Efforts at genome reduction can be lumped into two broad categories—‘top-down’ removal of putative nonessential elements and ‘bottom-up’ assembly of essential components into a synthetic genome. Most groups active in this field have pursued the former approach, chipping away at the genomes of workhorse microbes like *Escherichia coli* and *Bacillus subtilis*.

Some sequences are obviously dispensable, artifacts of millions of years of evolution. György Pósfai of the Hungarian Academy of Sciences identified such elements in *E. coli* by comparing genome sequences generated from multiple strains by his collaborator Frederick Blattner. “There were specific genomic islands that were unique to particular strains,” says Pósfai. “These mostly contain parasitic DNA or prophage or insertion sequences, which are really not important for the basic machinery of the cell.”

The next stage of deletion typically targets genes that may be helpful for real-world survival but have little value in the laboratory. “You have all of these signal transduction mechanisms that produce biofilms or contribute to motility or sense compounds, and they’re all completely irrelevant when you have a cell that you want to program as a catalyst in a reactor,” says de Lorenzo, whose group is streamlining the genome of *Pseudomonas putida*. In many cases, he notes, such deletions can free up considerable metabolic resources in the cell.

As the go-to bacteria for many



Scanning electron micrograph of JCVI-syn3.0 bacteria.

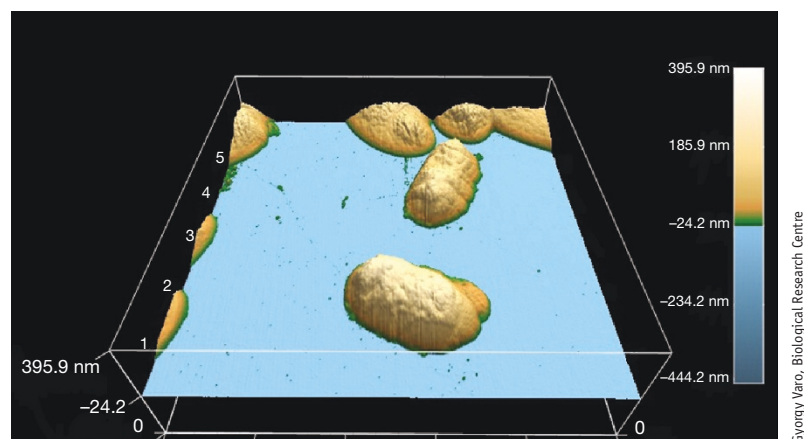
Tom Deerinck, University of California at San Diego  
National Center for Imaging and Microscopy Research  
(NCMIR)



György Pósfai

laboratories, most genome manipulation tools have been developed specifically for *E. coli*. One of the most powerful tools is a technique developed by George Church’s group at Harvard University known as multiplex automated genome engineering (MAGE)<sup>2</sup>. MAGE combines a bacteriophage-derived recombinase enzyme with oligonucleotide sequences targeted to specific sites in the genome, allowing researchers to rapidly introduce numerous alterations in a high-throughput fashion. The site-specific genome-editing capabilities of CRISPR–Cas9 have also been adapted for *E. coli*. “We’ve used MAGE and CRISPR to basically eliminate all the insertion sequences genome wide in a single step,” says Pósfai. Unfortunately, these tools are not instantly transferrable across species—for example, the recombinase used in MAGE does not function efficiently in *B. subtilis*. “Classical genetic techniques are much easier and more straightforward with *Bacillus* than MAGE,” says Jörg Stülke of Germany’s Göttingen University, who is coordinating the multi-institutional ‘MiniBacillus’ effort<sup>3</sup>.

Their work has yielded considerable progress in bacterial genome streamlining. “We are now at a 42% level of genome reduction,



AFM image of the reduced-genome *E. coli* strain MDS42.

and I am quite confident that within reasonable time, we will reach 50%,” says Stülke. “The dream would be to go down to about 500 or 550 genes, but this might not be realistic.” He notes that other efforts to minimize *B. subtilis*, whose genome normally contains an estimated 4,100 protein-coding genes, ran into roadblocks after a 30% genome reduction, but his team has identified additional genome elements that must be preserved to enable the streamlined cell to function. In *E. coli*, Pósfai has generated numerous different deletion strains; the most heavily studied, MDS42, has shed roughly 15% of the genome<sup>4</sup>. However, his group has since tested additional deletion strains, and other groups have reported deletions of up to 30%.

### Basic bugs

After more than 16 years of effort, the J. Craig Venter Institute (JCVI) in La Jolla, California recently described the first true example of ‘bottom-up’ redesign of a microbial genome<sup>5</sup>. The project was born out of a comprehensive analysis of *M. genitalium*, the same microbe assessed by Koonin and Mushegian in 1996. This bacteria is considered a naturally ‘minimal’ organism, as it relies heavily on its human host for survival and has the smallest cellular genome identified to date.

In 1999, the JCVI team demonstrated the resilience of the *M. genitalium* genome against random disruption with self-inserting DNA elements known as transposons. “It had a lot of genes that were not necessary for growth in the lab,” says Clyde Hutchison III, the lead author on that study. “That meant you should be able to have a cell with a genome even smaller than that.” JCVI researchers subsequently developed numerous tools that enabled them to test how many of these genes could be disposed

of in parallel. One of the most critical was a strategy for assembling large fragments of synthetic DNA into even longer, megabase-scale chromosomes by transferring them into yeast and exploiting that organism’s natural DNA recombination machinery. Venter’s team subsequently demonstrated that they could safely transplant such newly assembled chromosomes into a genome-free bacterial host—essentially creating a fully ‘synthetic’ cell.

They eventually shifted their efforts from *M. genitalium* to its close relative, *Mycoplasma mycoides*, which has a faster doubling time and is thus more amenable to laboratory work. After making initial predictions of the minimal genome based on previously published work, the Venter team performed repeated rounds of transposon-mediated mutagenesis to empirically hone in on which genes were strictly essential, as well as genes that were ‘nonessential’ but nevertheless beneficial to robust growth. At each stage, they tested different combinations of rewritten and unmodified genome fragments to see how well their predictions had worked out. “It was a pretty good design right off the bat,” says John Glass, leader of the JCVI synthetic biology group. “Working with just ‘textbook knowledge’, the design was about 70% accurate as far as what genes were needed...and when we got the competitive growth assays going and really did the transposon bombardment, about 95% of the genes we determined to be essential were correct.”

The final organism, dubbed JCVI-syn3.0, contained an *M. mycoides* genome reduced by nearly half—from over a million base pairs containing 915 genes to 531,000 base pairs and 473 genes<sup>5</sup>. These deletions slowed the bacterium’s doubling time and notably

altered both cell shape and colony-formation behavior, but the cells remained viable and healthy. JCVI-syn3.0 may still retain some extraneous elements, but Hutchison is hesitant about pursuing further large-scale deletions, as these may yield diminishing returns in the way of biological insight. “When you start to decrease the growth rate of the organism past a certain point, it just becomes intractable to work with in a practical sense,” he says.

Critically, all of the DNA used in this process was manufactured rather than cloned. Although this has not historically been cost effective, the steadily falling cost of DNA synthesis promises to spare researchers the headache of devising specialized strategies for introducing serial targeted deletions in their species of interest. “Sooner rather than later, the answers to all of these questions about genome editing will rely on direct DNA synthesis,” says de Lorenzo.

### Brewing better yeast

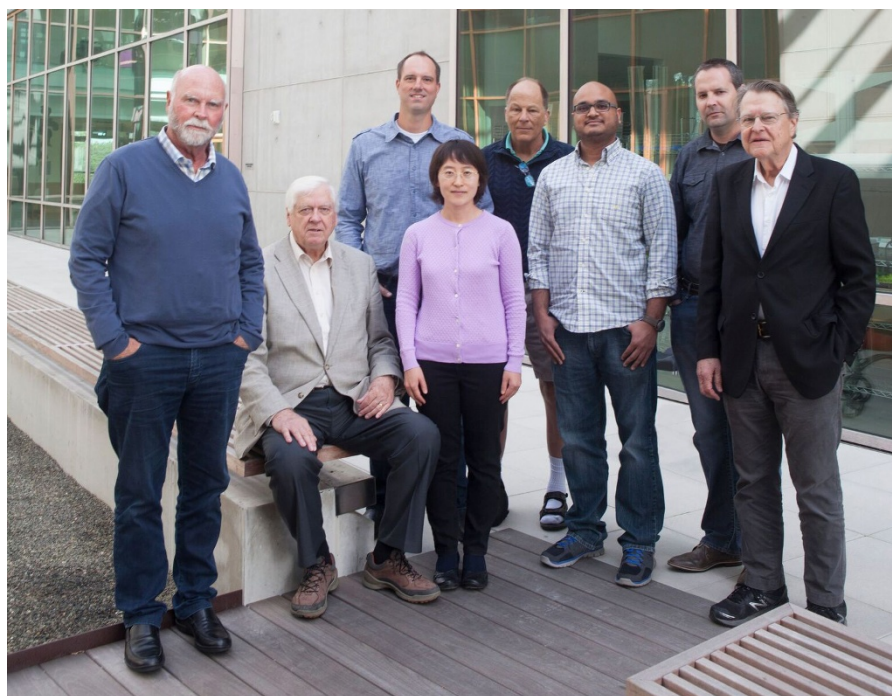
A handful of ambitious programs are now aspiring to similar feats in eukaryotic yeast cells. Furthest along is the *Sc2.0 initiative*, which originated from an effort by Jef Boeke at the New York University Langone Medical College and Srinivasan Chandrasegaran of the Johns Hopkins School of Public Health to build the first designer eukaryotic chromosome. *Sc2.0* aims to rebuild heavily engineered versions of all 16 chromosomes from *Saccharomyces cerevisiae*. In 2014, *Sc2.0* reached a major milestone with the completion of the first synthetic yeast chromosome<sup>6</sup>, *synIII*. “We totally redesigned it,” says Chandrasegaran of the Johns Hopkins School of Public Health, who coordinated the *synIII* team. “We removed the TAG stop codons, we removed all the transposons and subtelomeric sequences, and we put in artificial, universal telomeres.” The researchers also removed genes encoding transfer RNAs (tRNAs); these exist in many redundant copies in yeast and are potential sources of genomic instability. In the final genome, all tRNA genes will reside on an artificial 17th neochromosome. The engineered *synIII* chromosome retains most other genes, but those identified as nonessential based on previous work have been flanked with LoxP sequences, which means that they can later be subjected to targeted deletion via expression of the Cre recombinase enzyme. In this way, the *Sc2.0* researchers can eventually determine which genes and combinations of genes are essential for cell viability.



According to Chandrasegaran, *synIII* presented an invaluable opportunity for method development. For example, this pilot effort highlighted the necessity of careful computational design to ensure that sequence modification does not inadvertently disrupt essential chromosomal elements. It also demonstrated that chromosome-scale DNA fragments could efficiently be assembled in stepwise fashion from far smaller synthetic oligonucleotides—which were manufactured by a small army of undergraduates as part of a ‘Build-A-Genome’ class. Sc2.0-affiliated laboratories around the world are now constructing other yeast chromosomes to the same specifications as *synIII*, and Chandrasegaran predicts he will the project completed in the next five years. “Of course, you always have to take these predictions with a grain of salt,” he adds.

JCVI is embarking on its own ‘minimal yeast’ program with a pilot effort funded by the US Defense Advanced Research Programs Agency (DARPA) to manufacture a chromosome from the yeast *Kluyveromyces marxianus*. As with *M. mycoides*, this species was selected for its experimental tractability. “It divides as quickly as every 52 minutes—more than twice as fast as *S. cerevisiae*,” says Glass, “and you have access to pretty much the same set of genetic tools.” This team has applied a variant of the random mutagenesis strategy that worked so well in *Mycoplasma* to identify genes required for growth and survival in chromosome 7—the smallest of *K. marxianus*’ eight chromosomes. “We’ve identified nonessential genes and we’re resynthesizing the chromosome,” says Glass. “First we’ll produce something that’s around half a million base pairs, and then we’re also going to try and reorganize the chromosome as much as we can.” With luck, the rules identified in this initial effort will be sufficient to guide similar reduction of the remaining seven chromosomes—and perhaps of other eukaryotic genomes.

Indeed, some researchers are already envisioning a major leap forward. “We’re at a place now in science where it’s technologically believable that we could synthesize a genome the size of the human genome,” says Marc Lajoie, a postdoc in David Baker’s lab at the University of Washington. Lajoie is part of the ‘HGP-write’ initiative, an effort started by Church and Boeke to explore the possibility of moving beyond reading human cellular genomes to actively redesigning them. For now, however, the group is grappling with the broad strokes rather



Brett Shipe, JCVI

The J. Craig Venter Institute and Synthetic Genomics, Inc. team that created JCVI-syn3.0. Left to right: J. Craig Venter, Hamilton Smith, Dan Gibson, Lijie Sun, John Glass, Krishna Kannan, John Gill, and Clyde Hutchison III.

than drawing up blueprints. “My gut feeling is that it’s not a technological problem,” says Lajoie. “It’s a problem of imagination and biological knowledge.”

### Bare essentials

Honing in on what can stay and what can go remains a challenge. Comparative analysis of multiple microbial genomes has revealed certain gene functions that are indisputably critical—although not as many as one might think. “We did a comparative analysis of all the essentiality studies in all bacteria, and we only found about 60 proteins that are essential in every single bacteria,” says Luis Serrano, who studies *Mycoplasma pneumoniae* at the Centre for Genomic Regulation in Barcelona, Spain. Unsurprisingly, most proteins govern functions like RNA transcription, DNA replication and ribosomal translation.

High-throughput gene disruption strategies, such as MAGE or the transposon-based method used at JCVI, can efficiently distinguish additional genes that are essential or nonessential. However, the resulting list may be misleading. For example, individually dispensable genes can cause problems if deleted simultaneously, a phenomenon known as ‘synthetic lethality’. A microbe might obtain an essential amino acid by manufacturing it internally

or by absorbing it from the environment; in nutrient-rich environments, these functions are redundant, but one of the two must remain for the cell to survive. Synthetically lethal interactions can be challenging to predict, and remain a serious confounder for the assembly of minimal genomes.

Furthermore, the function of an experimentally determined essential gene may not be readily apparent, as demonstrated by JCVI-syn3.0. “The fact that 30% of the genes aren’t very well categorized but are nonetheless required for life was a surprise,” says Glass. He notes that his team is currently collaborating with external researchers who have devised theories as to what some of these mysterious essential functions may be. For example, a team led by Andrew Hanson at the University of Florida is working with JCVI to explore whether certain genes that appear to encode enzymes known as hydrolases may help cells process toxic intermediates of glucose metabolism.

However, Antoine Danchin of the Hôpital de la Pitié-Salpêtrière in Paris believes he has managed to get a good handle on many of the 149 ‘unknown unknowns’ from the Venter group’s minimal *Mycoplasma*<sup>7</sup>. “I identified functions for about half of them,” he says. “The other half is made of membrane proteins that are transporters,

but I cannot identify what they transport at this stage.” A former physicist and mathematician, Danchin considers the cell as a machine that should be approached as an engineering problem rather than just a bundle of genes. “An engineer would ask, ‘what are the cell’s master functions, and what are the helper functions that allow those master functions to work?’” By focusing on functions first, Danchin has tentatively characterized essential functions that might not be immediately obvious, like ‘nanoRNases’ that clean up potentially toxic scraps left behind by RNA degradation. Such function-focused perspectives are gaining a foothold in the genome-reduction world, as researchers come to grips with the fact that different



Victor de Lorenzo

Victor de Lorenzo.

Lorenzo. “That means we’re still missing a lot of the functional landscape.”

### Promoting productivity

Not all research groups are engaged in a race to the bottom. For some synthetic biology applications, the goal is to streamline rather than minimize. “With *Pseudomonas*, we have made in the range of 40 deletions in the genome,” says de Lorenzo. “If we go beyond that we start noticing that cells grow less or become more sensitive to stress, and we don’t want to go further in that direction.” This selective approach to genome deletion can greatly enhance an organism’s productivity, resulting in an up to 40% increase in recombinant protein yield<sup>8</sup>. Pósfai’s team has likewise found that reduced *E. coli* strains such as MDS42 might deliver a profound boost to the output of biomanufacturing efforts and his long-time collaborator Blattner is pursuing commercial applications of these strains through his company, [Scarab Genomics](#).

Researchers in the Church laboratory have devised a different approach to genome reduction to bolster *E. coli*’s usefulness as a synthetic biology chassis.

Multiple different three-nucleotide codons can trigger addition of the same amino acid during translation, and Church’s group has set about using MAGE to subtract some of these ‘redundant’ codons from the bacterial genome. A similar effort from Jason Chin’s team at Cambridge University<sup>9</sup> was able to achieve similar genome-wide codon replacement in *E. coli* using CRISPR–Cas9. Most recently, the Church group generated a heavily modified, fully synthetic *E. coli* genome in which 7 out of 64 codons were eliminated<sup>10</sup>. They are still examining whether this fully recoded genome can sustain a viable bacterium, but Lajoie, who helped spearhead the project as a student in Church’s lab, is optimistic. “You can really mess with codon usage a lot in *E. coli*,” he says. “It’s likely that it will be sick, but we’ve shown that we can improve fitness of sick strains.” The now-unused codons in these bacteria could be repurposed to code for the introduction of multiple unnatural amino acids, enabling bacteria to manufacture a more chemically diverse range of proteins.

No one organism can fill every synthetic biology need, as each microbe has its own strengths and limitations. For example, although *E. coli* is a well-established platform for protein production, *P. putida* is more robust against harsher industrial conditions. “It lives in places that have a history of pollution with chemicals, and those cells have high resistance to solvents and chemical stress,” says de Lorenzo. He anticipates that the field would benefit from a toolbox of perhaps a dozen different streamlined bugs with different specializations. Although JCVI-syn3.0 was not developed purely as a synthetic biology tool, its aggressively stripped-down genome could prove an asset for certain applications, and Venter’s team has made it commercially available so that the scientific community can test its mettle. “One of Craig’s ambitions is that it be distributed to students who want to use it for interesting things,” says Glass, “and industrial groups are already using it to see what happens when, for example, you drop metabolic pathways into the cell.”

In the longer term, efforts to distill out the essential genome could yield unprecedented insights into exactly what it means to be alive and make it possible to extrapolate aspects of the evolutionary origins of cellular life. Early achievements in this



Miriam Dornmeyer

Jörg Stülke

space are already offering some tantalizing clues. “I am convinced that the common core of all life is the information-processing machinery,” says Stülke, noting the persistent conservation of RNA transcrip-

tion and protein translation machinery across all cells examined to date. Kim Wise, part of the JCVI synthetic biology team, notes that their genome-reduced *Mycoplasma* has the surprising ability to reproduce robustly despite lacking cytoskeletal proteins with a known role in cell division. “It may be that we’ve come back to a sort of primitive biophysical process that allows cellular life forms to divide,” he says. “These are the type of biology problems we can use this cell for.”

The cellular blueprints from these studies could also allow scientists to computationally reconstruct sophisticated ‘virtual cells’. Serrano’s team has already made headway on this front as part of a multicenter European consortium to build a detailed simulation of *M. pneumoniae* based on a wealth of experimental data.

“We have done a full analysis of essentiality, metabolomics and transcriptomics and proteomics, with the idea of integrating everything into a model that will essentially allow the bug to live in the computer,” he says. Once the most raw basics of life can be simulated with reasonable fidelity, one can imagine either using those insights to guide experimental construction of fully synthetic cells or develop *in silico* systems that take the guesswork out of wet lab experiments. “It’s like a little boy who gets a car for Christmas, and wants to tear it apart and put it back together again to see how it works,” says Stülke. “We want to find out how life really functions.”

**Michael Eisenstein is a freelance science writer based in Philadelphia.**

1. Mushegian, A.R. & Koonin, E.V. *Proc. Natl. Acad. Sci. USA* **93**, 10268–10273 (1996).
2. Wang, H.H. *et al. Nature* **460**, 894–898 (2009).
3. Reuß, D.R., Commichau, F.M., Gundlach, J., Zhu, B. & Stülke, J. *Microbiol. Mol. Biol. Rev.* **80**, 955–987 (2016).
4. Pósfai, G. *et al. Science* **312**, 1044–1046 (2006).
5. Hutchison, C.A. III *et al. Science* **351**, aad6253 (2016).
6. Annaluru, N. *et al. Science* **344**, 55–58 (2014).
7. Danchin, A. & Fang, G. *Microb. Biotechnol.* **9**, 530–540 (2016).
8. Lieder, S., Nikel, P.I., de Lorenzo, V. & Takors, R. *Microb. Cell Fact.* **14**, 23 (2015).
9. Wang, K. *et al. Nature* **539**, 59–64 (2016).
10. Ostrov, N. *et al. Science* **353**, 819–822 (2016).