The sorcerer of synthetic genomes

Andrew Marshall

J. Craig Venter reflects on an effort spanning decades to create a living cell from chemically synthesized building blocks.

Dovetailed into J. Craig Venter's other scientific accomplishments—pioneering the use of novel sequencing approaches to decode tissue transcripts, microbial genomes and ultimately the human genome, not to mention his more recent exploits to catalog and sample the microbial diversity of the world's oceans—has been more than 15 years of work aimed at synthesizing a living organism from simple, chemical building blocks. This culminated with his most recent paper in *Science*[1], published together with his collaborators Hamilton Smith and Clyde Hutchison at the J. Craig Venter Institute (Rockville, MD, USA), that finally demonstrates the feasibility of transferring a genome from a prokaryote to yeast and then back into a different prokaryote. Venter and his team are now poised to take the last tantalizing step—constructing a genome synthetically and then rebooting that to life. *Nature Biotechnology* talked to him about the work and its implications for the future of biological engineering.

J. Craig Venter and his group at the JCVI are forging new ground in the field of synthetic genomics.

How did the synthetic genomics effort first come together?

J. Craig Venter: It started back in 1995 when we sequenced the first two genomes in history. The first genome was *Haemophilus influenzae* that had about 1,800 genes. After it was clear that our new method worked, we looked for a second genome to sequence that year. So the question came up: What would be the most interesting organism to sequence for the first genome comparison? Ham [Smith] and I got talking and we heard about Clyde Hutchison's work, where he'd been characterizing *Mycoplasma genitalium*, which he claimed had the smallest genome of any independently self-replicating organism, something that is still true today. And so after sequencing *Haemophilus*, we quickly sequenced the *Mycoplasma genitalium* genome. On doing the first-ever genome comparisons, we immediately started asking questions like: How small could a genome be and was there a minimal operating system?

What other kinds of questions centered around minimal genomes?

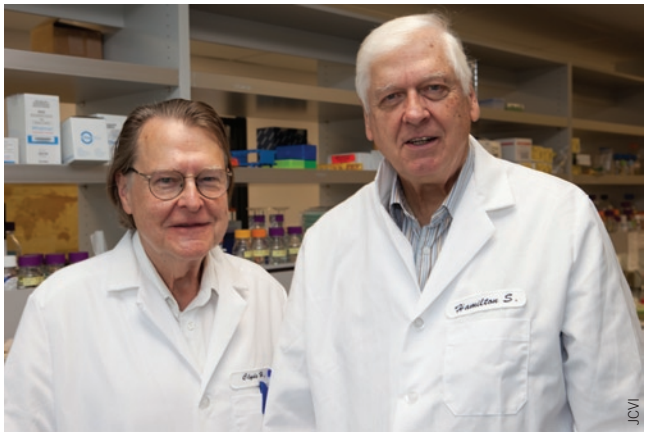
J.C.V.: To put our thinking into context: right after sequencing the *M. genitalium* genome, we started sequencing the third genome, *Methanococcus jannaschii*, the first Archaea genome that was published in 1996. At the same time, some NASA [National Aeronautics and Space Administration; Washington, DC, USA] scientists claimed they'd discovered these fossils of 'nanobacteria' in Martian meteorites. It turned out to be a complete artifact. But we sat down (like a lot of other scientists around the world) and did some calculations about the volume of those hypothetical nanobacteria and asked could they have any volume at all and could they even support small DNA molecules? The answer was no. As a result of all this, we had a great deal of thinking and discussion about minimal genomes.

How did the minimal genome work get underway?

J.C.V.: After we'd sequenced *M. genitalium*, we decided to start knocking out genes in the mycoplasma to see how many genes it could dispense with. That's one of those ideas that's very easy to say, but it's been extremely hard and frustrating to carry out, in part due to the lack of a genetics system in *M. genitalium*. So we (primarily Clyde Hutchison) developed this new approach that we called 'whole transposon mutagenesis' where we randomly inserted transposons into the

Andrew Marshall is Editor, Nature Biotechnology.

genome and then selected for cells that could survive the insertions in their genome. We then sequenced using a primer off the transposon to find out where it was inserted into the genome[2]. This was a whole new approach that you could only do if you had a sequenced genome. But a major limitation of this method was we could only knock out genes one at a time. While we collected more and more knockouts, we found that it did not tell us whether the genes could all be knocked out together, due to a lack of selectable markers. As a result, we decided the only way one could make a minimal genome would be to chemically synthesize the chromosome and then physically vary the gene content—and so that was the start of the field of synthetic genomics.



Clyde Hutchison (left) and Ham Smith (right), who have spearheaded the work by the teams at JCVI aiming to create a living organism from chemical building blocks.

What led you to ΦX174 as the first genome to synthesize?

J.C.V.: It actually had a very slow, difficult start. Clyde Hutchison was in Fred Sanger's lab when they sequenced ΦX174—the first DNA virus ever sequenced—and because of its historic import, we decided to synthesize that genome primarily as a test to see whether we could accurately synthesize genomes. That simple idea ended up being about a ten-year project—in part because we stopped for two years to sequence the human genome. But we just thought it would be simple and that we'd just make PCR primers that had sufficient overlap, anneal them together and then PCR copy the whole thing. When we did this, we obtained DNA molecules the right size (5 kb), but nothing worked. Even with selection by infectivity—where one molecule out of a million would have seen virus particles made—we got nothing. And it turns out that there are just too many errors in DNA synthesis.

After sequencing the human genome, Ham and I started back in on the project and then recruited Clyde up from North Carolina [Chapel Hill] to join us. As a result, we spent many years, particularly Ham and Clyde, working out error correction in the synthesis. That culminated with our report when it took two weeks going from the DNA sequence in the computer to synthesizing the ΦX174 genome, which was activated by injecting it into *E. coli.* The *E. coli* cellular machinery read the synthetic genetic DNA and produced the viral proteins, which self assembled to form the active virus[3].

Could you talk a little more about error correction?

J.C.V.: What we described in the ΦX174 paper[3] were some nice elegant methods for doing repair in real time off of a correct strand, but we still had to select clones and sequence them to ensure the correct order of bases. What ΦX174 gave us was the confidence that we could build accurate DNA units of 5 kb; our assumption was that we could assemble the smaller units using homologous recombination.

So we had a team of several scientists working on this problem. One of the early genomes we sequenced was *Deinococcus radiodurans,* which has a phenomenal DNA repair system that can take these huge doses of ionizing radiation (up to 3 mrads), blowing its chromosome apart with several hundred double-stranded DNA breaks, and then over 12 to 24 hours reassemble the genome as it was before. We spent years trying to isolate the DNA repair genes out of *Deinococcus*, and cloning them, to attempt to create an *in vitro* expression system to assemble our DNA fragments. But we never got it working outside of the intact cells.

It was about that same time that the team led by Dan Gibson discovered that we could assemble our synthetic DNA in yeast using its recombination system. This multi-year work culminated almost two years ago now with the complete synthesis of the *Mycoplasma genitalium* genome[4].

At every stage, we've had to develop new methodology and tools. Over 100 kb, the synthetic DNA segments were too big to clone in *Escherichia coli.* We were looking for another system and discovered if we just used an artificial yeast centromere we could convert bacterial genomes into yeast chromosomes. Work led by Gwyn Benders allows us to clone complete bacterial genomes in yeast artificial chromosomes just by adding a yeast centromere to them[1].

Tell us more about the most recent step of the work.

J.C.V.: The way we had originally envisioned it, we were just going to have the synthetic genome that we assembled on the lab bench and then we were going to transplant it into a recipient cell. But because we ended up doing the final genome assembly in yeast using homologous recombination, we now had to develop whole new methods for isolating our synthetic bacterial chromosome from yeast and then transplanting it.

In our original genome transplantation study, we isolated *Mycoplasma mycoides* genome and transplanted it into a closely related species[5]. But after cloning the *M. mycoides* chromosome in yeast and then isolating it, it would not transplant. It took 20 people two years to solve that little riddle of why it would not transplant from yeast but it would from *M. mycoides* cells[1]. We knew something was happening to the DNA in the *M. mycoides* cell that wasn't happening in yeast. It turns out the secret was DNA methylation.

What kind of approaches did you use to solve the riddle?

J.C.V.: While we were attempting to transfer and boot-up DNA derived from yeast, different members of the team worked out the methods for cloning bacterial chromosomes in yeast—something no one had ever done before. So those are pretty cool methods on their own. The methylation work involved the development of a number of new methods, including trying cellular extracts and using them to methylate the chromosome out of yeast. We purified and cloned all the specific methylases and used them to methylate the bacterial genomes cloned in and extracted from yeast. None of this is trivial as you cannot just pipette entire chromosomes and keep them intact as supercoiled DNA. We have to move and modify the genomes in gel blocks. All the enzymology has to take place in gel blocks. And it takes careful handling not to destroy the chromosomes. The team has done absolutely phenomenal work. As with all things in science, it's the little tiny breakthroughs on a daily basis that make for the big breakthrough.

Did you look at anything else other than methylation?

J.C.V.: We did all these studies to make sure that there were no proteins needed for transplantation. Because you could envision DNA-binding proteins—the equivalent of histones or some similar mechanism in our genomes—required to stabilize the genome. And so we used proteinases to digest all the proteins associated with the extracted DNA and we still found we could get the *M. mycoides* chromosome to transplant. We also worked out that it had to be supercoiled DNA, and if it wasn't supercoiled, clean DNA, it would not transplant.

What have we learned concerning the compatibility of a donor genome with a recipient cell?

J.C.V.: We learned very early on from our first genome sequencing—that of *Haemophilus*—that there are gaps in microbial sequences with the initial assemblies. Genes or sequences that might be toxic to *E. coli,* such as the *Haemophilus* ribosome genes, would delete or truncate when cloned. So it took a huge effort to totally close those genomes because we had to find ways to get clones so that we could sequence walk across the deleted sequence. One of several advantages of the new sequencing techniques is that we don't need cloning in *E. coli*.

When we had the synthetic quarter molecules of the synthetic *Mycoplasma genitalium* genome (when we had pieces of 175,000 base pairs), we got two of the four to clone initially in *E. coli*. All four have now been grown in *E. coli*, but for some reason, passing them through yeast first made them clonable, which we don't understand. We have to solve each of these riddles one at a time. That's why it's so slow. The good news is instead of just gee whiz quickly getting a synthetic cell, we're really learning the processes of life and now being able to move what we call the software of life across the branches of life. So, in a way, it's good that it's taken us 15 years to do because we've just learned so much that's really critical for the next stages by struggling to get through it. We just lucked out with one system and it worked. It just as easily could have led us down a blind alley for a long time.

What about evolutionary divergence and chromosome compatibility?

J.C.V.: The difference between *M. mycoides* and *M. capricolum* eventually design a system to be a universal recipient. One where you have the right tRNAs and have the ability to start reading the genetic code. It's a fundamental aspect of life that you see when you throw the ΦX174 genome into *E. coli* and it just starts reading the DNA and making the viral proteins and they self-assemble. I think the key thing is probably just having the protection against the restriction enzymes in the cell first and foremost, and then having the right machinery to be able to read that genetic code and express those genes. But we don't know how far afield we can go. I'm betting quite far as long as we stick to those fundamental rules. That's the kind of thing we're just starting to test right now.

What role does codon usage play in stability of the clones?

J.C.V.: Did the different codon usage in *M. mycoides* and lack of transcription in *E. coli* facilitate the stable cloning of assembled mycoplasma fragments in *E. coli*? Was it just the UGA codon that yeast didn't recognize so a lot of the proteins weren't translated? With *E. coli,* you always assume that when you have proteins that are expressed that could interfere with *E. coli* biology, that the bacteria essentially goes back and deletes them. But we're doing some work now that suggests it may not be an issue at all. It might just be a matter of selecting for stable clones. We have a different mycoplasma genome that we've cloned in now with basically the same codon usage as yeast, and it is totally stable as well.

But presumably, recoding starting genomes is going be very important?

J.C.V.: What works with one cell is probably going to take some engineering for each new one we take on.

What about the final step of rebooting a synthetic genome? How close are you to that?

J.C.V.: Every time we've tried to reboot a synthetic genome, we've come up with a new set of challenges. Our view now is that we've solved them all, but we'll only know that when we actually have the cell totally controlled by a completely chemically made genome. As yet we do not have that. There's still a chance it will happen this year—which I somewhat optimistically remind people I've said for the past two years now.

Will everything in the future be made from DNA synthesizers?

J.C.V.: No. We'll start with our repertoire of 20 to 40 million genes, and some of them will need to be synthesized and the rest we'll make PCR copies. We'll have 40 million bottles of genes and we'll pull those down for assembling genomes in the future. But I think for proof of concept, it's important for us to start with 4 bottles of chemicals and to watermark the genome—to make it absolutely foolproof that it's really the synthetic chromosome that is controlling the cells. It is essentially important as a theoretical concept, but we're not quite there yet.

Do you anticipate ethical controversy once rebooting of a synthetic genome is demonstrated?

J.C.V.: We have asked and driven the ethical discussion from the beginning. We've been trying to bring the community along with us every step of the way. We think once we do activate a genome that yes, it probably will impact people's thinking about life. But I think it already has, as we've progressed in a logical fashion with each step of these studies. Perhaps one of the good things about it taking so much time for us to do all this work is that we've had time to have the in-depth ethical discussions before we get to that key experiment. For example, the Sloan Foundation report [http://www.synbioproject.org/library/publications/archive/synbio3/], the NSABB [National Science Advisory Board for Biosecurity; http://oba.od.nih.gov/biosecurity/pdf/Final_NSABB_Report_on_Synthetic_Genomics.pdf], the Fink Report out of the US National Academy of Sciences and the report out of the Royal Academy in the UK [http://www.raeng.org.uk/news/publications/list/reports/Synthetic_biology.pdf]. We've watched the mistakes of others, for example with the issues that GMOs [genetically modified organisms] have had. We've worked really hard to bring the world along with us as we develop each step of the technology.

How far are we from understanding how to regulate complex genetic circuits in a synthetic system?

J.C.V.: Using the principle of in-the-lab evolution, if you have a minimal chassis you should be able to replicate billions of years of evolution by adding back components. When you look at what the various

if it could be replicated. I think our findings, that what you need for life is the DNA information molecule and the ability to read the information to produce proteins that self assemble into living cells, are very important. We are starting with living cells and reprogramming them with new DNA software of life but we are not creating life from basic elements. I think it is very surprising to many that we can reprogram cells into new species simply by changing out the software.

I've defined synthetic genomics in a very precise way compared with synthetic biology—which can be anything from molecular biology to genetic engineering to gene circuits. But with synthetic genomics, the goal is to start in the computer in the digital world from digitized biology and make new DNA constructs for very specific purposes. That's why the proof of concept of being able to do that is so critical. We're not there yet, but we are close. It can mean that as we learn the rules of life we will be able to develop robotics and computational systems that are self learning systems. By doing combinatorial genomics using the 20 million genes in our databases, single robotic systems can learn more biology than in the previous decade. It's the beginning of the new era of very rapid learning. If science moves forward in a linear fashion, we've all failed. There's not a single aspect of human life that doesn't have the potential to be totally transformed by these technologies in the future.

1.