 doi:10.1093/jmcb/mjq015

Research Highlight

Synthetic Genomes for Synthetic Biology

Harris H. Wang*

Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA
* Correspondence to: Harris H. Wang, Tel: +1-617-955-9575; E-mail: hhwang@genetics.med.harvard.edu

A genome synthesized entirely from scratch has been used to replace the native genome of a living cell, thus creating a new cell. This achievement marks a new frontier in synthetic biology to design and create genomes for organisms with few genetic tools and for applications in areas of energy, health care and the environment.

Biologists revel in the excitement of discovery, and engineer in the art of creation. In a fusion of two cultures, synthetic biologists are dissecting the inner workings of life by attempting to recreate it in the laboratory, piece by piece. Since the dawn of civilization, humans have been builders and engineers, constructing houses from bricks, machines from metals and now genomes from nucleotides. As the blue print of life, the genome encodes all the necessary heritable information to allow a cell to survive and replicate. Genomes of all living organisms replicate based on a preexisting copy. By contrast, *de novo* synthesis is a new paradigm and a powerful approach to create genomes of any sequence, architecture or design (Figure 1).

Now for the first time, a genome made entirely from chemically synthesized pieces has been successfully booted up in a living cell at the J. Craig Venter Institute in a culminating effort that has stretched over the past decade. In a technical tour de force, Gibson et al. synthesized and assembled a 1.08-Mb *Mycoplasma mycoides* genome *de novo* and successfully transplanted it into a *Mycoplasma capricolum* recipient to create a new *M. mycoides* cell (Gibson et al., 2010). This effort highlights a new breed of synthetic biology based on *de novo* synthesis and engineering for creating synthetic genomes (Carr and Church, 2009).

Mycoplasmas are small commensal or parasitic bacteria that can cause human respiratory and inflammatory diseases. These microorganisms lack a cell wall and are unique for their altered genetic code (UGA encodes tryptophan instead of a stop codon). The design of the synthetic genome (JCVI-syn1.0) was based on the sequence of an *M. mycoides* strain the JCVI group previously used for genome transplantation (Lartigue et al., 2009) with four additional non-disruptive synthetic watermark sequences. This particular genome was chosen for synthesis because of its modest genome size and fast doubling time (80 min). To construct the designed genome, a commercial gene synthesis vendor first generated a sequence-verified library of ~1.1 kb DNA fragments from chemically synthesized oligonucleotides using a strategy first described in the 1970s (Khorana et al., 1972). Then in three hierarchical stages, Gibson et al. assembled the library of 1 kb fragments, which contained 80 bp homologous overlaps, first into 10 kb fragments and then into 100 kb fragments mostly using *in vivo* homologous recombination in yeast. Finally, the 11-second-stage fragments were assembled into the whole *M. mycoides* genome, which propagated as a yeast centromeric plasmid in a yeast clone. DNA sequencing, multiplex PCR reactions and restriction digests were used to sequence-verify the assembled fragments at each step.

To address the bottleneck in rapid screening of non-functional assemblies, Gibson et al. made semi-synthetic genomes to test functionality of the 100 kb intermediates. Through yeast homologous recombination of the native genome with pieces of the intermediate constructs, the method identified a non-viable 100 kb assembly caused by a frameshift error in an essential replication gene (*dnaA*), which was corrected.

To boot up the synthetic genome, Gibson et al. utilized a genome transplantation technique in which the native genome is entirely replaced with a new genome in a living host. Prior work showed that fully intact *M. mycoides* genomic DNA could be transferred into an *M. capricolum* cell by a polyethylene glycol (PEG)-based chemical transformation method (Lartigue et al., 2007). However, because the synthetic *M. mycoides* genome that propagated as a yeast centromeric plasmid was unmethylated, it was degraded quickly by the restriction endonuclease system of the recipient *M. capricolum* cell upon genome transplantation. To overcome this issue, *in vitro* methylation of the synthetic genome using an *M. capricolum* cell extract or transplantation into a restriction endonuclease-deficient *M. capricolum* strain proved to be two viable solutions. Gibson et al. adopted the latter approach to isolate tetracycline-resistant blue-colored clones that were putatively former *M. capricolum* cells that now contained a transplanted *M. mycoides* genome.

While the exact mechanism of how the synthetic *M. mycoides* genome and the native *M. capricolum* genome resolve in the recipient cell remains unknown, PCR genotyping, restriction pattern analysis and DNA sequencing of the genome from the self-replicating transplanted cell suggested the presence of the designed *M. mycoides* genome. Eight unintended single nucleotide polymorphisms and two mutations disrupting nonessential genes were found through whole-genome

© The Author (2010). Published by Oxford University Press on behalf of *Journal of Molecular Cell Biology*, IBCB, SIBS, CAS. All rights reserved.

Downloaded from jmcb.oxfordjournals.org at Harvard University on March 1, 2011

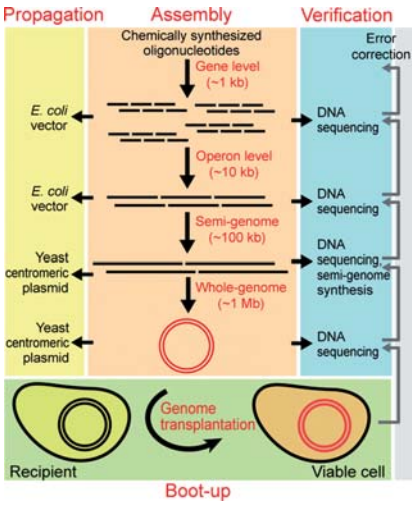

Figure 1 Genome synthesis requires the assembly and propagation of synthetic DNA fragments, accurate sequence verification and error correction methods, and the ability to jumpstart the synthetic genome in a living cell through transplantation.

sequencing, including an 85 bp duplication and an *E. coli* IS1 transposon element (likely from 10 kb fragment cloning in *E. coli*). Morphological studies by electron microscopy and proteomic analysis by 2D gel-electrophoresis of the transplanted cell provided further evidence that the strain was an *M. mycoides*. Based on these results, the group concluded that the transplanted cell contained the synthetic JCVI-syn1.0 genome.

The study highlights the important considerations for genome synthesis. As with any construction project, errors must be minimized and readily addressed when encountered. Modes of failure for a synthetic genome fall mainly into three categories: in design, in synthesis or in boot-up. First, poorly designed genomes will have a high likelihood of failure in many ways, such as presence of toxic genes, absence of essential genes, or improper genetic regulation. Currently, design errors are the most troubling because risk of failure in new designs is still very high as our ability to design a genome remains in its infant stages. Furthermore, experimental indication of a bad design only surfaces well after the synthesis steps. Second, errors can be introduced during genome synthesis at one of many steps. The frequency of encountering sequence errors in chemically synthesized oligonucleotides can be >40%, which can increase to >95% when assembled into 1 kb gene fragments. Gibson et al. found an error frequency of 90% for some assemblies of 10 kb fragments. For assemblies of 100 kb, the error frequency was 75% with small deletions dominating the *in vitro* (e.g. ligation or amplification based) or *in vivo* (e.g. yeast homologous recombination) methods. For the 1 Mb assemblies, the success rate was only ~2%. Thus, errors during genome synthesis can potentially accumulate to yield only one error-free genome in $\sim 10^5$ or more assembly reactions. A rigorous DNA sequencing step after each stage of assembly enables efficient removal of error-containing products, however, at the expense of time and additional resources. Finally, upon producing a genome free of synthesis errors, experimental failures in booting up the genome can arise through a variety of causes such as transplantation failure or donor/recipient incompatibility. Error detection and correction pipelines are thus crucial for any large-scale genome synthesis endeavors and require further improvement.

While many important technical limitations for constructing genomes *de novo* have been resolved by Gibson et al., barriers to successfully design and create new and functional genomes from scratch still remain quite high. Even though the cost of gene synthesis is dropping precipitously (Carlson, 2009), the present day construction cost is still too high for genome synthesis to be practical. To be able to design a whole genome *de novo*, we need to have deeper understanding of how life's essential components function and how they interact with one another. Progress towards creating a minimal genome may shed light on the design principles for creating simpler self-replicating life forms (Forster and Church, 2006) that can grow in complexity with further engineering over time. For now, we will continue to rely on mixing and matching genetic sequences from natural organisms as a way forward because our ability to design complex biological circuits or new proteins *de novo* is still limited. Innovations that enable rapid prototyping of libraries of new genetic designs (Wang et al., 2009) coupled with *in silico* predictions (Feist et al., 2009) of physiology, metabolism and regulation of a synthetic cell will play a crucial role for creating useful synthetic genomes. These genomes may contain entirely new properties (e.g. reassigned genetic codes, rewired regulation, reorganized operons) with new phenotypic traits (e.g. resistance to viruses, modular genome structure).

As the synthetic biologist's toolbox continues to grow, a new budding branch on the Tree of Life is taking shape, emerging from new organisms designed, synthesized and created by engineers. These endeavors will require the careful development of ethical frameworks around the construction of synthetic life and its potential risk, utility and impact on society.

References

Carlson, R. (2009). The changing economics of DNA synthesis. Nat. Biotechnol. *27*, 1091–1094.

Carr, P.A., and Church, G.M. (2009). Genome engineering. Nat. Biotechnol. *27*, 1151–1162.

Feist, A.M., Herrgard, M.J., Thiele, I., Reed, J.L., and Palsson, B.Ø. (2009). Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol. *7*, 129–143.

Forster, A.C., and Church, G.M. (2006). Towards synthesis of a minimal cell. Mol. Syst. Biol. *2*, 45.

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. Science. *329*, 52–56.

Khorana, H.G., Agarwal, K.L., Büchi, H., Caruthers, M.H., Gupta, N.K., Kleppe, K., Kumar, A., Otsuka, E., RajBhandary, U.L., Van de Sande, J.H., et al. (1972). Studies on polynucleotides. CIII. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. J. Mol. Biol. *72*, 209–217.

Lartigue, C., Glass, J.I., Alperovich, N., Pieper, R., Parmar, P.P., Hutchison, C.A., III, Smith, H.O., and Venter, J.C. (2007). Genome transplantation in bacteria: changing one species to another. Science *317*, 632–638.

Lartigue, C., Vashee, S., Algire, M.A., Chuang, R.-Y., Benders, G.A., Ma, L., Noskov, V.N., Denisova, E.A., Gibson, D.G., Assad-Garcia, N., et al. (2009). Creating bacterial strains from genomes that have been cloned and engineered in yeast. Science *325*, 1693–1696.

Wang, H.H., Isaacs, F.J., Carr, P.A, Sun, Z.Z., Xu, G., Forest, C.R., and Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature *460*, 894–898.

Downloaded from jmcb.oxfordjournals.org at Harvard University on March 1, 2011