Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome

Daniel G. Gibson,1 John I. Glass,1 Carole Lartigue,1 Vladimir N. Noskov,1 Ray-Yuan Chuang,1 Mikkel A. Algire,1 Gwennedd A. Benders,1 Michael G. Montague,1 Li Ma,2 Monzia M. Moodie,1 Chuck Merriman,1 Sanjay Vashee,1 Radha Krishnakumar,1 Nacyra Assad-Garcia,1 Cynthia Andrews-Pfannkoch,1 Evgeniya A. Denisova,1 Lei Young,2 Zhi-Qing Qi,1 Thomas H. Segall-Shapiro,1 Christopher H. Calvey,1 Prashanth P. Parmar,1 Clyde A. Hutchison III,2 Hamilton O. Smith,2,3 J. Craig Venter1,2,3

We report the design, synthesis, and assembly of the 1.08-mega–base pair Mycoplasma mycoides JCVI-syn1.0 genome starting from digitized genome sequence information and its transplantation into a M. capricolum recipient cell to create new M. mycoides cells that are controlled only by the synthetic chromosome. The only DNA in the cells is the designed synthetic DNA sequence, including "watermark" sequences and other designed gene deletions and polymorphisms, and mutations acquired during the building process. The new cells have expected phenotypic properties and are capable of continuous self-replication.

In 1977, Sanger and colleagues determined the complete genetic sequence of phage ϕX174 (1), the first DNA genome to be completely sequenced. Eighteen years later, in 1995, our team was able to read the first complete genetic sequence of a self-replicating bacterium, Haemophilus influenzae (2). Reading the genetic sequence of a wide range of species has increased exponentially from these early studies. The ability to rapidly digitize genomic information has increased by more than eight orders of magnitude over the past 25 years (3). Efforts to understand all this new genomic information have spawned numerous new computational and experimental paradigms, yet our genomic knowledge remains very limited. No single cellular system has all of its genes understood in terms of their biological roles. Even in simple bacterial cells, do the chromosomes contain the entire genetic repertoire? If so, can a complete genetic system be reproduced by chemical synthesis starting with only the digitized DNA sequence contained in a computer?

Our interest in synthesis of large DNA molecules and chromosomes grew out of our efforts over the past 15 years to build a minimal cell that contains only essential genes. This work was inaugurated in 1995 when we sequenced the genome of Mycoplasma genitalium, a bacterium with the smallest complement of genes of any known organism capable of independent growth in the laboratory. More than 100 of the 485 protein-coding genes of M. genitalium are dispensable when disrupted one at a time (4–6).

We developed a strategy for assembling viral-sized pieces to produce large DNA molecules that enabled us to assemble a synthetic M. genitalium genome in four stages from chemically synthesized DNA cassettes averaging about 6 kb in size. This was accomplished through a combination of in vitro enzymatic methods and in vivo recombination in Saccharomyces cerevisiae. The whole synthetic genome [582,970 base pairs (bp)] was stably grown as a yeast centromeric plasmid (YCP) (7).

Several hurdles were overcome in transplanting and expressing a chemically synthesized chromosome in a recipient cell. We needed to improve methods for extracting intact chromosomes from yeast. We also needed to learn how to transplant these genomes into a recipient bacterial cell to establish a cell controlled only by a synthetic genome. Because M. genitalium has an extremely slow growth rate, we turned to two faster-growing mycoplasma species, M. mycoides subsp. capri (GM12) as donor, and M. capricolum subsp. capricolum (CK) as recipient.

To establish conditions and procedures for transplanting the synthetic genome out of yeast, we developed methods for cloning entire bacterial chromosomes as centromeric plasmids in yeast, including a native M. mycoides genome (8, 9). However, initial attempts to extract the M. mycoides genome from yeast and transplant it into M. capricolum failed. We discovered that the donor and recipient mycoplasmas share a common restriction system. The donor genome was methylated in the native M. mycoides cells and was therefore protected against restriction during the transplantation from a native donor cell (10). However, the bacterial genomes grown in yeast are unmethylated and so are not protected from the single restriction system of the recipient cell.

We overcame this restriction barrier by methylating the donor DNA with purified methylases or crude M. mycoides or M. capricolum extracts, or by simply disrupting the recipient cell’s restriction system (8).

We now have combined all of our previously established procedures and report the synthesis, assembly, cloning, and successful transplantation of the 1.08-Mbp M. mycoides JCVI-syn1.0 genome, to create a new cell controlled by this synthetic genome.

Synthetic genome design. Design of the M. mycoides JCVI-syn1.0 genome was based on the highly accurate finished genome sequences of two laboratory strains of M. mycoides subspecies capri GM12 (8, 9, 11). One was the genome donor used by Lartigue et al. [GenBank accession CP001661] (10). The other was a strain created by transplantation of a genome that had been cloned and engineered in yeast, YCPMmyn1.1-AtypellIres [GenBank accession CP001668] (8). This project was critically dependent on the accuracy of these sequences. Although we believe that both finished M. mycoides genome sequences are reliable, there are 95 sites at which they differ. We began to design the synthetic genome after both sequences were finished. Consequently, most of the cassettes were designed and synthesized based on the CP001621 sequence (11). When it was finished, we chose the sequence of the genome successfully transplanted from yeast (CP001668) as our design reference (except that we kept the intact typeIII res gene). All differences that appeared biologically significant between CP001668 and previously synthesized cassettes were corrected to match it exactly (11). Sequence differences between our synthetic cassettes and CP001668 that occurred at 19 sites appeared harmless and so were not corrected. These provide 19 polymorphic differences between our synthetic genome (JCVI-syn1.0) and the natural (nonsynthetic) genome (YCPMmyn1.1) that we have cloned in yeast and use as a standard for genome transplantation from yeast (8). To further differentiate between the synthetic genome and the natural one, we designed four watermark sequences (fig. S1) to replace one or more cassettes in regions experimentally demonstrated [watermarks 1 (1246 bp) and 2 (1081 bp)] or predicted [watermarks 3 (1109 bp) and 4 (1222 bp)] to not interfere with cell viability. These watermark sequences encode unique identifiers while limiting their translation into peptides. Table S1 lists the differences between the synthetic genome and this natural standard. Figure S2 shows a map of the M. mycoides JCVI-syn1.0 genome. Cassette and assembly intermediate boundaries, watermarks, deletions, insertions, and genes of the M. mycoides JCVI syn1.0 are shown in fig. S2, and the sequence of the transplanted mycoplasma clone sMmYCP235-1 has been submitted to GenBank (accession CP000207).

Synthetic genome assembly strategy. The designed cassettes were generally 1080 bp with 80-bp overlaps to adjacent cassettes (11). They were all produced by assembly of chemically

1The J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA. 2The J. Craig Venter Institute, 10355 Science Center Drive, San Diego, CA 92121, USA. 3To whom correspondence should be addressed. E-mail: jcventer@jcvi.org
synthesized oligonucleotides by Blue Heron (Bothell, Washington). Each cassette was individually synthesized and sequence-verified by the manufacturer. To aid in the building process, DNA cassettes and assembly intermediates were designed to contain Not I restriction sites at their termini and recombinated in the presence of vector elements to allow for growth and selection in yeast (7, 11). A hierarchical strategy was designed to assemble the genome in three stages by transformation and homologous recombination in yeast from 1078 1-kb cassettes (Fig. 1) (12, 13).

Assembly of 10-kb synthetic intermediates. In the first stage, cassettes and a vector were recombined in yeast and transferred to Escherichia coli (11). Plasmid DNA was then isolated from individual E. coli clones and digested to screen for cells containing a vector with an assembled 10-kb insert. One successful 10-kb assembly is represented (Fig. 2A). In general, at least one 10-kb assembled fragment could be obtained by screening 10 yeast clones. However, the rate of success varied from 10 to 100%. All of the first-stage intermediates were sequenced. Nineteen out of 111 assemblies contained errors. Alternate clones were selected, sequence-verified, and moved on to the next assembly stage (11).

Assembly of 100-kb synthetic intermediates. The pooled 10-kb assemblies and their respective cloning vectors were transferred into yeast as above to produce 100-kb assembly intermediates (11). Our results indicated that these products cannot be stably maintained in E. coli, so recombinant DNA had to be extracted from yeast. Multiplex polymerase chain reaction (PCR) was performed on selected yeast clones (fig. S3 and table S2). Because every 10-kb assembly intermediate was represented by a primer pair in this analysis, the presence of all amplicons would suggest an assembled 100-kb intermediate. In general, 25% or more of the clones screened contained all of the amplicons expected for a complete assembly. One of these clones was selected for further screening. Circular plasmid DNA was extracted and sized on an agarose gel alongside a supercoiled marker. Successful second-stage assemblies with the vector sequence are ~105 kb in length (Fig. 2B). When all amplicons were produced following multiplex PCR, a second-stage assembly intermediate of the correct size was usually produced. In some cases, however, small deletions occurred. In other instances, multiple 10-kb fragments were assembled, which produced a larger second-stage assembly intermediate. Fortunately, these differences could easily be detected on an agarose gel before complete genome assembly.

Complete genome assembly. In preparation for the final stage of assembly, it was necessary to isolate microgram quantities of each of the 11 second-stage assemblies (11). As reported (14), circular plasmids the size of our second-stage assemblies could be isolated from yeast spheroplasts after an alkaline-lysis procedure. To further purify the 11 assembly intermediates, they were treated with exonuclease and passed through an anion-exchange column. A small fraction of the total plasmid DNA (1/100) was digested with Not I and analyzed by field-inversion gel electrophoresis (FIGE) (Fig. 2C). This method produced ~1 μg of each assembly per 400 ml of yeast culture (~1011 cells).

The method above does not completely remove all of the linear yeast chromosomal DNA, which we found could substantially decrease the yeast transformation and assembly efficiency. To further enrich for the 11 circular assembly intermediates, ~200 ng samples of each assembly were pooled and mixed with molten agarose. As the agarose solidifies, the fibers thread through and topologically “trap” circular DNA (15). Untrapped linear DNA can then be separated out of the agarose plug by electrophoresis, thus enriching for the trapped circular molecules. The 11 circular assembly intermediates were digested with Not I so that the inserts could be released. Subsequently, the fragments were extracted from the agarose plug, analyzed by FIGE (Fig. 2D), and transformed into yeast spheroplasts (11). In this third and final stage of assembly, an additional vector sequence was not required because the yeast cloning elements were already present in assembly 811-900.

To screen for a complete genome, multiplex PCR was carried out with 11 primer pairs,

![Image](https://www.sciencemag.org)
designed to span each of the 11 100-kb assembly junctions (table S3). Of 48 colonies screened, DNA extracted from one clone (sMmYCp235) produced all 11 amplimers. PCR of the wild-type positive control (YCpMmyc1.1) produced an indistinguishable set of 11 amplimers (Fig. 3A). To further demonstrate the complete assembly of a synthetic M. mycoides genome, intact DNA was isolated from yeast in agarose plugs and subjected to two restriction analyses: Asc I and BssH II (17). Because these restriction sites are present in three of the four watermark sequences, this choice of digestion produces restriction patterns that are distinct from that of the natural M. mycoides genome (Figs. 1 and 3B). The sMmYCp235 clone produced the restriction pattern expected for a completely assembled synthetic genome (Fig. 3C).

Synthetic genome transplantation. Additional agarose plugs used in the gel analysis above (Fig. 3C) were also used in genome transplantation experiments (17). Intact synthetic M. mycoides genomes from the sMmYCp235 yeast clone were transplanted into restriction-minus M. capricolum recipient cells, as described (8). Results were scored by selecting for growth of blue colonies on SP4 medium containing tetracycline and X-gal at 37°C. Genomes isolated from this yeast clone produced 5 to 15 tetracycline-resistant blue colonies in all transplantation experiments was produced by the YCpMmyc1.1 control. Recovery of colonies in all transplantation experiments was dependent on the presence of both M. capricolum recipient cells and an M. mycoides genome.

Semisynthetic genome assembly and transplantation. To aid in testing the functionality of each 100-kb synthetic segment, semisynthetic genomes were constructed and transplanted. By mixing natural pieces with synthetic ones, the successful construction of each synthetic 100-kb assembly could be verified without having to sequence intermediates. We cloned 11 overlapping natural 100-kb assemblies in yeast by using a previously described method (16). In 11 parallel reactions, yeast cells were cotransformed with fragmented M. mycoides genomic DNA (YCpMmyc 1.1) that averaged ~100 kb in length and a PCR-amplified vector designed to overlap the ends of the 100-kb inserts. To maintain the appropriate overlaps so that natural and synthetic fragments could be recombined, the PCR-amplified vectors were produced via primers with the same 40-bp overlaps used to clone the 100-kb synthetic assemblies. The semisynthetic genomes that were constructed contained 100 bp of the 11 100-kb synthetic subassemblies (Table 1). The production of viable colonies produced after transplantation confirmed that the synthetic fraction of each genome contained no lethal mutations. Only one of the 100-kb subassemblies, 811-900, was not viable.

Initially, an error-containing 811-820 clone was used to produce a synthetic genome that did not transplant. This was expected because the error was a single-base pair deletion that creates a frameshift in dnaA, an essential gene for chromosomal replication. We were previously unaware of this mutation. By using a semisynthetic genome construction strategy, we pinpointed 811-900 as the source for failed synthetic transplantation experiments. Thus, we began to reassemble an error-free 811-900 assembly, which was used to produce the sMmYCp235 yeast strain. The dnaA-mutated genome differs by only one nucleotide from the synthetic genome in sMmYCp235. This genome served as a negative control in our transplantation experiments. The dnaA mutation was also repaired at the 811-900 level by genome engineering in yeast (17). A repaired 811-900...
assembly was used in a final-stage assembly to produce a yeast clone with a repaired genome. This yeast clone is named sMmYCP142 and could be transplanted. A complete list of genomes that have been assembled from 11 pieces and successfully transplanted is provided in Table 1.

Characterization of the synthetic transplants.

To rapidly distinguish the synthetic transplants from *M. capricolum* or natural *M. mycoides*, two analyses were performed. First, four primer pairs that are specific to each of the four watermarks were designed such that they produce four amplicons in a single multiplex PCR reaction (Table S4). All four amplicons were produced by transplants generated from sMmYCP235, but not YCpMmyc1.1 (Fig. 4A). Second, the gel analysis with Ascl and BssH II, described above (Fig. 3C), was performed. The restriction pattern obtained was consistent with a transplant produced from a synthetic *M. mycoides* genome (Fig. 4B).

Table 1. Genomes that have been assembled from 11 pieces and successfully transplanted. Assembly 2-100, 1; assembly 101-200, 2; assembly 201-300, 3; assembly 301-400, 4; assembly 401-500, 5; assembly 501-600, 6; assembly 601-700, 7; assembly 701-799, 8; assembly 811-900, 9; assembly 901-1000, 10; assembly 1001-1104, 11. WM, watermarked assembly.

<table>
<thead>
<tr>
<th>Genome assembly</th>
<th>Synthetic fragments</th>
<th>Natural fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconstituted natural genome</td>
<td>None</td>
<td>1–11</td>
</tr>
<tr>
<td>2/11 semisynthetic genome with one watermark</td>
<td>5 WM, 10</td>
<td>1–4, 6–9, 11</td>
</tr>
<tr>
<td>8/11 semisynthetic genome without watermarks</td>
<td>1–4, 6–8, 11</td>
<td>5, 9, 10</td>
</tr>
<tr>
<td>9/11 semisynthetic genome without watermarks</td>
<td>1–4, 6–8, 10–11</td>
<td>5, 9</td>
</tr>
<tr>
<td>9/11 semisynthetic genome with three watermarks</td>
<td>1, 2 WM, 3 WM, 4, 6, 7 WM, 8, 10–11</td>
<td>5, 9</td>
</tr>
<tr>
<td>10/11 semisynthetic genome with three watermarks</td>
<td>1, 2 WM, 3 WM, 4, 5 WM, 6, 7 WM, 8, 10–11</td>
<td>9</td>
</tr>
<tr>
<td>11/11 synthetic genome, 811-820 correction of dnaA</td>
<td>1, 2 WM, 3 WM, 4, 5 WM, 6, 7 WM, 8, 9–11</td>
<td>None</td>
</tr>
<tr>
<td>11/11 synthetic genome, 811-900 correction of dnaA</td>
<td>1, 2 WM, 3 WM, 4, 5 WM, 6, 7 WM, 8, 9–11</td>
<td>None</td>
</tr>
</tbody>
</table>

A single transplant originating from the sMmYCP235 synthetic genome was sequenced. We refer to this strain as *M. mycoides* JCVI-syn1.0. The sequence matched the intended design with the exception of the known polymorphisms, eight new single-nucleotide polymorphisms, an *E. coli* transposon insertion, and an 85-bp duplication (Table S1). The transposon insertion exactly matches the size and sequence of IS1, a transposon in *E. coli*. It is likely that IS1 infected the 10-kb subassembly following its transfer to *E. coli*. The IS1 insert is flanked by direct repeats of *M. mycoides* sequence, suggesting that it was inserted by a transposition mechanism. The 85-bp duplication is a result of a nonhomologous end joining event, which was not detected in our sequence analysis at the 10-kb stage. These two insertions disrupt two genes that are evidently nonessential. We did not find any sequences in the synthetic genome that could be identified as belonging to *M. capricolum*. This indicates that there was a complete replacement of the *M. capricolum* genome by our synthetic genome during the transplant process.

The cells with only the synthetic genome are self-replicating and capable of logarithmic growth. Scanning and transmission electron micrographs (EMs) of *M. mycoides* JCVI-syn1.0 cells show small, ovoid cells surrounded by cytoplasmic membranes (Fig. 5, C to F). Proteomic analysis of *M. mycoides* JCVI-syn1.0 and the wild-type control (YCpMmyc1.1) by two-dimensional gel electrophoresis revealed almost identical patterns of protein spots (Fig. S4) that differed from those previously reported for *M. capricolum* (10). Fourteen genes are deleted or disrupted in the *M. mycoides* JCVI-syn1.0 genome; however, the rate of appearance of colonies on agar plates and the colony morphology are similar (compare Fig. 5, A and B). We did observe slight differences in the growth rates in a color-changing unit assay, with the JCVI-syn1.0 transplants growing slightly faster than the MmcyYCp1.1 control strain (Fig. S6).

**Discussion.** In 1995, the quality standard for sequencing was considered to be one error in 10,000 bp, and the sequencing of a microbial genome required months. Today, the accuracy is substantially higher. Genome coverage of 30 to 50× is not unusual, and sequencing only requires a few days. However, obtaining an error-free genome that could be transplanted into a recipient cell to create a new cell controlled only by the synthetic genome was complicated and required many quality-control steps. Our success was thwarted for many weeks by a single-base pair deletion in the essential gene *dnaA*. One wrong base out of more than 1 million in an essential gene rendered the genome inactive, whereas major genome insertions and deletions in nonessential parts of the genome had no observable effect on viability. The demonstration that our synthetic genome gives rise to transplants with the characteristics of *M. mycoides* cells implies that the DNA sequence on which it is based is accurate enough to specify a living cell with the appropriate properties.

Our synthetic genomic approach stands in sharp contrast to various other approaches to genome engineering that modify natural genomes by introducing multiple insertions, substitutions, or deletions (18–22). This work provides a proof of principle for producing cells based on computer-designed genome sequences. DNA sequencing of a cellular genome allows storage of the genetic instructions for life as a digital file. The synthetic genome described here has only limited modifications from the naturally occurring *M. mycoides* genome. However, the approach we have developed should be applicable to the synthesis and transplantation of more novel genomes as genome design progresses (23).

We refer to such a cell controlled by a genome assembled from chemically synthesized pieces of DNA as a "synthetic cell," even though the cytoplasm of the recipient cell is not synthetic. Phenotypic effects of the recipient cytoplasm are diluted with protein turnover and as cells carrying only the
transplanted genome replicate. Following transplantation and replication on a plate to form a colony (>30 divisions or >10^2-fold dilution), progeny will not contain any protein molecules that were present in the original recipient cell (10, 24). This was previously demonstrated when we first described genome transplantation (10). The properties of the cells controlled by the assembled genome are expected to be the same as if the whole cell had been produced synthetically (the DNA software builds its own hardware).

The ability to produce synthetic cells renders it essential for researchers making synthetic DNA constructs and cells to clearly watermark their work to distinguish it from naturally occurring DNA and cells. We have watermarked the synthetic chromosome in this and our previous study (7).

If the methods described here can be generalized, design, synthesis, assembly, and transplantation of synthetic chromosomes will no longer be a barrier to the progress of synthetic biology. We expect that the cost of DNA synthesis will follow what has happened with DNA sequencing and continue to exponentially decrease. Lower synthesis costs combined with automation will enable broad applications for synthetic genomics.

We have been driving the ethical discussion concerning synthetic life from the earliest stages of this work (25, 26). As synthetic genomic applications expand, we anticipate that this work will continue to raise philosophical issues that have broad societal and ethical implications. We encourage the continued discourse.

References and Notes
11. Supporting material is available on Science Online.
24. A mycoplasma cell, with a mass of about 10^{-11} g, contains fewer than 10^8 molecules of protein. (If it contains 20% protein, this is equivalent to 2 \times 10^{-14} g of protein per cell. At a molecular mass of 120 daltons per amino acid residue, each cell contains (2 \times 10^{-14} g/120) \times 6 \times 10^23 = 1 \times 10^8 residues per cell. If the average size of a protein is 300 residues, then a cell contains about 3 \times 10^5 protein molecules.) After 20 cell divisions the number of progeny exceeds the total number of protein molecules in the recipient cell. So, following transplantation and replication to form a colony on a plate, most cells will contain no protein molecules that were present in the original recipient cell.
28. We thank Synthetic Genomics, Inc. for generous funding of this work. We thank J. B. Hostetler, D. Radune, N. B. Fedorova, M. D. Kim, B. J. Szczypinski, I. K. Singh, J. R. Miller, S. Kausch, R. M. Friedman, and J. Mulligan for their contributions to this work. Electron micrographs were generously provided by T. Deerinck and M. Ellisman of the National Center for Microscopy and Imaging Research at the University of California at San Diego.
29. J.C.V. is chief executive officer and co-chief scientific officer of SGI. H.O.S. is co-chief scientific officer and on the Board of Directors of SGI. C.A.H. is chairman of the SGI Scientific Advisory Board. All three of these authors and JCVI hold SGI stock. JCVI has filed patent applications on some of the techniques described in this paper.
Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome


*Science* 329 (5987), 52-56. [doi: 10.1126/science.1190719]
originally published online May 20, 2010

Editor's Summary

**Let There Be Life**

The DNA sequence information from thousands of genomes is stored digitally as ones and zeros in computer memory. Now, *Gibson et al.* (p. 52, published online 20 May; see the cover; see the Policy Forum by *Cho and Relman*) have brought together technologies from the past 15 years to start from digital information on the genome of *Mycoplasma mycoides* to chemically synthesize the genomic DNA as segments that could then be assembled in yeast and transplanted into the cytoplasm of another organism. A number of methods were also incorporated to facilitate testing and error correction of the synthetic genome segments. The transplanted genome became established in the recipient cell, replacing the recipient genome, which was lost from the cell. The reconstituted cells were able to replicate and form colonies, providing a proof-of-principle for future developments in synthetic biology.