RESEARCH ARTICLE SUMMARY

SYNTHETIC BIOLOGY

Design and synthesis of a minimal bacterial genome

Clyde A. Hutchison III,*† Ray-Yuan Chuang,† Vladimir N. Noskov, Nacyra Assad-Garcia, Thomas J. Deerinck, Mark H. Ellisman, John Gill, Krishna Kannan, Bogumil J. Karas, Li Ma, James F. Pelletier, Zhi-Qing Qi, R. Alexander Richter, Elizabeth A. Strychalski, Lijie Sun, Yo Suzuki, Billyana Tsvetanova, Kim S. Wise, Hamilton O. Smith, John I. Glass, Chuck Merryman, Daniel G. Gibson, J. Craig Venter*

INTRODUCTION: In 1984, the simplest cells capable of autonomous growth, the mycoplasmas, were proposed as models for understanding the basic principles of life. In 1995, we reported the first complete cellular genome sequences (*Haemophilus influenza*, 1815 genes, and *Mycoplasma genitalium*, 525 genes). Comparison of these sequences revealed a conserved core of about 250 essential genes, much smaller than either genome. In 1999, we introduced the method of global transposon mutagenesis and experimentally demonstrated that *M. genitalium* contains many genes that are nonessential for growth in the laboratory, even though it has the smallest genome known for an autonomously replicating cell found in nature. This implied that it should be possible to produce a minimal cell that is simpler than any natural one. Whole genomes can now be built from chemically synthesized oligonucleotides and brought to life by installation into a receptive cellular environment. We have applied whole-genome design and synthesis to the problem of minimizing a cellular genome.

RATIONALE: Since the first genome sequences, there has been much work in many bacterial models to identify nonessential genes and define core sets of conserved genetic functions, using the methods of comparative genomics. Often, more than one gene product can perform a particular essential function. In such cases, neither gene will be essential, and neither will necessarily be conserved. Consequently, these approaches cannot, by themselves, identify a set of genes that is sufficient to constitute a viable genome. We set out to define a minimal cellular genome experimentally by designing and building one, then testing it for viability. Our goal is a cell so simple that we can determine the molecular and biological function of every gene.

RESULTS: Whole-genome design and synthesis were used to minimize the 1079–kilobase pair (kbp) synthetic genome of *M. mycoides* JCVI-syn1.0. An initial design, based on collective knowledge of molecular biology in combination

with limited transposon mutagenesis data, failed to produce a viable cell. Improved transposon mutagenesis methods revealed a class of quasi-essential genes that are needed for robust growth, explaining the failure of our initial design. Three more cycles of design, synthesis, and testing, with retention of quasi-essential genes, produced JCVI-syn3.0 (531 kbp, 473 genes). Its genome is smaller than that of any autonomously replicating cell found in nature. JCVI-syn3.0 has a doubling time of ~180 min, produces colonies that are morphologically similar to those of JCVI-syn1.0, and appears to be polymorphic when examined microscopically.

CONCLUSION: The minimal cell concept appears simple at first glance but becomes more complex upon close inspection. In addition to essential and nonessential genes, there are many quasi-essential genes, which are not absolutely critical for viability but are nevertheless required for robust growth. Consequently, during the process of genome minimization, there is a trade-off between genome size and growth rate. JCVI-syn3.0 is a working approximation of a minimal cellular genome, a compromise between small genome size and a workable growth rate for an experimental organism. It retains almost all the genes that are involved in the synthesis and processing of macromolecules. Unexpectedly, it also contains 149 genes with unknown biological functions, suggesting the presence of undiscovered functions that are essential for life. JCVI-syn3.0 is a versatile platform for investigating the core functions of life and for exploring whole-genome design. ∎
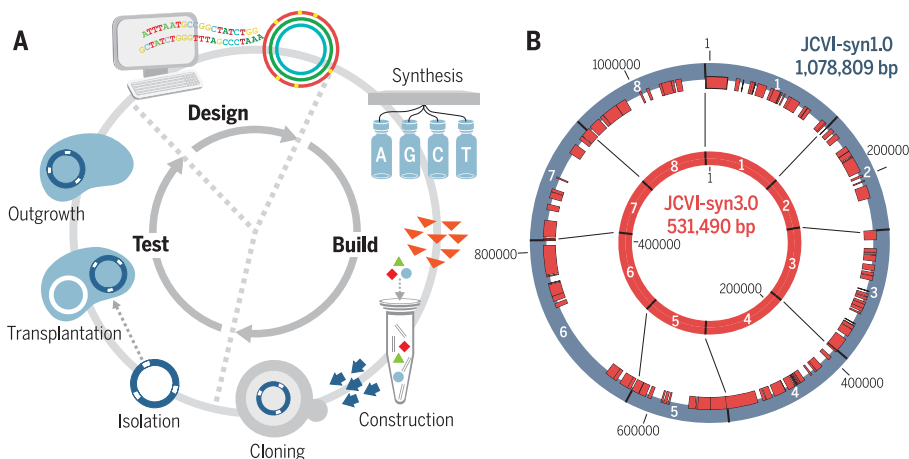
Four design-build-test cycles produced JCVI-syn3.0. (**A**) The cycle for genome design, building by means of synthesis and cloning in yeast, and testing for viability by means of genome transplantation. After each cycle, gene essentiality is reevaluated by global transposon mutagenesis. (**B**) Comparison of JCVI-syn1.0 (outer blue circle) with JCVI-syn3.0 (inner red circle), showing the division of each into eight segments. The red bars inside the outer circle indicate regions that are retained in JCVI-syn3.0. (**C**) A cluster of JCVI-syn3.0 cells, showing spherical structures of varying sizes (scale bar, 200 nm).
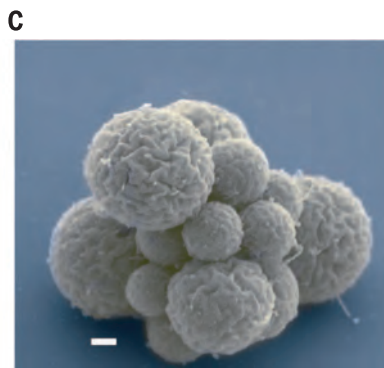
SYNTHETIC BIOLOGY

Design and synthesis of a minimal bacterial genome

Clyde A. Hutchison III,[1]*† Ray-Yuan Chuang,[1]†‡ Vladimir N. Noskov,[1]
Nacyra Assad-Garcia,[1] Thomas J. Deerinck,[2] Mark H. Ellisman,[2] John Gill,[3]
Krishna Kannan,[3] Bogumil J. Karas,[1] Li Ma,[1] James F. Pelletier,[4]§ Zhi-Qing Qi,[3]
R. Alexander Richter,[1] Elizabeth A. Strychalski,[4] Lijie Sun,[1]|| Yo Suzuki,[1]
Billyana Tsvetanova,[3] Kim S. Wise,[1] Hamilton O. Smith,[1,3] John I. Glass,[1]
Chuck Merryman,[1] Daniel G. Gibson,[1,3] J. Craig Venter[1,3]*

We used whole-genome design and complete chemical synthesis to minimize the
1079–kilobase pair synthetic genome of *Mycoplasma mycoides* JCVI-syn1.0. An initial
design, based on collective knowledge of molecular biology combined with limited
transposon mutagenesis data, failed to produce a viable cell. Improved transposon
mutagenesis methods revealed a class of quasi-essential genes that are needed for robust
growth, explaining the failure of our initial design. Three cycles of design, synthesis, and
testing, with retention of quasi-essential genes, produced JCVI-syn3.0 (531 kilobase pairs,
473 genes), which has a genome smaller than that of any autonomously replicating cell
found in nature. JCVI-syn3.0 retains almost all genes involved in the synthesis and
processing of macromolecules. Unexpectedly, it also contains 149 genes with unknown
biological functions. JCVI-syn3.0 is a versatile platform for investigating the core functions
of life and for exploring whole-genome design.

Cells are the fundamental units of life. The genome sequence of a cell may be thought of as its operating system. It carries the code that specifies all of the genetic functions of the cell, which in turn determine the cellular chemistry, structure, replication, and other characteristics. Each genome contains instructions for universal functions that are common to all forms of life, as well as instructions that are specific to the particular species. The genome is dependent on the functions of the cell cytoplasm for its expression. In turn, the properties of the cytoplasm are determined by the instructions encoded in the genome. The genome can be viewed as a piece of software; DNA sequencing allows the software code to be read. In 1984, Morowitz proposed the simplest cells capable of autonomous growth, the mycoplasmas, as models for understanding the basic principles of life (1). A key early step in his proposal was the sequencing of a mycoplasma genome, which we accomplished for *Mycoplasma genitalium* in 1995 (2). Even with the sequence in hand,

deciphering the operating system of the cell was a daunting task.

We have long been interested in simplifying the genomic software of a bacterial cell by eliminating genes that are nonessential for cell growth under ideal conditions in the laboratory. This facilitates the goal of achieving an understanding of the molecular and biological function of every gene that is essential for life. To survive in nature, most bacterial cells must be capable of adapting to numerous environments. Typical well-studied bacteria such as *Bacillus subtilis* and *Escherichia coli* carry 4000 to 5000 genes. They are highly adaptable, because many of their genes provide functions that are needed only under certain growth conditions. Some bacteria, however, grow in restricted environments and have undergone genome reduction over evolutionary time. They have lost genes that are unnecessary in a stable environment. The mycoplasmas, which typically grow in the nutrient-rich environment of animal hosts, have the smallest known genomes of any autonomously replicating cells. A comparison of the first two available genome sequences, *Haemophilus influenzae* [1815 genes (3)] and *M. genitalium* [the smallest known mycoplasma genome; 525 genes (2)], revealed a common core of only 256 genes, much smaller than either genome. This was proposed to be the minimal gene set for life (4).

In 1999, to put this comparative study to an experimental test, we introduced the method of global transposon mutagenesis (5), which allowed us to catalog 150 nonessential genes in *M. genitalium* (6) and predict a set of 375 essen-

tial genes. These results showed that it should be possible to produce a minimal genome that is smaller than any found in nature, but that the minimal genome would be larger than the common set of 256 genes. At that time, we proposed to create and test a cassette-based minimal artificial genome (5). We have been working since then to produce the tools needed to accomplish this. We developed methods to chemically synthesize the *M. genitalium* genome (7). However, *M. genitalium* grows very slowly, so we turned to the faster-growing *M. mycoides* genome as our target for minimization. We developed the method of genome transplantation, which allowed us to introduce *M. mycoides* genomes, as isolated DNA molecules, into cells of a different species, *M. capricolum* (8, 9). In this process, the *M. capricolum* genome is lost, resulting in a cell containing only the transplanted genome. In 2010, we reported the complete chemical synthesis and installation of the genome of *M. mycoides* JCVI-syn1.0 [1,078,809 base pairs (bp) (10); hereafter abbreviated syn1.0]. This genome was an almost exact copy of the wild-type *M. mycoides* genome, with the addition of a few watermark and vector sequences.

Genome reduction in bacteria such as *E. coli* and *B. subtilis* has previously been achieved by a series of sequential deletion events (11, 12). After each deletion, viability, growth rate, and other phenotypes can be determined. In contrast to this approach, we set out to design a reduced genome, then build and test it. We initially designed a hypothetical minimal genome (HMG) based on a combination of existing transposon mutagenesis and deletion data (13) and cumulative knowledge of molecular biology from the literature (14).

We designed the genome to be built in eight segments, each of which could be independently tested for viability in the context of a seven-eighths syn1.0 genome (i.e., a syn1.0 genome that is seven-eighths complete). Initially, only one of the designed HMG segments produced a viable genome. Improvements to our global transposon mutagenesis method allowed us to reliably classify genes as essential or nonessential and to identify quasi-essential genes that are needed for robust growth, though not absolutely required [figs. S1 to S4 (9); a similar result in *M. pneumoniae* is presented in (15)]. We also established rules for removing genes from our genome design without disturbing the expression of the remaining genes. Methods that we developed in the course of building syn1.0 (10) provide a way to build a new genome as a centromeric plasmid in yeast and to test it for viability and other phenotypic traits after transplantation into an *M. capricolum* recipient cell. These methods, along with improvements described here, make up a design-build-test (DBT) cycle (Fig. 1).

Here we report a new cell, JCVI-syn3.0 (abbreviated syn3.0), that is controlled by a 531–kilobase pair (kbp) synthetic genome that encodes 438 proteins and 35 annotated RNAs. It is a working approximation to a minimal cell. Its genome is substantially smaller than that of

[1]J. Craig Venter Institute, La Jolla, CA 92037, USA. [2]National Center for Microscopy and Imaging Research, University of California–San Diego, La Jolla, CA 92037, USA. [3]Synthetic Genomics, La Jolla, CA 92037, USA. [4]National Institute of Standards and Technology, Gaithersburg, MD 20899, USA.
*Corresponding author. E-mail: chutchis@jcvi.org (C.A.H.); jcventer@jcvi.org (J.C.V.) †These authors contributed equally to this work. ‡Present address: American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110, USA. §Present address: Center for Bits and Atoms, Massachusetts Institute of Technology, Room E15-401, 20 Ames Street, Cambridge, MA 02139, USA. ||Present address: CBRITE, 11575 Sorrento Valley Road, Suite 204, San Diego, CA 92121, USA.
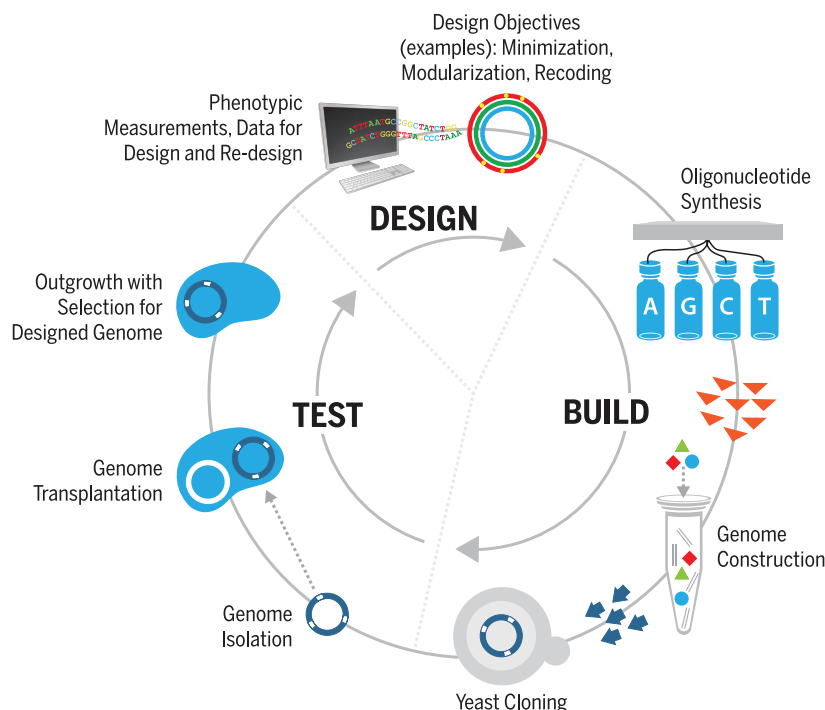
Fig. 1. The JCVI DBT cycle for bacterial genomes. At each cycle, the genome is built as a centromeric plasmid in yeast, then tested by transplantation of the genome into an *M. capricolum* recipient. In this study, our main design objective was genome minimization. Starting from syn1.0, we designed a reduced genome by removing nonessential genes, as judged by global Tn*5* gene disruption. Each of eight reduced segments was tested in the context of a seven-eighths syn1.0 genome and in combination with other reduced segments. At each cycle, gene essentiality was reevaluated by Tn*5* mutagenesis of the smallest viable assembly of reduced and syn1.0 segments that gave robust growth.

M. genitalium, and its doubling rate is about five times as fast.

Preliminary knowledge-based HMG design does not yield a viable cell

In our first attempt to make a minimized cell, we started with syn1.0 (*10*) and used information from the biochemical literature, as well as some transposon mutagenesis data, to produce a rational design. Genes that could be disrupted by transposon insertions without affecting cell viability were considered to be nonessential. Based on ~16,000 transposon 4001 (Tn*4001*) and Tn*5* insertions into the syn1.0 genome, we were able to find and delete a total of 440 apparently nonessential genes from the syn1.0 genome. The resulting HMG design was 483 kbp in size and contained 432 protein genes and 39 RNA genes (database S1 includes a detailed gene list).

In the course of designing the HMG, we developed a simple set of deletion rules that was used throughout the project. (i) Generally, the entire coding region of each nonessential gene was deleted, including start and stop codons. (Exceptions are described below.) (ii) When a cluster of more than one consecutive gene was deleted, the intergenic regions within the cluster were deleted also. (iii) Intergenic regions flanking a deleted gene or a consecutive cluster of deleted genes were retained. (iv) If part of a gene to be deleted overlapped a retained gene,

that part of the gene was retained. (v) If part of a gene to be deleted contained a ribosome binding site or promoter for a retained gene, that part of the gene was retained. (vi) When two genes were divergently transcribed, we assumed that the intergenic region separating them contained promoters for transcription in both directions. (vii) When a deletion resulted in converging transcripts, a bidirectional terminator was inserted, if one was not already present.

Because of the possibility of design flaws, we divided the genome into eight overlapping segments that could be independently synthesized and tested. We previously used this approach to identify a single lethal point mutation in our synthetic syn1.0 construct (*10*). As before, each of the eight designed synthetic segments had a corresponding syn1.0 DNA segment. This allowed untested pieces to be mixed and matched with viable syn1.0 pieces in one-pot combinatorial assemblies or to be purposefully assembled in any specified combination (*16*, *17*). Additionally, each of the eight target segments was moved into a seven-eighths syn1.0 background by recombinase-mediated cassette exchange [(RMCE; (*18*)] (fig. S10) (*9*). Unique restriction sites (NotI sites) flanked each HMG or syn1.0 segment in the resulting yeast strains (fig. S9 and table S12) (*9*). Upon transplantation, we obtained a mycoplasma strain carrying any viable HMG segment (flanked by NotI sites) and eight other strains, each carrying one syn1.0

segment (flanked by NotI sites). This facilitated the production of one-eighth genome segments, because they could be recovered from bacterial cultures, which produce much higher yields of better-quality DNA than yeast (*9*). All eight HMG segments were tested in a syn1.0 background, but only one of the segment designs produced viable colonies (HMG segment 2), and the cells grew poorly.

Perhaps the greatest value that we derived from the HMG work was the refinement of semi-automated DNA synthesis methods by means of error-correcting procedures. We had previously developed a variety of DNA synthesis and assembly methodologies that extend from oligonucleotides to whole chromosomes. In this work, we optimized the methodology to rapidly generate error-free large DNA constructs, starting from overlapping oligonucleotides in a semiautomated DNA synthesis pipeline. This was accomplished by developing robust protocols for (i) single-reaction assembly of 1.4-kbp DNA fragments from overlapping oligonucleotides, (ii) elimination of synthesis errors and facilitation of single-round assembly and cloning of error-free 7-kbp cassettes, (iii) cassette sequence verification to simultaneously identify hundreds of error-free clones in a single run, and (iv) rolling circle amplification (RCA) of large plasmid DNA derived from yeast. Together, these methods substantially increased the rate at which the DBT cycle could be carried out (*9*).

Figure 2 illustrates the general approach that we used for whole-genome synthesis and assembly, using HMG as an example. An automated genome synthesis protocol was established to generate overlapping oligonucleotide sequences, starting from a DNA sequence design (*9*). Briefly, the software parameters include the number of assembly stages, overlap length, maximum oligonucleotide size, and appended sequences to facilitate polymerase chain reaction (PCR) amplification or cloning and hierarchical DNA assembly. About 48 oligonucleotides were pooled, assembled, and amplified to generate 1.4-kbp DNA fragments in a single reaction (figs. S12 and S13) (*9*). The 1.4-kbp DNA fragments were then error corrected, re-amplified, assembled five at a time into a vector, and transformed into *E. coli*. Error-free 7-kbp cassettes were identified on a DNA sequencer (Illumina MiSeq), and as many as 15 cassettes were assembled in yeast to generate one-eighth molecules. Supercoiled plasmid DNA was prepared from positively screened yeast clones, and RCA was performed to generate microgram quantities of DNA for whole-genome assembly in yeast (figs. S14 to S16). This whole-genome synthesis workflow can be carried out in less than three weeks, which is about two orders of magnitude faster than the first reported synthesis of a bacterial genome (by our group) in 2008 (*7*).

Tn5 mutagenesis identifies essential, quasi-essential, and nonessential genes

It was clear from the limited success of our HMG design that we needed a better understanding of which genes are essential versus nonessential. To achieve this, we used Tn*5* mutagenesis (fig. S1).

An initial Tn*5* disruption map was generated by transforming JCVI-syn1.0 ΔRE ΔIS cells [in which all restriction enzyme (RE) genes and six insertion (IS) elements are deleted; table S6] with an activated form of a 988-bp miniature-Tn*5* puromycin resistance transposon (fig. S1) (*9*). Transformed cells were selected on agar plates containing 10 μg/ml of puromycin. About 80,000 colonies, each arising from a single Tn*5* insertion event, were pooled from the plates. A sample of DNA extracted from this "P0" pool was mechanically sheared and analyzed for the sites of Tn*5* insertion using inverse PCR and DNA sequencing. The P0 data set contained ~30,000 unique insertions. To remove slow-growing mutagenized cells, a sample of the pooled P0 cells was serially passaged for more than 40 generations, and DNA was prepared and sequenced to generate a "P4" data set containing ~14,000 insertions. (fig. S2).

Genes fell into three major groups: (i) Genes that were not hit at all, or that were sparsely hit in the terminal 20% of the 3′-end or the first few bases of the 5′-end, were classified as essential ("e-genes") (*5*). (ii) Genes that were hit frequently by both P0 and P4 insertions were classified as nonessential ("n-genes"). (iii) Genes hit primarily by P0 insertions but not P4 insertions were classified as quasi-essential, the deletion of which would cause growth impairments ("i-genes"). Cells with i-gene disruptions spanned a continuum of growth impairment, varying from minimal to severe. To differentiate this growth continuum, we designated i-genes with minimal growth disadvantage as "in-genes" and those with severe growth defect as "ie-genes." Of the 901 annotated protein and RNA coding genes in the syn1.0 genome, 432 were initially classified as n-, 240 as e-, and 229 as i-genes (Fig. 3, A and B, and fig S3).

In viewing a syn1.0 map of P4 insertions (fig. S4), it was evident that nonessential genes tended to occur in clusters far more often than expected by chance. We used deletion analysis to confirm that most of the n-gene clusters could be deleted without losing viability or substantively affecting growth rate (*9*). Individual gene clusters (or, in some cases, single genes) were replaced by the yeast *URA3* marker as follows. Fifty–base pair sequences flanking the gene(s) to be deleted were added to the ends of the *URA3* marker by PCR, and the DNA was introduced into yeast cells carrying the syn1.0 genome. Yeast clones were selected on plates not containing uracil, confirmed by PCR, and transplanted to determine viability. Deletions fell into three classes: (i) those resulting in no successful transplants, indicating deletion of an essential gene; (ii) those resulting in transplants with normal or near-normal growth rates, indicating deletion of nonessential genes; and (iii) those resulting in transplants with slow growth, indicating deletion of quasi-essential genes.

A large number of deletions, including all of the HMG deletions, were individually tested for viability and yielded valuable information for subsequent reduced-genome designs. The transposon insertion data that were available at the time of the HMG design were all collected from passage
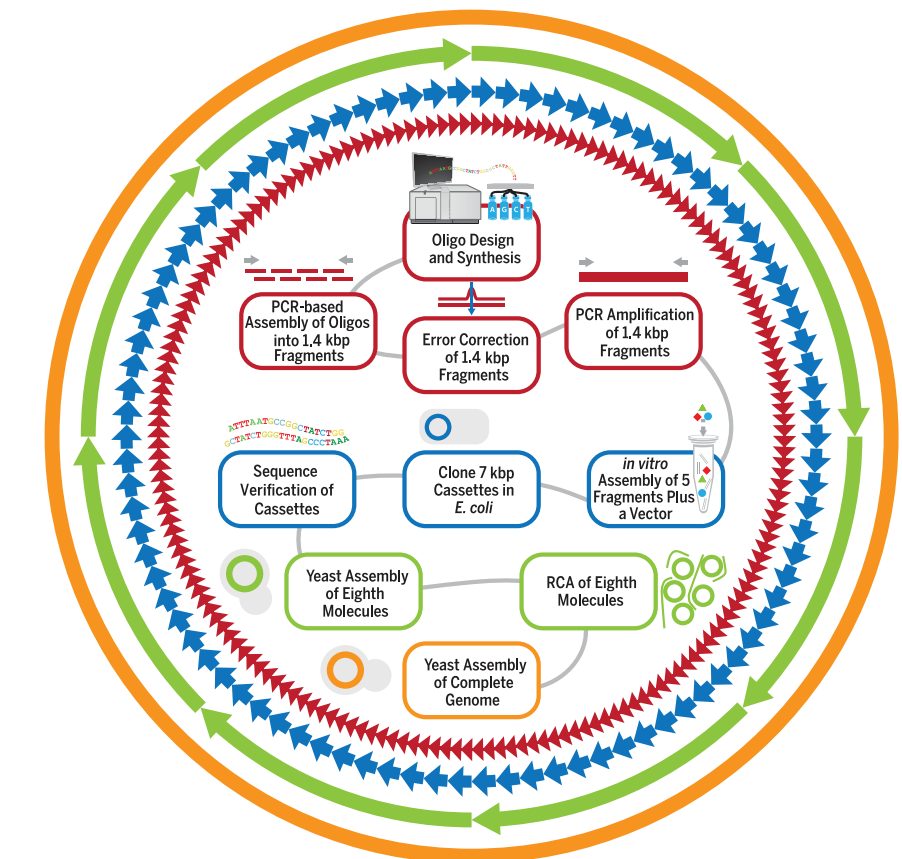

Fig. 2. Strategy for whole-genome synthesis. Overlapping oligonucleotides (oligos) were designed, chemically synthesized, and assembled into 1.4-kbp fragments (red). After error correction and PCR amplification, five fragments were assembled into 7-kbp cassettes (blue). Cassettes were sequence-verified and then assembled in yeast to generate one-eighth molecules (green). The eight molecules were amplified by RCA and then assembled in yeast to generate the complete genome (orange).

P0. Consequently, genes with insertions included the genes that were subsequently characterized as quasi-essential i-genes, so some HMG deletions produced very small colonies or were nonviable.

In parallel with iterations of the DBT cycle, described below, we also took the traditional sequential deletion approach to genome reduction. We performed stepwise scarless deletions (fig. S8 and table S11) (*9*) of medium to large clusters to produce a series of strains with progressively greater numbers of genes removed. Strain D22, with 255 genes and 357 kbp of DNA removed, grew at a rate similar to syn1.0 (table S6). We discontinued this approach when it became clear that the synthesis of redesigned segments at each DBT cycle would yield a minimal cell more quickly. These deletion studies also helped to validate our simple set of deletion rules.

Retaining quasi-essential genes yields viable segments but no viable complete genome

To improve on the design of the HMG, we redesigned a reduced genome using the Tn*5* and deletion data described above. This reduced genome design (RGD1.0) achieved a 50% reduction of syn1.0 by removing ~90% of the n-genes (table S1). In a few cases, n-genes were retained—

specifically, if their biochemical function appeared essential, or if they were singlet n-genes separating two large e- or i-gene clusters. To preserve the expression of genes upstream and downstream of deleted regions, we followed the same design rules that we used in the HMG design.

The eight segments of RGD1.0 were chemically synthesized as described above, and each synthetic reduced segment was inserted into a seven-eighths syn1.0 background in yeast by means of RMCE (fig. S10 and table S13) (*9*). Each one-eighth RGD–seven-eighths syn1.0 genome was then transplanted out of yeast to test for viability. Each of the eight reduced segments produced a viable transplant; however, segment 6 produced only a very small colony in the first 6 days. On further growth over the next 6 days, sectors of faster-growing cells developed (fig. S18). Several isolates of the faster-growing cells were sequenced and found to have destabilizing mutations in a transcription terminator that had been joined to an essential gene when the nonessential gene preceding it was deleted (figs. S19 and S21). Another mutation produced a consensus TATAAT box in front of the essential gene (fig. S20). This illustrates the potential for expression errors when genes are deleted, but it shows that these errors can sometimes be corrected by
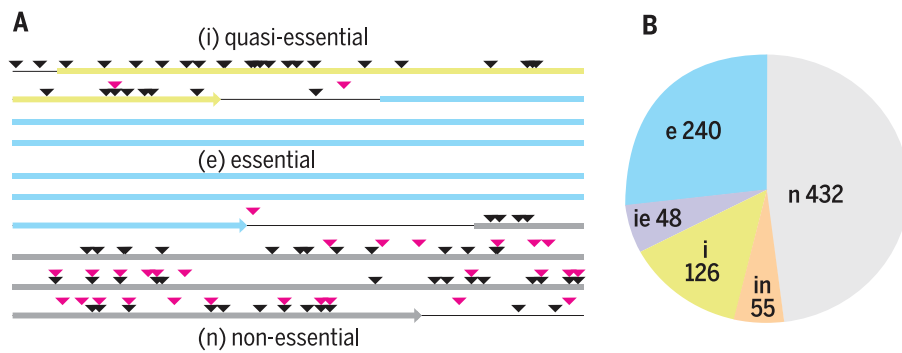
Fig. 3. Classification of gene essentiality by transposon mutagenesis. (**A**) Examples of the three gene classifications, based on Tn*5* mutagenesis data. The region of syn1.0 from sequence coordinates 166,735 to 170,077 is shown. The gene *MMSYN1_0128* (lime arrow) has many P0 Tn*5* inserts (black triangles) and is an i-gene (quasi-essential). The next gene, *MMSYN1_0129* (light blue arrow), has no inserts and is an e-gene (essential). The last gene, *MMSYN1_0130* (gray arrow), has both P0 (black triangles) and P4 (magenta triangles) inserts and is an n-gene (nonessential). Intergenic regions are indicated by black lines. (**B**) The number of syn1.0 genes in each Tn*5*-mutagenesis classification group. The n- and in-genes are candidates for deletion in reduced genome designs.
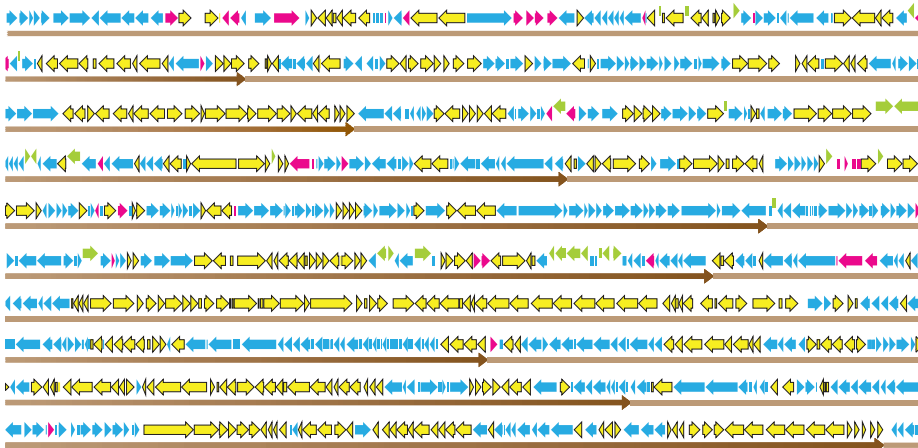

Fig. 4. The three DBT cycles involved in building syn3.0. This detailed map shows syn1.0 genes that were deleted or added back in the various DBT cycles leading from syn1.0 to syn2.0 and finally to syn3.0 (compare with fig. S7). The long brown arrows indicate the eight NotI assembly segments. Blue arrows represent genes that were retained throughout the process. Genes that were deleted in both syn2.0 and syn3.0 are shown in yellow. Green arrows (slightly offset) represent genes that were added back. The original RGD1.0 design was not viable, but a combination of syn1.0 segments 1, 3, 4, and 5 and designed segments 2, 6, 7, and 8 produced a viable cell, referred to as RGD2678. Addition of the genes shown in green resulted in syn2.0, which has eight designed segments. Additional deletions, shown in magenta, produced syn3.0 (531,560 bp, 473 genes). The directions of the arrows correspond to the directions of transcription and translation.

subsequent spontaneous mutation. Ultimately, we identified a promoter that had been overlooked and erringly deleted. When this region was resupplied in accordance with the design rules, cells containing designed segment 6 grew rapidly. This solution was incorporated in later designs.

Despite the growth of cells containing each designed segment, combining all eight reduced RGD1.0 segments, including the self-corrected segment 6, into a single genome did not produce a viable cell when transplanted into *M. capricolum* (*9*). We then mixed the eight RGD1.0 segments with the eight syn1.0 segments to perform combinatorial assembly of genomes in yeast (*9*). A number of completely

assembled genomes were obtained in yeast that contained various combinations of RGD1.0 segments and syn1.0 segments. When transplanted, several of these combinations gave rise to viable cells (table S7). One of these (RGD2678)—containing RGD1.0 segments 2, 6, 7, and 8 and syn1.0 segments 1, 3, 4, and 5, with an acceptable growth rate (105-min doubling time, compared with 60 min for syn1.0)—was analyzed in more detail.

To obtain a viable genome, avoid deleting pairs of redundant genes for essential functions

In bacteria, it is common for certain essential (or quasi-essential) functions to be provided by

more than one gene. The genes may or may not be paralogs. Suppose gene A and gene B each supplies the essential function, E1. Either gene can be deleted without loss of E1, so each gene by itself in a single knockout study is classed as nonessential. However, if both are deleted, the cell will be dead because E1 is no longer provided. Such a lethal combination of mutations is called a "synthetic lethal pair" (*19*). Redundant genes for essential functions are common in bacterial genomes, although less so in genomes that have undergone extensive evolutionary reduction, such as the mycoplasmas. Our biggest design challenge has been synthetic lethal pairs in which gene A has been deleted from one segment and gene B from another segment. Each segment is viable in the context of a seven-eighths syn1.0 background, but when combined, the resulting cell is nonviable, or, in the case of a shared quasi-essential function, grows more slowly. We do not know how many redundant genes for essential functions are present in each of the eight segments, but when RGD1.0 segments 2, 6, 7, and 8 were combined, the cell was viable.

We subjected RGD2678 to Tn*5* mutagenesis and found that some n-genes in the syn1.0 segments 1, 3, 4, and 5 had converted to i- or e-genes in the genetic context of RGD2678 (table S2). This was presumably because these genes encoded essential or quasi-essential functions that were redundant with a gene that had been deleted in RGD2678.

In addition, we examined 39 gene clusters and single genes that had been deleted in the design of RGD1.0 segments 1, 3, 4, and 5 (table S8). These were deleted one at a time in an RGD2678 background (tables S8 and S14) and tested for viability by transplantation. In several cases, this resulted in slow-growth transplant or no transplants, suggesting that they included one or more genes that are functionally redundant with genes that had been deleted in segments 2, 6, 7, or 8.

The combined Tn*5* and deletion data identified 26 genes (tables S2 and S9) as candidates for adding back to RGD1.0 segments 1, 3, 4, and 5 to produce a new RGD2.0 design for these segments (fig. S5 and tables S1 and S2). An assembly was carried out in yeast using the newly designed and synthesized RGD2.0 segments 1, 3, 4, and 5, together with RGD1.0 segments 2, 6, 7, and 8 (tables S7 and S15). This assembly was not viable initially, but we found that substituting syn1.0 segment 5 for RGD2.0 segment 5 resulted in a viable transplant. Working with this strain, we deleted a cluster of genes (*MMSYN1_0454* to *MMSYN1_0474*) from syn1.0 segment 5 and replaced another cluster of genes (*MMSYN1_0483* to *MMSYN1_0492*) with gene *MMSYN1_0154* (figs. S6 and S11 and table S10) (*9*). Gene *MMSYN1_0154* was originally deleted from segment 2 in the RGD1.0 design but was found to increase growth rate when added back to RGD2678. The described revision of syn1.0 segment 5 in the RGD2.0 genetic context yielded a viable cell, which we refer to as JCVI-syn2.0 (abbreviated syn2.0; Fig. 4). With syn2.0, we achieved for the first time a minimized cell with a genome smaller than that of the smallest
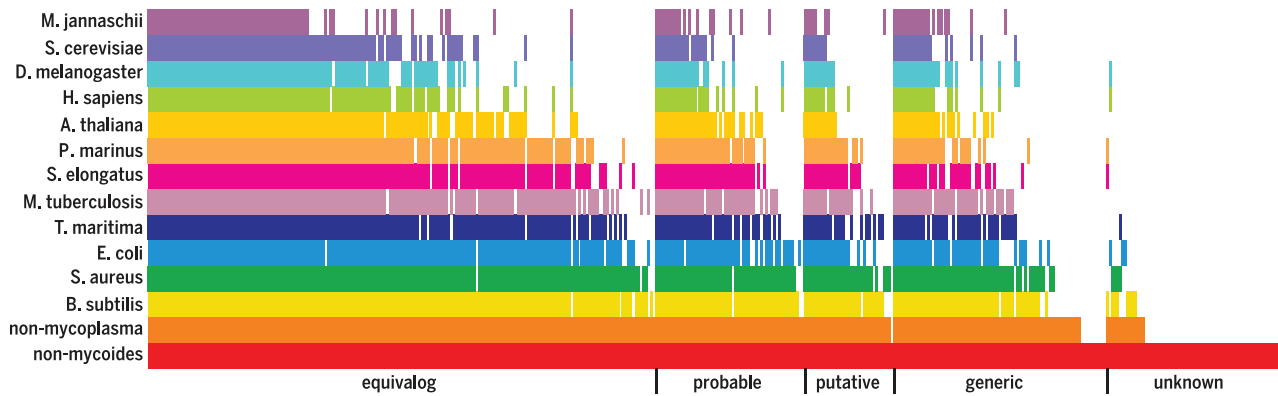
Fig. 5. Map of proteins in syn3.0 and homologs found in other organisms. Searches using BLASTP software were performed for all syn3.0 protein-coding genes against a panel of 14 organisms ranging from non-*Mycoides* mycoplasmas to Archaea. A score of $1e^{-5}$ was used as the similarity cutoff. From left to right, five classes (equivalog, 232 genes; probable, 58 genes; putative, 34 genes; generic, 84 genes; and unknown, 65 genes) proceed from nearly complete certainty about a gene's activity (equivalog) to no functional information (unknown). White space indicates no homologs to syn3.0 in that organism.

known natural bacterium, *M. genitalium*. Syn2.0 doubles in laboratory culture every 92 min. Its total genome size is 576 kbp. It contains 478 protein coding genes and 38 RNA genes from *M. mycoides*, with 12 kbp of vector sequences for selection of the genome and for propagation in yeast and *E. coli*.

Removing 42 additional genes yields an approximately minimal cell, syn3.0

We performed a new round of Tn5 mutagenesis on syn2.0. In this new genetic background, transition of some i-genes to apparent n-genes was expected. At this point, the composition of the P4 serial passage population was depleted of original n-genes; the faster-growing i-gene knockouts predominated and were classified as n-genes by our rules. We classified 90 genes as apparently nonessential. These were subdivided into three groups. The first group contained 26 genes that were frequently classed as i- or e-genes in previous rounds of mutagenesis. The second group contained 27 genes that were classified as i- or borderline i-genes in some of the previous Tn5 studies. The third group contained 37 genes that had previously been classified as nonessential in several iterations of Tn5 mutagenesis involving various genome contexts. To create the new RGD3.0 design, these 37 genes were selected for deletion from syn2.0, along with two vector sequences, *bla* and *lacZ*, and the ribosomal RNA (rRNA) operon in segment 6 (Fig. 4 and table S3).

The eight newly designed RGD3.0 segments were synthesized and propagated as yeast plasmids. These plasmids were amplified in vitro by RCA (*9*). All eight segments were then reassembled in yeast to obtain several versions of the RGD3.0 genome as yeast plasmids (*9*). These assembled RGD3.0 genomes were transplanted out of yeast into an *M. capricolum* recipient cell. Several were viable. One of these, RGD3.0 clone g-19 (table S4), was selected for detailed analysis and named JCVI-syn3.0.

A final round of Tn5 mutagenesis was performed on syn3.0 to determine which genes continue to show Tn5 insertions after serial passaging (P4). Nonessential vector genes and intergenic sequences are the most frequent insertion sites. As expected, cells with insertions in genes that were originally classified as quasi-essential make up almost the whole population of P4 cells that have insertions in mycoplasma

Table 1. Syn1.0 genes listed by functional category and whether they were kept or deleted in syn3.0. Categories with asterisks are mostly kept in syn3.0, whereas those without are depleted in syn3.0. Vector sequences, for selection of the genome and for propagation in other hosts, are not included in these gene tallies.

Functional category	Kept	Deleted
Glucose transport and glycolysis*	15	0
Ribosome biogenesis*	14	1
Protein export*	10	0
Transcription*	9	0
RNA metabolism*	7	0
DNA topology*	5	0
Chromosome segregation*	3	0
DNA metabolism*	3	0
Protein folding*	3	0
Translation*	89	2
RNA (rRNAs, tRNAs, small RNAs)*	35	4
DNA replication*	16	2
Lipid salvage and biogenesis*	21	4
Cofactor transport and salvage*	21	4
rRNA modification*	12	3
tRNA modification*	17	2
Efflux*	7	3
Nucleotide salvage	19	8
DNA repair	6	8
Metabolic processes	10	10
Membrane transport	31	32
Redox homeostasis	4	4
Proteolysis	10	11
Regulation	9	10
Unassigned	79	134
Cell division	1	3
Lipoprotein	15	72
Transport and catabolism of nonglucose carbon sources	2	34
Acylglycerol breakdown	0	4
Mobile elements and DNA restriction	0	73
Total	**473**	**428**

genes. The genes in syn3.0 are predominantly e- or i-genes, based on the original syn1.0 classification. Of these, only the i-genes can tolerate Tn5 insertions without producing lethality. The most highly represented in-, i-, and ie-genes are shown in table S5. It might be possible to remove a few of these in a fourth DBT cycle, but there would probably be further erosion of growth rate. In addition, a dozen genes that were originally classified as nonessential continue to retain that classification (table S5 and database S1).

In syn3.0, 149 genes cannot be assigned a specific biological function

Syn3.0 has 438 protein- and 35 RNA-coding genes. We organized the 473 genes into five classes, based on our level of confidence in their precise functions: equivalog, probable, putative, generic, and unknown (Fig. 5 and database S1). Many of these genes have been studied exhaustively, and their primary biological functions are known.

We used the TIGRfam equivalog family of hidden Markov models (20) to annotate equivalog genes (~49% of the genes). The less certain classes were defined in a stepwise manner (Fig. 5). The probable class included genes that scored well against unambiguous TIGRfam mathematical models but that nevertheless scored below the trusted cutoff. These genes were consistently supported by other lines of evidence. Genomic context and threading alignment to crystal structures both agreed with the assignment. The putative class included genes that were similarly supported by multiple lines of evidence; at the same time, either their scores, genomic context, or alignment to structures with known activities were not convincing. The generic class included genes encoding clearly identifiable proteins (e.g., kinase) but lacking consistent clues as to their substrates or biological role. Unknown genes were those that could not be reliably categorized with regard to a putative activity.

Thus, biological functions could not be assigned for the ~31% of the genes that were placed in the generic and unknown classes. Nevertheless, potential homologs for a number of these were found in diverse organisms. Many of these genes probably encode universal proteins whose functions are yet to be characterized. Each of the five sectors has homologs in species ranging from mycoplasma to humans. However, some of each annotation class is blank, indicating that no homologs for these genes were found among the organisms chosen for display in Fig. 5. Because mycoplasmas evolve rapidly, some of the white space in Fig. 5 corresponds to sequences that have diverged so as to align poorly with representatives from other organisms.

In Table 1, we have assigned syn1.0 genes to 30 functional categories and indicated how many were kept or deleted in syn3.0. Of the 428 deleted genes, the largest group is the functionally unassigned genes; 134 out of 213 have been deleted. All of the 73 mobile element and DNA modification and restriction genes have been removed, as well as most genes encoding lipoproteins (72 out of 87). These three categories alone account for

65% of the deleted genes. In addition, because of the rich growth medium that supplies almost all of the necessary small molecules, many genes involved in transport, catabolism, proteolysis, and other metabolic processes have become dispensable. For example, because glucose is plentiful in the medium, most genes for transport and catabolism of other carbon sources have been deleted (34 out of 36), whereas all 15 genes involved in glucose transport and glycolysis have been retained.

In contrast, almost all of the genes involved in the machinery for reading and expressing the genetic information in the genome and in ensuring the preservation of genetic information across generations have been retained. The first of these two fundamental life processes, the expression of genetic information as proteins, requires the retention of 195 genes in the categories of transcription, regulation, RNA metabolism, translation, protein folding, RNA (rRNA, tRNA, and small RNAs), ribosome biogenesis, rRNA modification, and tRNA modification. The second of these two fundamental processes, the preservation of genome sequence information, requires the retention of 34 genes in the categories of DNA replication, DNA repair, DNA topology, DNA metabolism, chromosome segregation, and cell division. These two processes together require 229 (48%) of the 473 total genes in syn3.0 (Fig. 6).

In addition to the two vital processes just described, another major component of living cells is the cell membrane that separates the outer medium from the cytoplasm and governs molecular traffic into and out of the cell. It is an isolatable structure, and many of the syn3.0 genes code for its protein constituents. Because our minimal cell is largely lacking in biosynthesis of amino acids, lipids, nucleotides, and vitamins, it depends on the rich medium to supply almost all of these required small molecules. This necessitates numerous transport systems within the membrane. In addition, the membrane is rich in lipoproteins. Membrane-related genes account for 84 (18%) of the 473 total syn3.0 genes. Included categories from Table 1 are lipoproteins, cofactor transport, efflux systems, protein transport, and other membrane transport systems.

Lastly, 81 genes (17%) that are primarily involved in cytosolic metabolism are retained in the categories of nucleotide salvage, lipid salvage and biogenesis, proteolysis, metabolic processes, redox homeostasis, transport and catabolism of nonglucose carbon sources, and glucose transport and glycolysis (Fig. 6).

We presume that most of the 79 genes that are not assigned to a functional category belong to one or another of these same four major groups (gene expression, genome preservation, membrane structure and function, and cytosolic metabolism). Among these 79 genes, 65 have completely unknown functions and 24 are classified as generic, such as in the case of a hydrolase for which neither the substrate nor the biological role is discernible. The other 60 of the 84 genes in the generic class were assigned to a functional category on the basis of their generic classification. For example, an ABC transporter is assigned to membrane transport, even though the substrate is unknown. Some of the unassigned essential genes match domains of unknown function that have been found in a wide variety of organisms.

Syn3.0 has a doubling time of 3 hours and is polymorphic in appearance

Comparison of syn3.0 with the starting cell, syn1.0 (Fig. 7A) (9), showed that they have similar colony morphologies, which are characteristic of the natural, wall-less *Mycoplasma mycoides* subspecies *capri* on which the synthetic syn1.0 genome was originally based (10). The smaller colony size of syn3.0 suggested a slower growth rate and perhaps an altered colony architecture on the solid medium. A corresponding reduction in the growth rate of syn3.0 in static liquid culture (Fig. 7B), from a doubling time of ~60 min for syn1.0 to ~180 min, confirmed the lower intrinsic rate of propagation for syn3.0. This rate, however, greatly exceeds the 16-hour doubling time of *M. genitalium* (21).

In contrast to the anticipated reduction in growth rate, we found unexpected changes in macro- and microscopic growth properties of syn3.0 cells. Whereas syn1.0 grew in static culture as nonadherent planktonic suspensions of

Fig. 6. Partition of genes into four major functional groups. Syn3.0 has 473 genes. Of these, 79 have no assigned functional category (Table 1). The remainder can be assigned to four major functional groups: (i) expression of genome information (195 genes); (ii) preservation of genome information (34 genes); (iii) cell membrane structure and function (84 genes); and (iv) cytosolic metabolism (81 genes). The percentage of genes in each group is indicated.
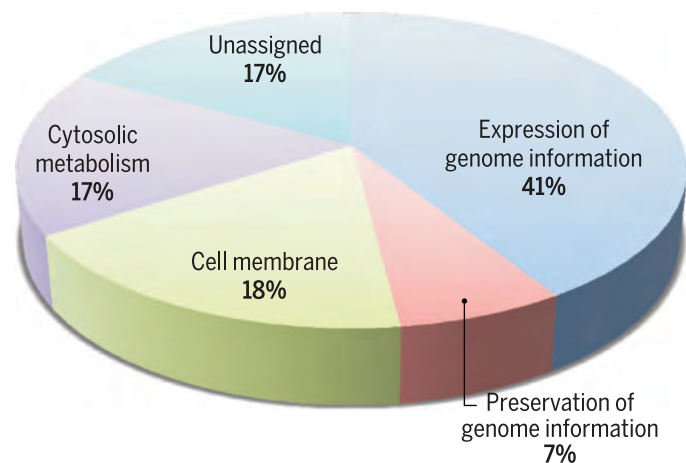
predominantly single cells with a diameter of ~400 nm (*10*), syn3.0 cells formed matted sediments under the same conditions. Microscopic images of these undisturbed cells revealed extensive networks of long, segmented filamentous structures, along with large vesicular bodies (Fig. 7C), which were particularly prevalent at late stages of growth. Both of these structures were easily disrupted by physical agitation, yet such suspensions contained small replicative forms that passed through 0.2-μm filters to render colony-forming units (CFU). This same procedure retained 99.9% of the CFU in planktonic syn1.0 cultures.

Exploring the design of reorganized genomes

To further refine our genome-design rules, we also investigated prospects for logically organizing genomes and recoding them at the nu-

cleotide level. This effort was meant to clarify whether gene order and gene sequence are major contributors to cell viability. The ability to invert large sections of DNA in many genomes has demonstrated that overall, gene order is not critical. We showed that fine-scale gene order is also not a major factor in cell survival. About an eighth of the genome was reconfigured into seven contiguous DNA cassettes, six of which represented known biological systems; the seventh cassette contained genes whose system-level assignment was somewhat equivocal. The vertical bar on the right side of Fig. 8 specifies the biological systems. Individual genes (colored horizontal lines) and intergenic regions (black lines) can be traced from their native location to their new positions by following a line from left to right. Intersecting lines represent a change in the relative position

of two genetic elements. Despite extensive reorganization, the resulting cell grew about as fast as syn1.0, as judged by colony size. Thus, the details of genetic organization impinge upon survival in hypercompetitive natural environments, but the finer details are apparently not critical for life.

Recoding and rRNA gene replacement provide examples of genome plasticity

Our DBT cycle for bacterial genomes allows us to assess the plasticity of gene content in terms of sequence and functionality. This includes testing modified versions of genes that are fundamentally essential for life. We tested whether an altered 16*S* rRNA gene (*rrs*), which is essential, could support life (Fig. 9A). The single copy of the syn3.0 *rrs* gene was designed and synthesized to include seven single-nucleotide changes
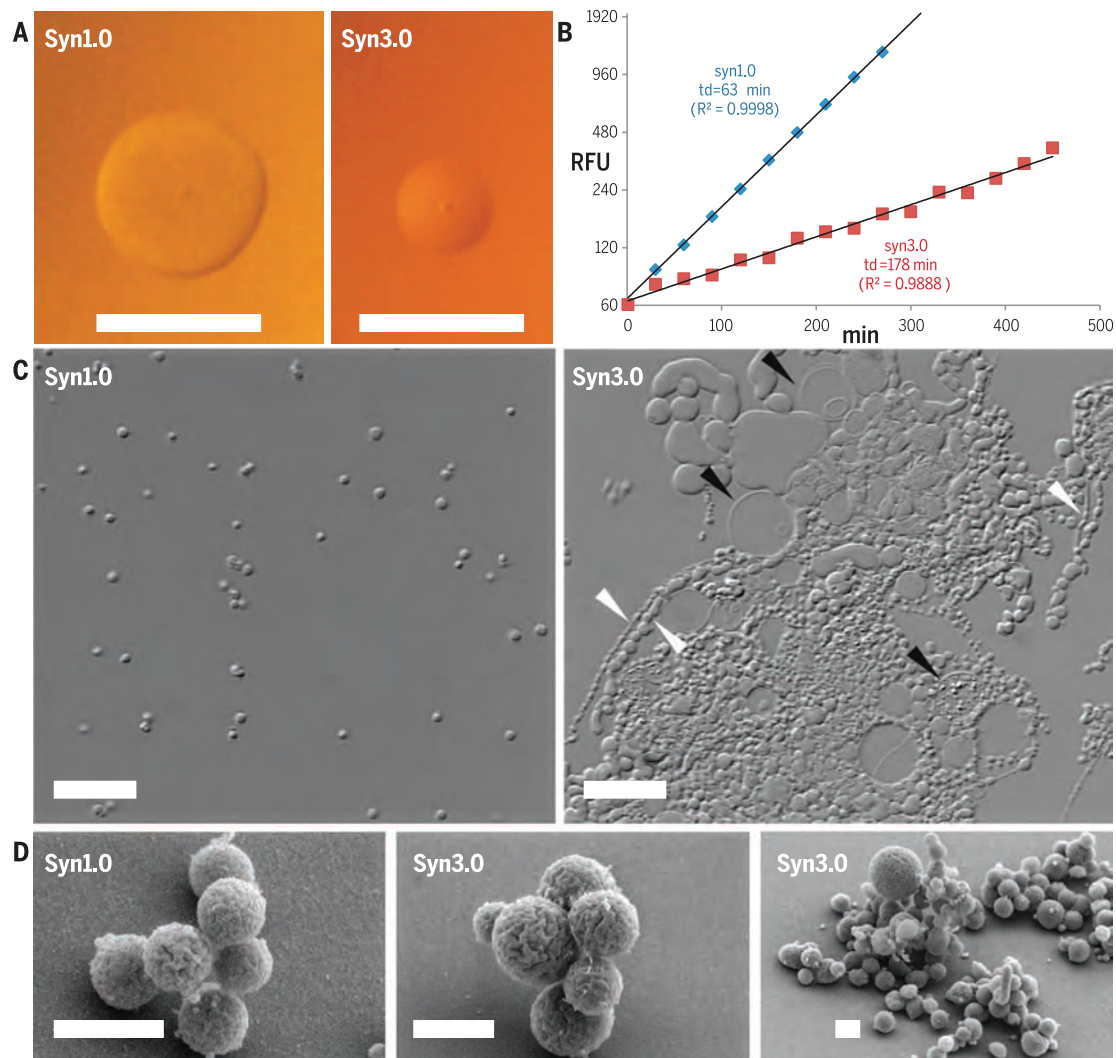

Fig. 7. Comparison of syn1.0 and syn3.0 growth features. (**A**) Cells derived from 0.2 μm–filtered liquid cultures were diluted and plated on agar medium to compare colony size and morphology after 96 hours (scale bars, 1.0 mm). (**B**) Growth rates in liquid static culture were determined using a fluorescent measure (relative fluorescent units, RFU) of double-stranded DNA accumulation over time (minutes) to calculate doubling times (td). Coefficients of determination (R^2) are shown. (**C**) Native cell morphology in liquid culture was imaged in wet mount preparations by means of differential interference contrast microscopy (scale bars, 10 μm). Arrowheads indicate assorted forms of segmented filaments (white) or large vesicles (black). (**D**) Scanning electron microscopy of syn1.0 and syn3.0 (scale bars, 1 μm). The picture on the right shows a variety of the structures observed in syn3.0 cultures.

corresponding to those contained in the *rrs* gene of *M. capricolum*. In addition, we replaced helix h39 (35 nucleotides) with that from a phylogenetically distant *E. coli rrs* counterpart. This unique 16*S* gene was successfully incorporated into syn3.0 without noticeably affecting growth rate. Some other variants of the *rrs* gene were constructed but proved nonviable. This demonstrates our ability to test the plasticity of a gene sequence and, at the same time, provides a watermark by which to quickly identify this strain.

We also tested the underlying codon usage principles in the *M. mycoides* genome, which has extremely high adenine and thymine (AT) content. *M. mycoides* uses TGA as a codon for the amino acid tryptophan, instead of a stop codon, and occasionally uses nonstandard start codons; in addition, the codon usage is heavily biased toward high-AT content. We modified this uncommon codon usage in a 5-kbp region containing three essential genes (*era*, *recO*, and *glyS*) to determine its role. Specifically, we modified this region to include (i) *M. mycoides* codon adaptation index (CAI), but with the unusual start codons recoded and tryptophan encoded by the TGG codon, instead of by TGA; (ii) *E. coli* CAI, with tryptophan still encoded by TGA; or (iii) *E. coli* CAI with standard codon usage (TGG encoding tryptophan) (Fig. 9B). Unexpectedly, we found that all three versions were functional and resulted in *M. mycoides* cells without noticeable growth differences. However, large-scale changes in codon usage may need to accompany modifications in the tRNA dosage levels to ensure efficient translation.

Discussion and conclusions

Genomics is moving from a descriptive phase, in which genomes are sequenced and analyzed, to a synthetic phase, in which whole genomes can be built by chemical synthesis. As the detailed genetic requirements for life are discovered, it will become possible to design whole genomes from first principles, build them by chemical synthesis, and then bring them to life by installation into a receptive cellular environment. We have applied this whole-genome design and synthesis approach to the problem of minimizing a cellular genome.

A minimal cell is usually defined as a cell in which all genes are essential. This definition is incomplete, because the genetic requirements for survival, and therefore the minimal genome size, depend on the environment in which the cell is grown. The work described here has been conducted in medium that supplies virtually all the small molecules required for life. A minimal genome determined under such permissive conditions should reveal a core set of environment-independent functions that are necessary and sufficient for life. Under less permissive conditions, we expect that additional genes will be required.

There is a large body of literature concerning the minimal cell concept and minimal sets of essential genes in a number of organisms [for a review, see (*22*)]. Work in the area has focused on comparative genomic analyses and on experiments in which genes are individually knocked out or disrupted by transposon insertion. Such studies identify a core of essential genes, often about 250 in number. But this is not a set of genes that is sufficient to constitute a viable cellular genome, because redundant genes for essential functions are scored as nonessential in these studies.

In contrast, we set out to construct a minimal cellular genome in order to experimentally determine a core set of genes for an independently replicating cell. We designed a genome using genes from *M. mycoides* JCVI-syn1.0 (*10*). This mycoplasma cell has several advantages for this purpose. First, the mycoplasmas already have very small genomes. They have evolved from gram-positive bacteria with larger genomes by losing genes that are unnecessary in their niche as mammalian parasites. They are already far along an evolutionary pathway to a minimal genome, and consequently they are likely to have fewer functionally redundant genes than other bacteria. We also have a highly developed set of tools for building this genome and for assembling and manipulating the genome as an extra chromosome in yeast.

Our initial attempt to design a minimal genome was based on the current collective knowledge of molecular biology, in combination with limited data concerning transposon disruption of genes, which provided additional information about gene essentiality. This information was particularly valuable with respect to the genes of unknown function. Specific experimental proposals for minimal genome construction have been made solely on the basis of accumulated knowledge concerning the genes that are involved in fundamental biological processes (*14*). Our HMG was assembled from eight segments
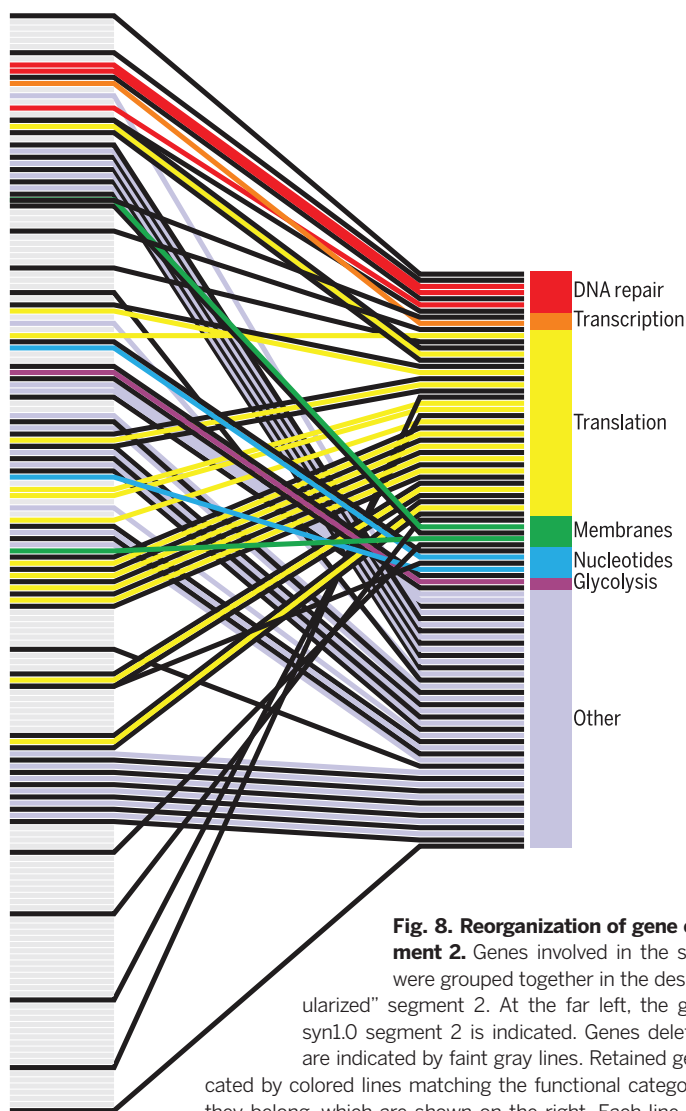

Fig. 8. Reorganization of gene order in segment 2. Genes involved in the same process were grouped together in the design for "modularized" segment 2. At the far left, the gene order of syn1.0 segment 2 is indicated. Genes deleted in syn3.0 are indicated by faint gray lines. Retained genes are indicated by colored lines matching the functional categories to which they belong, which are shown on the right. Each line connects the position of the corresponding gene in syn1.0 with its position in the modularized segment 2. Black lines represent intergenic sequences containing promoters or transcriptional terminators.

and proved nonviable, although one of the segments (segment 2) was functional when tested in the context of the other seven syn1.0 segments. These results convinced us that initially, we did not have sufficient knowledge to design a functional minimal genome from first principles. Therefore, to obtain better information concerning gene essentiality, we made major improvements in our transposon mutagenesis methods.

To produce a genome containing all of the essential and quasi-essential genes, we developed a DBT cycle for bacterial genomes (Fig. 1).

Any design, viable or not, can be built in yeast and then tested to determine whether it can function as the genome of a viable bacterium. After four DBT cycles (genome designs HMG, RGD1.0, RGD2.0, and RGD3.0), we obtained a viable genome with all eight segments reduced,
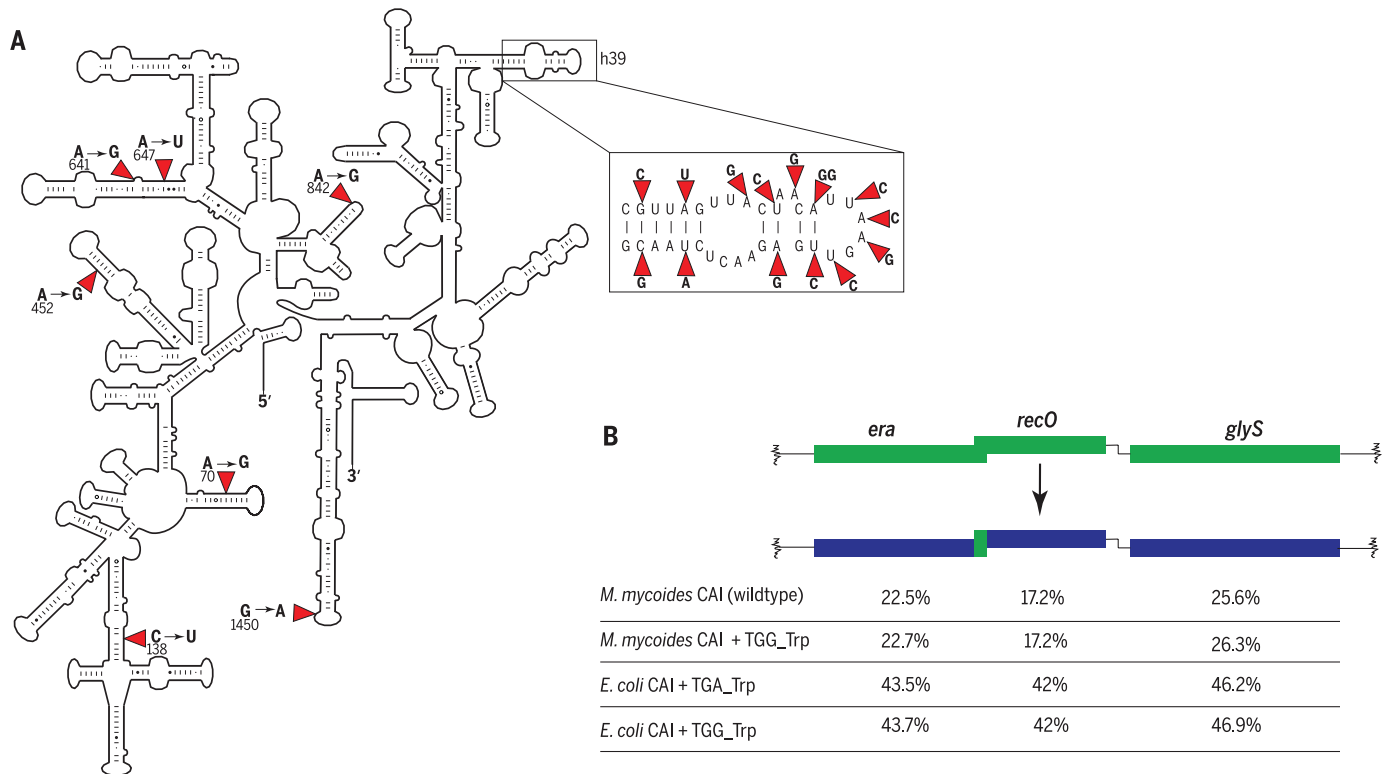
Fig. 9. Gene content and codon usage principles, tested using the DBT cycle. (**A**) Secondary structure of the modified *rrs* gene that was successfully incorporated into the syn3.0 genome; this gene was carrying *M. capricolum* mutations and had its h39 (inset) swapped with that of *E. coli*. Positions with nucleotide changes are indicated by red arrows, and *E. coli* numbering is used to indicate the position of *M. capricolum* mutations. (**B**) The sequences of the essential genes *era*, *recO*, and *glyS* were modified in three different ways: using *M. mycoides* CAI with TGG encoding tryptophan, *E. coli* CAI with TGG encoding tryptophan, or *E. coli* CAI with TGA encoding tryptophan. GC content of the wild-type and modified genes is noted. The JCat codon adaptation tool was used for this exercise (www.jcat.de) to optimize the three open reading frames, with the exception of the overlapping gene fragment. Green and purple indicate wild-type and codon-optimized sequences, respectively.

Table 2. Reduced genomes resulting from the DBT cycles, ultimately leading to syn3.0. Column 1 indicates the round of genome design (dashes indicate the starting genome, syn1.0), column 2 gives the size of the designed genome (in kilobase pairs), and column 3 gives the number of mycoplasma genes in the design. Column 4 shows the genome composition for key viable cell strains; for nonviable designs, a viable strain with the highest number of segments from the design is shown, as well as a more robust alternative for RGD1.0 (fourth row) and a smaller derivative for RGD2.0 (sixth row, syn2.0). Column 5 gives the size of the genome corresponding to column 3, and column 6 shows a quantitative or qualitative estimate of the growth rate of cells with the genome described in column 4.

1. Genome design	2. Design size	3. Number of design genes	4. Cellular genome segment composition for key viable strains	5. Cellular genome size	6. Growth rate
-		901	syn1.0: all eight syn1.0 segments	1079 kbp	Doubling time, td = 60 min
HMG	483 kbp	460	HMG segment 2 + 7/8 syn1.0	1003 kbp	Slow-growing
RGD1.0	544 kbp	483	RGD1.0 segments 2,6,7,8 + syn1.0 segments 1,3,4,5	758 kbp	Slow-growing
"	"		RGD1.0 segments 1,2,4,6,8 + syn1.0 segments 3,5,7	718 kbp	Slow-growing
RGD2.0	575 kbp	512	RGD2.0 segments 1,2,3,4,6,7,8 + syn1.0 segment 5	617 kbp	?
"	"		syn2.0: RGD2.0 segments 1,2,3,4,6,7,8 + syn1.0 segment 5 with genes *MMSYN1_0454* to *-0474* and *MMSYN1_0483* to *-0492* deleted	576 kbp	td = 92 min
RGD3.0	531 kbp	473	syn3.0: all eight segments of RGD3.0	531 kbp	td = 180 min

syn3.0. Table 2 summarizes the process leading to syn3.0. The first three designs did not yield complete viable cellular genomes. But in each case, one or more of the eight segments yielded a viable genome when combined with syn1.0 segments for the remainder of the genome. The composition of several of these intermediate strains is listed in the table. The viable syn3.0 cell is our best approximation to a minimal cell. We obtained another strain with a genome smaller than any free-living cell found in nature (syn2.0). This cell has a genome consisting of seven RGD2.0 segments plus a syn1.0 segment 5 with 31 genes deleted.

Syn3.0 has a 531-kbp genome that encodes 473 gene products. It is substantially smaller than *M. genitalium* (580 kbp), which has the smallest genome of any naturally occurring cell that has been grown in pure culture. The syn3.0 genome contains the core set of genes that are required for cellular life, but it is only half the size of syn1.0 (*10*). In comparing the HMG to the subsequently derived viable syn3.0 genome, we found agreement among 329 deleted genes and 365 retained genes. However, 111 genes that were kept in syn3.0 were deleted from HMG, and 100 genes deleted in syn3.0 were kept in HMG. The discrepancies were primarily due to the sparseness and quality of the initial transposon data, which resulted in incorrect identification of a number of essential or nonessential genes and did not identify quasi-essential genes that affect growth rate (discussed below). One example of the importance of the quasi-essential gene classification was the case of four genes (*MMSYN_0008* to *MMSYN_0011*) that make up the transport system for nucleosides. The original annotation of the system was as a ribose/galactose ABC transporter, which led us to target it for deletion in the HMG. Our initial transposon data showed that all four genes were hit heavily and appeared to confirm that the genes were nonessential. However, in later transposon mapping experiments, P0 transposon data confirmed that they were hit heavily, but after serial passage to deplete slow-growing cells, all four genes received zero hits, confirming that they were quasi-essential and should have been retained.

Gene content of syn3.0

Syn3.0 is a working approximation of a minimal cell. Our first synthetic cell, syn1.0, contained 901 mycoplasma genes plus some watermarks and vector sequences. Of these, 428 have been removed in syn3.0, leaving 438 protein-coding genes and 35 RNA genes. More genes could probably be removed while retaining viability, but it seems likely that growth rate would be compromised. The slower growth rate of syn3.0 is not due to the removal of one of the rRNA operons. We also constructed a strain with the same gene complement, except that it retained both rRNA operons, and this strain grew at close to the same rate as syn3.0.

The largest group of genes retained in syn3.0 is involved in gene expression (195 genes, 41%). Approximately equal numbers of genes are involved in the cell membrane (84 genes, 18%) and in metabolism (81 genes, 17%). During reductive evolution as a mycoplasma, many biosynthetic genes were lost and replaced by transporters residing in the membrane, resulting in a trade-off between these two categories. A relatively small number of genes function in the replication of the genome and the preservation of genomic information through cell division (36 genes, 7%). Unexpectedly, there are 79 genes (17%) that we have been unable to assign to a functional category. Of these, 19 are in the essential category (e-genes), 36 are needed for rapid growth (i- or ie-genes), and 24 are nonessential or nearly so (n- or in-genes). We presume that most of these will fall into one of the four major categories described above (gene expression, membrane structure and function, cytosolic metabolism, and genome preservation), but it seems likely that some of them may perform previously undescribed biological functions. In particular, 13 of the 19 functionally unassigned e-genes are of completely unknown function. Some of these match genes of unknown function in other bacteria or even in eukaryotes, and these are prime candidates for proteins with novel functions. Genes of unknown function that are required in syn3.0 and present in most organisms must represent nearly universal functions and thus can provide biological insights. Likewise, unknown genes without homologs may be novel, or they may represent unusual sequences but well-understood functions. In contrast to the wholly unknown genes, it is easy to oversimplify a gene's putative role in cell survival if it has a generic functional assignment. For example, some of the numerous hydrolases and kinases will undoubtedly contribute to processes such as nucleotide or cofactor salvage. The question is, will all of the generic functions of the unknown genes be so commonplace, or do some of them represent fundamentally new processes? There are genes whose generic annotations are perplexing even though they are needed for survival. For example, there are six different efflux systems, encoded by the genes *MMSYN1_0034*, *MMSYN1_0371* and *MMSYN1_0372*, *MMSYN1_0399*, *MMSYN1_0531*, *MMSYN1_0639*, and *MMSYN1_0691*. Except for the heterodimeric pair, *MMSYN1_0371* and *MMSYN1_0372*, which may be a flippase, the substrates and functions of these proteins are unclear. It is somewhat disconcerting to imagine that all of these exclude or remove toxic substances. Similarly, a rather complex pathway (*23*) for producing and exporting glycoglycerolipids was required. Although there is some evidence that galactofuranose residues are important for membrane integrity (*24*), a detailed explanation for the biological role fulfilled by glycoglycerolipids remains obscure.

Phenotype of the syn3.0 cell

The replication of genomic information and its coordinated distribution into segregated membrane-bound cellular compartments are hallmarks of extant living systems that are commonly considered to be among the attributes that define cellular life (*25*). The minimal requirements for this process are not known, but evidence from disparate fields of study suggests that mechanisms far simpler than the complex division apparatus in most eubacteria may be sufficient.

First, several types of bacterial cells, both natural (*26*) and experimentally manipulated (*27*, *28*), have been shown to divide in the absence of key cytoskeletal components, most notably the FtsZ cytoskeletal scaffolding and force-generating component. Through our empirically based design process, a nonessential gene cluster present in syn1.0 (*MMYSYN1_0520* through *MMYSYN1_0522*) was removed during construction of syn3.0 cells. This contained orthologs of *ftsZ* and *sepF* [encoding a membrane-anchoring component that interacts with FtsZ (*29*)]. An adjacent gene, *ftsA*, which is reported to share some redundant functions with *sepF* in other systems (*30*), remained essential in progressively reduced constructs that lacked *ftsZ*.

Second, completely synthetic lipid vesicles have been shown to spontaneously segregate without the involvement of macromolecular scaffolding or catalysis (*31*). In propagating cell wall–deficient bacteria, alteration of the lipid content and properties of the plasma membrane have been shown to elicit analogous membrane vesicle segregation (*32*). In several wall-less mycoplasma species, filamentous and large-vesicle morphotypes similar to those in syn3.0 have long been observed under certain growth conditions, depending in part on the nature of lipid precursors available to these cells (*33*). Ultimately, understanding the genetic and mechanistic basis for the phenotype of syn3.0 propagation may shed light on the minimal requirements for segregation of the membrane-bounded cellular compartment that is essential for a living cell.

Use of the DBT cycle for applications other than genome minimization

Our main focus here has been the application of the whole-genome DBT cycle to a specific problem, the construction of a minimal cellular genome. However, the approach we describe can be applied to the construction of a cell with any desired properties. For example, a cell could be designed with added metabolic pathways (*34*), an altered genetic code (*35*), or dramatically altered gene arrangements. We have begun to design genomes with modified 16S rRNA sequences and to assess the effects of dramatic alterations in codon usage. Application of our DBT cycle is limited only by our ability to produce designs with a reasonable chance of success. With increasing knowledge of the functions of essential genes that are presently unknown, and with increasing experience in reorganizing the genome, we expect that our design capabilities will strengthen. The ability to design cells in which the function of every gene is known should facilitate complete computational modeling of the cell (*36*). This would make it possible to calculate the consequences of adding pathways for the production of useful products, such as drugs or industrial chemicals, and would lead to greater efficiency in development.

Methods summary

Our methods for the identification of nonessential genes by global Tn*5* mutagenesis, manipulation

of bacterial genomes in yeast by the scarless TREC (tandem repeat coupled with endonuclease cleavage) deletion method, synthesis and assembly of reduced genomes, genome transplantation, microscopic analysis of cells with reduced genomes, and observation of their growth characteristics are described in detail in the supplementary materials. General information about our methods, accompanied by specific references to the supplementary materials, is included throughout the text.

REFERENCES AND NOTES

1. H. J. Morowitz, The completeness of molecular biology. *Isr. J. Med. Sci.* **20**, 750–753 (1984). pmid: 6511349
2. C. M. Fraser *et al.*, The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–404 (1995). doi: 10.1126/science.270.5235.397; pmid: 7569993
3. R. D. Fleischmann *et al.*, Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995). doi: 10.1126/science.7542800; pmid: 7542800
4. A. R. Mushegian, E. V. Koonin, A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10268–10273 (1996). doi: 10.1073/pnas.93.19.10268; pmid: 8816789
5. C. A. Hutchison III *et al.*, Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**, 2165–2169 (1999). doi: 10.1126/science.286.5447.2165; pmid: 10591650
6. J. I. Glass *et al.*, Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 425–430 (2006). doi: 10.1073/pnas.0510013103; pmid: 16407165
7. D. G. Gibson *et al.*, Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008). pmid: 18218864
8. C. Lartigue *et al.*, Genome transplantation in bacteria: Changing one species to another. *Science* **317**, 632–638 (2007). doi: 10.1126/science.1144622; pmid: 17600181
9. Materials and methods are available as supplementary materials on *Science* Online.
10. D. G. Gibson *et al.*, Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010). pmid: 20488990
11. M. Juhas, D. R. Reuß, B. Zhu, F. M. Commichau, *Bacillus subtilis* and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiology* **160**, 2341–2351 (2014). doi: 10.1099/mic.0.079376-0; pmid: 25092907
12. G. Pósfai *et al.*, Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**, 1044–1046 (2006). doi: 10.1126/science.1126439; pmid: 16645050
13. Y. Suzuki *et al.*, Bacterial genome reduction using the progressive clustering of deletions via yeast sexual cycling. *Genome Res.* **25**, 435–444 (2015). doi: 10.1101/gr.182477.114; pmid: 25654978
14. A. C. Forster, G. M. Church, Towards synthesis of a minimal cell. *Mol. Syst. Biol.* **2**, 45 (2006). doi: 10.1038/msb4100090; pmid: 16924266
15. M. Lluch-Senar *et al.*, Defining a minimal cell: Essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. *Mol. Syst. Biol.* **11**, 780 (2015). pmid: 25609650

16. A. Ramon, H. O. Smith, Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. *Biotechnol. Lett.* **33**, 549–555 (2011). doi: 10.1007/s10529-010-0455-x; pmid: 21107654
17. C. Merryman, D. G. Gibson, Methods and applications for assembling large DNA constructs. *Metab. Eng.* **14**, 196–204 (2012). doi: 10.1016/j.ymben.2012.02.005; pmid: 22629570
18. V. N. Noskov, L. Ma, S. Chen, R. Y. Chuang, Recombinase-mediated cassette exchange (RMCE) system for functional genomics studies in *Mycoplasma mycoides*. *Biol. Proced. Online* **17**, 6 (2015). doi: 10.1186/s12575-015-0016-8; pmid: 25774095
19. T. Dobzhansky, Genetics of natural populations. XIII. Recombination and variability in populations of *Drosophila pseudoobscura*. *Genetics* **31**, 269–290 (1946).
20. D. H. Haft, J. D. Selengut, O. White, The TIGRFAMs database of protein families. *Nucleic Acids Res.* **31**, 371–373 (2003). doi: 10.1093/nar/gkg128; pmid: 12520025
21. J. S. Jensen, H. T. Hansen, K. Lind, Isolation of *Mycoplasma genitalium* strains from the male urethra. *J. Clin. Microbiol.* **34**, 286–291 (1996). pmid: 8789002
22. R. Gil, F. J. Silva, J. Peretó, A. Moya, Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* **68**, 518–537 (2004). doi: 10.1128/MMBR.68.3.518-537.2004; pmid: 15353568
23. J. Romero-García, C. Francisco, X. Biarnés, A. Planas, Structure-function features of a *Mycoplasma* glycolipid synthase derived from structural data integration, molecular simulations, and mutational analysis. *PLOS ONE* **8**, e81990 (2013). doi: 10.1371/journal.pone.0081990; pmid: 24312618
24. E. Schieck *et al.*, Galactofuranose in *Mycoplasma mycoides* is important for membrane integrity and conceals adhesins but does not contribute to serum resistance. *Mol. Microbiol.* **99**, 55–70 (2016). pmid: 26354009
25. J. C. Xavier, K. R. Patil, I. Rocha, Systems biology perspectives on minimal and simpler cells. *Microbiol. Mol. Biol. Rev.* **78**, 487–509 (2014). doi: 10.1128/MMBR.00050-13; pmid: 25184563
26. R. Bernander, T. J. Ettema, FtsZ-less cell division in archaea and bacteria. *Curr. Opin. Microbiol.* **13**, 747–752 (2010). doi: 10.1016/j.mib.2010.10.005; pmid: 21050804
27. R. Mercier, Y. Kawai, J. Errington, Excess membrane synthesis drives a primitive mode of cell proliferation. *Cell* **152**, 997–1007 (2013). doi: 10.1016/j.cell.2013.01.043; pmid: 23452849
28. M. Lluch-Senar, E. Querol, J. Piñol, Cell division in a minimal bacterium in the absence of ftsZ. *Mol. Microbiol.* **78**, 278–289 (2010). doi: 10.1111/j.1365-2958.2010.07306.x; pmid: 20735775
29. S. Gola, T. Munder, S. Casonato, R. Manganelli, M. Vicente, The essential role of SepF in mycobacterial division. *Mol. Microbiol.* **97**, 560–576 (2015). doi: 10.1111/mmi.13050; pmid: 25943244
30. R. Duman *et al.*, Structural and genetic analyses reveal the protein SepF as a new membrane anchor for the Z ring. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E4601–E4610 (2013). doi: 10.1073/pnas.1313978110; pmid: 24218584
31. J. C. Blain, J. W. Szostak, Progress toward synthetic cells. *Annu. Rev. Biochem.* **83**, 615–640 (2014). doi: 10.1146/annurev-biochem-080411-124036; pmid: 24606140

32. R. Mercier, P. Domínguez-Cuevas, J. Errington, Crucial role for membrane fluidity in proliferation of primitive cells. *Cell Reports* **1**, 417–423 (2012). doi: 10.1016/j.celrep.2012.03.008; pmid: 22832271
33. S. Razin, B. J. Cosenza, M. E. Tourtellotte, Filamentous growth of mycoplasma. *Ann. N. Y. Acad. Sci.* **143**, 66–72 (1967). doi: 10.1111/j.1749-6632.1967.tb27645.x; pmid: 4861145
34. M. J. Smanski *et al.*, Functional optimization of gene clusters by combinatorial design and assembly. *Nat. Biotechnol.* **32**, 1241–1249 (2014). doi: 10.1038/nbt.3063; pmid: 25419741
35. M. J. Lajoie *et al.*, Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013). pmid: 24136966
36. J. R. Karr *et al.*, A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401 (2012). doi: 10.1016/j.cell.2012.05.044; pmid: 22817898

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/351/6280/aad6253/suppl/DC1
Materials and Methods
Figs. S1 to S21
Tables S1 to S15
References (37–59)
Database S1

13 October 2015; accepted 5 February 2016
10.1126/science.aad6253

Design and synthesis of a minimal bacterial genome

Clyde A. Hutchison III *et al.*
Science **351**, (2016);
DOI: 10.1126/science.aad6253

Science
AAAS

This copy is for your personal, non-commercial use only.