



Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome

Daniel G. Gibson, *et al.*

Science **319**, 1215 (2008);

DOI: 10.1126/science.1151721

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 20, 2008):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/319/5867/1215>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/1151721/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/319/5867/1215#related-content>

This article **cites 29 articles**, 16 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/319/5867/1215#otherarticles>

This article has been **cited by** 1 article(s) on the ISI Web of Science.

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome

Daniel G. Gibson, Gwynedd A. Benders, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Holly Baden-Tillson, Jayshree Zaveri, Timothy B. Stockwell, Anushka Brownley, David W. Thomas, Mikkel A. Algire, Chuck Merryman, Lei Young, Vladimir N. Noskov, John I. Glass, J. Craig Venter, Clyde A. Hutchison III, Hamilton O. Smith*

We have synthesized a 582,970–base pair *Mycoplasma genitalium* genome. This synthetic genome, named *M. genitalium* JCVI-1.0, contains all the genes of wild-type *M. genitalium* G37 except MG408, which was disrupted by an antibiotic marker to block pathogenicity and to allow for selection. To identify the genome as synthetic, we inserted “watermarks” at intergenic sites known to tolerate transposon insertions. Overlapping “cassettes” of 5 to 7 kilobases (kb), assembled from chemically synthesized oligonucleotides, were joined by in vitro recombination to produce intermediate assemblies of approximately 24 kb, 72 kb (“1/8 genome”), and 144 kb (“1/4 genome”), which were all cloned as bacterial artificial chromosomes in *Escherichia coli*. Most of these intermediate clones were sequenced, and clones of all four 1/4 genomes with the correct sequence were identified. The complete synthetic genome was assembled by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*, then isolated and sequenced. A clone with the correct sequence was identified. The methods described here will be generally useful for constructing large DNA molecules from chemically synthesized pieces and also from combinations of natural and synthetic DNA segments.

Mycoplasma genitalium is a bacterium with the smallest genome of any independently replicating cell that has been grown in pure culture (1, 2). Approximately 100 of its 485 protein-coding genes are nonessential under optimal laboratory conditions when individually disrupted (3, 4). However, it is not known which of these 100 genes are simultaneously dispensable. We proposed that one approach to this question would be to produce reduced genomes by chemical synthesis and introduce them into cells to test their capacity to provide the essential genetic

functions for life (4, 5). This paper describes a necessary step toward these goals—the complete chemical synthesis of a mycoplasma genome.

The actual synthesis and assembly of this genome presented a formidable technical challenge. Although chemical synthesis of genes has become routine, the only completely synthetic genomes so far reported have been viral (6–8). The largest previously published synthetic DNA that we are aware of is a 32-kb polyketide gene cluster (9). To accomplish assembly of the 582,970–base pair (bp) *M. genitalium* JCVI-1.0

genome, we needed to establish convenient and reliable methods for the assembly and cloning of much larger synthetic DNA molecules.

Strategy for synthesis and assembly. The native 580,076-bp *M. genitalium* genome sequence (*Mycoplasma genitalium* G37 ATCC 33530 genomic sequence; accession no. L43967) (3) was partitioned into 101 cassettes of approximately 5 to 7 kb in length (Fig. 1) that were individually synthesized, verified by sequencing, and then joined together in stages. In general, cassette boundaries were placed between genes so that each cassette contained one or several complete genes. This will simplify the future deletion or manipulation of the genes in individual cassettes. Most cassettes overlapped their adjacent neighbors by 80 bp; however, some segments overlapped by as much as 360 bp. Cassette 101 overlapped cassette 1, thus completing the circle.

Short “watermark” sequences were inserted in cassettes 14, 29, 39, 55 and 61. Watermarks are inserted or substituted sequences used to identify or encode information into DNA. This information can be either in noncoding or coding sequences (10–12). Most commonly, watermarking has been used to encrypt information within coding sequences without altering the amino acid sequences (10, 11). We opted to insert watermark sequences at intergenic sites because synonymous codon changes may have substantial biological effects. Our watermarks are located at sites known to tolerate transposon insertions, so we expect minimal biological effects. They allow us to easily differentiate the synthetic genome from the native genome (2, 13).

In addition to the watermarks, a 2514-bp insertion in gene MG408 (*msrA*), which includes an aminoglycoside resistance gene, was placed in

The J. Craig Venter Institute, Rockville, MD 20850, USA.

*To whom correspondence should be addressed. E-mail: hsmith@jcv.org

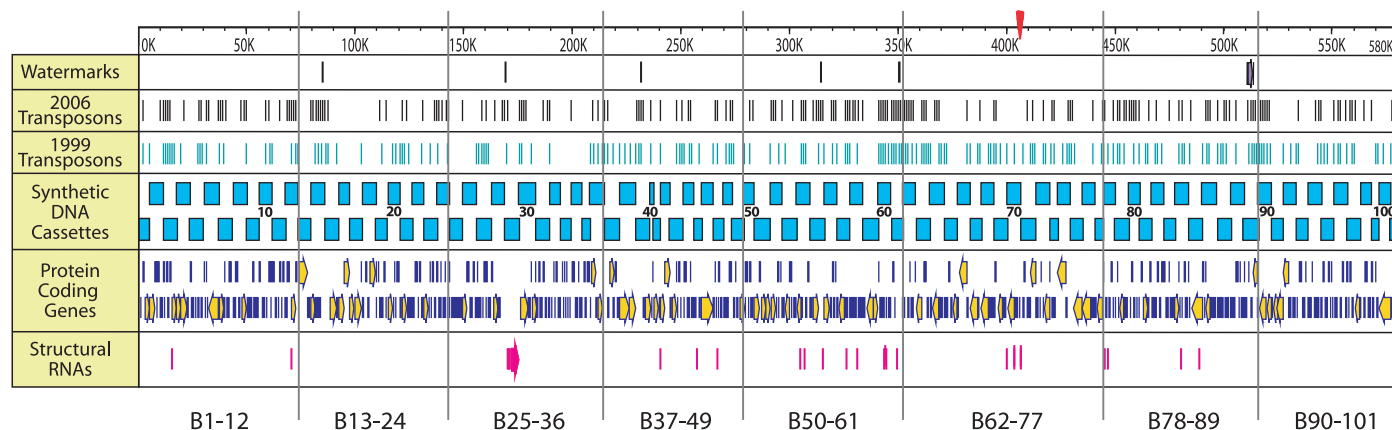


Fig. 1. Linear GenomBench (Invitrogen) representation of the circular 582,970-bp *M. genitalium* JCVI-1.0 genome. Features shown include locations of watermarks and the aminoglycoside resistance marker, viable Tn4001 transposon insertions determined in our 1999 and 2006 studies (3, 4), overlapping synthetic DNA cassettes that comprise the whole genome sequence, 485 *M. genitalium* protein-

coding genes, 43 *M. genitalium* rRNA, tRNA, and structural RNA genes, and B-series assemblies (Fig. 2). The red dagger on the genome coordinates line shows the location of the yeast/*E. coli* shuttle vector insertion. Table S1 lists cassette coordinates; table S2 has FASTA files for all 101 cassettes; table S3 lists watermark coordinates; table S4 lists the sequences of the watermarks.

cassette 89. It has been shown that a strain with this specific defect in this virulence factor cannot adhere to mammalian cells, thus eliminating pathogenicity in the best available model systems (14). The synthetic genome with all of the above insertions is 582,970 bp in length. Figure 1 is a map of the *M. genitalium* JCVI-1.0 genome showing various features such as genes, ribosomal and tRNAs, transposon insertions (3, 4), watermark locations, and cassette positions.

Synthesis of DNA the size of our cassettes has become a commodity, so we opted to outsource their production, principally to Blue Heron Technology, but also to DNA2.0 and GENEART. The main challenges in this project were the assembly and cloning of synthetic DNA molecules larger than those previously reported. We planned a five-stage assembly as diagrammed in Fig. 2. In the first stage, sets of four neighboring cassettes were assembled by in vitro recombination and joined to a bacterial artificial chromosome (BAC) vector DNA to form circularized recombinant plasmids with ~24-kb inserts. For example, cassettes 1 to 4 were joined together to form the A1-4 assembly, cassettes 5 to 8 were assembled to form A5-8, and so forth. In the second stage, the 25 A-series assemblies were taken three at a time to form B-series assemblies. For example, B1-12 was constructed from A1-4, A5-8, and A9-12. This reduced the 25 A-assemblies to only 8 B-assemblies, each about 1/8 of a genome in size (~72 kb). In the third stage, the 1/8-genome B-assemblies were taken two at a time to make four C-assemblies, each approximately 1/4-genome (~144 kb) in size. These first three stages of assembly were done by in vitro recombination and cloned into *E. coli*. We encountered difficulties in carrying out the planned assembly and cloning of the half and whole synthetic genomes in *E. coli*. For this reason, the final assemblies were carried out in *S. cerevisiae* by transformation-associated recombination (TAR) cloning.

Assembly of synthetic cassettes by in vitro recombination. Figure 3 illustrates the reaction used for the first stage of assembly of the overlapping cassettes. Recombinant plasmids bearing the individual cassette DNA inserts were cleaved with the appropriate type IIS restriction enzymes, which cleave outside of their recognition site to one side, to release the insert DNA. After phenol-chloroform extraction and ethanol precipitation, the cassettes were used without removing vector DNA. The essential steps of the reaction are (i) the overlapping DNA molecules are digested with a 3' exonuclease to expose the overlaps, (ii) the complementary overlaps are annealed, and (iii) the joints are repaired. Polymerase chain reaction (PCR) amplification was used to produce a unique BAC vector for the cloning of each assembly, with terminal overlaps to the ends of the assembly. Each PCR primer includes an overlap with one end of the BAC, a Not I restriction site, and an overlap with one end of the cassette assembly. Cassettes were assembled, four at a time, in the presence of the appropriate BAC vector.

Because the *M. genitalium* JCVI-1.0 genome does not contain a Not I site, all of the assemblies can be released intact from the BAC.

For example, the assembly A66-69 was constructed by mixing together equimolar amounts of the four cassette DNAs and the linear PCR-

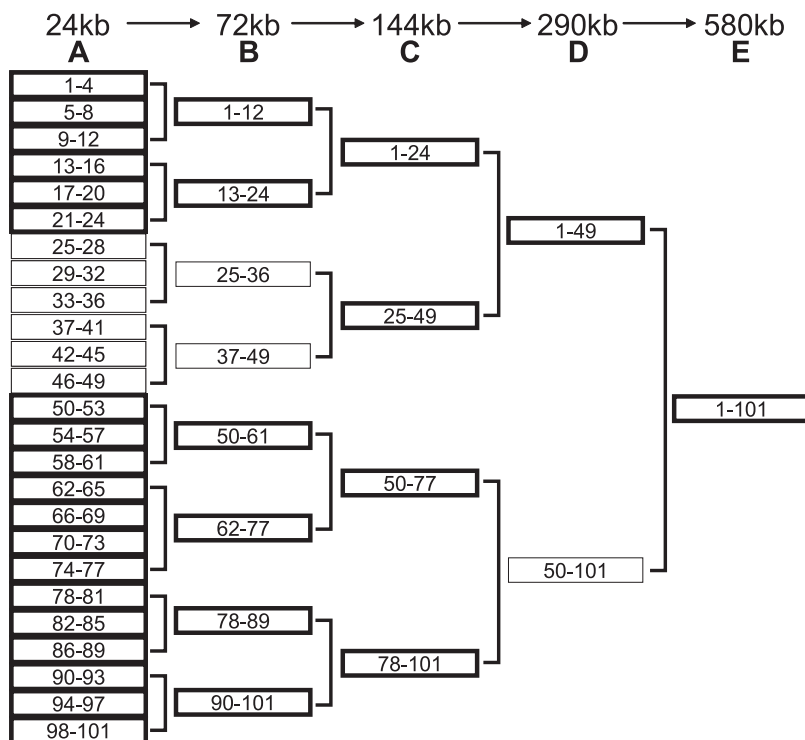


Fig. 2. A plan for the five-stage assembly of the *M. genitalium* chromosome. In the first stage of assembly, four cassettes are joined to make an A-series assembly approximately 24 kb in length (assembly 37-41 contained five cassettes). In the next stage, three A-assemblies are joined together to make a total of eight ~72-kb B-series assemblies (assembly B62-77 contained four A-series assemblies). The eighth-genome B-assemblies are taken two at a time to make quarter-genome C-series assemblies. These assemblies were all made by in vitro recombination (see Fig. 3) and cloned into *E. coli* using BAC vectors. Half-genome and whole-genome assemblies were made by in vivo yeast recombination. Assemblies in bold boxes were sequenced to verify their correctness. For the final molecule, the D-series half molecules were not employed. Rather, we assembled the whole molecule from the four C-series quarter molecules.

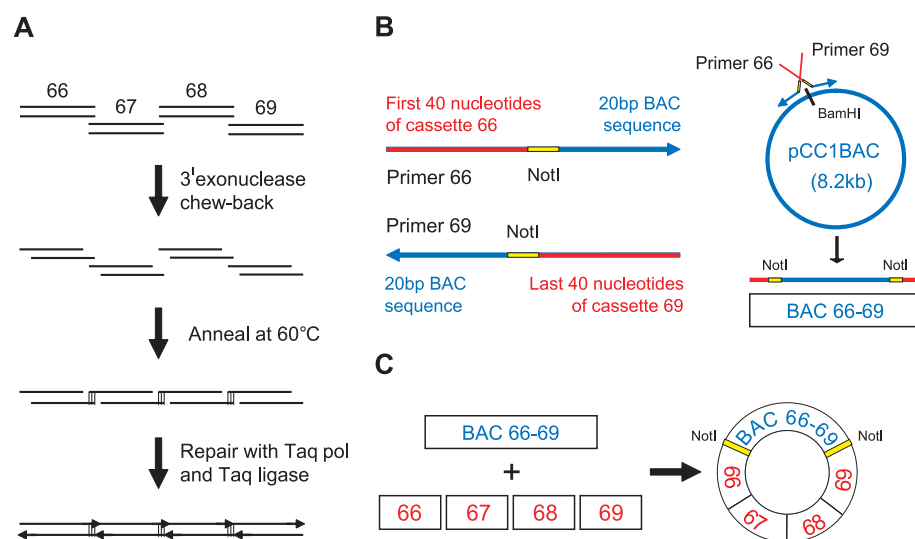


Fig. 3. Assembly of cassettes by in vitro recombination. (A) Diagram of steps in the in vitro recombination reaction, using the assembly of cassettes 66 to 69 as an example. (B) BAC vector is prepared for the assembly reaction by PCR amplification using primers as illustrated. The linear amplification product, after gel purification, is included in the assembly reaction of (A), such that the desired assembly is circular DNA containing the four cassettes and the BAC DNA as depicted in (C).

amplified BAC vector specific for this assembly, BAC 66-69, as described above (Fig. 3) (13). The 3' ends of the mixture of duplex vector and cassette DNAs were then digested to expose the overlap regions using T4 polymerase in the absence of 2'-deoxyribonucleoside-5'-triphosphates (dNTPs). The T4 polymerase was inactivated by incubation at 75°C, followed by slow cooling to anneal the complementary overlap regions. The annealed joints were repaired using Taq polymerase and Taq ligase at 45°C in the presence of all four dNTPs and nicotinamide adenine dinucleotide (NAD). [See the supporting online material for details of the assembly reaction (13)].

Samples of the assembly reactions were subjected to field inversion gel electrophoresis (FIGE) to evaluate the success of the assembly (Fig. 4) (13). Additional samples were electroporated into *E. coli* EPI300 (Epicentre) or DH10B (Invitrogen) cells and plated on LB agar plates containing 12.5 µg/ml chloramphenicol. Colonies appeared after 24 to 48 hours. A-series assembly reactions generally yielded several thousand colonies. B- and C-series assembly reactions generally yielded several hundred colonies. Colonies were picked and BAC DNA was prepared from cultures using an alkaline lysis procedure. The DNA was then cleaved with Not I and analyzed by FIGE to verify the correct sizes of the assemblies. Typically, more than 90% of the A-series and 50% of the B- and C-series clones contained a BAC with the correct insert size. Clones with the correct size were preserved as frozen glycerol stocks. Some of the cloned assemblies were sequenced to ascertain the accuracy of the synthesis as indicated by bold boxes in Fig. 2.

The 25 A-series assemblies and all the larger assemblies were cloned in the pCC1BAC vector from Epicentre (Fig. 3). The pCC1BAC clones could be propagated at the single-copy level in EPI300 cells and then induced to 10 copies per cell according to the Epicentre protocol. Induced 100-ml cultures yielded up to 200 µg of BAC DNA. The assembly inserts in the BACs were immediately flanked on each side by a Not I site such that cleavage efficiently yielded the insert DNA with part of the Not I site attached at each end (the *M. genitalium* genome has no Not I sites). When the Not I-flanked assemblies were used in higher assemblies, the 3' portion of the Not I site (2 nucleotides) was removed by the chew-back reaction. The 5' portion of the Not I site produced a 6-nucleotide overhang after annealing, but the overhang was removed during repair by the Taq polymerase 5' exonuclease activity (Fig. 5).

B-series assemblies were constructed from Not I-digested A-series clones, and C-series assemblies were constructed from Not I-digested B-series assemblies. It was generally not necessary to gel-purify the inserts from the cleaved vector DNA because, without complementary overhangs, they were inactive in subsequent reactions. FIGE analyses of the assembly reactions for A66-69, B50-61, and C25-49 are shown in Fig. 4, A to C. Figure 4D shows a FIGE analysis of the sizes of these cloned inserts.

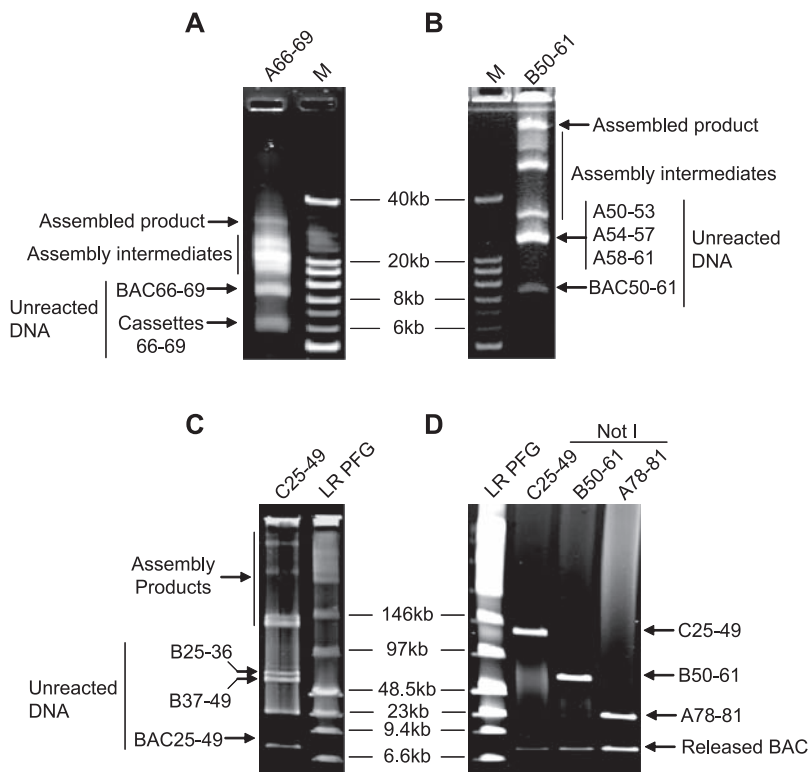


Fig. 4. Gel electrophoretic analyses of selected examples of A-, B-, and C-series assembly reactions and their cloned products. (A to C) A 10-µl sample of the chew-back assembly reactions for A66-69 (A), B50-61 (B), and C25-49 (C) was loaded onto a 0.8% Invitrogen E-gel (A and B) or onto a 1% BioRad Ready Agarose Mini Gel (C), then subjected to FIGE using the U-5 program (A and B) or the U-9 program (13) (C). See (13) for FIGE parameters. (D) Sizes of the Not I-cleaved assemblies were determined by FIGE analysis as in (C). The DNA size standards were the 1-kb extension ladder (M; Invitrogen) and the low-range PFG marker (LR PFG; NEB). Bands were visualized with a BioRad Gel Doc (A and B) or using an Amersham Typhoon 9410 Fluorescence Imager (C and D). Unreacted cassette, A-series, B-series, and BAC DNA, incomplete assembly products, and full-length assembly products are indicated.

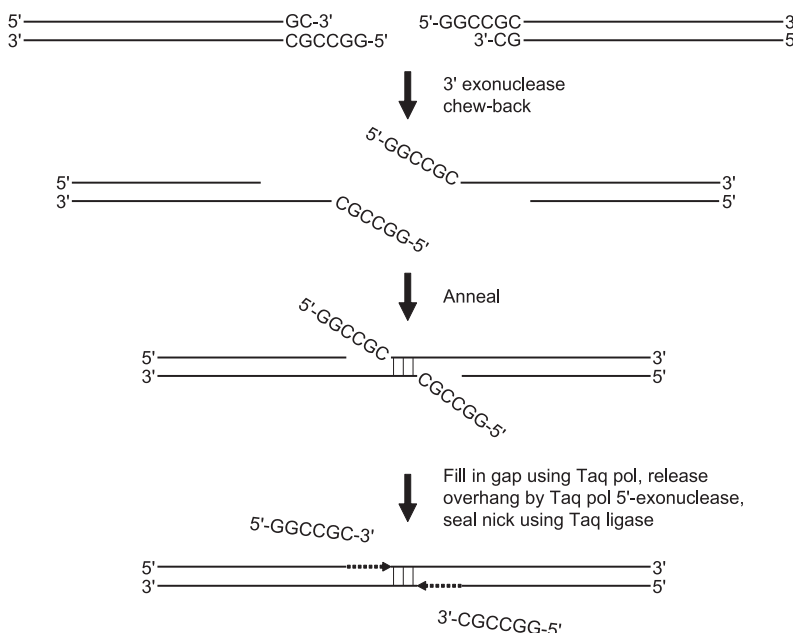


Fig. 5. Repair of annealed junctions containing nonhomologous 3' and 5' Not I sequences. The 3' GC nucleotides are removed during the chew-back reaction. In the repair reaction the 5'-GGCCGC Not I overhangs are removed by the 5'-exonuclease activity contained in the Taq polymerase.

Assembly by in vivo recombination in yeast.

We were unable to obtain half-genome clones in *E. coli* by the in vitro recombination procedure described above. We suspected that larger assemblies were simply not stable in *E. coli*. We had already experienced difficulty in maintaining the C78-101 clone except in Stb14 *E. coli* cells (Invitrogen). Thus, we turned to *S. cerevisiae* as a cloning host. Yeast will support at least 2 Mb of DNA in a linear centromeric yeast artificial chromosome (YAC) (15) and has been used to clone sequences that are unstable in *E. coli* (16).

Linear YAC clones are usually constructed by ligation of an insert into a restriction enzyme cloning site (17). An improvement on this method uses cotransformation of overlapping insert

and vector DNAs into yeast spheroplasts, where they are joined by homologous recombination (Fig. 6A). This produces circular clones and is known as TAR cloning (18). A TAR clone, like a linear YAC, contains a centromere and thus is maintained at chromosomal copy number along with the native yeast genome. However, unlike linear YACs, circular TAR clones can be readily separated from the linear yeast chromosomes.

To assemble quarter genomes into halves and wholes in yeast, we used the pTARBAC3 vector (19). This vector contains both YAC and BAC sequences (Fig. 6B). The vector was prepared using a strategy similar to the one described above for BAC vectors, but longer, 60 bp, overlaps were generated at the termini (20). In TAR

cloning, recombination is stimulated by a factor of about 20 at double-stranded breaks (21). Thus, we integrated the vector at the cleaved intergenic BsmB I site in C50-77. This resulted in the elimination of the four bases of the BsmB I 5' overhang. The DNA to be transformed consisted of six pieces (one vector, two fragments of quarter 3, and quarters 1, 2, and 4). To obtain a full-sized genome as an insert in pTARBAC3, a single yeast cell must take up all six pieces and assemble them by homologous recombination.

Transformation of the yeast cells was performed using a published method (22). Vector and inserts were transformed at approximately equimolar amounts. Transformants were screened first by PCR and then by Southern blot with mycoplasma-specific probes (13). Positive clones were tested for stability by Southern blotting of subclones. Based on these assays, at least 17 out of 94 transformants screened carried a complete synthetic genome. One of these clones, sMgTARBAC37, was selected for sequencing.

TAR cloning was also performed with each of the four sets of two adjacent quarter genomes, as well as with a mixture of C1-24, C25-49, and C50-77. DNAs from transformants of these various experiments were isolated and electroporated into *E. coli* (23). In this way, we obtained BAC clones of the sizes expected for D1-49, D50-101, and assemblies 25-77 and 1-77. Of these, D1-49 was chosen for sequencing, and it was correct. Our lack of success in obtaining these clones directly by in vitro recombination may have been due to inefficient circularization of large DNA molecules or to breakage during the handling of the DNA before transforming *E. coli*.

Recovery of the synthetic *M. genitalium* genome from yeast and confirmation of its sequence. A 600-kb YAC is about 5% of the total DNA in a yeast cell. To enrich sMgTARBAC37 for sequencing, we used a strategy of total DNA isolation in agarose, selective restriction digestion of yeast host chromosomes, and electrophoretic separation of these linear fragments from the large, relatively electrophoretically immobile circular molecules (13). Figure 6 shows the size and purity of the sMgTARBAC37 DNA that was used to prepare a library for sequencing. The sMgTARBAC37 DNA was sequenced by the random shotgun method to ~7X coverage. The sequence exactly matched our designed genome and can be accessed at GenBank accession number CP000925.

Error management. Our objective was to produce a cloned synthetic genome 582,970 bp in length with exactly the sequence we designed. This was not trivial, because differences (errors) between the actual and designed sequence can arise in several ways. An error could be present in the sequence that was supplied to the contractors. The contractors could produce cassettes with errors. Errors could occur during repair of the assembly junctions. Propagation of assemblies in *E. coli* or yeast could lead to errors. In the latter two instances, errors could occur at a late stage of the assembly. At various points during the ge-

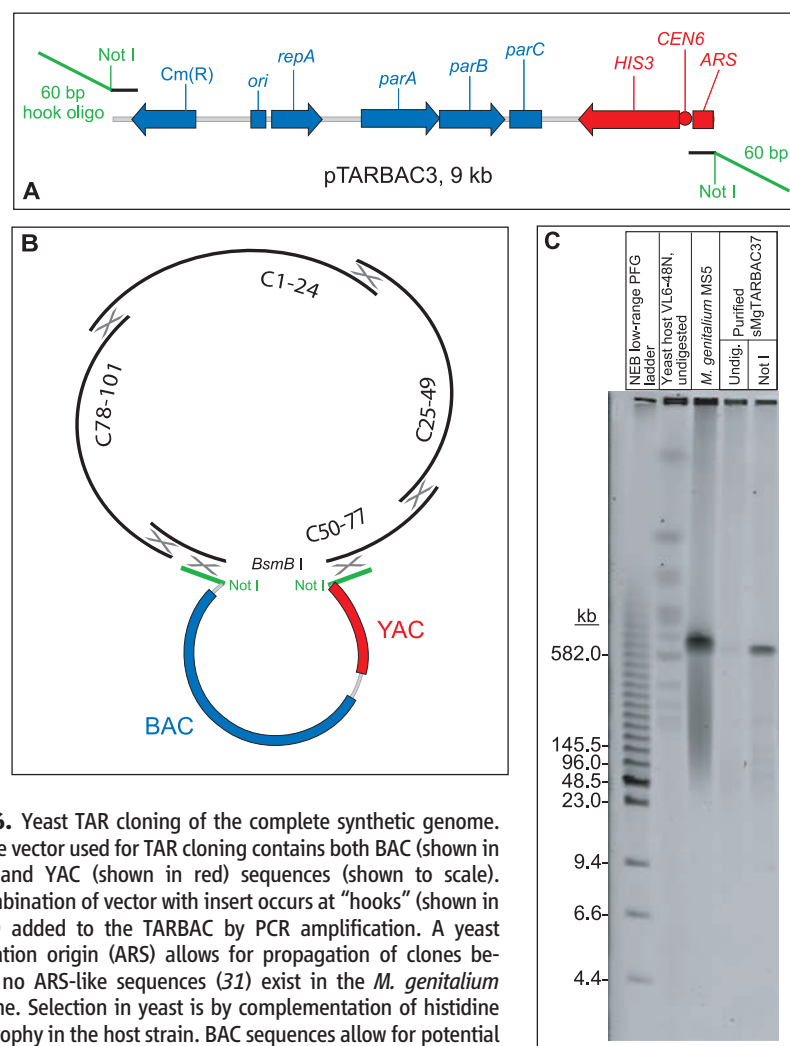


Fig. 6. Yeast TAR cloning of the complete synthetic genome. (A) The vector used for TAR cloning contains both BAC (shown in blue) and YAC (shown in red) sequences (shown to scale). Recombination of vector with insert occurs at “hooks” (shown in green) added to the TARBAC by PCR amplification. A yeast replication origin (ARS) allows for propagation of clones because no ARS-like sequences (31) exist in the *M. genitalium* genome. Selection in yeast is by complementation of histidine auxotrophy in the host strain. BAC sequences allow for potential electroporation into *E. coli* of clones purified from yeast. (B) *M. genitalium* JCVI-1.0 quarter genomes were purified from *E. coli*, Not I-digested, and mixed with a TARBAC vector for cotransformation into *S. cerevisiae*, where recombination at overlaps from 60 to 264 bp combined the six fragments into a single clone. The TARBAC was inserted into the BsmB I site in C50-77. (C) CHEF gel analysis of the complete synthetic genome clone sMgTARBAC37. Size markers are the low-range pulsed field gel marker (NEB), the host yeast strain VL6-48N (32), undigested, and the native *M. genitalium* MS5 (14) genome, which contains an insertion disrupting the MG408 gene. Purified sMgTARBAC37 from the preparation used for sequencing is shown both undigested and Not I digested. The Not I digest releases the 583-kb synthetic *M. genitalium* genome from the vector. The undigested sample confirms the circularity of the clone, because a 592-kb circle was too large to electrophorese into the gel. A small fraction of the clone was broken, and these linear molecules were detected by a faint signal.

nome assembly, clones were sequenced (Fig. 2). Most of the assemblies were exactly correct; however, in our *E. coli* clones, we encountered at least one example of each of the error types described above. Several errors were repaired by rebuilding assemblies, but in some cases other methods were used.

During sequence verification of the C50-77 quarter molecule, two single-bp deletions were detected. One was traced back to a synthesis error in cassette 65, and a corrected version was supplied by the contractor. An error in cassette 55 resulted from an incorrect sequence transmitted to the contractor. This cassette was corrected by replacing a restriction fragment containing the error with a newly synthesized fragment. C50-77 was then reassembled and sequenced. The two errors were corrected, but two new single-base substitution errors appeared. Taq polymerase misincorporation in a joint region likely caused one of these errors. The other remains unexplained but could have arisen during propagation in *E. coli*. One final reassembly yielded the correct quarter molecule that was used to assemble the whole chromosome.

Concluding remarks. We designed, chemically synthesized, and assembled the entire *M. genitalium* JCVI-1.0 chromosome, which is based on *M. genitalium* G37, and cloned it in yeast. This construct is more than an order of magnitude larger than previously reported chemically synthesized DNA products (9). The final product is built from $\sim 10^4$ synthetic oligonucleotides, each ~ 50 nucleotides in length, and is the largest chemically synthesized molecule of defined structure of which we are aware.

Very large nonsynthetic constructs have previously been produced from bacterial genomic DNA using *in vivo* methods. Itaya *et al.* (24) developed a method for cloning megabase-sized segments of DNA into the *Bacillus subtilis* genome using the natural transformation system of this bacterium. They cloned almost all of the *Synechocystis* PCC6803 genome as a set of four separate ~ 800 - to 900 -kb fragments into the *B. subtilis* chromosome by a reiterated “inch worm” process to generate a composite genome. Using a similar approach, this group recently reported the assembly and cloning of PCR products into an extrachromosomal vector (25). Holt *et al.* (26) have described how one might reassemble a fragmented donor genome from *Haemophilus influenzae* piecewise into *E. coli* using, for example, lambda Red recombination. All these methods used sequential stepwise addition of segments to reconstruct a donor genome within a recipient bacterium. The sequential nature of these constructions makes such methods slower than the purely hierarchical scheme that we employed (Fig. 2). Other approaches have been proposed that could use hierarchical assembly strategies (27).

The Itaya (24) and Holt (26) groups found that the bacterial recipient strains were unable to tolerate some portions of the donor genome to be cloned, for example, ribosomal RNA (rRNA)

operons. In contrast, we found that the *M. genitalium* rRNA genes could be stably cloned in *E. coli* BACs. We were able to clone the entire *M. genitalium* genome, and also to assemble the four quarter-genomes in a single step, using yeast as a recipient host. However, we do not yet know how generally useful yeast will be as a recipient for bacterial genome sequences.

For the assembly of our synthetic genome, we used both *in vitro* and *in vivo* recombination methods. The efficiency of our *in vitro* procedure declined as the assemblies became larger. We were able to obtain quarter-genome, but not half-genome, clones using the *in vitro* methods described above. Some of the larger products in the half-genome reactions appeared to be concatamers that formed in preference to circles. In addition, large BACs (>100 kb) transform *E. coli* less efficiently. Sheng *et al.* (28) found that a 240-kb BAC transformed less efficiently by a factor of 30 than an 80-kb BAC in the same recipient strain of *E. coli*.

To complete the assembly, we turned to *in vivo* yeast recombination. Previous work had established that relatively large segments (>100 kb) of the human genome can be cloned in a circular yeast vector if the vector carries terminal homologies (“hooks”) that flank the human genome segment (18). If yeast is cotransformed with a mixture of vector and high molecular weight human DNA, clones containing the human DNA segment are obtained. Recombination is stimulated by breaks at the point of homology. We surmised that our overlapping pieces, each of which has terminal 80-bp homologies to adjacent pieces, might be efficiently assembled and then joined to overlapping vector DNA by the transformation-associated recombination mechanism in yeast (20). We found that two quarters could be efficiently cloned to produce half genomes in the yeast vector. More surprisingly, four quarters, one of which had been cleaved at the vector insertion point, could be recombined and cloned to yield whole genomes. This implies that some of the competent yeast cells are capable of taking up as many as six separate DNA molecules and recombining them into a circular DNA molecule. This raises the question: How many pieces can be assembled in yeast in a single step? The ability to assemble many pieces of DNA in a single reaction could be very useful for generation of combinatorial genome libraries. In the future, it may be advantageous to make greater use of yeast recombination to assemble chromosomes.

We are currently using a TARbac vector to propagate the synthetic chromosome in yeast. We do not know whether this vector might interfere with the production of viable cells by transplantation (5), nor do we know whether the genomic location of the vector could affect viability. It may be necessary to alter the vector sequences or even to excise the vector before transplantation.

The methods described here have advantages compared with those previously described for constructing large DNA molecules, either chem-

ically synthesized or natural. Large *in vitro* DNA assemblies (>30 kb) have used type IIS restriction enzymes to generate unique sticky ends on the components of the assembly, which are then joined by ligation [for example, see (9, 29)]. As the pieces to be assembled grow larger, it becomes increasingly difficult to find a type IIS enzyme that does not cleave within the piece. Our method is not limited to type IIS enzymes. We can use enzymes that cleave infrequently, for example type II enzymes with eight base recognition sites (e.g., Not I; see Figs. 3 and 5) or enzymes with even greater specificity [e.g., homing endonucleases; see New England Biolabs (NEB) catalog]. Instead of type IIS sticky end ligation, our method uses *in vitro* recombination of overlaps between the ends of the fragments to be assembled. A chew-back and anneal method (Fig. 3) similar to the first step of the assembly reaction described here was used to simultaneously assemble and clone up to nine small overlapping DNA fragments (275 to 980 bp) into a plasmid vector (30). The second-step repair reaction included in our method (13) greatly increases the efficiency of cloning of large assemblies (>50 kb).

Nothing in our methodology restricts its use to chemically synthesized DNA. It should be possible to assemble any combination of synthetic and natural DNA segments in any desired order by designing PCR primers to generate appropriate overlaps between them.

In closing, we wonder whether use of the UGA codon to code for tryptophan in mycoplasmas, rather than for termination as in the “universal” code, contributed to our success in cloning the synthetic *M. genitalium* JCVI-1.0 genome. This may make cloning in *E. coli* and other organisms less toxic because most *M. genitalium* proteins will be truncated. If so, then it should be possible to synthesize other genome constructions using this same code. The genome would then need to be installed, for example, by transplantation (5), in a cytoplasm that can properly translate the UGA to tryptophan. To generalize on this phenomenon, it might be possible to use other codon changes as long as there is a receptive cytoplasm with appropriate codon usage.

Note added in proof: While this paper was in press, we realized that the TARbac vector in our sMgTARbac37 clone interrupts the gene for the RNA subunit of RNase P (mpB). This confirms our speculation that the vector might not be at a suitable site for subsequent transplantation experiments.

References and Notes

1. S. D. Colman, P. C. Hu, W. Litaker, K. F. Bott, *Mol. Microbiol.* **4**, 683 (1990).
2. C. M. Fraser *et al.*, *Science* **270**, 397 (1995).
3. J. I. Glass *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 425 (2006).
4. C. A. Hutchison *et al.*, *Science* **286**, 2165 (1999).
5. C. Lartigue *et al.*, *Science* **317**, 632 (2007).
6. K. J. Blythe, A. A. Kolykhalov, C. M. Rice, *Science* **290**, 1972 (2000).
7. J. Cello, A. V. Paul, E. Wimmer, *Science* **297**, 1016 (2002).
8. H. O. Smith, C. A. Hutchison 3rd, C. Pfannkuch, J. C. Venter, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15440 (2003).

9. S. J. Kodumal *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15573 (2004).
10. M. Arita, Y. Ohashi, *Biotechnol. Prog.* **20**, 1605 (2004).
11. D. Heider, A. Barnekow, *BMC Bioinformatics* **8**, 176 (2007).
12. B. Shimanovsky, J. Feng, M. Potkonjak, in *Revised Papers from the 5th International Workshop on Information Hiding* (Springer-Verlag, Berlin, 2003).
13. Materials and methods are available as supporting material on Science Online.
14. S. Dhandayuthapani, M. W. Blaylock, C. M. Bebear, W. G. Rasmussen, J. B. Baseman, *J. Bacteriol.* **183**, 5645 (2001).
15. P. Marshall, N. Malik, Z. Larin, *Gene Ther.* **6**, 1634 (1999).
16. N. Kouprina *et al.*, *EMBO Rep.* **4**, 257 (2003).
17. D. T. Burke, G. F. Carle, M. V. Olson, *Science* **236**, 806 (1987).
18. V. Larionov, N. Kouprina, J. Graves, M. A. Resnick, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13925 (1996).
19. C. Zeng *et al.*, *Genomics* **77**, 27 (2001).
20. V. N. Noskov *et al.*, *Nucleic Acids Res.* **29**, E32 (2001).
21. S. H. Leem *et al.*, *Nucleic Acids Res.* **31**, e29 (2003).
22. N. Kouprina, V. N. Noskov, V. Larionov, *Methods Mol. Biol.* **349**, 85 (2006).
23. G. A. Silverman, *Methods Mol. Biol.* **54**, 65 (1996).
24. M. Itaya, K. Tsuge, M. Koizumi, K. Fujita, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15971 (2005).
25. M. Itaya, K. Fujita, A. Kuroki, K. Tsuge, *Nat. Methods* **5**, 41 (2008).
26. R. A. Holt, R. Warren, S. Flibotte, P. I. Missirlis, D. E. Smailus, *Bioessays* **29**, 580 (2007).
27. T. Knight, "Idempotent Vector Design for Standard Assembly of Biobricks" (MIT, Cambridge, MA, 2003).
28. Y. Sheng, V. Mancino, B. Birren, *Nucleic Acids Res.* **23**, 1990 (1995).
29. B. Yount, M. R. Denison, S. R. Weiss, R. S. Baric, *J. Virol.* **76**, 11065 (2002).
30. M. Z. Li, S. J. Elledge, *Nat. Methods* **4**, 251 (2007).
31. C. S. Newlon, J. F. Theis, *Curr. Opin. Genet. Dev.* **3**, 752 (1993).
32. V. Larionov, N. Kouprina, G. Solomon, J. C. Barrett, M. A. Resnick, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7384 (1997).
33. We thank J. Mulligan for his interest in our work and expediting cassette synthesis by Blue Heron Technologies, S. Vashee and R.-Y. Chuang for many helpful discussions about the manuscript, and J. Johnson and T. Davidsen for

assistance with GenBank submissions. Additionally, we thank the Larionov laboratory at NIH for their gifts of yeast strains and TAR cloning expertise. The bulk of the work was supported by Synthetic Genomics, Inc. J.C.V. is Chief Executive Officer and Co-Chief Scientific Officer of Synthetic Genomics, Inc., a privately held entity that develops genomic-driven strategies to address global energy and environmental challenges. H.O.S. is Co-Chief Scientific Officer and on the Board of Directors of Synthetic Genomics, Inc. C.A.H. is Chairman of the Synthetic Genomics, Inc., Scientific Advisory Board. All three of these authors hold Synthetic Genomics, Inc., stock.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1151721/DC1

Materials and Methods

Fig. S1

Tables S1 to S4

References

15 October 2007; accepted 11 January 2008

Published online 24 January 2008;

10.1126/science.1151721

Include this information when citing this paper.

REPORTS

Asphericity in Supernova Explosions from Late-Time Spectroscopy

Keiichi Maeda,^{1,2,3*} Koji Kawabata,⁴ Paolo A. Mazzali,^{2,5,6} Masaomi Tanaka,⁷ Stefano Valenti,^{8,9} Ken'ichi Nomoto,^{1,6,7} Takashi Hattori,¹⁰ Jinsong Deng,¹¹ Elena Pian,⁵ Stefan Taubenberger,² Masanori Iye,¹² Thomas Matheson,¹³ Alexei V. Filippenko,¹⁴ Kentaro Aoki,¹⁰ George Kosugi,¹⁵ Youichi Ohya, ¹⁶ Toshiyuki Sasaki,¹⁰ Tadamfumi Takata¹⁷

Core-collapse supernovae (CC-SNe) are the explosions that announce the death of massive stars. Some CC-SNe are linked to long-duration gamma-ray bursts (GRBs) and are highly aspherical. One important question is to what extent asphericity is common to all CC-SNe. Here we present late-time spectra for a number of CC-SNe from stripped-envelope stars and use them to explore any asphericity generated in the inner part of the exploding star, near the site of collapse. A range of oxygen emission-line profiles is observed, including a high incidence of double-peaked profiles, a distinct signature of an aspherical explosion. Our results suggest that all CC-SNe from stripped-envelope stars are aspherical explosions and that SNe accompanied by GRBs exhibit the highest degree of asphericity.

Massive stars (≥ 10 solar masses) end their lives when the nuclear fuel in their innermost region is consumed; lacking sufficient internal pressure support, they can no longer withstand the pull of gravity. Their core then collapses to a neutron star or a black hole. Gravitational energy from the collapse produces an explosion that expels the rest of the star in what is observed as a supernova (SN).

CC-SNe are classified (1) by how much of the stellar envelope is present at the time of the explosion (2). Stars that retain their H envelope produce SNe with a H-rich spectrum, classified as type II. On the other hand, stars that have lost all or most of the H envelope produce CC-SNe that are known as stripped-envelope (or stripped) SNe. These include, in a sequence of increasing degrees of envelope stripping, type IIb (He-rich but still showing

some H), type Ib (He-rich, no H), and type Ic (deprived of both H and He). Some type Ic SNe (hereafter broad-lined SNe Ic) show very broad absorption features in optical spectra obtained within a few weeks after the explosion; these features are produced by material moving at a velocity $> 0.1 c$ (c is the speed of light), probably as the result of an explosion characterized by a kinetic energy (E_K) larger than the canonical value of $\sim 10^{51}$ erg (3). The most energetic broad-lined SNe Ic can reach $E_K \geq 10^{52}$ erg [hereafter hypernovae (HNe) or gamma-ray burst HNe (GRB-HNe)] (4) and can be associated with GRBs (5). Figure 1 summarizes the relation between E_K and the mass of ^{56}Ni that powers the optical light of stripped CC-SNe (6).

An important unsolved question concerns how the gravitational energy of the collapse is turned into the outward motion of the SN explo-

sion. Most recently proposed scenarios involve aspherical explosions (7–11). Therefore, mapping the explosion geometry can be illuminating. It is especially critical to establish whether the explosion geometry is similar for all CC-SNe or at least for different subclasses (GRB-HNe, broad-lined SNe Ic, normal-energy SNe, or stripped versus type II SNe).

Some evidence for asymmetric explosions was obtained from the polarization detected in several SNe II (12) and SNe Ib/c (13) and in the broad-lined SN Ic 2002ap (14–16). However, because there are still few such detections, no systematic study exists.

¹Institute for the Physics and Mathematics of the Universe (IPMU), University of Tokyo, Kashiwa-no-ha 5-1-5, Kashiwa City, Chiba 277-8582, Japan. ²Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, 85741 Garching, Germany. ³Department of Earth Science and Astronomy, College of Arts and Science, University of Tokyo, Tokyo 153-8902, Japan. ⁴Hiroshima Astrophysical Science Center, Hiroshima University, Hiroshima 739-8526, Japan. ⁵National Institute for Astrophysics-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy. ⁶Research Center for the Early Universe, School of Science, University of Tokyo, Tokyo 113-0033, Japan. ⁷Department of Astronomy, School of Science, University of Tokyo, Tokyo 113-0033, Japan. ⁸Department of Physics, University of Ferrara, I-44100 Ferrara, Italy. ⁹European Organization for Astronomical Research in the Southern Hemisphere, Karl-Schwarzschild-Strasse 1, 85741 Garching, Germany. ¹⁰Subaru Telescope, National Astronomical Observatory of Japan (NAOJ), 650 North A'ohoku Place, Hilo, HI 96720, USA. ¹¹National Astronomical Observatory, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100012, China. ¹²Division of Optical and Infrared Astronomy, NAOJ, Osawa 2-21-1, Mitaka, Tokyo 181-8588, Japan. ¹³National Optical Astronomy Observatory, Tucson, AZ 85719, USA. ¹⁴Department of Astronomy, University of California, Berkeley, CA 94720-3411, USA. ¹⁵Atacama Large Millimeter/Submillimeter Array Project, NAOJ, Mitaka, Tokyo 181-8588, Japan. ¹⁶Department of Infrared Astrophysics, Institute of Space and Astronomical Science, Japan Aerospace Exploration Agency (JAXA), 3-1-1 Yoshinodai, Sagami-hara, Kanagawa 229-8510, Japan. ¹⁷Astronomy Data Center, NAOJ, Mitaka, Tokyo 181-8588, Japan.

*To whom correspondence should be addressed. E-mail: maeda@ea.c.u-tokyo.ac.jp