We have synthesized a 582,970 bp *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 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 (BACs) 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 (TAR) 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.

**M. 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 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 significant 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 2,514 bp insertion in gene MG408 (*msrA*), which includes an aminoglycoside resistance
gene, was placed in 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. Fig. 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 4 neighboring cassettes were assembled by in vitro recombination as described below and joined to a 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 3-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 2-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 TAR cloning.

**Assembly of synthetic cassettes by in vitro recombination.** Fig. 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 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. 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. Since 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 4 cassette DNAs and the linear PCR 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 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 Materials & Methods 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 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 since, 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. Fig. 4D shows a FIGE analysis of the sizes of these cloned inserts.

Assembly 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 Stbl4 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 upon this method uses co-transformation 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 (see below).

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 about twenty-fold 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 6 pieces (1 vector, 2 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 6 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 prior to transforming E. coli.

Recovery of the synthetic M. genitalium genome from yeast and confirmation of its sequence. A 600 kb YAC is about five percent 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 approximately 7X coverage. The sequence exactly matched our designed genome, and can be accessed as GenBank sequencing project 25337.

Error management. Our objective was to produce a cloned synthetic genome 582,970 base pairs 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 genome 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 base pair 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. The final product is built from ~10^10 synthetic oligonucleotides, each ~50 nucleotides in length, and is the largest chemically synthesized molecule of defined structure that we are aware of.

Very large non-synthetic 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 4 separate approximately 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* piece-wise 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 operons. In contrast, we found that the *M. genitalium* ribosomal RNA 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 30X less efficiently 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 co-transformed 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 2 quarters could be efficiently cloned to produce half genomes in the yeast vector. More surprisingly, 4 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 6 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 prior to transplantation.

The methods described here have advantages compared to those previously described for constructing large DNA molecules, either chemically 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 8 base recognition sites (e.g. Not I, see Figs. 3 and 5) or enzymes with even greater specificity (e.g. homing endonucleases, see 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 9 small overlapping DNA fragments (275 – 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 TARBACK vector in our sMgTARBAC37 clone interrupts the gene for the RNA subunit of RNase P (rnpB). This confirms our speculation that the vector might not be at a suitable site for subsequent transplantation experiments.

References and Notes
12. B. Shimanovsky, J. Feng, M. Potkonjak, in Revised Papers from the 5th International Workshop on Information Hiding (Springer-Verlag, 2003).
13. Detailed descriptions of materials and methods are posted in Supporting Online Material.
19. C. Zeng et al., Genomics 77, 27 (Sep, 2001).
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. Davidson 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.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 S5
Fig. S1

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.

Fig. 2. A plan for the five-stage assembly of the M. genitalium chromosome. In the first stage of assembly, 4 cassettes are joined to make an A-series assembly approximately 24 kb in length (assembly 37-41 contained 5 cassettes). In the next stage, 3 A-assemblies are joined together to make a total of 8 approximately 72 kb B-series assemblies (assembly B62-77 contained 4 A-series assemblies). The eighth-genome B-assemblies are taken 2-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 4 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 66-69 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 4 cassettes and the BAC DNA as depicted in (C).

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 SOM Materials & Methods 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 non-homologous 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.

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, as 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 electroformation 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 co-transformation into S. cerevisiae, where recombination at overlaps from 60-264 bp combined the 6 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, as 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.
A. Assembly of cassettes 66 to 69 by *in vitro* recombination

1. 3' exonuclease chew-back
2. Anneal at 60°C
3. Repair with Taq pol and Taq ligase

B. PCR amplification of pCC1BAC vector using primers with Not I sites flanking the BamH I cloning site

- Primer 66
- Primer 69

First 40 nucleotides of cassette 66

20bp BAC sequence

20bp BAC sequence

Last 40 nucleotides of cassette 69

C. Simultaneous assembly with PCR-amplified BAC

\[ \text{BAC} 66-69 + 66 \ 67 \ 68 \ 69 \]
3' exonuclease chew-back

Anneal

Fill in gap using Taq pol, release overhang by Taq pol 5'-exonuclease, seal nick using Taq ligase