

Enzymatic assembly of DNA molecules up to several hundred kilobases

Daniel G Gibson¹, Lei Young¹, Ray-Yuan Chuang¹, J Craig Venter^{1,2}, Clyde A Hutchison III² & Hamilton O Smith²

We describe an isothermal, single-reaction method for assembling multiple overlapping DNA molecules by the concerted action of a 5' exonuclease, a DNA polymerase and a DNA ligase. First we recessed DNA fragments, yielding single-stranded DNA overhangs that specifically annealed, and then covalently joined them. This assembly method can be used to seamlessly construct synthetic and natural genes, genetic pathways and entire genomes, and could be a useful molecular engineering tool.

For nearly 40 years, scientists have had the ability to join DNA sequences and produce combinations that are not present in nature. This 'recombinant DNA technology' was initiated soon after the discovery of DNA ligase¹ and restriction endonucleases². Since then, multiple approaches for joining DNA molecules through the use of restriction enzymes^{3,4} and PCR^{5–8} have been adapted. Ligation-independent cloning strategies have also been developed^{9,10}. We had recently described an *in vitro* recombination method that we had used to join 101 DNA cassettes into four quarter molecules of the *Mycoplasma genitalium* genome, each 136–166 kilobases (kb)¹¹ (Supplementary Figs. 1–3 and Supplementary Results online). Because we performed this recombination method in a thermocycler (to which we refer as 'thermocycled' here), individual reactions were carried out in only two steps. Here we improved this two-step thermocycled method by using exonuclease III and antibody-bound *Taq* DNA polymerase, which allow for one-step thermocycled *in vitro* recombination (Supplementary Fig. 4 and Supplementary Results online).

We now present an *in vitro* recombination system that differs from the ones above by its capacity to assemble and repair overlapping DNA molecules in a single isothermal step. This approach can be used to join DNA molecules that are as large as 583 kb and to clone joined products in *Escherichia coli* that are as large as 300 kb. All reagents and enzymes are commercially available, and all that is required for DNA assembly is for the reagent-enzyme mix (which can be stored at –20 °C until needed) to be combined with

overlapping DNA molecules and then incubated at 50 °C for as few as 15 min (Online Methods). This approach dramatically simplifies the construction of large DNA molecules from constituent parts.

Exonucleases that recess double-stranded DNA from 5' ends will not compete with polymerase activity. Thus, all enzymes required for DNA assembly can be simultaneously active in a single isothermal reaction. Furthermore, circular products can be enriched as they are not processed by any of the three enzymes in the reaction. We optimized a 50 °C isothermal assembly system using the activities of the 5' T5 exonuclease (Epicentre), Phusion DNA polymerase (New England Biolabs (NEB)) and *Taq* DNA ligase (NEB) (Fig. 1). *Taq* DNA polymerase (NEB) can be used in place of Phusion DNA polymerase (data not shown), but the latter is preferable as it has inherent proofreading activity for removing noncomplementary sequences (for example, partial restriction sites) from assembled molecules.

To test this system, we cleaved two restriction fragments that overlapped by ~450 base pairs (bp) from the 6-kb pRS415 vector and then reassembled them into a circle (Fig. 2a). After 6–8 min at 50 °C, the linear substrate DNA was completely reacted, and the major product was the 6-kb circle, which migrated just below the 4-kb linear position on an agarose gel. T5 exonuclease actively degrades linear DNA molecules, but closed circular DNA molecules are not degraded¹². We confirmed the circularity of this assembled product by treating it with additional T5 exonuclease (Fig. 2a). To demonstrate that this assembled product was the predicted 6-kb circle, we digested it with *Not* I (a single-cutter) and observed the

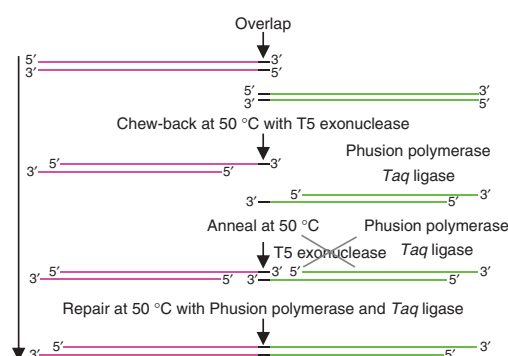
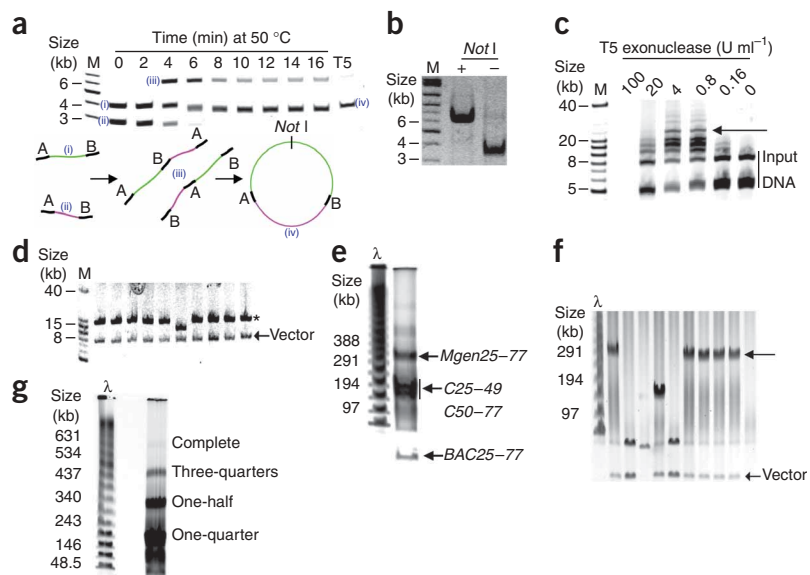


Figure 1 | One-step isothermal *in vitro* recombination. Two adjacent DNA fragments (magenta and green) sharing terminal sequence overlaps (black) were joined into a covalently sealed molecule in a one-step isothermal reaction. T5 exonuclease removed nucleotides from the 5' ends of double-stranded DNA molecules, complementary single-stranded DNA overhangs annealed, Phusion DNA polymerase filled the gaps and *Taq* DNA ligase sealed the nicks. T5 exonuclease is heat-labile and is inactivated during the 50 °C incubation.

¹The J. Craig Venter Institute, Synthetic Biology Group, Rockville, Maryland, USA, and ²San Diego, California, USA. Correspondence should be addressed to D.G.G. (dgibson@jcv.org).

Figure 2 | Examples of the one-step isothermal assembly method. **(a)** Two DNA molecules (4,024 bp, green (i) and 2,901 bp, magenta (ii)), overlapping by ~450 bp at the termini (black, overlaps labeled A and B), were reacted for 0–16 min to form a 6-kb circle, pRS415 (iv). Linear assembly products (iii) were then removed by incubation with additional tenfold excess T5 exonuclease after the 16-min incubation (T5). **(b)** *Not* I digest of the assembled circles (pRS415) shown in **a**. **(c)** Fragments F1–F3 were assembled into an 8-kb PCR-amplified pCC1BAC (Epicentre) containing 40-bp overlaps to F1 and F3, using the indicated amounts of T5 exonuclease. The ~23-kb circular assembly products are indicated by the arrow. **(d)** *Not* I digestion of BACs purified from ten *E. coli* clones after electroporation of the 4 U ml⁻¹ T5 exonuclease reaction shown in **c**. *, correct 15-kb insert. **(e)** Assembly of quarter *M. genitalium* genomes C25–49 and C50–77 with BAC25–77 (ref. 11) to produce *Mgen25–77*. **(f)** *Not* I digestion of BACs purified from ten *E. coli* clones after electroporation of the assembly reaction shown in **e**. The correct insert size (310 kb) is indicated by the arrow. **(g)** Assembly of quarter *M. genitalium* genomes C1–24, C25–49, C50–77 and C78–101 (ref. 11) to produce a complete *M. genitalium* genome. DNA products were analyzed by conventional gel electrophoresis (**a,b**) and by field-inversion gel electrophoresis (**c–g**). M, 1-kb DNA extension markers; λ, lambda markers.



6-kb linear fragment (Fig. 2b). We concluded that DNA molecules can be assembled and repaired in a single isothermal step using this method.

We next determined whether DNA molecules with overlaps of only 40 bp could be joined. We accomplished this when we reduced the concentration of T5 exonuclease (Fig. 2c). Three 5-kb DNA fragments, F1–F3, were efficiently assembled into an 8-kb bacterial artificial chromosome (BAC). Furthermore, when we transformed these assembled DNA molecules into *E. coli*, we obtained 4,500 colonies, and nine out of ten colonies tested had the predicted 15-kb insert (Fig. 2d).

During the construction of the synthetic *M. genitalium* genome, we could not use our two-step thermocycled *in vitro* recombination method to clone assembled DNA molecules larger than ~150 kb in *E. coli*¹¹. To determine whether the isothermal assembly method could be used to join and clone DNA fragments of larger size, we reacted two synthetic *M. genitalium* quarter DNA molecules, C25–49 (144 kb) and C50–77 (166 kb), with BAC25–77 (8 kb), a cloning vector specific for the assembly of these two DNA molecules. The 318-kb *Mgen25–77* product was efficiently produced, so we conclude that DNA fragments this size can be joined by this method (Fig. 2e). To determine whether this method could be used to clone assembled DNA fragments this size, we transformed a fraction of this assembly reaction into *E. coli*. We obtained several hundred clones, and 5 out of 10 colonies screened had the correct insert size (310 kb; Fig. 2f). Thus, this DNA assembly system can be used to join and clone DNA molecules up to several hundred kilobases in length in *E. coli*, the approximate upper limit for transformation into this bacterium¹³. In a direct comparison of all our assembly methods, we found that only the one-step *in vitro* recombination methods could be used to clone assembled DNA fragments this size (Supplementary Fig. 5 online).

During *in vitro* recombination, errors may be introduced in the assembled DNA. However, sequencing of 30 cloned DNA

molecules (210 repaired junctions) after two-step thermocycled assembly revealed only 4 errors (Supplementary Table 1 online). This equates to only about 1 error per 50 DNA molecules joined. Therefore, if our hierarchical scheme to assemble the *M. genitalium* genome was used¹¹ without sequence verification at intermediate steps, 3–4 errors would likely be present. We expect that the number of mutations would be even lower with the isothermal assembly system because gaps are filled in by Phusion DNA polymerase, which has higher fidelity than *Taq* polymerase.

Our isothermal method can be used to assemble DNA molecules of unprecedented sizes, and we used it to assemble the complete synthetic 583-kb *M. genitalium* genome (Fig. 2g). The size limit for *in vitro* DNA assembly is not known, but products as large as 900 kb have been observed (Supplementary Fig. 6 online). Of the three *in vitro* recombination methods, we prefer the one-step-isothermal system because of its simplicity. This approach could be very useful for cloning multiple inserts into a vector without relying on the availability of restriction sites and for rapidly constructing large DNA molecules. For example, regions of DNA too large to be amplified by PCR can be divided into multiple overlapping PCR amplicons and then assembled into one piece. The one-step thermocycled method could be used to generate linear assemblies as the exonuclease is inactivated during the reaction (Supplementary Figs. 4 and 5).

Synthetic biologists are engineering genetic pathways for the production of biofuels, pharmaceuticals and industrial compounds^{14,15}. Here we provide efficient methods for constructing these pathways, from natural or synthetic DNA.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturemethods/>.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

This work was supported by the Office of Science (Biological and Environmental Research) United States Department of Energy grant number DE-FG02-02ER63453, and Synthetic Genomics, Inc.

AUTHOR CONTRIBUTIONS

D.G.G., L.Y., R.-Y.C., J.C.V., C.A.H. and H.O.S. designed research; D.G.G., L.Y., R.-Y.C., C.A.H. and H.O.S. performed research; D.G.G., L.Y., R.-Y.C., J.C.V., C.A.H. and H.O.S. analyzed data; and D.G.G., C.A.H. and H.O.S. wrote the paper.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturemethods/>.

Published online at <http://www.nature.com/naturemethods/>
Reprints and permissions information is available online at
<http://npg.nature.com/reprintsandpermissions/>

1. Gellert, M. *Proc. Natl. Acad. Sci. USA* **57**, 148–155 (1967).

2. Smith, H.O. & Wilcox, K.W. *J. Mol. Biol.* **51**, 379–391 (1970).
3. Shetty, R.P., Endy, D. & Knight, T.F. Jr. *J. Biol. Eng.* **2**, 5 (2008).
4. Yount, B., Denison, M.R., Weiss, S.R. & Baric, R.S. *J. Virol.* **76**, 11065–11078 (2002).
5. Horton, R.M., Cai, Z.L., Ho, S.N. & Pease, L.R. *Biotechniques* **8**, 528–535 (1990).
6. Horton, R.M. *Mol. Biotechnol.* **3**, 93–99 (1995).
7. Bang, D. & Church, G.M. *Nat. Methods* **5**, 37–39 (2008).
8. Geu-Flores, F., Nour-Eldin, H.H., Nielsen, M.T. & Halkier, B.A. *Nucleic Acids Res.* **35**, e55 (2007).
9. Aslanidis, C. & de Jong, P.J. *Nucleic Acids Res.* **18**, 6069–6074 (1990).
10. Li, M.Z. & Elledge, S.J. *Nat. Methods* **4**, 251–256 (2007).
11. Gibson, D.G. *et al. Science* **319**, 1215–1220 (2008).
12. Sayers, J.R. & Eckstein, F. *J. Biol. Chem.* **265**, 18311–18317 (1990).
13. Sheng, Y., Mancino, V. & Birren, B. *Nucleic Acids Res.* **23**, 1990–1996 (1995).
14. Endy, D. *Nature* **438**, 449–453 (2005).
15. Drubin, D.A., Way, J.C. & Silver, P.A. *Genes Dev.* **21**, 242–254 (2007).

ONLINE METHODS

Preparation of DNA molecules for *in vitro* recombination. The DNA molecules used in the assembly analyses were derived from several sources including (i) the assembly intermediates of the synthetic *M. genitalium* genome¹¹, (ii) PCR products derived from plasmids (F6 and F8), *Clostridium cellulolyticum* genomic DNA (F1–F4) and *Mycoplasma gallisepticum* genomic DNA (F5 and F7) and (iii) pRS415 restriction fragments. *E. coli* strains carrying each of *M. genitalium* cassettes 66–69 (contained in pENTR223), each of *M. genitalium* cassettes 78–85 (contained in pBR322), C1–24, C25–49, C50–77 and C78–101 (each contained in pCC1BAC) or pRS415 were propagated in LB medium containing the appropriate antibiotic and incubated at 30 °C or 37 °C for 16 h. The cultures were collected and the DNA molecules were purified using Qiagen's HiSpeed Plasmid Maxi Kit according to the manufacturer's instructions, with the exception of C1–24, C25–49, C50–77 and C78–101, which were not column-purified. Instead, after neutralization of the lysed cells, these DNA molecules were centrifuged then precipitated with isopropanol. DNA pellets were dissolved in Tris-EDTA (TE) buffer (pH 8.0) then RNase treated, phenol-chloroform extracted and ethanol precipitated. DNA pellets were dissolved in TE buffer. Cassettes 66–69 and 78–85 were excised from the vectors by restriction digestion with either *Fau* I or *Bsm* BI and C1–24, C25–49, C50–77 and C78–101 were excised by digestion with *Not* I. To generate the 4,024-bp and 2,901-bp overlapping fragments of pRS415, DNA was digested with *Pvu* II and *Sca* I or *Psi* I, respectively. Restriction digestions were terminated by phenol-chloroform extraction and ethanol precipitation. DNA was dissolved in TE buffer, then quantified by gel electrophoresis with known DNA standards. Fragments F1–F8 were generated by PCR using the Phusion Hot Start High-Fidelity DNA polymerase with HF buffer (NEB) according to the manufacturer's instructions. PCR products were extracted from agarose gels after electrophoresis and purified using the QIAquick Gel Extraction kit (Qiagen) according to the manufacturer's instructions, except DNA was eluted from the columns with TE buffer pH 8.0. Fragments F1–F4 were amplified from *Clostridium cellulolyticum* genomic DNA using primers F1-For and F1-Rev, F2-For and F2-Rev, F3-For and F3-Rev, and F4-For and F4-Rev, respectively. F5 and F7 were amplified from *Mycoplasma gallisepticum* genomic DNA using primers F5-For and F5-Rev, and F7-For and F7-Rev, respectively. F6 and F8 were amplified from pRST2 (ref. 16) using primers F6-For and F6-Rev, and F8-For and F8-Rev, respectively. Primer sequences are listed in **Supplementary Table 2** online.

Two-step thermocycled DNA assembly. A 4× chew-back and anneal (CBA) reaction buffer (20% PEG-8000, 800 mM Tris-HCl pH 7.5, 40 mM MgCl₂, 4 mM DTT) was used for thermocycled DNA assembly. DNA molecules were assembled in 20-μl reactions consisting of 5 μl 4× CBA buffer, 0.2 μl of 10 mg ml⁻¹ BSA (NEB) and 0.4 μl of 3 U μl⁻¹ T4 polymerase (NEB). T7 polymerase can be substituted for T4 polymerase (data not shown). Approximately 10–100 ng of each ~6 kb DNA segment was added in equimolar amounts. For larger DNA segments, proportional amounts of DNA were added (for example, 250 ng of each 150 kb DNA segment). Assembly reactions were prepared in 0.2 ml PCR tubes and cycled as follows: 37 °C from 0 to 18 min as indicated in the text, 75 °C for 20 min, 0.1 °C s⁻¹ to 60 °C, held at 60 °C for 30 min, then cooled to 4 °C at a rate of 0.1 °C s⁻¹. In general, a

chew-back time of 5 min was used for overlaps less than 80 bp and 15 min for overlaps greater than 80 bp. Ten microliters of the CBA reactions were then added to 25.75 μl of *Taq* repair buffer (TRB), which consisted of 5.83% PEG-8000, 11.7 mM MgCl₂, 15.1 mM DTT, 311 μM each of the four dNTPs and 1.55 mM NAD. Four microliters of 40 U μl⁻¹ *Taq* DNA ligase and 0.25 μl of 5 U μl⁻¹ *Taq* polymerase were added and the reactions were incubated at 45 °C for 15 min. For the T4 polymerase fill-in assembly method, 10 μl of the CBA reaction was mixed with 0.2 μl of 10 mM dNTPs and 0.2 μl of 3 U μl⁻¹ T4 polymerase. This reaction was carried out at 37 °C for 30 min.

One-step thermocycled DNA assembly. A 4× chew-back, anneal and repair (CBAR) reaction buffer (20% PEG-8000, 600 mM Tris-HCl pH 7.5, 40 mM MgCl₂, 40 mM DTT, 800 μM each of the four dNTPs and 4 mM NAD) was used for one-step thermocycled DNA assembly. DNA molecules (added in amounts described above for CBA reactions) were assembled in 40 μl reactions consisting of 10 μl 4× CBAR buffer, 0.35 μl of 4 U μl⁻¹ ExoIII (NEB), 4 μl of 40 U μl⁻¹ *Taq* DNA ligase and 0.25 μl of 5 U μl⁻¹ Ab-*Taq* polymerase (Applied Biosystems). ExoIII was diluted 1:25 from 100 U μl⁻¹ in its stored buffer (50% glycerol, 5 mM KPO₄, 200 mM KCl, 5 mM 2-mercaptoethanol, 0.05 mM EDTA and 200 μg ml⁻¹ BSA, pH 6.5). DNA assembly reactions are prepared in 0.2 ml PCR tubes and cycled using the following conditions: 37 °C for 5 or 15 min as indicated in the text, 75 °C for 20 min, 0.1 °C s⁻¹ to 60 °C, then held at 60 °C for 1 h. In general, a chew-back time of 5 min was used for overlaps less than 80 bp and 15 min for overlaps greater than 80 bp. ExoIII is less active on 3' protruding termini¹⁷, which can result from digestion with certain restriction enzymes. This can be overcome by removing the overhangs to form blunt ends with the addition of T4 polymerase and dNTPs, as described above, before assembly (data not shown).

One-step isothermal DNA assembly. A 5× isothermal reaction buffer (25% PEG-8000, 500 mM Tris-HCl pH 7.5, 50 mM MgCl₂, 50 mM DTT, 1 mM each of the four dNTPs and 5 mM NAD) was used for one-step DNA assembly at 50 °C. DNA molecules (added in amounts described above for CBA reactions) were assembled in 40 μl reactions consisting of 8 μl 5× isothermal buffer, 0.8 μl of 0.2 U μl⁻¹ or 1.0 U μl⁻¹ T5 exonuclease, 4 μl of 40 U μl⁻¹ *Taq* DNA ligase and 0.5 μl of 2 U μl⁻¹ Phusion DNA polymerase. T5 exonuclease was diluted 1:50 or 1:10 from 10 U μl⁻¹ in its stored buffer (50% glycerol, 50 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 1 mM DTT, 0.1 M NaCl and 0.1% Triton X-100) depending on the overlap size. For overlaps shorter than 150 bp, 0.2 U μl⁻¹ T5 exonuclease is used. For overlaps larger than 150 bp, 1.0 U μl⁻¹ T5 exonuclease was used. All isothermal assembly components can be stored at -20 °C in a single mixture at 1.33× concentration for more than one year. The enzymes are still active after more than ten freeze-thaw cycles. To constitute a reaction, 5 μl DNA was added to 15 μl of this mixture. Incubations were carried out at 50 °C for 15 to 60 min, with 60 min being optimal.

One-step isothermal DNA assembly protocol. Six milliliters of 5× isothermal reaction buffer were prepared by combining 3 ml of 1 M Tris-HCl pH 7.5, 150 μl of 2 M MgCl₂, 60 μl of 100 mM dGTP, 60 μl of 100 mM dATP, 60 μl of 100 mM dTTP, 60 μl of 100 mM dCTP, 300 μl of 1 M DTT, 1.5 g PEG-8000 and



300 μ l of 100 mM NAD. This buffer can be aliquoted and stored at -20°C . An assembly master mixture was prepared by combining 320 μ l $5\times$ isothermal reaction buffer, 0.64 μ l of 10 U μl^{-1} T5 exonuclease, 20 μ l of 2 U μl^{-1} Phusion DNA polymerase, 160 μ l of 40 U μl^{-1} Taq DNA ligase and water up to a final volume of 1.2 ml. Fifteen microliters of this reagent-enzyme mix were aliquoted and stored at -20°C . This mixture can tolerate numerous freeze-thaw cycles and remains stable even after one year. The exonuclease amount is ideal for the assembly of DNA molecules with 20–150 bp overlaps. For DNA molecules overlapping by greater than 150 bp, 3.2 μ l of 10 U μl^{-1} T5 exonuclease was used to prepare the assembly master mixture above. Frozen 15 μ l assembly mixture aliquots were thawed and then kept on ice until ready to be used. Five microliters of the DNA to be assembled were added to the master mixture in equimolar amounts. Between 10 and 100 ng of each ~ 6 kb DNA fragment was added. For larger DNA segments, proportional amounts of DNA were added (for example, 250 ng of each 150 kb DNA segment). Incubations were performed at 50°C for 15 to 60 min (60 min was optimal).

Rolling-circle amplification (RCA) of assembled products. RCA was carried out as previously described¹⁸. One microliter of the repaired or unrepaired reaction was mixed with 1 μ l of 100 mM NaOH and incubated at room temperature (18 – 22°C) for 5 min to denature the double-stranded DNA. One microliter of this alkaline-treated mixture was then added to 19 μ l of RCA components in a 0.2 ml PCR tube. The final reaction concentrations for RCA are as follows: 37 mM Tris-HCl pH 7.5, 50 mM KCl, 10 mM MgCl_2 , 5 mM $(\text{NH}_4)_2\text{SO}_4$, 100 $\mu\text{g ml}^{-1}$ BSA, 1 mM DTT, 3.25 mM random hexamers (Fidelity System), 1 U ml^{-1} yeast pyrophosphatase (United States Biochemical) and 250 U ml^{-1} phi29 DNA polymerase (NEB). The reaction was incubated at 30°C for 20 h then terminated by incubation at 65°C for 10 min.

Cloning the DNA assembly products. To clone assembled products, reactions were carried out in the presence of PCR-amplified BACs containing 40 bp of overlapping sequence to the ends of the assembled product. *Not* I restriction sites were also included to allow release of the vector¹¹. To produce BAC-F1/F3, primers

BACF1 For and BACF3 Rev were used in PCR. To produce BAC66-69, primers BAC66 For and BAC69 Rev were used in PCR. To produce BAC25-77, primers BAC25 For and BAC77 Rev were used in PCR. Primer sequences are listed in **Supplementary Table 2**. In general, pCC1BAC was used as DNA template. However, for cloning *Mgen25-77*, a version of pCC1BAC, named KanBAC, was constructed that contains the kanamycin resistance gene in place of the chloramphenicol resistance gene. Samples (up to 1 μ l) of the assembly reactions were transformed into 30 μ l TransforMax EPI300 (Epicentre) electrocompetent *E. coli* cells in a 1-mm cuvette (BioRad) at 1,200 V, 25 μF and 200 Ω using a Gene Pulser Xcell Electroporation system (BioRad). Cells were allowed to recover at 30°C or 37°C for 2 h in 1 ml SOC medium then plated onto LB medium containing 12.5 $\mu\text{g ml}^{-1}$ chloramphenicol or LB medium containing 25 $\mu\text{g ml}^{-1}$ kanamycin. After incubation at 30°C or 37°C for 24–48 h, individual colonies were selected and grown in 3 ml LB medium with 12.5 $\mu\text{g ml}^{-1}$ chloramphenicol or 25 $\mu\text{g ml}^{-1}$ kanamycin overnight at 30°C or 37°C . DNA was prepared from these cells by alkaline lysis using the P1, P2 and P3 buffers (Qiagen) followed by isopropanol precipitation. DNA pellets were dissolved in TE buffer containing RNase and then digested with *Not* I to release the insert from the BAC.

Agarose gel analyses of assembled DNA molecules and cloned products. U-5 field-inversion gel electrophoresis analysis was performed on 0.8% E-gels (Invitrogen) and the parameters were forward 72 V, initial switch 0.1 s, final switch 0.6 s, with linear ramp and reverse 48 V, initial switch 0.1 s, final switch 0.6 s, with linear ramp. U-2 field-inversion gel electrophoresis analysis was performed on 1% agarose gels (BioRad) in $1\times$ TAE buffer with 0.5 $\mu\text{g ml}^{-1}$ ethidium bromide without circulation, and the parameters were forward 90 V, initial switch 5.0 s, final switch 30 s, with linear ramp, and reverse 60 V, initial switch 5.0 s, final switch 30 s, with linear ramp. DNA bands were visualized with a BioRad Gel Doc or an Amersham Typhoon 9410 Fluorescence Imager.

16. Lartigue, C., Duret, S., Garnier, M. & Renaudin, J. *Plasmid* **48**, 149–159 (2002).

17. Henikoff, S. *Gene* **28**, 351–359 (1984).

18. Hutchison, C.A., III, Smith, H.O., Pfannkuch, C. & Venter, J.C. *Proc. Natl. Acad. Sci. USA* **102**, 17332–17336 (2005).

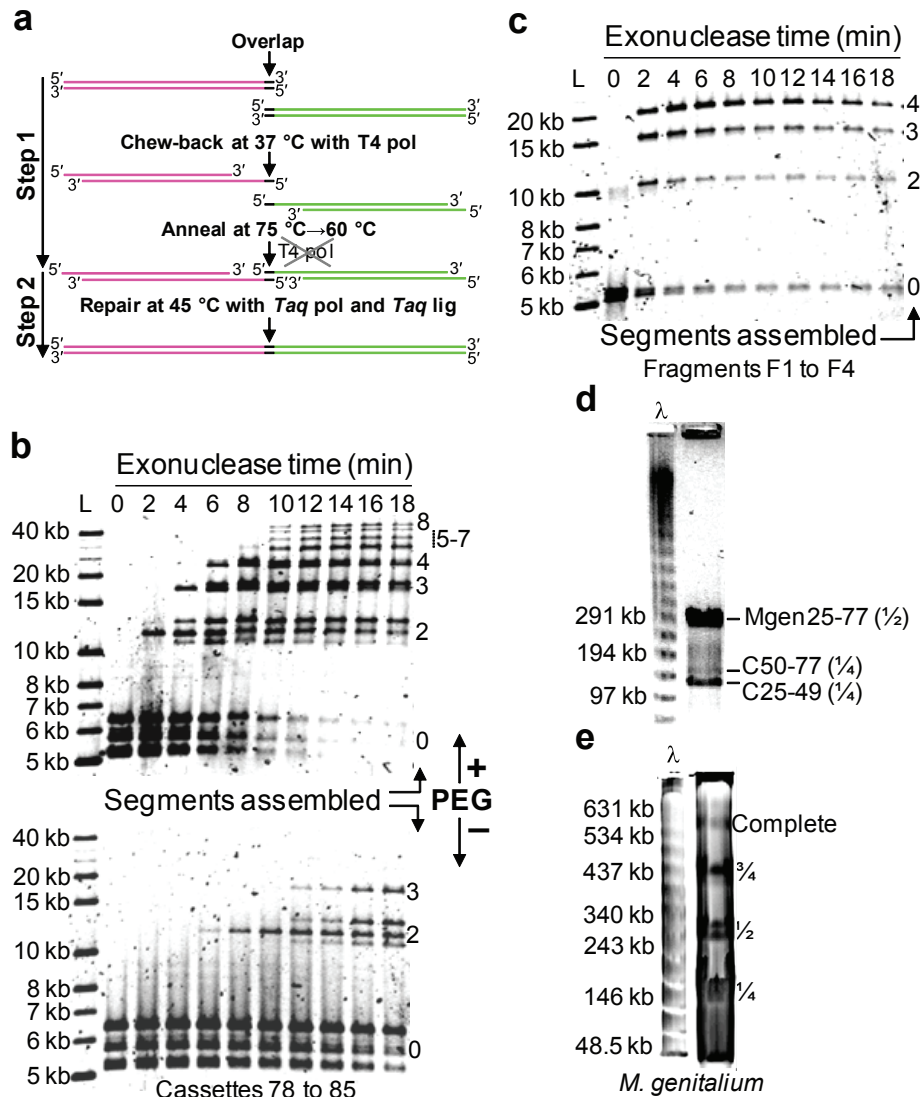
Enzymatic assembly of DNA molecules up to several hundred kilobases

Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison III & Hamilton O Smith

Supplementary figures and text:

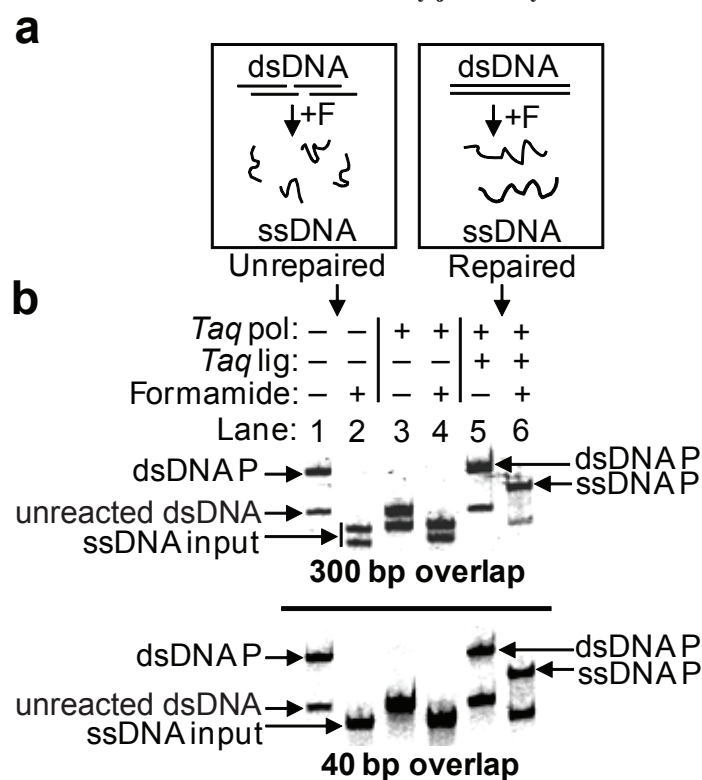
Supplementary Figure 1	Two-step-thermocycled <i>in vitro</i> recombination
Supplementary Figure 2	Assembled DNA molecules are covalently joined by the activities of <i>Taq</i> pol and <i>Taq</i> lig
Supplementary Figure 3	Rolling circle amplification of repaired assembly products
Supplementary Figure 4	One-step-thermocycled assembly of overlapping DNA segments
Supplementary Figure 5	Comparison of the assembly methods
Supplementary Figure 6	Assembly of DNA molecules up to at least 900 kb
Supplementary Table 1	Summary of variations observed following sequence verification of DNA molecules constructed by the two-step-thermocycled assembly method
Supplementary Table 2	PCR primers used in this study
Supplementary Results	

Supplementary Figure 1. Two-step-thermocycled *in vitro* recombination.



(a) Two adjacent DNA fragments (red and green) sharing terminal sequence overlaps (black) are joined into one covalently sealed molecule by a 2-step-thermocycled reaction. In this method, T4 pol is incubated with overlapping DNA segments at 37 °C in a buffer lacking dNTPs. Under these conditions, dsDNA is recessed (“chewed back”) at the 3′ ends, and 5′ ssDNA overhangs are produced. The DNA molecules are specifically annealed by shifting the reaction to 75 °C for 20 min, slow-cooling to 60 °C, then holding at 60 °C for 30 min. The annealed molecules are then repaired in a second step using *Taq* pol and dNTPs to fill-in the gapped molecules, and *Taq* lig to seal the nicks that are formed. T4 pol is inactivated at 75 °C. **(b)** DNA assembly of cassettes 78 to 85, each between 5.3 kb and 6.5 kb with 240 to 360 bp overlapping sequence¹, is carried out in the presence (+) or absence (-) of 5% PEG-8000. Cassettes were incubated at 37 °C in the presence of T4 pol for 0 to 18 min then annealed. Assembly efficiency was analyzed on 0.5% agarose gels following electrophoresis at 1.5 V/cm for 16 h and staining with SYBR Gold (Molecular Probes). The number of segments assembled (0 to 8) is indicated on the right side of the panel. L indicates the 1 kb DNA extension ladder (Invitrogen). **(c)** Fragments F1 to F4, each 5 kb with 40 bp overlaps, were reacted in the presence of PEG-8000 and analyzed as in **(b)**. **(d)** *M. genitalium* 1/4 genome assemblies, C25-49 (144 kb) and C50-77 (166 kb), which overlap by 257 bp, were annealed following incubation for 15 min at 37 °C to produce Mgen25-77. Assembly products were analyzed by U-2 FAGE. λ indicates the lambda DNA ladder (New England Biolabs [NEB]). **(e)** *M. genitalium* 1/4 genome assemblies C1-24, C25-49, C50-77, and C78-101, ~150 kb each with 80 to 257 bp overlaps¹, were annealed following a 15 min exonuclease reaction at 37 °C. “Complete”, “3/4”, “1/2”, and “1/4”, indicate the proportion of the *M. genitalium* genome that was assembled. Assembly efficiency was analyzed as in **(d)**.

Supplementary Figure 2. Assembled DNA molecules are covalently joined by the activities of *Taq* pol and *Taq* lig.



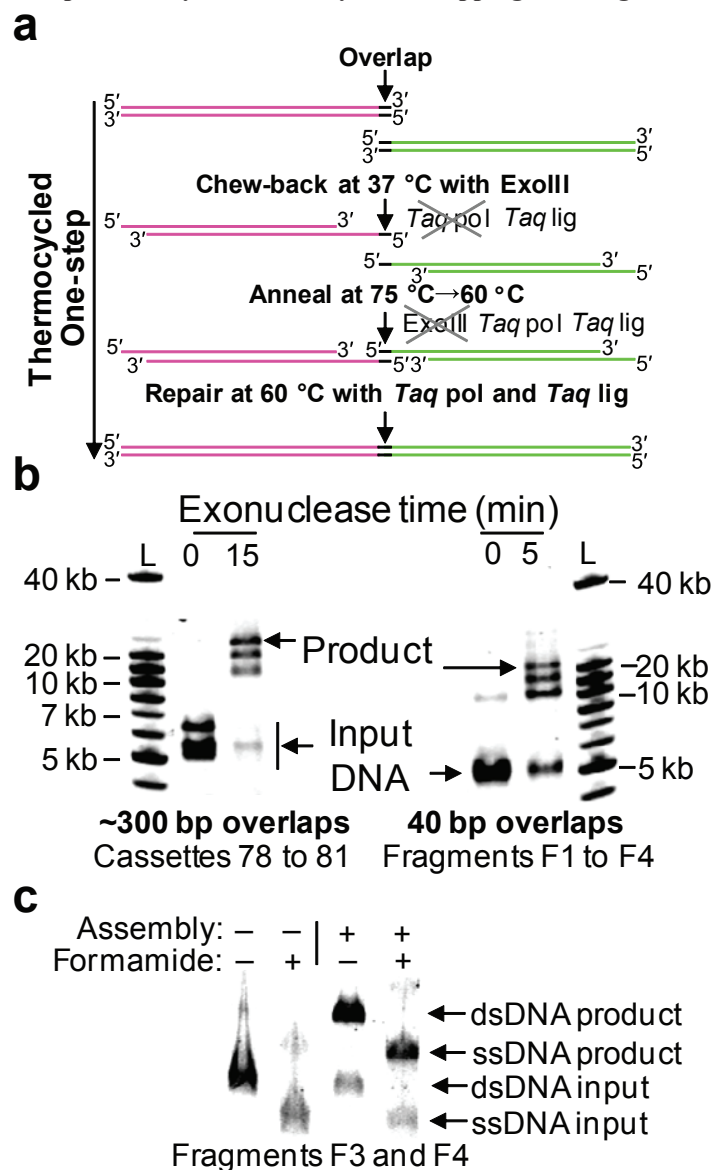
(a) Schematic outlining the strategy used to analyze the success of a repaired assembly reaction. Double-stranded DNA is denatured to ssDNA by incubating the reactions at 94 °C for 2 min in the presence of formamide (+F). If repair has occurred, the resulting ssDNA product remains intact and has a higher molecular weight than the assembled but unrepaired products treated in the same way. These differences are analyzed by gel-electrophoresis on 0.8% E-gels (Invitrogen) and SYBR Gold staining. **(b)** Cassettes 83 and 84 (300 bp overlap) and F3 and F4 (40 bp overlap) were reacted with T4 pol for 15 min or 5 min respectively and annealed then incubated in the presence (+) or absence (-) of *Taq* pol and/or *Taq* lig. Reactions were left untreated (-) or subjected to denaturation conditions in the presence of formamide (+) and analyzed as described in **(a)**. The various forms of ssDNA and dsDNA are noted with arrows. P indicates product.

Supplementary Figure 3. Rolling circle amplification of repaired assembly products.



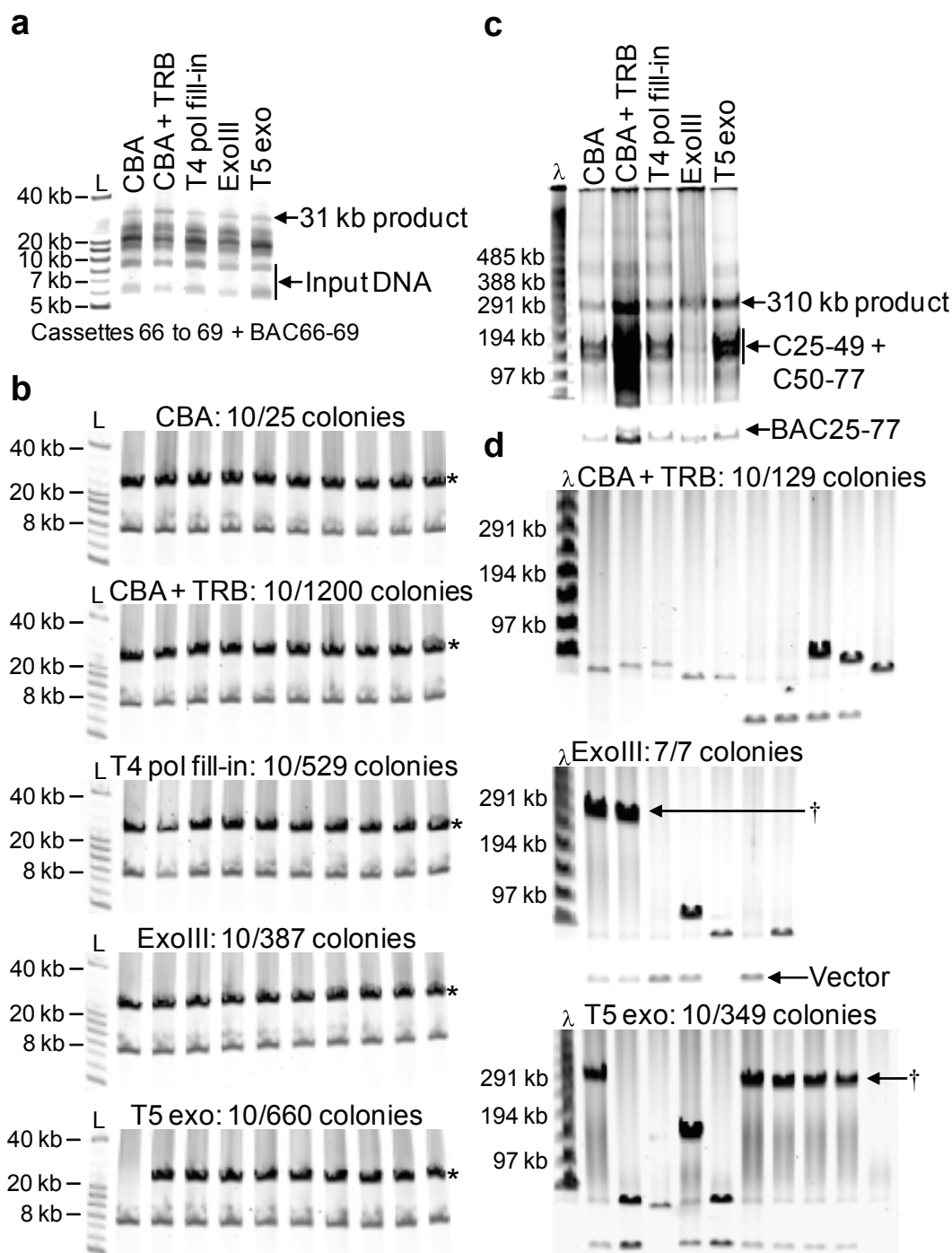
Four DNA fragments, F5 (1020 bp), F6, (1040 bp), F7 (2379 bp), and F8 (3246 bp), each with 40 bp overlaps, were joined into a 7,525 bp circle, as described in **Supplementary Figure 1a**, then repaired or left unrepaired. The unrepaired reaction contained all the components for repair; however, incubation took place at 0 °C instead of 45 °C. The repaired DNA product, but not the unrepaired DNA product, could be amplified by phi29 polymerase (RCA). RCA products were digested by *SphI*, which cuts once within the 7.5 kb product, and analyzed on a 1.2% E-gel. M indicates 1 kb DNA ladder (NEB).

Supplementary Figure 4. One-step-thermocycled assembly of overlapping DNA segments.



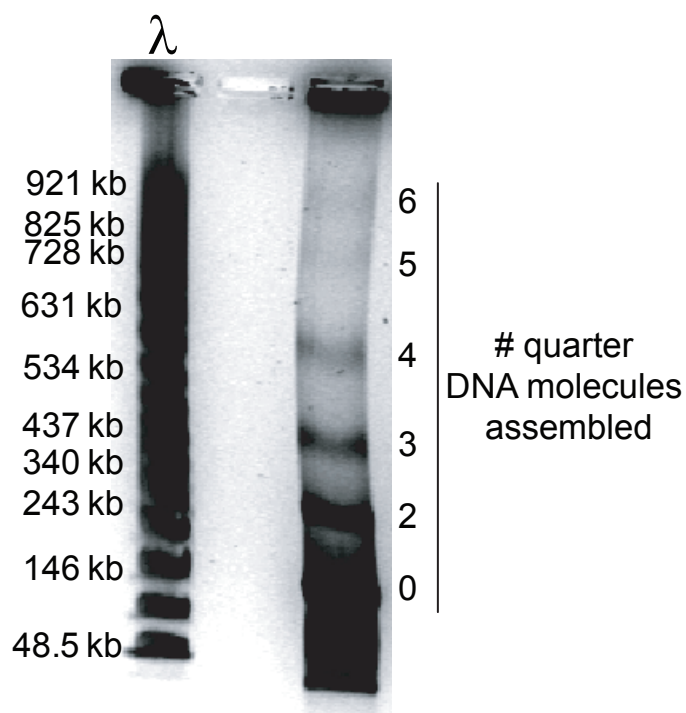
(a) Schematic outlining the approach used to assemble overlapping DNA segments in a single step, in a thermocycler. **(b)** Cassettes 78 to 81 (240 bp to 300 bp overlaps) and fragments F1 to F4 (40 bp overlaps) are assembled as in **(a)** then analyzed by U-5 FAGE. The completely assembled product and unreacted input DNA are indicated with arrows. **(c)** Fragments F3 and F4 were reacted as described in **(a)** in the presence (+ Assembly) or absence (- Assembly) of ExoIII. Repair is assessed by denaturation of the dsDNA molecules in the presence (+) or absence (-) of formamide as described in **Supplementary Figure 2**.

Supplementary Figure 5. Comparison of the assembly methods.



(a) Cassettes 66 to 69 (5.9 kb to 6.2 kb with 80 bp overlaps) were assembled into BAC66-69 (~8 kb with 40 bp overlaps), as described in **Supplementary Figure 1a**, without repair (CBA), with complete repair (CBA + TRB), or with gap fill-in repair with T4 pol but without ligation (T4 pol fill-in), and as described in **Supplementary Figure 4a** (ExoIII) and **Figure 1** (T5 exo). Equal amounts were analyzed by U-5 FGE then transformed into *E. coli*. **(b)** A 0.1 µl sample of the assembly reactions in **(a)** yielded the number of transformants noted. For each assembly method, DNA was extracted from 10 transformants and digested with *NotI* for determination of correct insert size (~23 kb, denoted by *). **(c)** Quarter *M. genitalium* genomes C25-49 and C50-77 were assembled into BAC25-77 (~8 kb with 40 bp overlaps) using the methods described in **(a)**. A fraction of each was analyzed by U-2 FGE. **(d)** Equal amounts were transformed into *E. coli*. A 1 µl sample of the assembly reactions in **(c)** yielded the number of transformants noted. No transformants were obtained for the CBA and T4 pol fill-in reactions so analysis ended at that step. DNA was prepared from 7 to 10 transformants of each assembly method, then digested with *NotI* for determination of correct insert size (~310 kb, denoted by †).

Supplementary Figure 6. Assembly of DNA molecules up to at least 900 kb.



Quarter *M. genitalium* genomes C25-49 and C50-77 were assembled into BAC25-77, and then analyzed by FIGE. This assembly reaction resulted in the concatemerization of at least 6 quarter molecules. Since the quarter molecules are ~150 kb each, the largest assembled products are at least 900 kb.

Supplementary Table 1. Summary of variations observed following sequence verification of DNA molecules constructed by the two-step-thermocycled assembly method.

Assembly	# Joints Repaired Before Getting Sequenced	# New Variations Since Last Sequenced	Variation Type	Likely Cause
A1-4	5	0	None	////////
A5-8	5	0	None	////////
A9-12	5	0	None	////////
A13-16	5	0	None	////////
A17-20	5	0	None	////////
A21-24	5	0	None	////////
A25-28	Not Sequenced	Not Sequenced	Not Sequenced	Not Sequenced
A29-32	Not Sequenced	Not Sequenced	Not Sequenced	Not Sequenced
A33-36	Not Sequenced	Not Sequenced	Not Sequenced	Not Sequenced
A37-41	Not Sequenced	Not Sequenced	Not Sequenced	Not Sequenced
A42-45	Not Sequenced	Not Sequenced	Not Sequenced	Not Sequenced
A46-49	Not Sequenced	Not Sequenced	Not Sequenced	Not Sequenced
A50-53	5	0	None	////////
A54-57	5	1	Δ C	Miscommunication
A58-61	5	0	None	////////
A62-65	5	1	Δ T	Synthesis
A66-69	5	0	None	////////
A70-73	5	0	None	////////
A74-77	5	0	None	////////
A78-81	5	0	None	////////
A82-85	5	1	G to A	Gap Fill-in
A86-89	5	0	None	////////
A90-93	5	0	None	////////
A94-97	5	0	None	////////
A98-101	5	0	None	////////
B1-12	4	0	None	////////
B13-24	4	0	None	////////
B25-36	Not Sequenced	Not Sequenced	Not Sequenced	Not Sequenced
B37-49	Not Sequenced	Not Sequenced	Not Sequenced	Not Sequenced
B50-61	4	0	Δ C	Miscommunication
B62-77	5	0	Δ T	Synthesis
B78-89	4	0	G to A	Gap Fill-in
B90-101	4	0	None	////////
C1-24	3	0	None	////////
C25-49	41	0	None	////////
C50-77-R1	22	2	C to T, T to C	Gap Fill-in, <i>E. coli</i>
C50-77-R2	12	0	None	////////
C78-101	12	1	Δ CA	BAC PCR

During *in vitro* recombination, errors may be introduced in the assembled DNA because of incorrect nucleotide insertion of DNA polymerase during the fill-in reaction (Gap Fill-in), from the primers used to PCR-amplify a BAC (BAC PCR), and from inaccurate DNA replication in *E. coli* (*E. coli*). All of these error types were observed when 30 assembled molecules (210 repaired junctions) were cloned and sequenced during the synthesis of the *M. genitalium* genome¹. One additional variation was due to an incorrect sequence getting transmitted to the cassette manufacturer (miscommunication), and another was due to an incorrect sequence getting synthesized by the cassette manufacturer (synthesis).

Supplementary Table 2. PCR primers used in this study.

Primer Name	Primer Sequence
F1-For	GCAGCTTCAAGTCCTGCAAACAAGGTGTACCAGGATCGTT
F1-Rev	GATTTCAAGTGTAGTTAGGGCCAGTTGAATTCAAACCTGCC
F2-For	GGCAGGTTTGAATTCAACTGGCCCTAACTACACTGAAATC
F2-Rev	CTTGGTGCCATCAGCATTGTTCTCTGTACCGCCCACTGTC
F3-For	GACAGTGGGCGGTACAGAGAACAATGCTGATGGCACCAAG
F3-Rev	CAGTTGAATAATCATGTGTTCTGCGGCAAATGCAGTACC
F4-For	GGTACTGCATTTGCCGCAGGAACACATGATTATTCAACTG
F4-Rev	TTATTTACCAAGAACCTTTGCCTTTAACATTGCAAAGTCA
F5-For	GCTTGCATGCATCCTGTTTATTCATCACAAACATTGAAC
F5-Rev	AATTCTGCAGTTTTTATTTCTTAACAGAACATTTTTCTAGTATAGC
F6-For	TAGAAAAAATGTTCTGTTAGGAAATAAAAACTGCAGAATTAAAAGTTAGTGAACAAGAAAAAC
F6-Rev	AGCCTCAAAAAGATAAAGAAAGGCTATTTATCTAGAGTCGACCTGCAGTTCAGATC
F7-For	CGACTCTAGATAAATAGCCTTTCTTTATCTTTTTGAGGC
F7-Rev	CCGGGGATCCCTTTCTCAATTGTCTGCTCCATATATGTT
F8-For	TGTTCAATGTTTGTGATGAATAAACAGGATGCATGCAAGCTTTTGTTCCTTTAG
F8-Rev	AAACATATATGGAGCAGACAATTGAGAAAGGGATCCCCGGGTACCGAGCTC
BACF1 For	AAC GATCCTGGTACACCTTGTTCGAGGACTTGAAGCTGCgcggccgcgacacctctagagtcgacctg
BACF3 Rev	GGTACTGCATTTGCCGCAGGAACACATGATTATTCAACTGgcggccgcgggtaccgagctcgaattc
BAC66 For	TAAAACAACTTTAATTAGCACTTTTAGTGTTTGAGTGCTgcggccgcgacacctctagagtcgacctg
BAC69 Rev	ATAATTGCAAAATATATAGATAAGGATACTTACCCAAGTGgcggccgcgggtaccgagctcgaattc
BAC25 For	CCATTTTTTAAAATACCTAATAAAATTTCAAAATCAAGGTgcggccgcgacacctctagagtcgacctg
BAC77 Rev	AAGTTAAAAACGCTCTTTTAAAAATTAATCAAAGTCCTTAgcggccgcgggtaccgagctcgaattc

SUPPLEMENTARY RESULTS

Assembly of overlapping DNA by a 2-step-thermocycled reaction

Step 1: Chew-back and anneal. Our previously described 2-step *in vitro* recombination method for assembling overlapping DNA molecules makes use of the 3'-exonuclease activity of T4 DNA polymerase (T4 pol) to produce ssDNA overhangs, and a combination of *Taq* DNA polymerase (*Taq* pol) and *Taq* DNA ligase (*Taq* lig) to repair the annealed joints¹ (**Supplementary Fig. 1a**). To better understand the kinetics of this reaction, 8 DNA molecules, each ~6 kb and overlapping by ~300 bp, were exposed to T4 pol at 37 °C for up to 18 min. Samples were removed every 2 min and annealed. Following a 10 min exonuclease reaction, the majority of the input DNA was annealed, and the predicted ~48 kb full-length product is observed. These reactions require the presence of PEG-8000, a reagent that induces macromolecular crowding (**Supplementary Fig. 1b**)².

Assembling DNA molecules with significantly smaller overlaps than 300 bp would have several advantages. When synthetic DNA fragments are joined, smaller overlaps would reduce the overall cost of synthesis. Additionally, small overlaps can be added to PCR primers. For these reasons, we determined whether DNA molecules with only 40 bp overlaps could be assembled. The assembly reaction in **Supplementary Fig. 1b** was performed using 4 DNA molecules, each 5 kb in length, and overlapping by 40 bp. Following a 2 min exposure to T4 pol, all 4 DNA molecules were efficiently assembled into the full-length 20 kb product (**Supplementary Fig. 1c**).

We next determined whether significantly larger DNA molecules could be joined by this method. Two ¼ molecules of the synthetic *M. genitalium* genome, C25-49 (144 kb) and C50-77 (166 kb), with a 257 bp overlap¹, were reacted with T4 pol for 15 min and annealed, then analyzed by field-inversion gel electrophoresis (FIGE) (**Supplementary Fig. 1d**). They were efficiently assembled into the 310 kb product (Mgen25-77). Further, when all 4 quarter molecules were reacted under the same conditions, the full-length synthetic *M. genitalium* genome (~583 kb) is assembled (**Supplementary Fig. 1e**).

Step 2: Repairing the assembled molecules. *Taq* pol is our preferred gap-filling enzyme since it does not strand-displace, which would lead to disassembly of the joined DNA fragments. It also has inherent 5'-exonuclease activity (or nick translation activity)³, which eliminates the need to phosphorylate the input DNA (a requirement for DNA ligation). This is because 5'-phosphorylated ends are created following nick translation. Further, this activity removes any non-complementary sequences (e.g. partial restriction sites), which would otherwise end up in the final joined product.

To verify that assembled DNA molecules have been successfully repaired, dsDNA products can be denatured at 94 °C in the presence of formamide and analyzed by agarose gel electrophoresis (**Supplementary Fig. 2a**). Repair was assessed for 2 pairs of ~5-6 kb DNA molecules with 40 bp or 300 bp overlaps. In each case, similar results were obtained (**Supplementary Fig. 2b**). Assembled, but unrepaired DNA molecules (lane 1) are denatured to ssDNA input in the presence of formamide (lane 2). In the absence of *Taq* lig, the nicks are not sealed and the 5'-exonuclease activity of *Taq* pol eliminates the overlapping DNA sequence, leading to disassembly of the DNA molecules (compare lanes 3 and 4). In the presence of *Taq* lig (lane 5), the nicks are sealed and a higher molecular weight ssDNA product is observed (lane 6). Thus, we conclude that

dsDNA molecules, with as little as 40 bp overlaps, are covalently joined by this assembly method.

Rolling circle amplification (RCA) of joined DNA molecules using phi29

polymerase. Assembly methods that employ a repair step to produce covalently sealed circular DNA molecules allow for the possibility of RCA⁴. This is not the case for assembly methods that omit a repair step. To demonstrate this, 4 fragments from ~1-3 kb and overlapping by 40 bp were assembled into a circle then repaired or left untreated. As expected, only repaired assembled products could be amplified by phi29 polymerase (**Supplementary Fig. 3**).

One-step-thermocycled DNA assembly

A DNA assembly method that requires the absence of dNTPs to achieve exonuclease activity, such as the T4 pol-based system described above, can not be completed in one step. This is because dNTPs are required at a later point to fill-in the gapped DNA molecules. Exonuclease III (ExoIII), which removes nucleotides from the 3' ends of dsDNA, is fully functional even in the presence of dNTPs so it is a candidate for a 1-step reaction. However, it will compete with polymerase for binding to the 3' ends. To eliminate this competition, and allow for 1-step DNA assembly, antibody-bound *Taq* pol (Ab-*Taq* pol) is used in combination with ExoIII (**Supplementary Fig. 4a**). In this assembly method, overlapping DNA fragments and all components necessary to covalently join the DNA molecules (i.e. ExoIII, Ab-*Taq* pol, dNTPs, and *Taq* lig) are added in a single tube, and placed in a thermocycler. At 37 °C, ExoIII is active (but Ab-*Taq* pol remains inactive) and recesses the 3' ends of the dsDNA molecules. The reaction is then shifted to 75 °C, which inactivates ExoIII. Annealing of the DNA molecules commences and the antibody dissociates from *Taq* pol, thus activating this enzyme. Further annealing, extension, and ligation is then carried out at 60 °C. **Supplementary Figure 4b** demonstrates the efficient assembly of four 5 to 7 kb DNA molecules with 40 bp overlaps or ~300 bp overlaps. To demonstrate that the joined DNA molecules are repaired by this method, assembly products were denatured in the presence of formamide and analyzed on agarose gels. The DNA molecules are efficiently assembled and repaired (**Supplementary Fig. 4c**).

Comparison of 5 DNA assembly methods

Cloning of assembled DNA molecules is a common application of our methods. Thus, it is important to determine which assembly method is best for cloning. We first compared the joining efficiencies of synthetic *M. genitalium* cassettes 66 to 69 (~6 kb each and with 80 bp overlaps) into a BAC with 40 bp overlaps to the ends of the assembly¹. The joining efficiencies of the methods described in **Supplementary Figure 1a**, **Supplementary Figure 4a**, and **Figure 1** were compared. We also included a comparison with 2 additional DNA assembly systems that omit fill-in and ligation steps from the method described in **Supplementary Figure 1a**. Each of these 5 methods efficiently and similarly assembled cassettes 66 to 69 into BAC66-69 as determined by FIGE (**Supplementary Fig. 5a**). Equal amounts of these DNA molecules were then transformed into *E. coli*. Ten randomly selected clones from each method were analyzed following *NotI* digestion, which released the vector from the ~23 kb insert (**Supplementary Fig. 5b**). For each method, 90 to 100% of the clones had the correct

insert. Omitting both DNA polymerase and ligase yields only 2% of the number of colonies achieved with complete repair. This emphasizes the importance of a repair step. Leaving the nicks unsealed but filling in the gaps increases the cloning efficiency to 44% of the complete reaction, suggesting that gaps can significantly influence cloning efficiencies in *E. coli*.

During the construction of the synthetic *M. genitalium* genome, we could not use the DNA assembly strategy shown in **Supplementary Figure 1a** to clone $\frac{1}{2}$ genomes from $\frac{1}{4}$ molecules in *E. coli*¹. We repeated the analysis presented in **Supplementary Figures 5a and 5b** to determine if either of these assembly methods could be used to clone Mgen25-77 (310 kb) from C25-49 and C50-77. Each method efficiently joined the 310 kb, half *M. genitalium* genome (**Supplementary Fig. 5c**). As expected from our previous study, this DNA molecule can not be cloned in *E. coli* using the strategy outlined in **Supplementary Figure 1a**. Filling in the gaps with T4 pol, but leaving the nicks unsealed, does not produce transformants. However, we find that the 1-step ExoIII- and T5 exo-based systems can be used to clone these large DNA molecules (**Supplementary Fig. 5d**). Thus, we have identified 2 DNA assembly systems that can be used to efficiently join and clone DNA molecules up to several hundred kb in length in *E. coli*, the approximate upper limit for transformation into this bacterium⁵.

SUPPLEMENTARY REFERENCES

1. D. G. Gibson, G. A. Benders, C. Andrews-Pfannkoch et al., *Science* **319** (5867), 1215 (2008).
2. S. B. Zimmerman and A. P. Minton, *Annu Rev Biophys Biomol Struct* **22**, 27 (1993).
3. A. Chien, D. B. Edgar, and J. M. Trela, *J Bacteriol* **127** (3), 1550 (1976).
4. F. B. Dean, J. R. Nelson, T. L. Giesler et al., *Genome Res* **11** (6), 1095 (2001).
5. Y. Sheng, V. Mancino, and B. Birren, *Nucleic Acids Res* **23** (11), 1990 (1995).