nature | methods

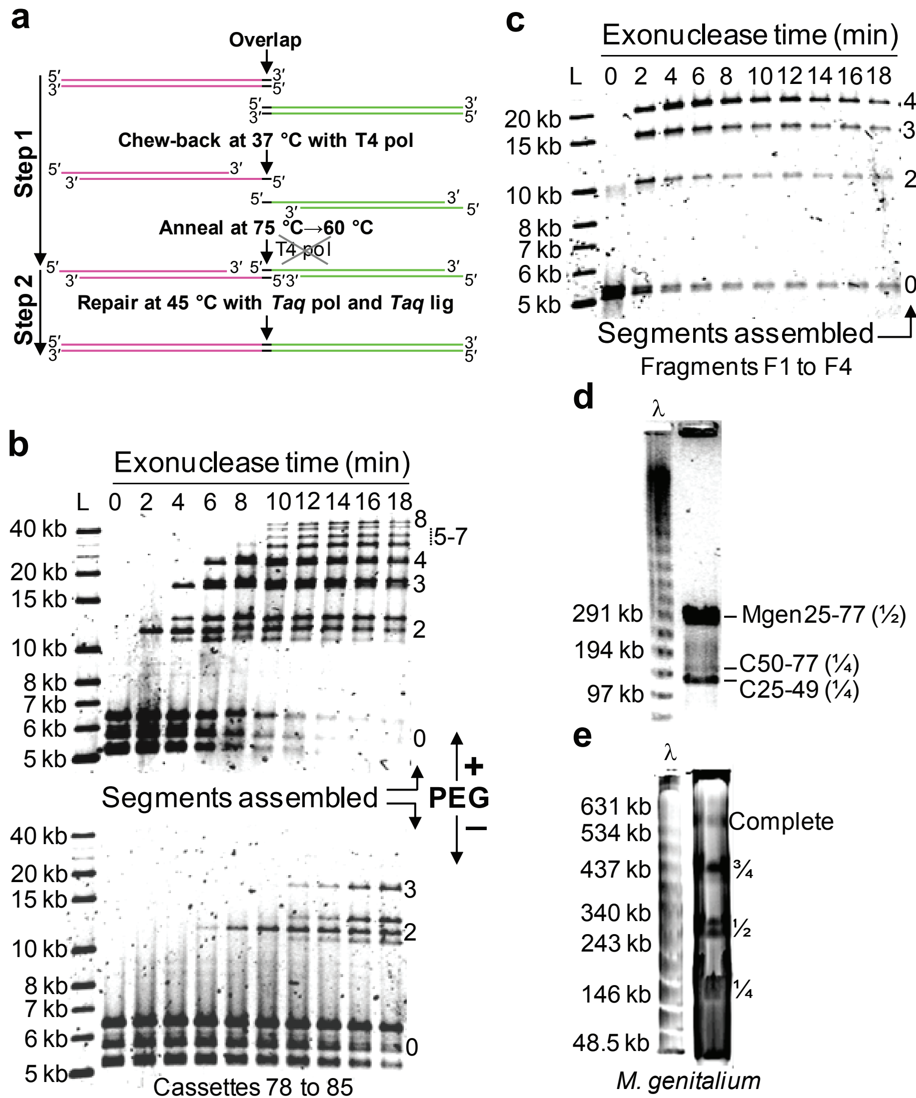
Enzymatic assembly of DNA molecules up to several hundred kilobases

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

Supplementary figures and text:

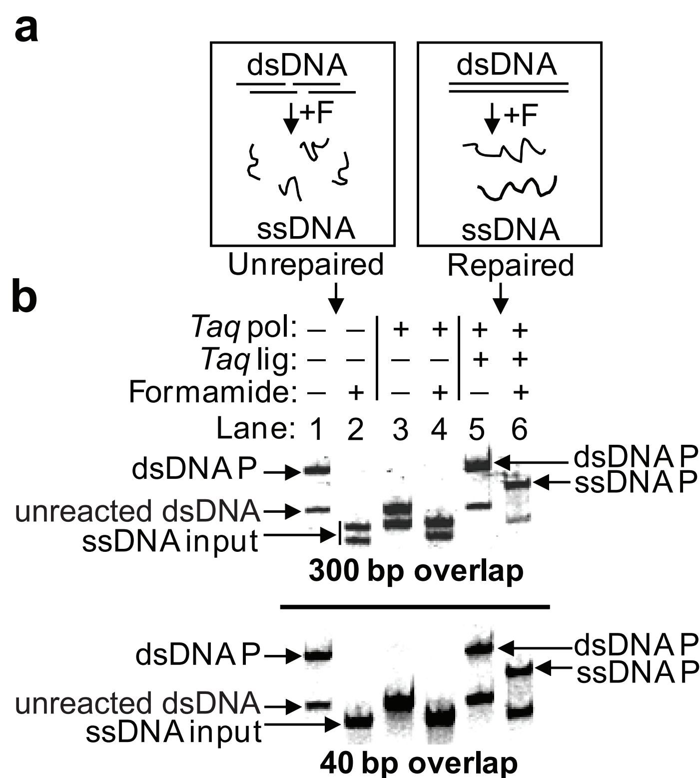
Supplementary Figure 1	Two-step-thermocycled *in vitro* recombination
Supplementary Figure 2	Assembled DNA molecules are covalently joined by the activities of *Taq* pol and *Taq* lig
Supplementary Figure 3	Rolling circle amplification of repaired assembly products
Supplementary Figure 4	One-step-thermocycled 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 sequence[1], 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* ¼ 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 FIGE. λ indicates the lambda DNA ladder (New England Biolabs [NEB]). **(e)** *M. genitalium* ¼ genome assemblies C1-24, C25-49, C50-77, and C78-101, ~150 kb each with 80 to 257 bp overlaps[1], were annealed following a 15 min exonuclease reaction at 37 °C. "Complete", "¾", "½", and "¼", 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.

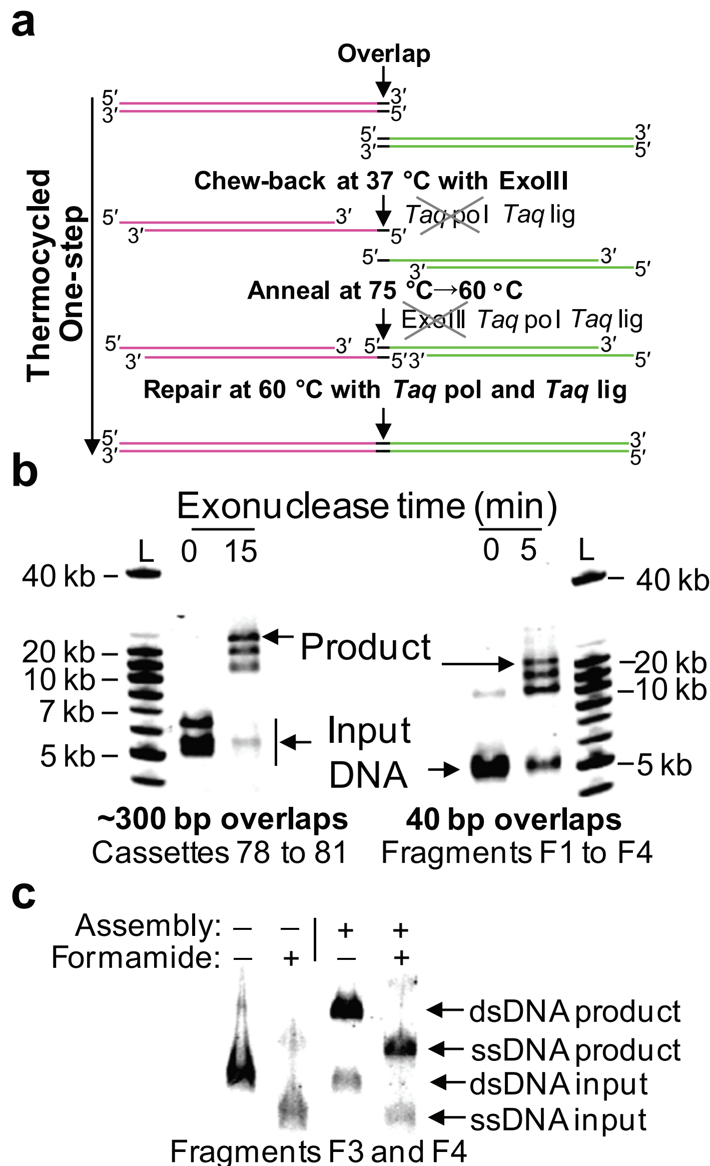


(a) Schematic outlining the strategy used to analyze the success of a repaired assembly reaction. Double-stranded DNA is denatured to ssDNA by incubating the reactions at 94 °C for 2 min in the presence of formamide (+F). If repair has occurred, the resulting ssDNA product remains intact and has a higher 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.

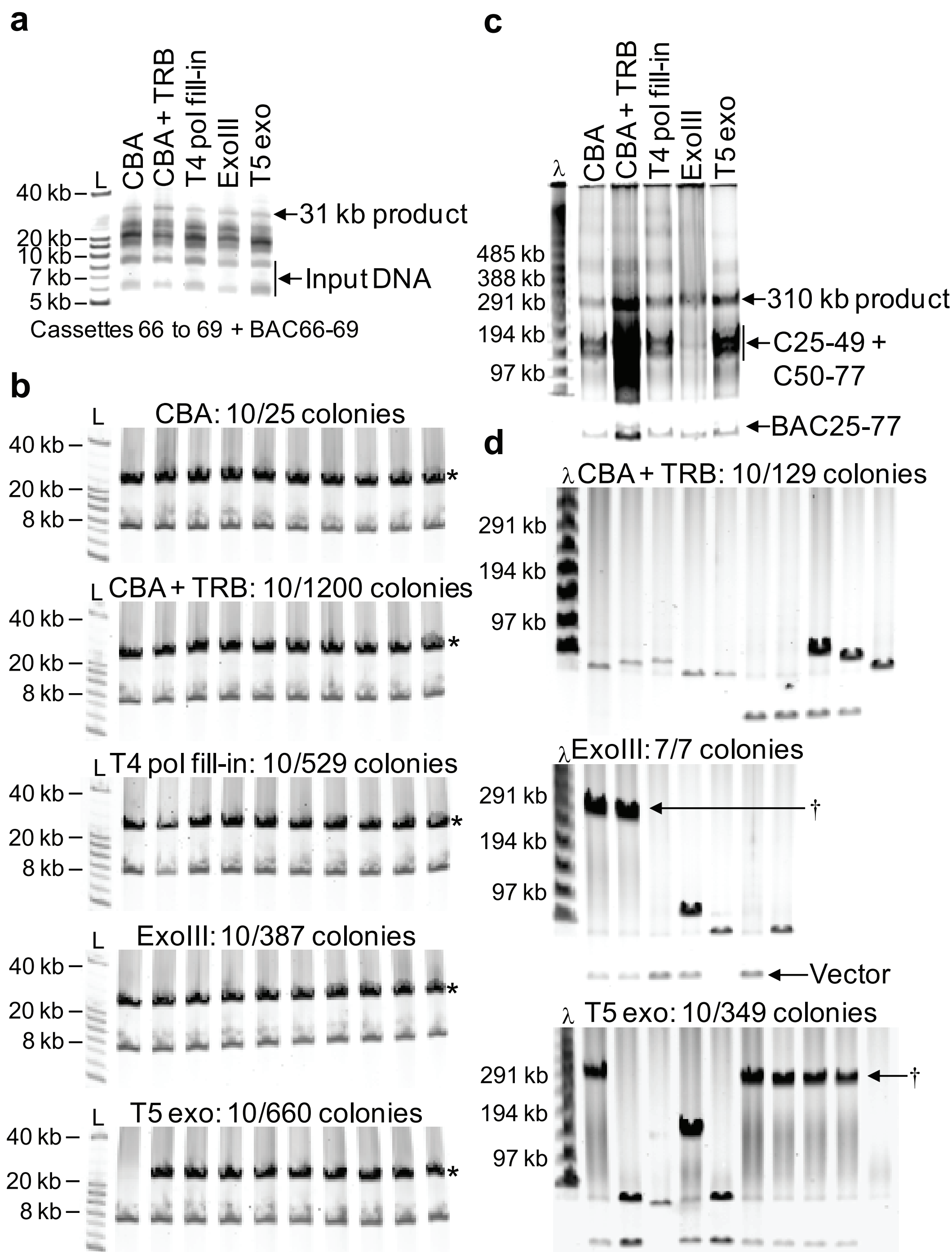
Four DNA fragments, F5 (1020 bp), F6, (1040 bp), F7 (2379 bp), and F8 (3246 bp), each with 40 bp overlaps, were joined into a 7,525 bp circle, as described in **Supplementary Figure 1a**, then repaired or left unrepaired. The unrepaired reaction contained all the components for repair; however, incubation 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 *Sph*I, 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.



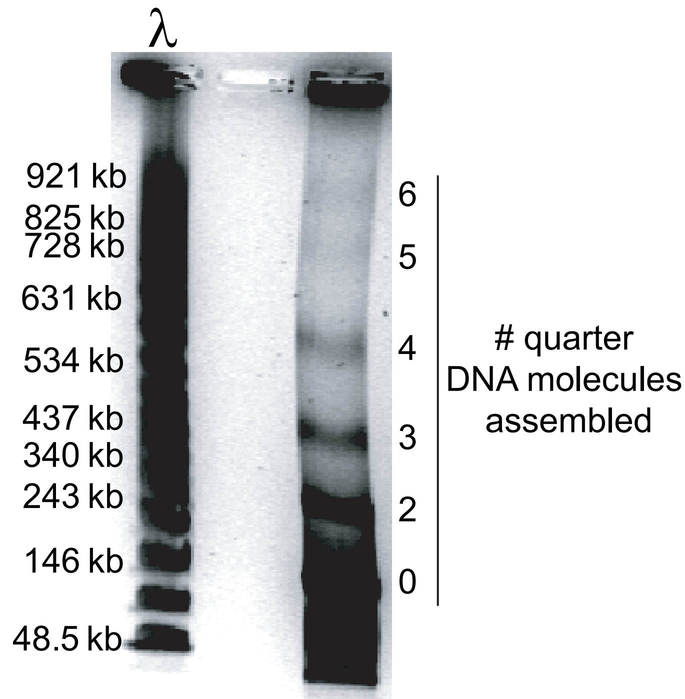
(a) Schematic outlining the approach used to assemble overlapping DNA segments in a single step, in a thermocycler. **(b)** Cassettes 78 to 81 (240 bp to 300 bp overlaps) and fragments F1 to F4 (40 bp overlaps) are assembled as in **(a)** then analyzed by U-5 FIGE. The completely assembled product and 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 FIGE 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 *Not*I 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 FIGE. **(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 *Not*I 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, *E. coli*
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* genome[1]. 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	GATTTCAGTGTAGTTAGGGCCAGTTGAATTCAAACCTGCC
F2-For	GGCAGGTTTGAATTCAACTGGCCCTAACTACACTGAAATC
F2-Rev	CTTGGTGCCATCAGCATTGTTCTCTGTACCGCCCACTGTC
F3-For	GACAGTGGGCGGTACAGAGAACAATGCTGATGGCACCAAG
F3-Rev	CAGTTGAATAATCATGTGTTCCTGCGGCAAATGCAGTACC
F4-For	GGTACTGCATTTGCCGCAGGAACACATGATTATTCAACTG
F4-Rev	TTATTTACCAAGAACCTTTGCCTTTAACATTGCAAAGTCA
F5-For	GCTTGCATGCATCCTGTTTATTCATCACAAACATTGAAC
F5-Rev	AATTCTGCAGTTTTTATTTCCTAACAGAACATTTTTTCTAGTATAGC
F6-For	TAGAAAAAATGTTCTGTTAGGAAATAAAAACTGCAGAATTAAAAGTTAGTGAACAAGAAAAC
F6-Rev	AGCCTCAAAAAGATAAAGAAAGGCTATTTATCTAGAGTCGACCTGCAGTTCAGATC
F7-For	CGACTCTAGATAAATAGCCTTTCTTTATCTTTTTGAGGC
F7-Rev	CCGGGGATCCCTTTCTCAATTGTCTGCTCCATATATGTT
F8-For	TGTTCAATGTTTGTGATGAATAAACAGGATGCATGCAAGCTTTTGTTCCCTTTAG
F8-Rev	AAACATATATGGAGCAGACAATTGAGAAAGGGATCCCCGGGTACCGAGCTC
BACF1 For	AAC GATCCTGGTACACCTTGTTTGCAGGACTTGAAGCTGCgcggccgcgatcctctagagtcgacctg
BACF3 Rev	GGTACTGCATTTGCCGCAGGAACACATGATTATTCAACTGgcggccgccgggtaccgagctcgaattc
BAC66 For	TAAAACAAACTTTAATTAGCACTTTTAGTGTTTGAGTGCTgcggccgcgatcctctagagtcgacctg
BAC69 Rev	ATAATTGCAAAATATATAGATAAGGATACTTACCCAAGTGgcggccgccgggtaccgagctcgaattc
BAC25 For	CCATTTTTTAAAATACCTAATAAAATTTCAAAATCAAGGTgcggccgcgatcctctagagtcgacctg
BAC77 Rev	AAGTTAAAAACGCTCTTTTAAAAATTAATCAAAGTCCTTAgcggccgccgggtaccgagctcgaattc

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 joints[1] (**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 overlap[1], 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

Nature Methods: doi: 10.1038/nmeth.1318

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 RCA[4]. 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 assembly[1]. 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 *Not*I 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 ½ genomes from ¼ 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 bacterium[5].

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).