Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries

Jiayuan Quan[1,2] & Jingdong Tian[1,2]

[1]Department of Biomedical Engineering, Duke University, Durham, North Carolina, USA. [2]Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina, USA. Correspondence should be addressed to J.T. (jtian@duke.edu).

Published online February 3 2011; doi:10.1038/nprot.2010.181

High-throughput genomics, proteomics and synthetic biology studies require ever more efficient and economical strategies to clone complex DNA libraries or variants of biological modules. In this paper, we provide a protocol for a sequence-independent approach for cloning complex individual or combinatorial DNA libraries, and routine or high-throughput cloning of single or multiple DNA fragments. The strategy, called circular polymerase extension cloning (CPEC), is based on polymerase overlap extension and is therefore free of restriction digestion, ligation or single-stranded homologous recombination. CPEC is highly efficient, accurate and user friendly. Once the inserts and the linear vector have been prepared, the CPEC reaction can be completed in 10 min to 3 h, depending on the complexity of the gene libraries.

INTRODUCTION

Circular polymerase extension cloning (CPEC) is a simple, efficient and economical circular DNA assembly and cloning method developed to meet the ever-increasing demand from high-throughput genomics, proteomics and synthetic biology. In addition to routine single-gene cloning, CPEC is ideal for a wide variety of other applications, including complex gene library cloning, high-throughput expression cloning and multiway assembly of genetic pathways[1].

CPEC is a single-tube, one-step reaction that normally takes 5–10 min to complete for everyday laboratory cloning. The method is directional, sequence independent and ligase free. It uses the polymerase extension mechanism[2] to join overlapping DNA fragments into a double-stranded circular form, such as a plasmid. In a typical CPEC reaction, linear double-stranded insert(s) and vector are first heat-denatured; the resulting single strands then anneal with their overlapping ends and extend using each other as a template to form double-stranded circular plasmids. In CPEC, all overlapping regions between insert(s) and the vector are unique and carefully designed to have very similar and high melting temperatures (T_m), which eliminates vector reannealing and concatenation of inserts and makes CPEC very efficient and accurate. The low concentrations of fragments in the reaction favor plasmid circularization and effectively prevent plasmid concatenation. After the CPEC reaction, the perfectly formed double-stranded circular plasmids, with one nick in each strand, can be directly transformed into competent host cells.

CPEC has its root in assembly PCR, also called polymerase cycling assembly (PCA), which has typically been used to assemble double-stranded linear DNA constructs[3–7]. An earlier attempt by Stemmer *et al.*[4] to assemble a circular plasmid using a pool of overlapping oligonucleotides resulted in the formation of linear concatemers of the plasmid unit, which had to be cut by a unique restriction enzyme and then circularized by ligation. Since then, there have been very few reports on using overlap extension as a cloning method[8], until our laboratory recently developed and optimized the CPEC procedure not only for single-gene cloning but also for high-efficiency library cloning, multiway cloning and combinatorial library cloning[1]. CPEC differs from PCA mainly in that CPEC forms abundant circular products of distinct lengths whereas PCA forms a smear of linear products, and the correct-sized product can only be obtained by further PCR amplification of PCA products.

Compared with existing cloning strategies, either sequence dependent or independent, CPEC offers significant benefits by combining simplicity, efficiency, versatility and cost-effectiveness in one method[9]. For example, sequence-dependent cloning methods, including restriction digestion/ligation[10,11] and homologous recombination[12,13], require the presence of specific sequences (e.g., restriction sites or recombination sites) at the ends of both the vector and the insert and often leave unwanted sequences in the final products, which limits their usage in library cloning and pathway construction. On the other hand, the existing sequence-independent cloning methods, including ligase-independent cloning (LIC)[14], sequence and ligase-independent cloning (SLIC)[15], uracil DNA glycosylase cloning[16,17] and In-Fusion cloning (Clontech), require special steps, reagents, kits or enzymes (e.g., exonuclease, uracil DNA glycosylase, polymerases, etc.) to generate single-stranded overhangs at the ends of the vector and the insert(s). LIC and uracil DNA glycosylase cloning processes also require the presence or absence of certain types of nucleotides at specified locations in the overlapping regions, making these methods less 'sequence independent'.

Complex library cloning and multiway pathway assembly require high cloning efficiency and accuracy. Although other relevant cloning methods only allow the overlapping fragments to anneal or recombine once, CPEC allows multiple annealing-extension cycles that not only increase the chance of hybridization but also permanently join the fragments through polymerase extension, thereby maximizing the cloning efficiency. Whereas the other relevant cloning methods perform the critical annealing/incubation step under ambient temperature, which tends to cause nonspecific hybridization and leads to compromised cloning efficiency and accuracy, CEPC designs the overlapping ends to have very similar T_m (±2 °C) and performs the annealing step at high, stringent temperatures

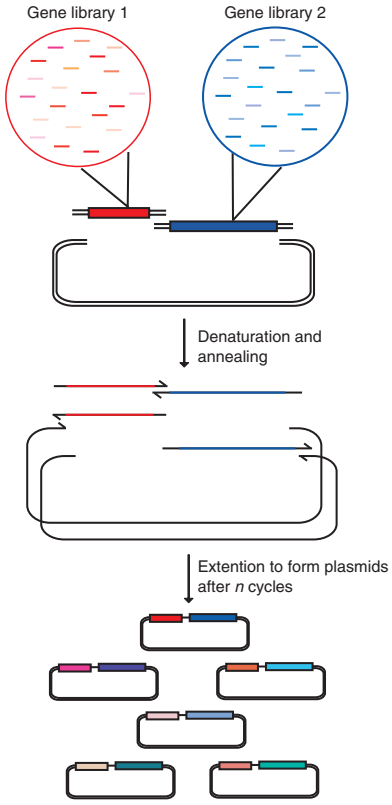

Figure 1 | Schematic diagram of CPEC cloning of combinatorial gene libraries. Two gene libraries are cloned in frame into a vector. The vector and the inserts share overlapping regions at the ends. In each CPEC cycle, after denaturation and annealing, the hybridized inserts and vector extend using each other as a template until they complete a full circle. The assembled plasmid library can be directly used for transformation into competent cells.

(typically in the range of 55–65 °C) to ensure highest accuracy in multiway assembly and complex library cloning. Unlike PCR, CPEC does not amplify sequences and therefore does not propagate errors with an increased number of thermal cycles.

In conclusion, the CPEC cloning strategy can accomplish all the tasks that other *in vitro* cloning methods can perform and compares favorably in almost every aspect, including accuracy, efficiency, speed and cost[1]. However, because of the intrinsic limitations of DNA polymerase and the overlap extension method, we expect that there will be limits to the size of the plasmid and the total number of fragments that can be assembled by CPEC, although no exhaustive experiments have been done to determine those limits. So far, we have comfortably assembled an 8.4-kb plasmid using four PCR fragments of 3,280, 2,959, 2,047 and 171 bp[1]. With careful optimization, linear products as long as 20 kb have been constructed using overlap extension PCR[5], which could indicate an approximate upper limit for CPEC. To assemble bigger constructs with more fragments, *in vivo* homologous recombination methods using yeast as a host organism may be attempted[18,19].

We believe that CPEC has a real advantage in complex and combinatorial library cloning because of its exceptionally high efficiency[1]. The combinatorial library cloning strategy using CPEC is illustrated in **Figure 1**. In this example, two libraries are cloned simultaneously into a single vector for expression or functional screens to identify the best combinatorial sequences. It is anticipated that such screens will be performed more and more frequently in synthetic biology applications to construct and identify the optimal macromolecular complexes or gene networks. So far, CPEC is the only *in vitro* method that works well in our hands for combinatorial library cloning[1].

Experimental design

Design of overlapping sequences between vector and insert(s). The key to successful CPEC library construction and multiway CPEC is to carefully select and design the overlapping sequences between the vector and the insert(s) so that all overlapping regions share very similar T_m. The T_m of the overlapping regions should be as high as possible (ideally between 60 and 70 °C) to maximize hybridization specificity. The T_m of all overlapping regions in the final CPEC assembly reaction should match each other as closely as possible, ideally with differences within ±2 to 3 °C. This will help eliminate mis-hybridization and ensure highest cloning efficiency and accuracy. The length of the overlapping region, typically between 15 and 35 bases, is of secondary consideration and is dictated by the T_m. Standard PCR primer selection rules and software can be applied to facilitate the design process[20]. If PCR is used to introduce overlapping regions with the vector or with adjacent fragments, primers should be designed to include at least two parts, each hybridizing to one end of the two neighboring fragments to be joined. If an additional short sequence needs to be inserted between two existing fragments, it can be simply included in the primer design between the two overlapping regions. CPEC primer design examples for single and multiple insert cloning are given in **Figure 2**. In case synthetic genes or libraries are used, it would be more convenient to directly add overlapping regions during synthesis (**Fig. 2**). No chemical modifications to the primers are required (e.g., phosphorylations).

Preparation of linear vector. The linear vector can be prepared most conveniently by PCR amplification using primers designed to introduce overlapping regions with the insert(s), as described below in the Procedure; this approach offers the most flexibility in selecting cloning sites. If a convenient restriction site is available on the vector that does not introduce unwanted sequences, restriction digestion can also be used to linearize the vector. To prevent carryover of undigested or intact circular vector templates, we recommend gel purification of the linear vector after PCR amplification or restriction digestion. In addition, to eliminate the effect of any residual carryover vector, we recommend using an empty vector as the starting material for PCR amplification or restriction digestion; this way, any carryover of the empty vector will not interfere with downstream functional assays or screens.

Preparation of inserts. The inserts can be a single gene, a gene library, multiple genes or even multiple libraries. They can be isolated from natural sources or synthesized on the basis of *in silico* designs. Irrespective of whether they are single sequences or libraries, ensure that they share overlapping regions with the vector or neighboring fragments, as described above. If PCR is used to prepare the inserts, as described in the Procedure below, a high-fidelity DNA polymerase (e.g., Phusion DNA polymerase) is preferred in order to minimize the introduction of mutations or addition of an extra nucleotide at the ends of amplified products. Gel purification is sometimes necessary to ensure purity of the products.

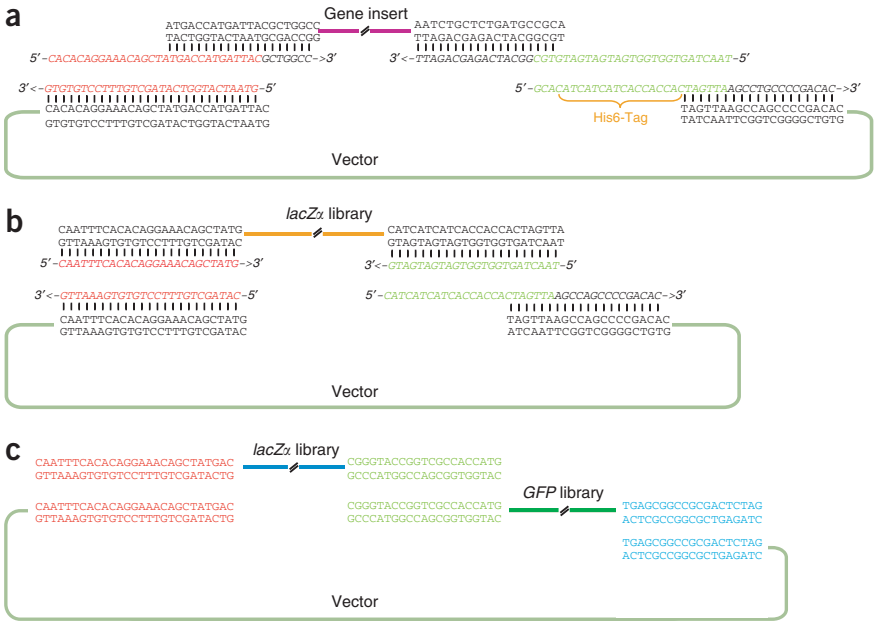

Figure 2 | Schematic diagrams of primer design in CPEC. Fragment ends are shown as double stranded. Overlapping regions between the insert(s) and the vector are shown in colors. The 5′→3′ direction of the PCR primers used to amplify the insert(s) or the vector is marked. (**a**) An example of a single insert cloning. The primers designed to amplify the gene insert introduce overlapping regions with the vector. Two primers are used to add an 18-base sequence encoding a His6-tag to the end of the gene. (**b**) Primer sequences used for *lacZα* gene library cloning as described in Anticipated Results. (**c**) Sequences of overlapping regions in combinatorial library cloning of the *lacZα* gene library and the *GFP* gene library as described in Anticipated Results. Overlapping regions have been added to the ends of the synthetic *lacZα* and *GFP* libraries. Therefore, there is no need to use bipartite PCR primers to add overlapping regions.

CPEC cloning. In the final CPEC assembly and cloning reaction, prepared linear vector and inserts are mixed together with the reaction cocktail, which includes dNTPs, 1× PCR buffer and a thermal-stable high-fidelity DNA polymerase. The composition of the CPEC reaction cocktail is almost identical to that of a standard PCR, except that no primers are added. The final vector concentration is normally in the range of 5–10 ng μl^{-1} and the insert-to-vector molar ratio is in the range of 1:1 to 2:1.

The thermal cycling conditions are also similar to those used for a standard PCR reaction, except that fewer cycles are needed. For example, 1–5 cycles are sufficient to clone a single insert or a less complex library; 15–30 cycles may be needed to assemble multiple fragments or clone complex or combinatorial libraries. Depending on the number of cycles used, the total reaction time can be anywhere from 10 min to a few hours. A stringent annealing temperature should be used for thermal cycling, the value of which is determined by the T_m of the overlapping regions and recommendations for the particular DNA polymerase used. Extension time is calculated by the size of the construct and the extension rate of the polymerase. For single-fragment or single-library cloning, shorter extension times can be used.

To quickly assess whether a CPEC reaction is successful before proceeding to the transformation step, a small aliquot of the reaction can be separated by electrophoresis with an agarose gel. An appropriate molecular marker and the amount of insert/vector DNA present in the initial CPEC reaction should be loaded side by side onto the gel as controls. If the reaction has been successful, a new high-molecular-weight band corresponding to the total linear length of the final construct should be visible on the gel. The presence of intermediate assembly products or unincorporated vector/insert(s) in the same lane is not a problem for downstream experiments and need not be removed as long as a band representing the full-length product is visible on the gel.

Transformation. The CPEC cloning product can be directly used for transformation without further purification. For library cloning, it is strongly recommended to use competent cells with transformation efficiencies greater than 1×10^8 c.f.u. (colony-forming units) per μg (e.g., GC5 competent cells, Genesee Scientific). A number of methods can be used to determine the presence of inserts in bacterial colonies grown on a culture plate, including colony PCR (as described in the PROCEDURE below), restriction mapping or direct sequencing.

MATERIALS

REAGENTS

- Cloning vectors. Commercial or custom designed vectors can be used.
- dNTP mix (dATP, dCTP, dGTP and dTTP; e.g., Bioline, cat. no. BIO-39043)
- Oligonucleotide primers. Custom DNA primer synthesis is available from commercial suppliers (primer sequences used to generate example data for the Anticipated Results are listed in **Supplementary Tables 1** and **2**).
- Phusion High-Fidelity DNA polymerase with 5× Phusion HF buffer (Finnzymes, cat. no. F-530) ▲ **CRITICAL** A high-fidelity DNA polymerase is required for CPEC cloning and the Phusion enzyme works very well in our experience.
- DNase/RNase-free water (Sigma-Aldrich, cat. no. W4502)
- Taq DNA polymerase with ThermoPol reaction buffer (New England Biolabs, cat. no. M0267)
- Agarose (Denville Scientific, cat. no. CA3510-8)
- Tris Acetate-EDTA buffer (10× TAE buffer; Sigma-Aldrich, cat. no. T9650)
- Ethidium bromide solution (Sigma-Aldrich, cat. no. E1510) ! **CAUTION** Ethidium bromide is a known mutagen and suspected carcinogen. When handling it, gloves should be worn at all times, and appropriate care should be taken to avoid skin contact.
- Gel loading dye, blue (6×; New England Biolabs, cat. no. B7021S)
- DNA ladder (1 kb; New England Biolabs, cat. no. N3232S)
- DNA ladder (1 kb; Bio-Rad, cat. no. 170-8355)
- DNA ladder (100 bp; New England Biolabs, cat. no. N3231S)
- LB agar (Sigma-Aldrich, cat. no. L3027-1KG)
- LB broth, Miller (BD, cat. no. 244620)
- GC5 competent cells (Genesee Scientific, cat. no. 42-653)

- S.O.C. medium (Cellgro, cat. no. 46-003-CR)
- Glycerol (Sigma-Aldrich, cat. no. G5516)
- GeneJET plasmid miniprep kit (Fermentas, cat. no. K0503)
- E.Z.N.A. gel extraction kit (Omega, cat. no. D2501)
- ExoSAP-IT for PCR cleanup (Affymetrix, cat. no. 78200)
- Distilled, deionized water (ddH_2O)

EQUIPMENT

- Thermal Cycler
- Microcentrifuge
- Electrophoresis apparatus for agarose gels (e.g., Gel XL Ultra V-2 electrophoresis system)
- UV transilluminator
- Gel imaging system (e.g., Alpha Innotech FluoChem multi-imaging system)
- NanoDrop spectrophotometer (ND-1000, Thermo Scientific)
- Shaker incubator at 37 °C
- Cabinet incubator at 37 °C
- Water bath at 42 °C
- Microwave
- Glassware, miscellaneous (flasks, bottles, and so on)
- Razor (for cutting gel bands)
- Parafilm

REAGENT SETUP

Oligonucleotide primers Prepare stock solutions of primers (e.g., 100 µM) using DNase/RNase-free water. Prepare aliquots of 10× working solution (e.g., 10 µM) and store at −20 °C to prevent contamination of stock and repeat freeze–thaw cycles. Under proper storage conditions, active working solutions of primers that are subject to freeze–thawing should be stable for several weeks, and reserve solutions of primers should be stable for at least a year.

1–1.5% (wt/vol) agarose gel Place the gel trays and selected combs into the gel casting stand on a level surface. Weigh the required amount of agarose powder into an appropriate glass flask containing a measured amount of 1× TAE buffer. Heat in a microwave until the agarose is completely dissolved. Allow the melted agarose to cool down to 70 °C or lower before pouring. Add ethidium bromide stock (10 mg ml^{-1}) to the cooled agarose to a final concentration of 0.5 µg ml^{-1}. Stir the solution to disperse the ethidium bromide and then pour it into the gel trays. Remove any air bubbles on the surface or inside the gel and allow the gel to solidify for ~20 min at room temperature (25 °C). **! CAUTION** Ethidium bromide is a known mutagen and suspected carcinogen. When handling, gloves should be worn at all times, and appropriate care should be taken to avoid skin contact.

DNA ladders Both 1-kb and 100-bp DNA ladders are diluted to 250 µg ml^{-1} with DNase/RNase-free water and 6× gel loading dye (supplied with the DNA ladder). For a 5-mm-wide lane, 2 µl of the mixture should be loaded onto the agarose gel. The amount of mixture should be scaled up or down, depending on the width of the lanes.

50% (vol/vol) sterile glycerol solution Pour equal amounts of glycerol and ddH_2O into a glass bottle. Mix well and autoclave. The glycerol solution can be stored at room temperature.

PROCEDURE

Insert preparation ● TIMING 2–2.5 h

1| Set up the PCR on ice as described below. Note that the amount of library DNA that is needed as template is generally larger than that required for standard PCR.

Initial concentration	Volume per 50 µl reaction	Final amount in 50 µl reaction
Phusion HF buffer (5×)	5 µl	1×
dNTP mix (40 mM)	1 µl	0.8 mM
Phusion high-fidelity DNA polymerase (2 U $µl^{-1}$)	0.5 µl	1 U
Forward primer (10 µM)	2.5 µl	0.5 µM
Reverse primer (10 µM)	2.5 µl	0.5 µM
DNA library (1–50 ng)	Variable	1–50 ng
ddH_2O	Up to 50 µl	—

2| Run the PCR under the following conditions. For the Phusion enzyme, the annealing temperature should be 3 °C higher than the lower T_m of the two primers; T_m should be calculated using the nearest-neighbor method[21].

Cycle number	Denature	Anneal	Extend
1	98 °C, 30 s		
2–31	98 °C, 10 s	(T_m + 3) °C, 30 s	72 °C, 15 s/kb
32			72 °C, 5 min

3| Combine 20 µl of the finished PCR with 4 µl of 6× DNA loading dye and run the product on a 1–1.5% agarose gel. The time required for electrophoresis is dependent on the length of the expected product.

4| Visualize the resolved product with a UV transilluminator or a gel imaging system. A prominent and unique band at the desired position of the length of the insert should be seen.

5| Cut the PCR product band from the gel with a clean sharp razor and purify the insert DNA using a commercial kit (e.g., E.Z.N.A. gel extraction kit). Measure the concentration of the amplified gene library DNA with the NanoDrop spectrophotometer.
■ **PAUSE POINT** Purified PCR product can be stored at −20 °C for at least 1 year if not used immediately.

Preparation of linear vector ● TIMING 2–3 h
6| Set up the PCR on ice as tabulated below:

Initial concentration	Volume per 50 μl reaction	Final amount in 50 μl reaction
Phusion PCR buffer (5×)	10 μl	1×
dNTP mix (40 mM)	1 μl	0.8 mM
Phusion High-Fidelity DNA polymerase (2 U μl^{-1})	0.5 μl	1 U
Forward primer (10 μM)	2.5 μl	0.5 μM
Reverse primer (10 μM)	2.5 μl	0.5 μM
Vector DNA template (1–10 ng)	Variable	1–10 ng
ddH_2O	Up to 50 μl	

7| Run the PCR with the following conditions. Again, the annealing temperature should be 3 °C higher than the lower T_m of the two primers.

Cycle number	Denature	Anneal	Extend
1	98 °C, 30 s		
2–31	98 °C, 10 s	(T_m + 3) °C, 30 s	72 °C, 15 s per kb
32			72 °C, 5 min

8| Combine 20 μl of the finished reaction with 4 μl of 6× DNA loading dye and run the product on a 1–1.5% (wt/vol) agarose gel. The time for electrophoresis is dependent on the length of the expected product.

9| Visualize the resolved product with a UV transilluminator or a gel imaging system. A prominent and unique band at the desired position of the length of the linear vector should be seen.

10| Cut the PCR product band from the gel with a clean sharp razor and purify DNA using a commercial kit (e.g., E.Z.N.A. gel extraction kit). Measure the concentration of the amplified gene library DNA with the NanoDrop spectrophotometer.
■ **PAUSE POINT** Purified linear vectors can be stored at −20 °C for at least 1 year if not used immediately.

CPEC of single or combinatorial gene libraries ● TIMING 1–3 h
11| Set up the cloning reaction on ice as follows. The amount of insert required will be dependent on its size and should be calculated to maintain an insert:vector molar ratio between 1:1 and 2:1.

Initial concentration	Volume per 20 μl reaction	Final amount per 20 μl reaction
Phusion HF buffer (5×)	4 μl	1×
dNTP mix (40 mM)	0.4 μl	0.8 mM
Phusion high-fidelity DNA polymerase (2 U μl^{-1})	0.2 μl	1 U
Vector DNA	Variable	100–200 ng
Insert gene library(-ies)	Variable	—
ddH_2O	Up to 20 μl	

▲ **CRITICAL STEP** The 20-μl reaction volume has been tested for cloning of at least two complex gene libraries. For cloning simpler inserts, such as a single fragment or one library, lower volumes can be used. The volume can also be scaled up to 50 μl, especially if the researcher wants to analyze the cloning product on an agarose gel.

12| Run the CPEC reaction with the following conditions:

Cycle number	Denature	Slow ramp anneal	Anneal	Extend
1	98 °C, 30 s			
2–26	98 °C, 10 s	70 to 55 °C (0.1 °C s^{-1}), 3 min	55 °C, 30 s	72 °C, 10–20 s per kb
27				72 °C, 5 min

▲ **CRITICAL STEP** Slow ramp annealing is optional if only one gene or gene library is involved in the cloning reaction.
▲ **CRITICAL STEP** The cycle number should be determined by the complexity and target efficiency of the cloning. If only one gene or gene library is involved, 1–10 cycles can be used. With multiple inserts or libraries, more cycles are generally required. We have found that up to 15–30 cycles are necessary for combinatorial library cloning[1]. For new users, we recommend testing a range of cycle numbers to determine the optimal number of cycles.
▲ **CRITICAL STEP** Extension time is determined by the polymerase used and may vary according to the complexity of the cloning. **Table 1** provides a guideline for using the Phusion enzyme. We recommend that new users test a range of extension times to find the most suitable one for their purposes.

13| Combine 10 µl of the finished reaction with 2 µl of 6× DNA loading dye and run the product on a 1–1.5% agarose gel to assess whether the CPEC reaction is successful. The time for electrophoresis is dependent on the length of the expected product.

14| Visualize the resolved products with a UV transilluminator or a gel imaging system.
▲ **CRITICAL STEP** This step is to confirm the success of the cloning process. A prominent band should be present on the agarose gel representing the total length of the vector plus inserts. Sometimes a high-molecular-weight smear and/or additional bands representing excessive or unincorporated vector or inserts may also be visible but should not affect the subsequent transformation step.
■ **PAUSE POINT** Cloning product can be stored at −20 °C for at least 1 year if not used immediately.
? TROUBLESHOOTING

Transformation of the cloning reaction ● TIMING 18–20 h (overnight)
15| Thaw the required number of tubes containing high-efficiency competent cells of your choice on wet ice. We use GC5 competent cells with efficiency ≥10^9 transformants per µg pUC19 DNA. We recommend following the manufacturer's recommended procedure for transformation; Steps 16–23 are based on the supplier's protocol for transformation of GC5 competent cells.

16| Add 1–5 µl of the cloning reaction to 50 µl of GC5 competent *Escherichia coli* cells in a microcentrifuge tube and gently tap the tube to ensure an even distribution of DNA in the solution.

17| Incubate the tubes on ice for 30 min.

18| Heat shock the cells for 45 s in a 42 °C water bath and then leave the tubes on ice for 2 min.
▲ **CRITICAL STEP** Follow the recommended incubation time and temperature precisely to ensure optimal transformation results.

19| Add 450 µl of room temperature S.O.C. medium to each transformation reaction.

20| Shake the tubes in a shaker incubator at 225 r.p.m. at 37 °C for 1 h.

21| Spread 50–100 µl of transformation reaction evenly on each 15-cm LB-agar plate containing appropriate antibiotics (depending on the vector used) and incubate at 37 °C overnight (16–18 h).

TABLE 1 | Recommended CPEC extension times for Phusion polymerase.

Single fragment or library as insert		Multiple fragments or libraries as inserts	
Length of the longer DNA fragment	Extension time	Length of the complete plasmid	Extension time
<2 kb	5 s per kb	<6 kb	12 s per kb
2–3 kb	5–10 s per kb	6–7 kb	14 s per kb
3–4 kb	10–15 s per kb	7–8 kb	16 s per kb
>4 kb	20 s per kb	>8 kb	18 s per kb

■ **PAUSE POINT** The plates can be sealed with Parafilm and stored at 4 °C if not used immediately.
? TROUBLESHOOTING

22| The next day, pick a desired number (varies according to the application) of isolated colonies from each plate and culture each colony in 3–10 ml of LB broth for 16–18 h (overnight).

23| Take 700 µl of the overnight bacterial culture and mix it well with 300 µl of 50% (vol/vol) sterile glycerol in a cryovial to prepare 15% (vol/vol) glycerol stocks and store at −80 °C.

Colony PCR (optional) ● **TIMING 2.5 h**
24| Colony PCR is performed to determine whether the picked colonies have insert(s) of the correct size. Colonies picked directly from the plate or from overnight cultures can be used as the template. Set up a PCR reaction on ice as follows:

Components	Final volume per 20 µl reaction	Final amount per 20 µl reaction
ThermoPol reaction buffer (10×)	2 µl	1×
dNTP mix (40 mM)	0.4 µl	0.8 mM
Taq DNA polymerase (5 U μl^{-1})	0.1 µl	0.5 U
Bacteria colony or overnight culture	0 µl/1 µl[a]	—
Vector forward primer (10 µM)	0.4 µl	0.2 µM
Vector reverse primer (10 µM)	0.4 µl	0.2 µM
ddH_2O	Up to 20 µl	

[a]If colony is used, gently tap the pipette tip on the colony and then dip it into the prepared PCR reaction mix.

25| Run the PCR using the following conditions. The T_m will depend on the vector primer pair.

Cycle number	Denature	Anneal	Extend
1	95 °C, 5 min		
2–26	95 °C, 30 s	Tm °C, 30 s	68 °C, 1 min/kb
27			68 °C, 5 min

■ **PAUSE POINT** PCR products can be stored at −20 °C for at least 1 year if not used immediately.

26| Combine 10 µl of the finished reaction with 2 µl of 6× DNA loading dye and run the product on a 1–1.5% (wt/vol) agarose gel. The time for electrophoresis is dependent on the length of the expected product.

27| Visualize the resolved product with a UV transilluminator and determine the presence of inserts and their size. Inserts of correct sizes can be further used for sequencing or restriction digestion if needed.
? TROUBLESHOOTING

Sequencing sample preparation (optional) ● **TIMING 1 h**
28| Prepare sequencing samples by first purifying the colony PCR reaction with ExoSAP-IT. Set up a reaction on ice as follows:

Components	Final volume per 7 µl reaction
Colony PCR reaction product	5 µl
ExoSAP-IT	2 µl

29| Incubate the mixture at 37 °C for 15 min to degrade remaining primers and nucleotides.

30| Incubate at 80 °C for 15 min to inactivate ExoSAP-IT. Purified colony PCR reactions are now ready for direct sequencing, either in-house or by commercial services.
■ **PAUSE POINT** Treated PCR products can be stored at −20 °C for at least 1 year if not used immediately.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
14	Little or no desirable cloning product	The CPEC conditions were not optimal; did not use a high-fidelity DNA polymerase; did not use high-quality library DNA	Optimize the cycle number and/or extension time; use a high-fidelity DNA polymerase (e.g., Phusion); use purified high-quality library DNA
21	Too few colonies after transformation	Competent cells used were not highly efficient	Use competent cells with library efficiency $>10^8$ c.f.u. per μg
		Cloning was not successful	Analyze the CPEC cloning product on the agarose gel to make sure that the reaction was successful
27	Many colonies yield PCR products of the wrong size	Did not run the cloning product on the gel to confirm	Analyze the CPEC cloning product on an agarose gel to make sure the cloning was successful
		Did not select colonies with positive reporter gene signals	If possible, select colonies with positive reporter gene signals for colony PCR
		Did not use high-quality library DNA	Check the quality of the library DNA used in the cloning. Make sure to use high-quality library DNA
		Did not cut the right band when gel purifying the library DNA	Cut the correct-sized band when extracting the gene library DNA from the agarose gel
	Many colonies with positive reporter signals yield PCR products of the wrong size	Did not use a negative control vector plasmid	Prepare a negative control vector plasmid with the reporter gene function abolished by either inserting a stop codon(s) in the MCS region or other methods before linearizing the vector
	Little or no PCR product	Did not use an appropriate DNA polymerase for colony PCR	Use an appropriate DNA polymerase for colony PCR (e.g., Taq DNA polymerase with ThermoPol buffer) and use an initial denaturation time of no less than 5 min

● TIMING

Steps 1–5, Insert preparation (e.g., adding overlapping sequences to a gene library by PCR): 2–2.5 h depending on the length of the inserted library
Steps 6–10, Preparation of linear vector (e.g., by PCR): 2–3 h depending on the length of the vector
Steps 11–14, CPEC of single or combinatorial gene libraries: 1–3 h depending on the complexity of the inserted library(s)
Steps 15–23, Transformation of the cloning reaction (e.g., for the expression and screening of gene libraries): overnight
Steps 24–27, Colony PCR (optional): 2.5 h
Steps 28–30, Sequencing sample preparation (optional): 1 h

ANTICIPATED RESULTS

By following the steps described in this protocol, after the linear vector and inserts are prepared, CPEC library cloning can be completed in less than 10 min to a few hours, depending on the complexity of the libraries. The cloning process is highly efficient with minimal unwanted products (<5%, ref. 1). For single-library cloning, we examined the cloning of a synthetic library containing $>10^{15}$ codon variants of a 306-bp modified *lacZα* gene construct into a modified pUC19 vector (pUC19stop, with a stop codon inserted in the multiple cloning site (MCS) region of the vector) (**Fig. 3a**). After CPEC cloning with five thermal cycles, the reaction was analyzed on an agarose gel. The upper band in Lane 4 indicates the

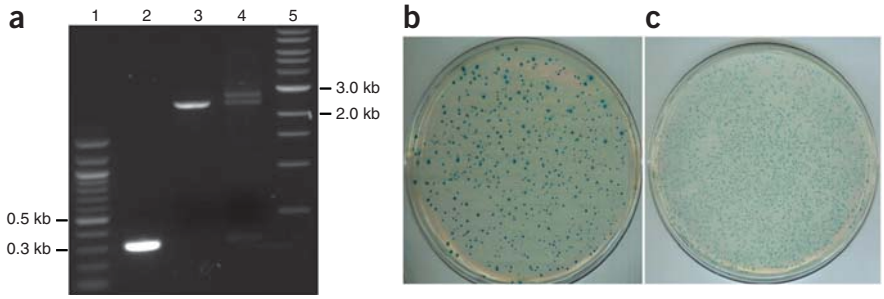

Figure 3 | CPEC cloning of the *lacZα* gene library. (**a**) Agarose gel examination of the gene library inserts (lane 2), the linearized vector (lane 3) and the CPEC cloning products after five thermal cycles (lane 4). The upper band in lane 4 represents the full-length cloning product; the lower band represents the empty vector. Lanes 1 and 5 are 100-bp and 1-kb DNA ladders (New England Biolabs), respectively. Amount of DNA loaded: 300 ng for the insert, 200 ng for the vector and 15 µl out of 25 µl of the cloning product. (**b**) *E. coli* colonies expressing a small portion of the library of *lacZα* codon variants show a wide range of variation in the intensity of blue color. (**c**) Control *E. coli* colonies expressing wild-type *lacZα* show very little variation in color intensity.

full-length cloning products. After transformation of the library of codon variants in GC5 competent cells and incubation overnight at 37 °C, bacterial colonies growing on the LB-agar plate showed a wide range of different shades of blue color, indicating different lacZα protein expression levels as a result of cloning different codon variants (**Fig. 3b**). In contrast, a control plate with colonies containing the CPEC cloning product of the wild-type *lacZα* gene showed almost uniform intensit ies of blue color (**Fig. 3c**). This functional test shows the complexity of the cloned library. We have sequenced a few hundred clonies and no identifical sequences have been found (data not shown).

The protocol described here has been used in our laboratory to clone a large number of gene libraries with minimum effort and satisfatory results (J.Q., I. Saaem, N. Tang, S. Ma, & J.T., unpublished data), which demonstrates the feasibility and advantage of using CPEC as a high-throughput cloning method. The result of cloning 14 gene libraries in a single run is shown in **Supplementary Figure 1**.

For combinatorial gene library cloning, in addition to published work[1] we examined the cloning of fusion libraries between codon variants of *lacZα* (306 bp) and *GFP* (741 bp) genes (**Fig. 4**). These two gene libraries were cloned simultaneously and in frame into a modified pAcGFP1 vector (pAcGFP1stop, with a stop codon added to the MCS region and the original *AcGFP* gene deleted from the linear vector). For combinatorial library cloning, more thermal cycles (10–25 cycles) are typically needed to generate a sufficiently large and complex library. As shown in **Figure 4**, the full-length cloning products steadly accumulated over 5–10 cycles (lanes 2–3) and reached a very high level after 20 cycles (lane 4). The upper arrow indicates the successful cloning product and lower arrow indicates the vector DNA. Refer to **Supplementary Tables 1** and **2** for primer information.

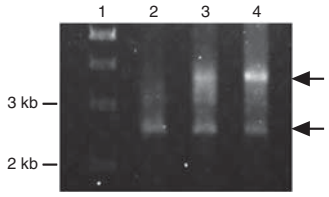

Figure 4 | CPEC cloning of combinatoral gene libraries. Lanes 2, 3 and 4 show the CPEC reaction products after 5, 10 and 20 thermal cycles. The upper arrow indicates the full-length cloning product of lacZα-AcGFP1 with the vector; the lower arrow indicates the remaining empty vector. Lane 1 is the 1-kb ladder from Bio-Rad.

Note: Supplementary information is available via the HTML version of this article.

ACKNOWLEDGMENTS We thank Z. Chen and P. Zuo for technical assistance. J.T. is a Beckman Young Investigator and a recipient of a Hartwell Individual Biomedical Research Award.

AUTHOR CONTRIBUTIONS J.T. conceived the CPEC strategies and supervised the project. J.Q. performed the experiments. Both authors contributed extensively to the experimental design, protocol development and manuscript preparation.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

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

1. Quan, J. & Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* **4**, e6441 (2009).
2. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. & Pease, L.R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61–68 (1989).
3. Prodromou, C. & Pearl, L.H. Recursive PCR: a novel technique for total gene synthesis. *Protein Eng.* **5**, 827–829 (1992).
4. Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M. & Heyneker, H.L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).
5. Shevchuk, N.A. *et al.* Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. *Nucleic Acids. Res.* **32**, e19 (2004).
6. Smith, H.O., Hutchison, C.A. III, Pfannkoch, C. & Venter, J.C. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA* **100**, 15440–15445 (2003).
7. Tian, J. *et al.* Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).
8. Garces, C. & Laborda, J. Single-step, ligase-free cloning of polymerase chain reaction products into any restriction site of any DNA plasmid. *Anal. Biochem.* **230**, 178–180 (1995).
9. McArthur, G.H.T. & Fong, S.S. Toward engineering synthetic microbial metabolism. *J. Biomed. Biotechnol.* **2010**, ID459760 (2010).
10. Cohen, S.N., Chang, A.C., Boyer, H.W. & Helling, R.B. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl. Acad. Sci. USA* **70**, 3240–3244 (1973).

11. Scharf, S.J., Horn, G.T. & Erlich, H.A. Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science* **233**, 1076–1078 (1986).
12. Walhout, A.J. *et al.* GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* **328**, 575–592 (2000).
13. Hartley, J.L., Temple, G.F. & Brasch, M.A. DNA cloning using *in vitro* site-specific recombination. *Genome Res.* **10**, 1788–1795 (2000).
14. Aslanidis, C. & de Jong, P.J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids. Res.* **18**, 6069–6074 (1990).
15. Li, M.Z. & Elledge, S.J. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256 (2007).
16. Rashtchian, A. Novel methods for cloning and engineering genes using the polymerase chain reaction. *Curr. Opin. Biotechnol.* **6**, 30–36 (1995).
17. Nisson, P.E., Rashtchian, A. & Watkins, P.C. Rapid and efficient cloning of Alu-PCR products using uracil DNA glycosylase. *PCR Methods Appl.* **1**, 120–123 (1991).
18. Gibson, D.G. *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. USA* **105**, 20404–20409 (2008).
19. Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids. Res.* **37**, e16 (2009).
20. Sambrook, J. & Russell, D.W. *Molecular Cloning: A Laboratory Manual* 3rd edn. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001).
21. Breslauer, K.J., Frank, R., Blocker, H. & Marky, L.A. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA* **83**, 3746–3750 (1986).