ACS SyntheticBiology

Technical Note

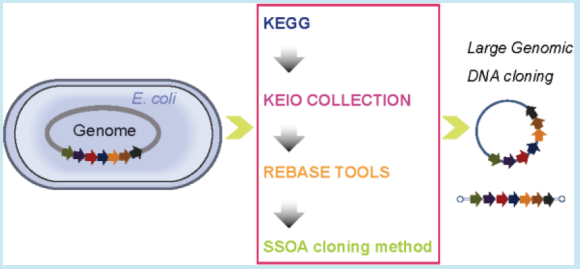
pubs.acs.org/synthbio

Cloning Large Gene Clusters from *E. coli* Using *in Vitro* Single-Strand Overlapping Annealing

Rui-Yan Wang,[†] Zhen-Yu Shi,[†] Jin-Chun Chen,[†] and Guo-Qiang Chen*[,†,‡]

[†]MOE Key Lab of Bioinformatics and Systems Biology, Department of Biological Science and Biotechnology, School of Life Sciences, Tsinghua-Peking Center for Life Sciences and [‡]Center for Nano and Micro Mechanics, Tsinghua University, Beijing 100084, China

ABSTRACT: Despite recent advances in genomic sequencing and DNA chemical synthesis, construction of large gene clusters containing DNA fragments is still a difficult and expensive task. To tackle this problem, we developed a gene cluster extraction method based on *in vitro* single-strand overlapping annealing (SSOA). It starts with digesting the target gene cluster in an existing genome, followed by recovering digested chromosome fragments. Subsequently, the single-strand DNA overhangs formed from the digestion process would be specifically annealed and covalently joined together with a circular and a linear vector, respectively. The SSOA method was successfully applied to clone a 18 kb DNA fragment encoding NADH:ubiquinone oxidoreductase. Genomic DNA fragments of different sizes including 11.86, 18.33, 28.67, 34.56, and 55.99 kb were used to test the cloning efficiency. Combined with genetic information from KEGG and the KEIO strain collection, this method will be useful to clone any specific region of an *E. coli* genome at sizes less than ~28 kb. The method provides a cost-effective way for genome assembly, alternative to chemically synthesized gene clusters.



KEYWORDS: *synthetic biology, homologous recombination, gene cluster, NADH:ubiquinone oxidoreductase, single-strand overlapping annealing, Escherichia coli*

Defined fragments of chromosomal DNA can be cloned to obtain metabolic pathways involving enzyme complexes encoded by a large gene cluster, to clarify detailed structures of genes, to express an enzyme complex, and to map and analyze alterations to the genome.[1] While chemical synthesis of DNA has become feasible, it is still expensive and time-consuming to synthesize large gene fragments. On the other hand, the cloning of large gene fragments is difficult using PCR amplification alone. The development of multiple-gene assembly methods has allowed regions of genomic DNA too large to be amplified by a single PCR event to be divided into multiple overlapping PCR products and assembled together.[2] Methods that have been widely used include library construction of genomic DNA fragments using cloning vectors such as λ phage, cosmids, BAC, YACs, or P1 and gene screening by hybridization with gene-specific probes or by PCR specific fragment amplifications.[3] Again, all of the above-mentioned methods are still not cost-effective and are time-consuming. *In vivo* homologous recombination has been used for large gene fragments removal, insertion, and cloning,[4] yet it includes a complicated *in vivo* recombination process with a recombination efficiency that is difficult to control.

When restriction enzymes digest a genome, the more frequently the enzymes cut the genome, the smaller the average size of the DNA fragments becomes. Some enzymes such as *Not*I that cuts an 8 bp DNA recognition sequences can be used to generate very large fragments. Prediction tools can be used to look for restriction enzymes that can produce DNA fragments with desired size ranges, e.g., a 20 kb fragment.[5]

A recombination method has been developed to assemble *in vitro* a completely chemically synthesized genome using a thermocycler to allow annealing of large synthetic DNA molecules via their homologous overhangs generated from exonuleases.[6] This principle was adopted to develop our approach of *in vitro* single-strand overlapping annealing (SSOA) between a target genomic DNA fragment and a predesigned vector with homologous arms (Figure 1).

To demonstrate the feasibility of the SSOA method, an 18 kb genomic DNA fragment from *E. coli* containing a 15 kb gene cluster *nuoABCDEFGHIJKLMN* termed *nuo*-genes or *nuo* operon was cloned. The operon encodes 13 different protein subunits that constitute the respiratory complex I, or NADH:ubiquinone oxidoreductase. All *nuo*-genes are required for the assembly and stability of a functional *E. coli* complex I.[7] To clone the large gene cluster, the *nuo* operon containing DNA fragment was first digested from *E. coli* genome via the predicted restriction enzyme *Nde*I. Either a replicon fragment from a circular pKD3 vector or a linear pJAZZ-OC vector (Table 1), with terminals overlapping to the adjacent ends of the digested *nuo* operon containing the DNA fragment, was used. Subsequently, the overlapping DNA strands from both genomic DNA and vector fragments were treated with an exonuclease from T4 DNA polymerase to generate single-

Special Issue: Synthetic Biology: Research Perspectives from China

Received: April 1, 2012
Published: June 11, 2012

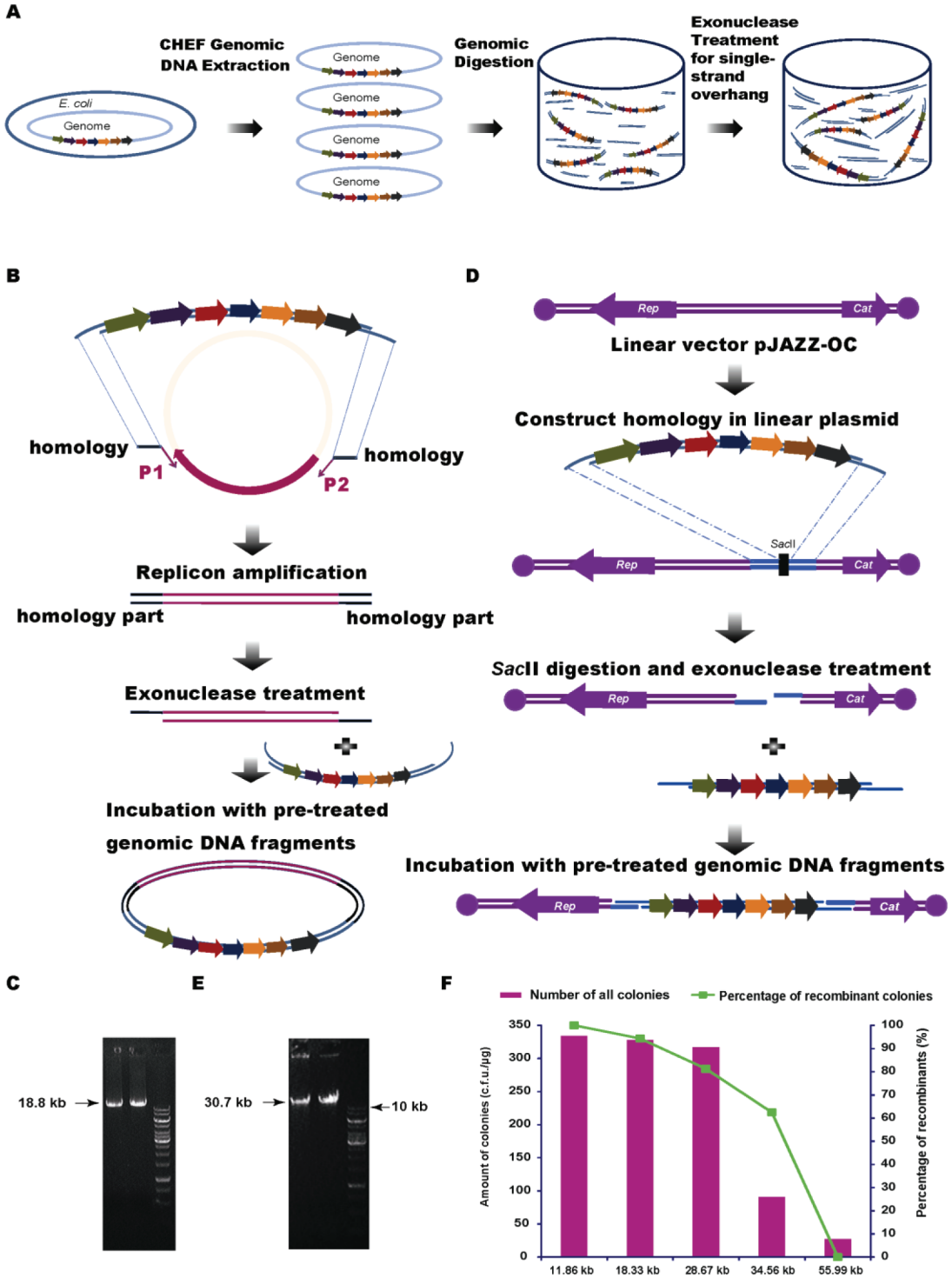


Figure 1. Single-strand overlapping annealing (SSOA) method for cloning large genomic DNA using both circular (left) and linear vectors (right) and the cloning efficiency. (A) Genome extraction, digestion, and exonuclease treatment. (B) SSOA large genomic DNA cloning using circular vector. (C) An 18.8 kb *nuo* operon contained R6K*γori* vector (left two lanes; 10 kb standard in right lane). (D) SSOA large genomic DNA cloning using linear vector. (E) A 30.7 kb pJAZZ-OC-nuo linear vector containing *nuo* operon (left two lanes; 10 kb standard in right lane). (F) Cloning efficiency of differently sized genomic DNA fragments. The amount of the colonies is shown in colony forming units per microgram of vector for the indicated regions of homology. Percentage of the recombinants was determined by dividing the number of colonies that contained the target genomic DNA fragment by the total number of transformants. Circular vector was used for this experiment.

stranded overhangs on both sites of the DNA fragments (Figure 1), followed by annealing the complementary overhangs. Lastly, the whole reaction mixture was electroporated into *E. coli* S17−1 *λpir* (for circular vector) or *E. coli* TSA cells (for linear vector).

E. coli JW2276 is an *E. coli* BW25113 Δ*nuo*I mutant containing a kanamycin gene inserted into the *nuo*I gene sequence. *Nde*I restriction enzyme was used to digest the *E. coli* JW2276 genome to produce the 18 kb fragment containing kanamycin resistant gene and the entire *nuo* operon. Intact

dx.doi.org/10.1021/sb300025d | *ACS Synth. Biol.* 2012, 1, 291−295

Table 1. Bacterial Strains, Plasmids, and Primers Used for This Study[a]

strain	description	ref
E. coli S17-1 λ*pir*	*Tp*R *Sm*R *rec*A, *thi*, *pro*, *hsd*R-M+RP4: 2-Tc: Mu: *Km* Tn7 λ*pir*; Constitutively expressed protein for replication of plasmids containing the R6Kγ origin of replication (R6Kγ*ori*); R6Kγ*ori* containing plasmids was maintained at approximately 15 copies per cell.	12
E. coli TSA	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80d*lac*ZΔM15 Δ*lac*X74 *end*A1 *rec*A1 *ara*D139 Δ(*ara*, *leu*)7697 *gal*U *gal*K *rps*L *nup*G λ- *ton*A *Amp*R *sop*AB *tel*N *ant*A; host strain for linear plasmid vector pJAZZ-OC.	8
E. coli JW2276	*E. coli* BW25113 Δ*nuo*I, *kan*.	9
plasmid	**description**	**ref**
pKD3	R6Kγ*ori*: requires the *pir* gene product for replication.	9
pJAZZ-OC	Linear plasmid vector based upon the phage N15, which has a linear dsDNA genome.	8
primer	**sequence**	
Primers used for SSOA based on circular vector cloning		
c-NdeIR6KF (for 18.33 kb DNA)	5′-GTACAAACCGAAACAGTCTCCGTTACCATA AACTAAGGAGGATATTCAT	
c-NdeIR6KR (for 18.33 kb DNA)	5′-TCCGGAACCCGGACGAAAGTAAAAATGCATA ATTGATTTAAAACTTCATT	
c-AsiSIf (for 28.67 kb DNA)	5′-TCGCTTGCCGGAATACCCAGCACATCGGCG AGGAGGATATTCATATGGACCAT	
c-AsiSIr (for 28.67 kb DNA)	5′-TTAATGACGAAGGCTGGTACGCTACGCGCGAT ATTAAGCATTGGTAACTGTCAGACC	
c-AscIf (for 34.56 kb DNA)	5′-GCTTTGCGGGTTGCCAAATAACTGCTCCGGCGCG AGGAGGATATTCATATGGACCAT	
c-AscIr (for 34.56 kb DNA)	5′-TTCCACCTGTGGACGCCAGACGTATACCAGGG ATTAAGCATTGGTAACTGTCAGACC	
c-NheIf (for 55.99 kb DNA)	5′-TTTGGCATTGCGCTTCATCCACAACGCTAG AGGAGGATATTCATATGGACCAT	
c-NheIr (for 55.99 kb DNA)	5′-GTGGGAAAGGGGATAATGAAAAAAATTGCG ATTAAGCATTGGTAACTGTCAGACC	
Primers used for SSOA based on linear vector cloning		
l-NdeIf (for 18.33 kb DNA)	5′-TATGCATTTTTACTTTCGTCCGGGTT*CCGCGG* TGCTTGAGGCCTGGGAAGAAC	
l-NdeIr (for 18.33 kb DNA)	5′-TATGGTAACGGAGACTGTTTCGGTTTGTACTGG*CCGCGG* TACCTGCCAGAAGTCATCGG	
l-AsiSIf (for 28.67 kb DNA)	5′-ATT*CGGCCG*CGCCGATGTGCTGGGTATTCCGGCAAGCGA*CCGCGG* AGGAGGATATTCATATGGACCAT	
l-AsiSIr (for 28.67 kb DNA)	5′-ATT*CGGCCG*ATCGCGCGTAGCGTACCAGCCTTCGTCATT*CCGCGG* CAAGATCCGCAGTTCAACCTG	
l-AscIf (for 34.56 kb DNA)	5′-ATT*CGGCCG*CGCGCCGGAGCAGTTATTTGGCAACCCGCA*CCGCGG* AGGAGGATATTCATATGGACCAT	
l-AscIr (for 34.56 kb DNA)	5′-ATT*CGGCCG*CGCGCCCTGGTATACGTCTGGCGTCCACAG*CCGCGG* CAAGATCCGCAGTTCAACCTG	
l-NheIf (for 55.99 kb DNA)	5′-ATT*CGGCCG*CTAGCGTTGTGGATGAAGCGCAATGCCAAA*CCGCGG* AGGAGGATATTCATATGGACCAT	
l-NheIr (for 55.99 kb DNA)	5′-ATT*CGGCCG*CGCAATTTTTTTCATTATCCCCTTTCCCAC*CCGCGG* CAAGATCCGCAGTTCAACCTG	

[a]All oligonucleotides were synthesized by AuGCT Biotech (Beijing, China). Restriction endonuclease digestion sites are in italic. Homology sequences are underlined. Primers used for the cloning of DNA fragments of different sizes are labelled. For the cloning of 11.86 kb fragment, primers of c-AsiSIf/c-AscIr and l-AsiSIf/l-AscIr were used.

genomic DNA from *E. coli* JW2276 was isolated using a CHEF Bacterial Genomic DNA Plug Kit (Bio-Rad Inc., USA), followed by *Nde*I restriction enzyme digestion at 37 °C (Figure 1A). All enzymes were from New England Biolabs (Ipswich, MA, USA) and used as recommended.

To use a circular vector for cloning the large gene fragment, R6Kγ origin of replication (R6Kγ*ori*) was amplified using Fast-pfu DNA polymerase (TRANSGEN Beijing, China) from vector pKD3 (Table 1). Thirty base pair overlapping DNA molecules from both sides of the predigested target genomic DNA were designed in the forward/reverse primers (Figure 1B). Then, R6Kγ*ori* was purified by the OMEGA E.ZNA Gel Extraction kit (Omega Bio-Tek, USA). Twenty units of *Dpn*I was added to the reaction mixture and incubated at 37 °C for 1 h to digest the template. Both the R6Kγ*ori* fragment and digested genome fragments were treated with 0.5 U of exonuclease from T4 DNA polymerase separately at 37 °C for 10 min (Figure 1B). The reaction was terminated using 0.1 vol of 10 mM 2′-deoxycytidine 5′-triphosphate (dCTP). The linear R6Kγ*ori* fragment and appropriate amounts of digested genomic fragments were mixed at a 1:1 or 2:1 molar ratio (Figure 1B). The annealing reaction was performed using 1x ligation buffer, subsequently incubated at 75 °C for 10 min, and then cooled down to 65 °C at a rate of 0.1 °C/min. The mixture was maintained at 65 °C for 30 min, followed by cooling down to 4 at 0.1 °C/min. The R6Kγ*ori* was added to the mixture at 10 min intervals up to 30 min. This annealing mixture was then transformed into 100 μL of electro-competent cells of *E. coli* S17−1 λ*pir*, which were then plated on kan^R Petri disks. The presence of the transformed vectors containing 18 kb target genomic fragment in the colonies was verified by PCR. The efficiency of the method was positively correlated to high concentrations of genomic DNA and by higher molar ratio of replicon to genomic DNA fragments, as well as the digestion efficiency of exonuclease derived from T4 DNA polymerase. There were fewer than 20 positive colonies found on the Petri disk, almost all of which contained the target genomic DNA fragment as verified by DNA sequencing. Thus, a 18 kb genomic DNA fragment was successfully cloned via a 460 bp R6Kγ*ori* to form a circular plasmid *in vitro* (Figure 1C).

Certain large gene fragments are difficult or impossible to clone using circular vectors, especially AT-rich fragments of up to 30 kb and short tandem repeats ones. To overcome this difficulty, linear cloning systems are commonly employed (Figure 1D). A linear cloning vector pJAZZ-OC (BIGEASY Lucigen, USA)[8] was used to study the feasibility of cloning the large *nuo* operon. Initially, 30 bp DNA homologous arms to each end of the *nuo* operon DNA fragment were constructed in linear plasmid pJAZZ-OC (Figure 1D). Restriction enzyme *Sac*II was used to digest the linear vector to form two separate strands, each possessing one homology sequence with the *nuo*-operon (Figure 1D). Subsequently, the same procedure used for the circular vector cloning described above (Figure 1B) was performed (Figure 1D). Bacterial cells were plated on *Cm*^{R},

dx.doi.org/10.1021/sb300025d | *ACS Synth. Biol.* 2012, 1, 291−295

Kan^R, and Cm^R+Kan^R Petri disks, respectively, for screening of recombinants containing the target genomic fragment, which possessed both resistant markers. Hundreds of colonies were found on Cm^R and Kan^R Petri disks. However, only fewer than 20 colonies were observed on Cm^R+Kan^R disks, revealing a similar efficiency between circular and linear vectors. The presence of the transformed linear vector in the colonies was verified by colony PCR using primers designed from both the linear vector and the genomic DNA fragment and their subsequent sequencing. The results suggested that all colonies from Cm^R+Kan^R disks contained the target 18 kb genomic DNA fragment, whereas less than 5% of the colonies on Cm^R and Kan^R plates did (Figure 1E).

To test the efficiency of this SSOA method, genomic DNA fragments with sizes of 11.86, 18.33, 28.67, 34.56, and 55.99 kb were cloned based on the circular vector. The cloning efficiency for the genomic DNA fragments shorter than 30 kb was similar (Figure 1F): 334, 328, and 317 colonies per microgram vector were found, respectively. However, the cloning of the 34.56 kb fragment resulted in only 91 colonies per microgram vector (Figure 1F). This size-dependent cloning efficiency may have been caused by limitation from the vector that could not accept an insert in excess of 30 kb. The percentage of recombinants that contained target genomic DNA fragments decreased with increased genomic DNA fragment size. For cloning genomic DNA fragment at a size of 55.99 kb, although there were several transformants found on the disk, no recombinant with the target genomic DNA fragment was found (Figure 1F). The possible reason included nonspecific recombination for the large genomic DNA fragment into the genomic DNA.

The SSOA method based on a linear vector showed the same DNA-size-relative constraint for cloning large genomic DNA fragments as described above. As the size of the linear vector was 13 kb, cloning efficiency for genomic DNA fragment larger than 18.33 kb was not very high. For the cloning of the 11.86 kb fragment, the transformants on Kan^R and Cm^R+Kan^R Petri disks, respectively, were all of the recombinants containing the target DNA fragment. However, less than 5% colonies from the Cm^R disk were indicated with the target DNA fragment. This result demonstrated that nonspecific recombination from the excess digested genomic DNA fragments resulted in an *in vitro* annealing efficiency lower than 5%, while a kanamycin-resistant gene in the cluster helped raise to a high accuracy.

These results showed that a drug-resistant marker inside the large target genomic DNA is needed for the efficient cloning. Since Keio Collection has systematically prepared mutants of *E. coli* K-12 BW25113 with precisely defined, single-gene deletions of all nonessential genes, each mutant contains a kanamycin resistance cassette in the mutated gene,[9] and their mutants provide valuable resources for genome-wide cloning of any large size genomic DNA fragment (Figure 2). Additionally, Restriction Enzyme Database (REBASE) Tools (New England Biolabs, Ipswich, USA)[10] can be used for the prediction of restriction sites on DNA fragments shorter than 1 Mbp (Figure 2). Thus, by combination of the Keio collection and the REBASE Tools, the SSOA method can be used to clone any specific genomic DNA fragment with sizes of at least ~28 kb, as demonstrated in this study (Figure 2). For certain gene clusters in which proper restriction sites are difficult to find, a proper restriction site may be designed and inserted into the genome by the one-step gene knockout method.[11] For AT-rich fragments, the linear vector may be more suitable compared with the circular one. The BAC vector can be used as a circular vector for cloning genomic DNA fragments larger than 50 kb.[6] The single-strand overlapping annealing (SSOA) method appears promising for cloning large gene clusters for synthetic biology applications.

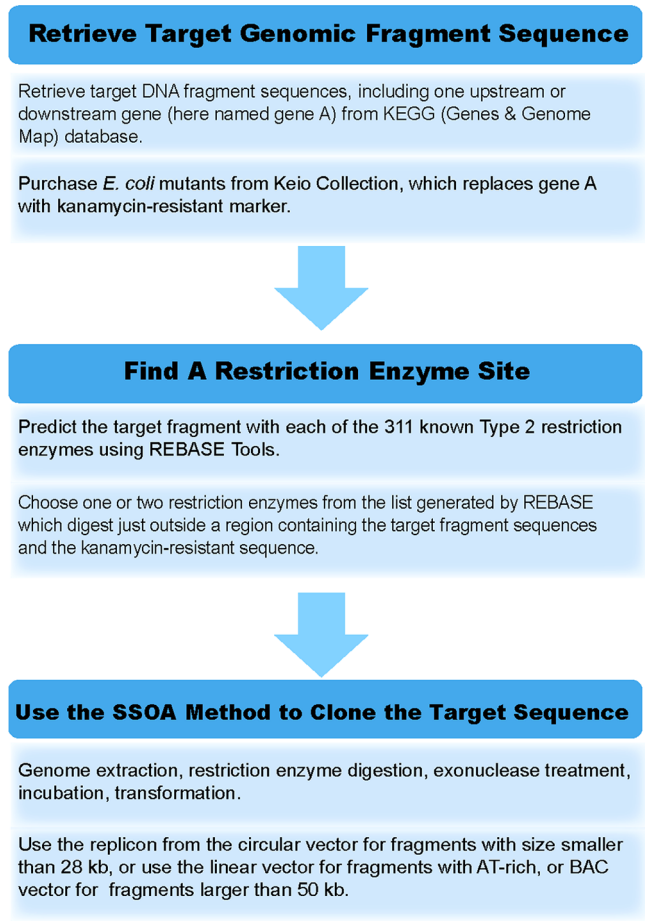


Figure 2. Flowchart of large genomic DNA cloning procedures in *E. coli* K-12 BW25113. Steps involved in large genomic DNA cloning: an appropriate gene was first selected on the basis of the target DNA using the KEGG gene database for the insertion of drug-resistant marker gene, then the restriction enzyme site was chosen using REBASE Tools, and finally the size of the target DNA determined the vector to be used.

AUTHOR INFORMATION

Corresponding Author

*Phone: +86-10-62783844. Fax: +86-10-62794217. E-mail: chengq@mail.tsinghua.edu.cn.

Author Contributions

R.Y.W. and Z.Y.S. contributed equally to this work. R.Y.W., Z.Y.S. and G.Q.C. designed the project. R.Y.W. performed the experiments and prepared the draft paper. G.Q.C. supervised the study.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the National BioResource Project *E. coli* strain at NIG in Japan for kindly donating the *E. coli* JW2276 mutant. This research was supported by the State Basic Science Foundation 973 (Grant No. 2012CB725200, 2012CB725201 and 2011CBA00807).

REFERENCES

(1) Kodumal, S. J., Patel, K. G., Reid, R., Menzella, H. G., Welch, M., and Santi, D. V. (2004) Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U.S.A. 101*, 15573−15578.

(2) Liang, X., Peng, L., Tsvetanova, B., Li, K., Yang, J. P., Ho, T., Shirley, J., Xu, L., Potter, J., Kudlicki, W., Peterson, T., and Katzen, F. (2012) Recombination-based DNA assembly and mutagenesis methods for metabolic engineering. *Methods Mol. Biol. 834*, 93−109.

(3) Farrar, K., and Donnison, I. S. (2007) Construction and screening of BAC libraries made from *Brachypodium* genomic DNA. *Nat. Protoc. 2*, 1661−1674.

(4) Wingler, L. M., and Cornish, V. W. (2011) Reiterative Recombination for the *in vivo* assembly of libraries of multigene pathways. *Proc. Natl. Acad. Sci. U.S.A. 108*, 15135−15140.

(5) Berlyn, M. K. (1998) Linkage map of *Escherichia coli* K-12, edition 10: the traditional map. *Microbiol. Mol. Biol. Rev. 62*, 814−984.

(6) Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., Stockwell, T. B., Brownley, A., Thomas, D. W., Algire, M. A., Merryman, C., Young, L., Noskov, V. N., Glass, J. I., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science 319*, 1215−1220.

(7) Schneider, D., Pohl, T., Walter, J., Dorner, K., Kohlstadt, M., Berger, A., Spehr, V., and Friedrich, T. (2008) Assembly of the *Escherichia coli* NADH:ubiquinone oxidoreductase (complex I). *Biochim. Biophys. Acta 1777*, 735−739.

(8) Godiska, R., Mead, D., Dhodda, V., Wu, C., Hochstein, R., Karsi, A., Usdin, K., Entezam, A., and Ravin, N. (2010) Linear plasmid vector for cloning of repetitive or unstable sequences in *Escherichia coli*. *Nucleic Acids Res. 38*, e88.

(9) Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol. 2*, 2006 0008.

(10) Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2010) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res. 38*, D234−236.

(11) Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A. 97*, 6640−6645.

(12) Simon, R., Priefer, U., and Pühler, A. (1983) A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in gram negative bacteria. *Nat. Biotech. 1*, 784−791.