



Original Article

The TelN/tos-assisted precise targeting of chromosome segments (TAPE)

You-Zhi Cui, Jian-Ting Zhou, Bing-Zhi Li*, Ying-Jin Yuan

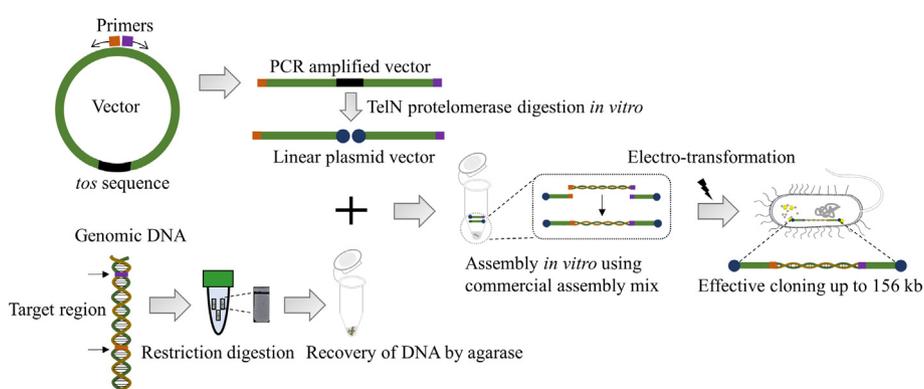


Frontiers Science Center for Synthetic Biology and Key Laboratory of Systems Bioengineering (Ministry of Education), School of Chemical Engineering and Technology, Tianjin University, 300072 Tianjin, China
 Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin University, 300072 Tianjin, China

HIGHLIGHTS

- We first apply the TelN/tos system to target genomic segments in *E. coli* host.
- We successfully cloned the targeted bacterial DNA fragment up to 156 kb.
- The TAPE method takes no more than five days to directly obtain the large DNA sequence.
- The TAPE method has no preferences on genome sequence.
- The results showed a considerable improvement of cloning efficiency.
- The TAPE method provides a powerful tool to support the study on synthetic biology.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 7 October 2021

Revised 12 January 2022

Accepted 29 January 2022

Available online 3 February 2022

Keywords:

Large genomic segments

Targeted cloning

TelN/tos system

Linear artificial chromosomes

ABSTRACT

Introduction: Performing genomic large segmentation experiments will promote the annotation of complex genomic functions and contribute to the synthesis of designed genomes. It is challenging to obtain and manipulate large or complex DNA sequences with high efficiency.

Objectives: This study aims to develop an effective method for direct cloning of target genome sequences from different species.

Methods: The TelN/tos system and a linear plasmid vector were first used to directly clone the large genomic segments in *E. coli*. For the *in vitro* cloning reaction, two telomeric sites were developed using TelN protelomerase at the end of the linear plasmid vector. The target DNA sequence can be easily hooked with the homology arms and maintained as a linear artificial chromosome with arbitrary restriction sites in a specific *E. coli* strain.

Results: Using the linear cloning strategy, we successfully cloned the bacterial DNA fragment of 156 kb, a yeast genomic fragment of 124 kb and mammalian mitochondrial fragment of 16 kb. The results showed a considerable improvement in cloning efficiency and demonstrated the important role of vector ratio in the cloning process.

Conclusion: Due to the high efficiency and stability, TAPE is an effective technique for DNA cloning and fundamental molecular biotechnology method in synthetic biology.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Cairo University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer review under responsibility of Cairo University.

* Corresponding author at: Frontiers Science Center for Synthetic Biology and Key Laboratory of Systems Bioengineering (Ministry of Education), School of Chemical Engineering and Technology, Tianjin University, 300072 Tianjin, China.

E-mail address: bzli@tju.edu.cn (B.-Z. Li).

<https://doi.org/10.1016/j.jare.2022.01.017>

2090-1232/© 2022 The Authors. Published by Elsevier B.V. on behalf of Cairo University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Next-generation sequencing technologies help to broaden knowledge on the life of genomes [1,2]. The exponential growth of DNA sequence data facilitates the annotation of complex

genomes, including non-coding DNA regions in a macroscopic view [3]. More novel biosynthetic gene clusters (BGCs) have also been uncovered for valuable bio-antibiotic, pharmaceutical and biofuel production [4–6]. Experts in synthetic biology need large DNA segments, even complete genomes, to study evolution or reprogramming [7–10]. To better understand the functions of genomic segments and facilitate the assembly of designed genomes, creative molecular techniques are needed to manipulate and engineer large genomes *in vitro*. Synthesis and cloning are the main methods to obtain target DNA sequences. However, long DNA synthesis is time-consuming as it uses yeast and bacteria hosts to assemble the designed sequence after multiple rounds [11–13]. Researchers always rely on BAC or YAC libraries to carry out studies on the functions of complete genes [14,15]. Nevertheless, it is difficult to acquire ideal genomic materials due to incomplete libraries [16]. Many cloning methods of genomic segments have emerged in recent years. However, transformation-associated recombination (TAR) is impeded by a low efficiency of chromosomal region capture [17]. For *E. coli* host, Cas9-Assisted Targeting of CHromosome Segments (CATCH), Cas12a-assisted precise targeted cloning using *in vivo* Cre-*lox* recombination (CAPTURE), Exonuclease Combined with RecET recombination (ExoCET) mainly focus on the excision of genome for target segments during cloning [18–20]. In addition, CAPTURE and CATCH technologies require an extra design of CRISPR/Cas system for a specific genomic segment that can clone multiple targets in one genomic library. The ExoCET takes advantage of the RecET homologous recombination's capacity, indicating that this method is limited to some DNA fragments due to repetitive or similar sequences with the *E. coli* host's genome [21].

The importance of vectors is often neglected in the cloning process. The circular plasmid is usually used to carry the cloned gene as the conventional vector, and it is widely believed that circular plasmid supercoiling produces cruciforms and other secondary structures that might cause DNA sequence deletions or rearrangements [22]. Linear plasmid can stably maintain a variety of inserts that are difficult to clone through the conventional circular plasmid, such as the skewed sequence, repetitive and AT-rich sequences [23,24]. TelN protelomerase, isolated from phage N15, was found to cleave the recognition sequence (*tos* sequence) into hairpin structures with covalently closed ends that are resistant to digestion by exonucleases [25]. In addition, linear plasmids have the potential to be used in DNA vaccines and reported a capacity of 100 kb DNA sequence without integration in mammalian cells [26–28].

In this study, we developed a facile cloning method named TelN/*tos*-assisted precise targeting of chromosome segments (TAPE). We used the TelN/*tos* system and flexible linear plasmid vector to target the DNA sequence from genomic libraries *in vitro*. A linear vector is designed and constructed with *tos* sequence to track the target DNA sequence *in vitro*, which was stably maintained with a linear chromosome in *E. coli*. Two primers were used with the target segments for the 40 bp homologous sequence in vectors. In a single step, we cloned the target segment of bacterial genome of up to 156 kb, a 124 kb genomic fragment containing yeast centromere and mammalian mitochondrial sequence in F9 cells from excised genomic libraries. With its stability, easy operability and high efficiency, the TAPE platform can be a practical tool in obtaining the target genome sequence without wasting time purchasing from companies or conducting lab synthesis.

Materials and methods

Materials

TelN protelomerase, β -Agarase I, SbfI, MluI, NotI and $2 \times$ NEBuilder[®] HiFi DNA Assembly Master Mix were purchased

from NEB (Ipswich, MA, USA). The zymolysase-20 T was bought from Nacalai Tesque (Kyoto, Japan). The Low Melting Point Agarose, Micropulser and CHEF Mapper XA System were purchased from Bio-Rad (Hercules, USA). Dulbecco's modified Eagle's medium (DMEM) was bought from Corning (NY, USA) and the 10% fetal bovine serum was bought from BI (Kibbutz Beit-Haemek, Israel). The phenylmethyl sulphonyl fluoride (PMSF), Proteinase K, Kanamycin, Spectinomycin and GlycoBlue[™] Coprecipitant were purchased from Sangon Biotech (Shanghai, China). The Azure Biosystems C150 was bought from Azure Biosystems (Dublin, CA, USA). The bacterial genomic segments for targeted cloning were from *E. coli* str. MG1655. The *E. coli* str. GB05RedTrfA (DH10B, *fhuA::IS2*, *Δ yhbcC*, *Δ recET*, *P_{RhasR}- γ β α A*, *P_{BAD}.trfA*) was obtained from the laboratory of Professor Youming Zhang of Shandong University [29]. The *tos* sequence (519 bp) and TelN protelomerase expression gene (1896 bp), primers were synthesized from GENEWIZ (Suzhou, China) (Supplementary Table S1 and Table S2).

Construction of the *E. coli* host strain and plasmid

The TelN protelomerase expression cassette was constructed with the J23100 promoter and RBS B0034, integrating in the *poxB* gene site (910849 to 910888) in the genome of GB05RedtrfA via CRISPR-Cas9 System [30]. The GB05RedTrfA::TelN positive colonies were verified using PCR and sequencing, as well as analyzing the plasmid linearization after transforming the plasmid carrying the *tos* sequence. The pBAC-*tos* plasmid was constructed by adding the *tos* sequence to the designated site based on pBAC plasmid. The plasmid backbone was amplified from pBAC plasmid flanking 40 bp homology with the *tos* sequence and transformed to the *E. coli* str. DH5a host after cloning with the $2 \times$ NEBuilder[®] HiFi DNA Assembly Master Mix. The pBAC-*tos* plasmid was inserted with the *tos* sequence between the *incC* and *sopA* elements in pBAC plasmid.

Genome digestion and isolation

The *E. coli* str. MG1655 was cultured overnight in 5 ml LB medium. The next morning, 1% bacteria liquid was transferred to fresh LB medium of up to OD 0.8–1.0 at a concentration of 5×10^8 cells ml^{-1} . Later, 3 ml bacteria liquid was centrifuged and resuspended in 250 μ l suspension buffer, then mixed and embedded in 250 μ l 2% low melting point agarose gel plugs [31]. These plugs were directly treated with Proteinase K solution for two days and washed four times using $1 \times$ wash buffer. During the third wash, 1 mM of phenylmethyl sulphonyl fluoride (PMSF) was added to inactivate the residual proteinase K in $1 \times$ wash buffer. The plugs were first washed with $0.1 \times$ wash buffer and $1 \times$ cutsmart buffer for 1 h. The bacterial genomic DNA in plugs were then immersed in 600 μ l $1 \times$ cutsmart buffer for 1 h with 25 units of the restriction enzyme NotI in ice bath. After digesting in 37 °C water bath for 2 h, about a quarter of the plugs were used to analyze the quality of genomic segments through PFGE. The remaining plugs were washed with $0.1 \times$ wash buffer for 10–30 mins and melted with 1/10 vol of $10 \times$ β -Agarase Buffer at 65 °C for 10 min. The molten agarose plugs were cooled for 10 min in 42 °C water bath and incubated overnight at the same temperature with 6 units of β -Agarase I per 250 μ l of plug. The digested genomic segments were added with 0.3 M sodium acetate and 5 μ l GlycoBlue Coprecipitant for 10 min, precipitated by 4 volumes of ethanol at room temperature for 2 h and resuspended in 20 μ l sterile water overnight. The genomic libraries can be stored at 4 °C for a week.

F9 cells were cultured in high-glucose Dulbecco's modified Eagle's medium and supplemented with 10% fetal bovine serum. The 2×10^7 F9 cells adopted the same procedures with *E. coli* to make low melting point agarose for constructing genomic

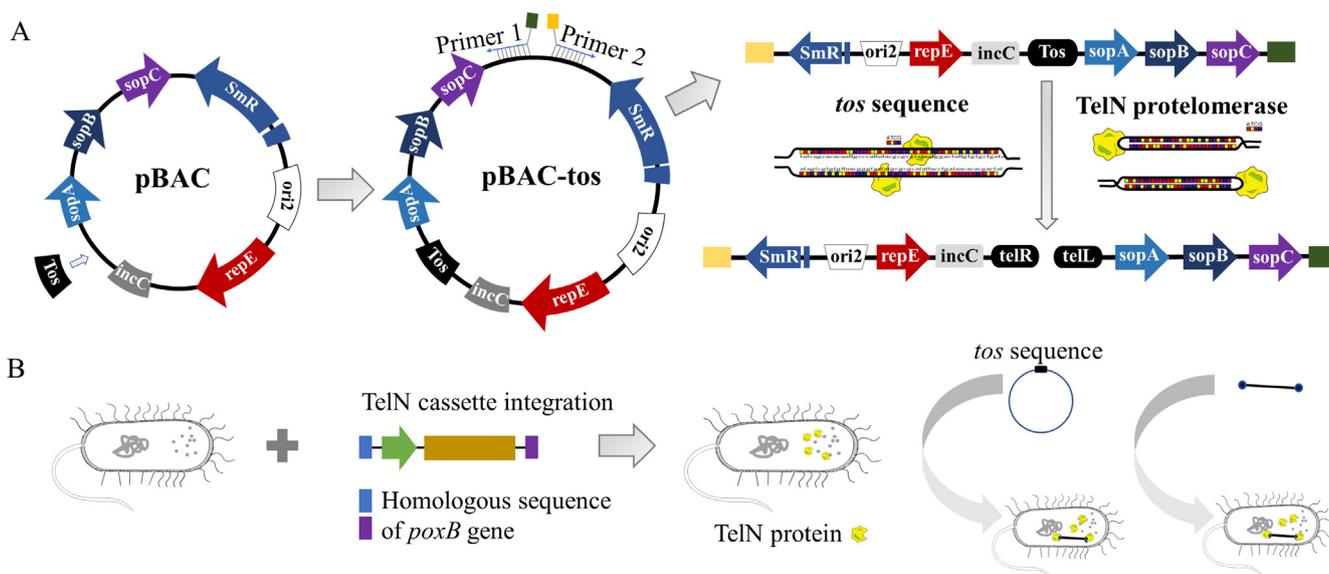


Fig. 1. The TAPE method. (A) The workflow of obtaining the linear plasmid vectors. The *tos* sequence was added to the pBAC plasmid. The cloning vector for TAPE was first amplified using PCR sharing 40 bp homology arms with target DNA fragment terminals, then digested into two parts in the *tos* sequence with covalently closed ends. (B) Construction of *E. coli* host. TelN expression cassette was integrated into the *E. coli* genome to express the TelN proteolomerase, which can cut the *tos* sequence and maintain the stability of linear plasmid in the host.

segments libraries, which use MluI for digestion. The genomic plugs of *Saccharomyces cerevisiae* BY4741 containing 1×10^9 cells were digested using SbfI to obtain the target segments. An additional procedure was performed to remove the yeast’s cell wall. Before washing the plugs, the yeast genomic plugs were digested in 1 ml lyticase solution (zymolyase-20 T, 2 mg/ml) at 37 °C for 2 h [32].

PFGE was used to verify the linear plasmid carrying the large inserts and quality of genomic segments with 1% agarose gel in $0.5 \times$ TBE using the CHEF Mapper XA System with fixed parameters (6 V cm^{-1} , 1–25 s, 16–18 h, 120°) and buffer circulation at 14 °C. After PFGE, the gel was stained with ethidium bromide and the DNA bands were visualized using Azure Biosystems C150.

Targeted cloning using linear plasmid vectors *in vitro*

In this study, a kanamycin resistance expression cassette (Kana cassette) was amplified from the pCas plasmid [30]. The complete vectors were first amplified using PCR from pBAC-tos plasmid flanking 40 bp homology with the target genomic terminal sequences introduced by primers (Supplementary Table S3). To obtain the linear plasmid vectors *in vitro*, three parallel tubes containing 3 ug of the recovery vectors were separately digested using 2 U TelN proteolomerase. Three tubes of digestion production were collected and recovered using one recovery column to obtain a high concentration of the linear plasmid vectors. The final linear plasmid vectors and prepared genomic segments were mixed with $2 \times$ NEBuilder® HiFi DNA Assembly Master Mix at a temperature of 50 °C for 1 h in the thermocycle instrument. During the cloning process, 3 ul was taken out and mixed with 100 ul electrocompetent *E. coli* host cells by end-cut wide bore tips in the 2 mm cuvette at 1250 V using Biorad Micropulser (Auto Ec2 Mode). The cells were recovered at 37 °C for 2 h in 900 ul LB medium without antibiotics and plated on LB agar medium containing antibiotics. The LB agar medium for cloning the Kana cassette had a final concentration of 100 ng/ul spectinomycin and 50 ng/ul kanamycin. The colonies carrying the target segments were screened on the LB agar plate with 100 ng/ul spectinomycin and further validated

by PCR covering the two intersections between vector and target genomic segments.

Results

Design and construction of TAPE cloning platform

The design of the TelN/*tos*-assisted precise targeting of chromosome segments (TAPE) is shown in Fig. 1. The TAPE platform used the TelN/*tos* system and a linear plasmid vector to improve the construction of BAC. The cloning vector constitutes two parts. One part was mainly constructed with replication elements and the other carried the partitioning protein genes. Both parts carried the homologous sequence of target genomic segments on one end while the other end was sealed with covalently closed end through TelN proteolomerase, similar to two matchsticks. The target segment in the genomic libraries were cloned with homologous sequence using assembly mix *in vitro*. The target DNA sequence was then transformed and maintained in the *E. coli* host with a constitutive expression of TelN proteolomerase protein.

In our experiment named pBAC-tos, the *tos* sequence was added to the pBAC plasmid between elements *incC* and *sopA*. To get the cloning vectors for TAPE, the vector was amplified from pBAC-tos and further digested into two parts where each end had a *tos* sequence with covalently closed ends through TelN proteolomerase *in vitro* (Fig. 1A). Meanwhile, the *E. coli* str. GB05RedTrfA::TelN host was constructed by integrating the TelN expression cassette in the *poxB* gene site. The TelN proteolomerase cleaved the *tos* sequence and maintained the plasmid as a linear chromosome *in vivo* when the circular or linear plasmids carrying *tos* sequence were transformed into the *E. coli* (Fig. 1B). In this cloning system, the target genomic segment and cloning vectors sharing 40 bp overlap sequences were ligated in $2 \times$ NEBuilder® HiFi DNA Assembly Master Mix, and electro-transformed into the *E. coli* host. The genomic libraries for different species were constructed through restriction enzyme digestion of the genome set in low melting point agarose. To reduce the tedious steps and DNA damage, extra steps were not taken to remove the background plasmid in recovery vectors or

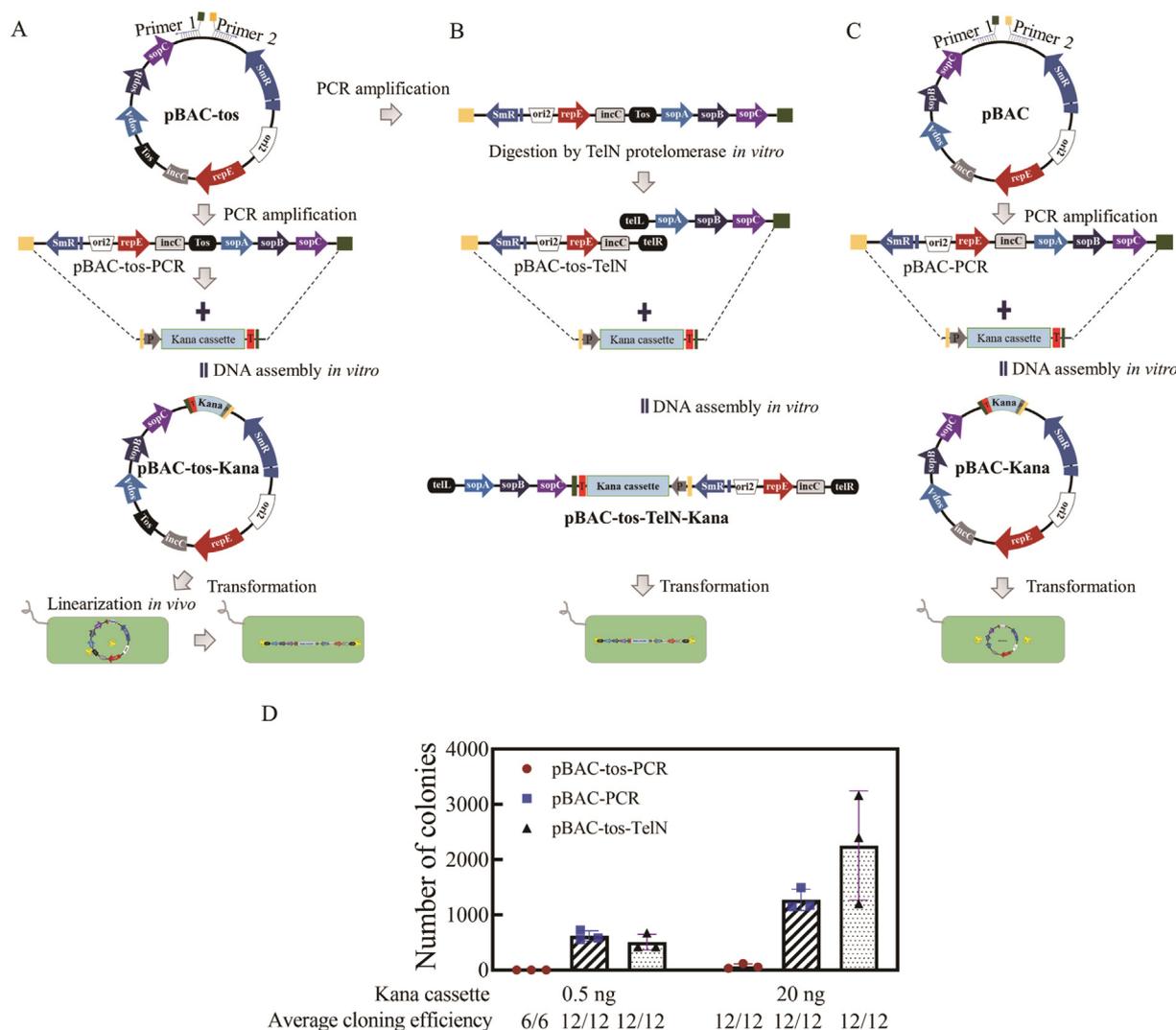


Fig. 2. Efficiency of cloning Kana cassette using different vectors. (A) The pBAC-tos-PCR cloning vector was directly amplified from pBAC-tos plasmid. (B) The pBAC-tos-TelN vector for TAPE was constructed by digestion of pBAC-tos-PCR vectors using TelN protelomerase. (C) The pBAC-PCR vector was amplified from pBAC plasmid. The 40 bp homology arms in vectors were introduced by primers with the Kana cassette. The cloning mix was transformed into the *E. coli* host and plated on the LB agar plate with kanamycin and spectinomycin antibiotics. (D) Direct cloning of Kana cassette using different vectors. Vectors with 50 ng and Kana cassette fragments of 0.5 ng and 20 ng were added into the cloning mix. The error bars indicate the standard deviations of all (n = 3) biological replicates.

procedures such as DNA desalt and high voltage for electrotransformation [33].

Effects of different cloning strategies for TAPE

To adopt a good vector strategy for constructing the linear plasmid with inserts, we designed and evaluated three vectors through two technical routes (Fig. 2). The first strategy used the untreated linear vector directly amplified from pBAC-tos plasmid to clone the DNA sequence, which was linearized in specific *E. coli* host *in vivo* (Fig. 2A). The second approach used the linear plasmid vector derived from the digestion of the PCR-amplified linear vector by TelN protelomerase in *tos* sequence site with covalent closed sequence *in vitro* on each end (Fig. 2B). Conventional vector control amplified from pBAC plasmid was used (Fig. 2C).

To test the cloning ability of the three vectors, a kanamycin resistance expression cassette was ligated with these vectors sharing 40 bp homology arms in the 2 × NEBuilder® HiFi DNA Assembly Master Mix. Through chemical transformation, 5 ul was taken out from the reaction mixtures separately to the *E. coli* strain GB05RedTrfA::TelN. PCR was used to check the

colonies grown on the double selection LB plates of spectinomycin and kanamycin (Supplementary Fig. S1A-C). The positive colonies were counted and some were randomly chosen to extract plasmids for further verification using restriction enzyme. The results showed that the different strategies had contrasting effects on cloning ability (Fig. 2D). The pBAC-tos-PCR vector was amplified from pBAC-tos plasmid and had low cloning efficiency compared to the other two vectors. This suggests that damage might have decreased the cloning efficiency during the plasmid linearization by TelN protelomerase *in vivo*. The pBAC-PCR vector (amplified from pBAC plasmid) and pBAC-tos-TelN vector (amplified from pBAC-tos plasmid and digested using TelN protelomerase) both reported efficient cloning with the addition of 20 ng Kana cassette. The two vectors had similar cloning ability with the same number of positive colonies following the addition of 0.5 ng Kana cassette. The restriction endonuclease pattern demonstrated that the linear plasmids were cleaved and maintained in the constructed *E. coli* str. GB05RedTrfA::TelN host (Supplementary Fig. S1D). Thus, cloning large DNA sequences is possible using the TelN protelomerase digested vectors based on pBAC-tos plasmid in a specific host.

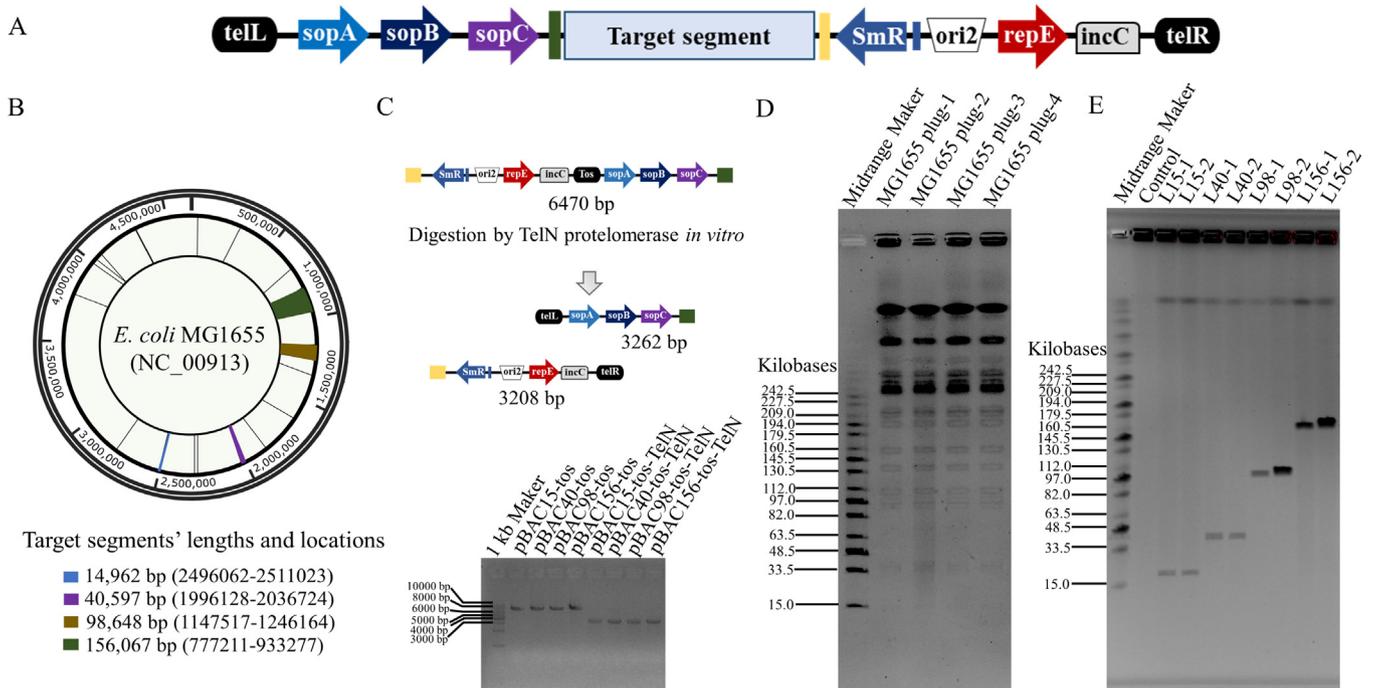


Fig. 3. Targeting genomic fragments from *E. coli* str. MG1655 using TAPE method. (A) Diagram of the linear plasmid carrying the bacterial genomic segment. (B) The locations of target sequence sites in *E. coli* str. MG1655. After digestion by NotI enzyme, the genomic segments with lengths of 14,962 bp (blue color), 40,957 bp (purple color), 98,648 bp (brown color), 156,067 bp (green color) were chosen as targets. (C) Verification of the cloning vectors using agarose gel electrophoresis. pBAC15-tos was the vector amplified from pBAC-tos plasmid to clone 14,962 bp DNA sequence. The pBAC15-tos-TelN was the linear plasmid vector that was digested from pBAC15-tos vector using telN protelomerase *in vitro*. Using the aforementioned strategies obtained from the pBAC-tos plasmid, the pBAC40-tos, pBAC98-tos, pBAC156-tos, pBAC156-tos-TelN represented the vectors targeting the sequences 40,957 bp, 98,648 bp, 156,067 bp, respectively. The length of complete linear vector was 6550 bp, which was further digested into two sections. Each section measured 3.2 kb. (D) The Pulsed Field Gel Electrophoresis (PFGE) map of genomic segments in low melting point plugs after digestion through NotI enzyme. The four plugs contained 1×10^8 cells of *E. coli*. (E) PFGE was used to analyze linear plasmids carrying the target sequences. Two positive colonies were randomly chosen to make the gel plugs. Control: MG1655; L15, L40, L98, L156 represented the linear plasmid carrying the bacteria segments of 14,962 bp, 40,957 bp, 98,648 bp, 156,067 bp, respectively. Number suffixes represent the parallel colonies.

Targeting the genomic segments from *E. coli* str. MG1655

Just like Kana cassette, our strategy can clone short DNA sequences. This method was adopted to clone the genomic segments from *E. coli* genome and determine whether the novel linear cloning platform can be applied in large DNA fragments. The cloned DNA fragments were maintained in a linear plasmid (Fig. 3A). The genomic libraries were constructed from *E. coli* str. MG1655 genome using NotI restriction enzyme (Fig. 3B). The four segments with lengths of 15 kb, 40 kb, 98 kb, 156 kb were chosen as the target. Cloning vectors for different segments amplified from pBAC-tos plasmid were further digested into two parts *in vitro* and verified through agarose gel electrophoresis (Fig. 3C). PFGE was also used to verify the quality of genomic segments set in low melting point agarose (Fig. 3D). The clear and visual genomic bands represented the high quality of the genomic segments in gel plugs and provided a good basis for cloning. The prepared vectors were ligated with the genomic libraries and assembled separately. PCR was then used to validate the colonies grown on the LB agar medium of spectinomycin at the intersection sites between vector and target sequence (Supplementary Fig. S2). To further validate the colonies, two positive colonies were randomly chosen to make low melting point agarose gel plugs for PFGE. The results indicated that the linear cloning system can clone long genomic segments of up to 156 kb (Fig. 3E).

We also targeted the chromosome segments using the vector control amplified from pBAC. The results showed that our method had a higher positive rate in cloning large segments compared to conventional vectors (Table 1). The vector amplified from pBAC

plasmid had a low positive rate for cloning the large segments in our experiments. Through PCR validation, the negative colonies were found to be self-ligated vectors and plasmid residues (Supplementary Fig. S3). The cloning positive rate showed that it was hard to clone segments larger than 98 kb using both vectors, but the linear plasmid vector achieved more efficient cloning. The linear plasmids with 156 kb insert were successfully obtained using the TAPE method at 5.5%–11.5% positive rate. No positive colonies for 156 kb insert were obtained using the conventional vector. The cloning vectors were found to be far shorter than target genomic segments, while linear plasmid vectors came into contact with target sequence terminals more easily and ligated to the target sequence. Moreover, the conventional vector had the same size as the template plasmid, which was difficult to eliminate during the recovery of vectors with high concentration. The failed cloning using circular plasmid vectors was repaired through self-ligation in the host, further improving the negative rate. Our linear plasmid vectors were obtained from the digestion of PCR-amplified vector and contained two parts with a lower background plasmid size, which significantly decreased plasmid residue.

For targeted cloning of large gene clusters, we believe that the most valuable parameter is the success rate, which is positive correlation with the number of positive colonies. In this experiment, different variables were taken into consideration to study their relationship with the cloning success rate (Fig. 4). The x-axis was set as the ratio of genomic segments, while the y-axis represented the number of positive colonies based on PCR screening, which offers a more realistic guide to achieve cloning (Fig. 4A). According to the cloning results in target segments of 15 kb and 40 kb, the

Table 1
Effects of different vectors on cloning bacterial segments.

Vector ^a	Target	Repeat	GB05redTrfA::TelN	
			Number of colonies	Positive rate
pBAC-15 kb	<i>E. coli</i> -14,962 bp	Repeat 1	849	5/24
		Repeat 2	419	14/24
		Repeat 3	629	4/24
pBAC-40 kb	<i>E. coli</i> -40,597 bp	Repeat 1	873	1/24
		Repeat 2	875	1/24
		Repeat 3	912	0/24
pBAC-98 kb	<i>E. coli</i> -98,648 bp	Repeat 1	582	0/24
		Repeat 2	569	0/24
		Repeat 3	483	1/24
pBAC-156 kb	<i>E. coli</i> -156,067 bp	Repeat 1	102	0/102
		Repeat 2	69	0/69
		Repeat 3	47	0/47
pBAC-tos-TelN-15 kb	<i>E. coli</i> -14,962 bp	Repeat 1	177	24/24
		Repeat 2	202	24/24
		Repeat 3	765	23/24
pBAC-tos-TelN-40 kb	<i>E. coli</i> -40,597 bp	Repeat 1	253	18/24
		Repeat 2	172	16/24
		Repeat 3	69	20/24
pBAC-tos-TelN-98 kb	<i>E. coli</i> -98,648 bp	Repeat 1	11	6/11
		Repeat 2	18	1/12
		Repeat 3	6	3/6
pBAC-tos-TelN-156 kb	<i>E. coli</i> -156,067 bp	Repeat 1	15	1/15
		Repeat 2	26	3/26
		Repeat 3	18	1/18

^a The pBAC-15 kb, pBAC-40 kb, pBAC-98 kb, pBAC-156 kb represented the vectors that were amplified from pBAC plasmid and targeted the bacterial segments of 14,962 bp, 40,957 bp, 98,648 bp, 156,067 bp, respectively. pBAC-tos-TelN-15 kb, pBAC-tos-TelN-40 kb, pBAC-tos-TelN-98 kb, pBAC-tos-TelN-156 kb were the linear plasmid vectors obtained from the digestion of amplified vectors based on pBAC-tos plasmid using TelN protelomerase that targeted the bacterial segments of 14,962 bp, 40,957 bp, 98,648 bp, 156,067 bp.

high number of positive colonies showed a downward trend at the x-axis (Fig. 4B). The cloning results of 98 kb and 156 kb indicated that it was easier to obtain positive colonies at a smaller x value. Due to the fixed electroporation volume and genomic concentration, the ratio of vector in the cloning mix was important to clone the long DNA sequence. Our cloning results demonstrated that a high ratio of cloning vectors help to increase the probability of cloning success. Thus, it is evident that our strategy can obtain a high concentration of linear plasmid vector to increase the ratio of vector in the cloning mix through PCR amplification.

Targeting genomic segments from yeast and mammalian mitochondria

This study also aimed to find out whether our cloning method can be applied on eukaryotic species. We targeted the 40 kb genomic segment in chromosome I and a genomic sequence with a length of 124 kb in chromosome XV from yeast genome digested by SbfI restriction enzyme that covered both centromere sequences (Fig. 5A, Supplementary Fig. S4A and S4B). In addition, we cloned the mitochondria sequence in the mouse F9 cells (Supplementary Fig. S5A-C). PFGE was used to verify the positive colonies with the correct size using bacterial gel plugs (Fig. 5B). NucleoBond Xtra Midi Plus Kit was also used to extract the linear plasmid with insert segments of 40 kb, 98 kb, 124 kb, 156 kb [34]. The results showed that it was effective to extract the linear plasmid with target DNA sequences of 40 kb and 98 kb, but the DNA band for linear plasmid carrying 124 kb fragment was weak and the linear plasmid carrying the 156 kb fragment could not be recognized (Supplementary Fig. S4C). Therefore, it is advisable to extract the linear plasmid below 124 kb through a plasmid kit, while the larger linear plasmid should be purified by gel for further

research. The cloned DNA segments were also validated through sequencing in the *E. coli* host (Supplementary Fig. S6). The results indicated that our method can be applied to clone genomic segments from prokaryotic and eukaryotic genomes.

Discussion

With the development of molecular biology, a simple and easy method is urgently needed to overcome gene size limitations [35]. To clone large DNA sequences, we developed a facile method named TAPE to directly hook the long DNA sequence from genomic libraries. The linear plasmid vectors were mixed and ligated with genomic fragments using 2 × NEBuilder® HiFi DNA Assembly Master Mix *in vitro*, while the DNA fragments were cloned and maintained in *E. coli* with linear chromosome. We used NEBuilder® HiFi DNA Assembly Master Mix as the cloning reagent, which circumvented the homologous recombination of *E. coli*. This is an ideal method for cloning repetitive sequences or high GC-content sequences that contribute to the synthesis of complex genomes. The linear cloning method showed an obvious improvement in efficiency and large capacity to carry and clone at least several hundred kilobase pairs. The genomic fragments from different species were successfully cloned in a linear chromosome. This means our method can clone all types of genome sequences. The cloning results in our experiments further demonstrated that the high ratio of cloning vectors increased the success rate when the genomic libraries are of high quality and at a suitable concentration. In the cloning reaction, the vector and genomic segment concentrations were the key factors that contributed to successful cloning. However, the high concentration of recovery genome was too viscous to pipette and affected ligation. Thus, the genomic libraries need to be optimized based on the characteristics of different

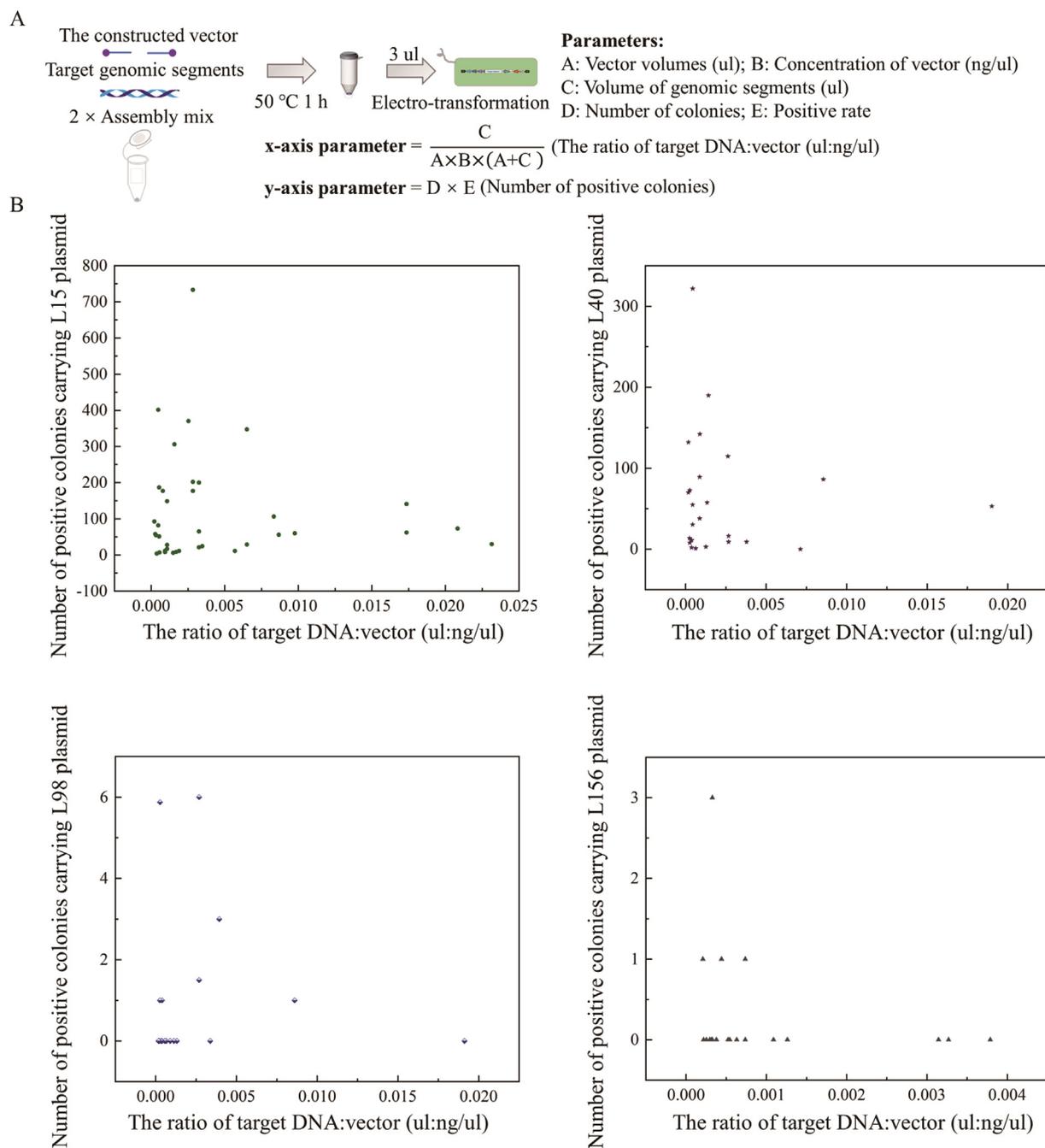


Fig. 4. Cloning efficiencies of different target sequences using TAPE method. (A) Parameters in the cloning process *in vitro*. (B) Results of cloning the bacterial DNA segments with different lengths in *E. coli* str. GB05RedTrfA::TeIN host. The L15, L40, L98, L156 plasmids represented the linear plasmid carrying the bacterial segments of 14,962 bp, 40,957 bp, 98,648 bp, 156,067 bp, respectively. PCR was used to validate the positive colonies.

genomes to support the target segments with high quality [36]. The success rate of cloning decreased when the plasmid size went beyond 98 kb, which is consistent with previous research findings [18,20]. In terms of reduced efficiency in cloning of large DNA sequences, we deduced that long DNA fragments *in vitro* are easily damaged and the large plasmids have low efficiency for electro-transformation. In addition, the mutual contact between linear plasmid vectors and large target segments is higher than conventional vectors, suggesting that it is easier to clone the target sequence. The unsuccessful cloning using circular plasmid vectors can be addressed by self-ligation to improve the negative rate. As a result, we managed to obtain the genomic

segment of 156 kb, which is reported to be the largest DNA sequence directly cloned in *E. coli* host to date (Supplementary Table S4). The direct cloning segments of 124 kb and 156 kb reflect the high efficiency of our technique compared to other cloning methods [18,20,29].

Furthermore, the genomic libraries contain various segments, which is cost-efficient and takes a shorter time to clone numerous target DNA sequences from one species. The linear plasmids can be purified in low melting point agarose or extracted using plasmid kit, which can be recovered and transformed to yeast that assemble genomes in megabase pairs. Meanwhile, the bacteria host also bypasses the yeast host to clone yeast genome. This further

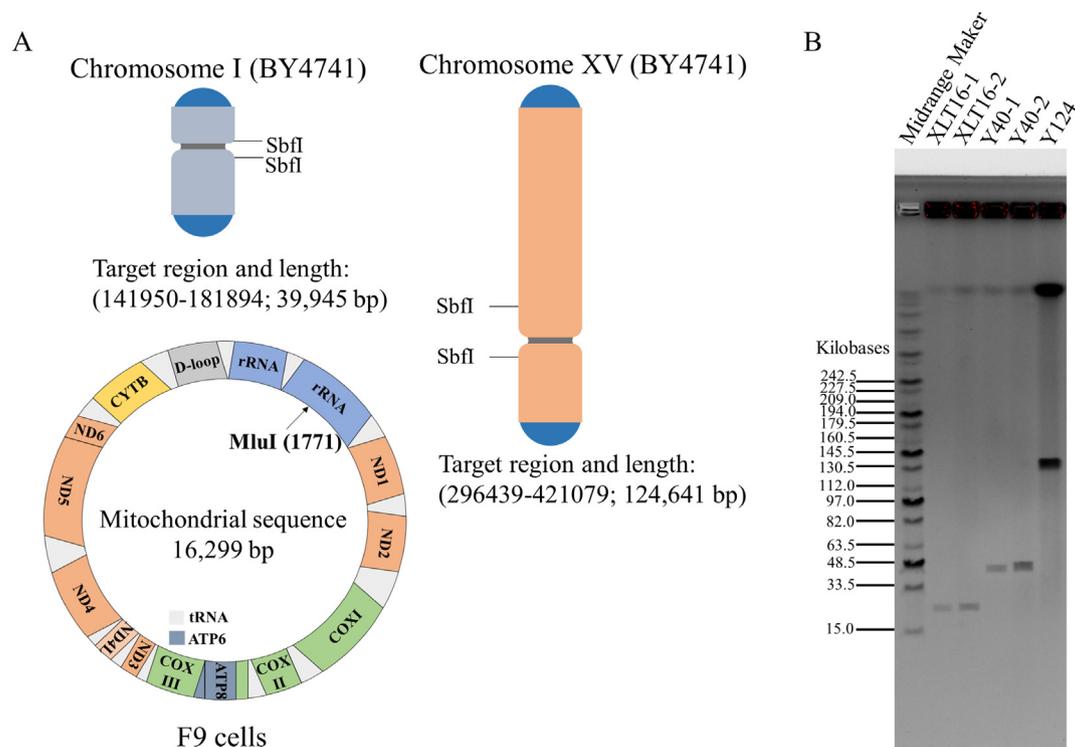


Fig. 5. Targeting genomic sequences of yeast and mammalian mitochondria. (A) Genome location and cleavage sites for target segments of yeast and mammalian mitochondria. The genome of *Saccharomyces cerevisiae* BY4741 was digested by SbfI enzyme, while a fragment of 39,945 bp in chromosome I and a fragment of 124,641 bp in chromosome XV were tested as the target DNA sequences. Mitochondria in F9 cells was also selected as the cloning target by digesting the genome using MluI enzyme. (B) Verification of the positive colonies through PFGE with bacterial plugs. XLT16: the linear plasmid with 16 kb mitochondria DNA sequence from F9 cells. Y40 and Y124 were the linear plasmids carrying 39,945 bp, 124,641 bp genomic segments of yeast, respectively.

demonstrates that it can assemble or edit the yeast genome in *E. coli*. In addition, arbitrary restriction sites can be introduced in vectors using primers. This means that our method can be applied to improve the conventional construction of BAC libraries to facilitate next-generation sequencing of complex genomes. The linear plasmids also contribute to functional expression of large genomic loci and linear viral genomes in natural linear forms.

Conclusion

In conclusion, an efficient method was developed to obtain the target genomic segments using the TelN/tos system and a linear plasmid vector. The TAPE method had a considerable improvement in cloning efficiency and could clone different types of genome sequences. This method took no more than five days to directly obtain the target DNA sequence of up to hundreds of thousands of base pairs. This cloning method is useful in DNA sequencing and cloning, and the annotation of complex genomes.

Compliance with ethics requirements

This article does not contain any studies with human or animal subjects.

CRedit authorship contribution statement

You-Zhi Cui: Methodology, Validation, Formal analysis, Investigation, Writing – original draft. **Jian-Ting Zhou:** Validation, Formal analysis, Investigation, Writing – review & editing, Resources. **Bing-Zhi Li:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing. **Ying-jin Yuan:** Conceptualization, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Professor Youming Zhang from Shandong University for providing the bacteria strains. This work was supported by grants from National Key Research and Development Program of China (2018YFA0900100), Tianjin Fund for Distinguished Young Scholars (19JCJQC63300) and the National Natural Science Foundation of China (21621004).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jare.2022.01.017>.

References

- [1] Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 2018;36(4):338–45. doi: <https://doi.org/10.1038/nbt.4060>.
- [2] Zhang J, Hou L, Zuo Z, Ji P, Zhang X, Xue Y, et al. Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. *Nat Biotechnol* 2021;39(7):836–45. doi: <https://doi.org/10.1038/s41587-021-00842-6>.
- [3] Gao L, Chang S, Xia W, Wang X, Zhang C, Cheng L, et al. Circular RNAs from BOULE play conserved roles in protection against stress-induced fertility decline. *Sci Adv* 2020;6(46). doi: <https://doi.org/10.1126/sciadv.abb7426>.
- [4] Guo Li, Winzer T, Yang X, Li Yi, Ning Z, He Z, et al. The opium poppy genome and morphinan production. *Science* 2018;362(6412):343–7.
- [5] Sugimoto Y, Camacho FR, Wang S, Chankhamjon P, Odabas A, Biswas A, et al. A metagenomic strategy for harnessing the chemical repertoire of the human microbiome. *Science* 2019;366(6471).

- [6] Kalkreuter E, Pan G, Cepeda AJ, Shen B. Targeting Bacterial Genomes for Natural Product Discovery. *Trends Pharmacol Sci* 2020;41(1):13–26. doi: <https://doi.org/10.1016/j.tips.2019.11.002>.
- [7] Tabebordbar M, Zhu K, Cheng JKW, Chew WL, Widrick JJ, Yan WX, et al. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* 2016;351(6271):407–11.
- [8] Gibson DG, Glass JL, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, et al. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* 2010;329(5987):52–6.
- [9] Fredens J, Wang K, de la Torre D, Funke LFH, Robertson WE, Christova Y, et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* 2019;569(7757):514–8. doi: <https://doi.org/10.1038/s41586-019-1192-5>.
- [10] Tagwerker C, Dupont CL, Karas BJ, Ma L, Chuang RY, Benders GA, et al. Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. *Nucleic Acids Res* 2012;40(20):10375–83. doi: <https://doi.org/10.1093/nar/gks823>.
- [11] Xie Z-X, Li B-Z, Mitchell LA, Wu Yi, Qi X, Jin Z, et al. “Perfect” designer chromosome V and behavior of a ring derivative. *Science* 2017;355(6329).
- [12] Wu Yi, Li B-Z, Zhao M, Mitchell LA, Xie Z-X, Lin Q-H, et al. Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* 2017;355(6329).
- [13] Zhou J, Wu R, Xue X, Qin Z. CasHRA (Cas9-facilitated Homologous Recombination Assembly) method of constructing megabase-sized DNA. *Nucleic Acids Res* 2016;44(14):e124. doi: <https://doi.org/10.1093/nar/gkw475>.
- [14] Li L, Blankenstein T. Generation of transgenic mice with megabase-sized human yeast artificial chromosomes by yeast spheroplast-embryonic stem cell fusion. *Nat Protoc* 2013;8(8):1567–82. doi: <https://doi.org/10.1038/nprot.2013.093>.
- [15] Li Y, Liu W, Oo TF, Wang L, Tang Yi, Jackson-Lewis V, et al. Mutant *LRKK2^{R144G}* BAC transgenic mice recapitulate cardinal features of Parkinson's disease. *Nat Neurosci* 2009;12(7):826–8. doi: <https://doi.org/10.1038/nn.2349>.
- [16] Kouprina N, Leem S-H, Solomon G, Ly A, Koriabine M, Otsot J, et al. Segments missing from the draft human genome sequence can be isolated by transformation-associated recombination cloning in yeast. *EMBO Rep* 2003;4(3):257–62. doi: <https://doi.org/10.1038/sj.embor.embor766>.
- [17] Lee NC, Larionov V, Kouprina N. Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. *Nucleic Acids Res* 2015;43(8):e55. doi: <https://doi.org/10.1093/nar/gkv112>.
- [18] Jiang W, Zhao X, Gabrieli T, Lou C, Ebenstein Y, Zhu TF. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat Commun* 2015;6:8101. doi: <https://doi.org/10.1038/ncomms9101>.
- [19] Enghiad B, Huang C, Guo F, Jiang G, Wang B, Tabatabaei SK, et al. Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination. *Nat Commun* 2021;12(1). doi: <https://doi.org/10.1038/s41467-021-21275-4>.
- [20] Wang H, Li Z, Jia R, Yin J, Li A, Xia L, et al. ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res* 2018;46(5):e28. doi: <https://doi.org/10.1093/nar/gkx1249>.
- [21] Thomason LC, Sawitzke JA, Li X, Costantino N, Court DL. Recombineering: Genetic Engineering in Bacteria Using Homologous Recombination. *Current Protocols in Molecular Biology* 2014;106(1). doi: <https://doi.org/10.1002/0471142727.mb0116s106>.
- [22] Ravin NV, Ravin VK. Use of a linear multicopy vector based on the mini-replicon of temperate coliphage N15 for cloning DNA with abnormal secondary structures. *Nucleic Acids Res* 1999;27(17):. doi: <https://doi.org/10.1093/nar/27.17.e13>.
- [23] Godiska R, Mead D, Dhodda V, Wu C, Hochstein R, Karsi A, et al. Linear plasmid vector for cloning of repetitive or unstable sequences in *Escherichia coli*. *Nucleic Acids Res* 2010;38(6):e88. doi: <https://doi.org/10.1093/nar/gkp1181>.
- [24] Ravin NV, Tolmasky M, Alonso JC. Replication and Maintenance of Linear Phage-Plasmid N15. *Microbiol Spectr* 2015;3(1). doi: <https://doi.org/10.1128/microbiolspec.PLAS-0032-2014>.
- [25] Liew PS, Chen Q, Ng AWR, Chew YC, Ravin NV, Sim EUH, et al. Phage N15 protelomerase resolves its tos recognition site into hairpin telomeres within mammalian cells. *Anal Biochem* 2019;583:113361. doi: <https://doi.org/10.1016/j.ab.2019.113361>.
- [26] Walters AA, Kinnear E, Shattock RJ, McDonald JU, Caproni LJ, Porter N, et al. Comparative analysis of enzymatically produced novel linear DNA constructs with plasmids for use as DNA vaccines. *Gene Ther* 2014;21(7):645–52. doi: <https://doi.org/10.1038/gt.2014.37>.
- [27] Liew PS, Tan TH, Wong YC, Sim EUH, Lee CW, Narayanan K. A Self-Replicating Linear DNA. *ACS Synth Biol* 2020;9(4):804–13. doi: <https://doi.org/10.1021/acssynbio.9b00478>.
- [28] Karda R, Counsell JR, Karbowiczek K, Caproni LJ, Tite JP, Waddington SN. Production of lentiviral vectors using novel, enzymatically produced, linear DNA. *Gene Ther* 2019;26(3–4):86–92. doi: <https://doi.org/10.1038/s41434-018-0056-1>.
- [29] Wang H, Li Z, Jia R, Hou Yu, Yin J, Bian X, et al. RecET direct cloning and Red $\alpha\beta$ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat Protoc* 2016;11(7):1175–90. doi: <https://doi.org/10.1038/nprot.2016.054>.
- [30] Jiang Yu, Chen B, Duan C, Sun B, Yang J, Yang S, et al. Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl Environ Microbiol* 2015;81(7):2506–14. doi: <https://doi.org/10.1128/AEM.04023-14>.
- [31] Zhang M, Zhang Y, Scheuring CF, Wu CC, Dong J, Zhang HB. Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nat Protoc* 2012;7(3):467–78. doi: <https://doi.org/10.1038/nprot.2011.455>.
- [32] Brown DM, Chan YA, Desai PJ, Grzesik P, Oldfield LM, Vashee S, et al. Efficient size-independent chromosome delivery from yeast to cultured cell lines. *Nat Protoc* 2017;45(7):e50. doi: <https://doi.org/10.1093/nar/gkw1252>.
- [33] Sheng Y, Mancino V, Birren B. Transformation of *Escherichia coli* with large DNA molecules by electroporation. *Nucleic Acids Res* 1995;23(11):1990–6. doi: <https://doi.org/10.1093/nar/23.11.1990>.
- [34] Chen Q, Narayanan K. Recombineering linear BACs. *Methods Mol Biol* 2015;1227:27–54. doi: https://doi.org/10.1007/978-1-4939-1652-8_2.
- [35] Mitchell LA, McCulloch LH, Pinglay S, Berger H, Bosco N, Brosh R, et al. De novo assembly and delivery to mouse cells of a 101 kb functional human gene. *Genetics* 2021;218(1):iyab038. doi: <https://doi.org/10.1093/genetics/iyab038>.
- [36] Jiang W, Zhu TF. Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat Protoc* 2016;11(5):960–75. doi: <https://doi.org/10.1038/nprot.2016.055>.