Original Research Article

Synthesis and assembly of full-length cyanophage A-4L genome

Ting Zhang [a,b,1], Bonan Xu [a,b,1], Jia Feng [a,b], Pingbo Ge [a,b], Guorui Li [a,b], Jiabao Zhang [a,b], Jianting Zhou [a,b,c,**], Jianlan Jiang [a,b,c,*]

[a] *School of Chemical Engineering and Technology, Tianjin University, Tianjin, 300072, China*
[b] *Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin, 300072, China*
[c] *Frontier Science Center for Synthetic Biology (Ministry of Education), Tianjin University, Tianjin, 300072, China*

ARTICLE INFO

ABSTRACT

Artificial cyanophages are considered to be an effective biological method to control harmful cyanobacterial bloom. However, no synthetic cyanophage genome has been constructed and where its obstacles are unclear. Here, we survey a stretch of 16 kb length sequence of cyanophage A-4L that is unclonable in *Escherichia coli*. We test 12 predicted promoters of cyanophage A-4L which were verified all active in *E. coli*. Next, we screen for eight ORFs that hindered the assembly of intermediate DNA fragments in *E. coli* and describe that seven ORFs in the 16 kb sequence could not be separately cloned in *E. coli*. All of unclonable ORFs in high-copy-number plasmid were successfully cloned using low-copy-number vector, suggesting that these ORFs were copy-number-dependent. We propose a clone strategy abandoned the promotor and the start codon that could be applied for unclonable ORFs. Last, we *de novo* synthesized and assembled the full-length genome of cyanophage A-4L. This work deepens the understanding of synthetic cyanophages studies.

1. Introduction

Synthetic biology has advanced to the point of synthesizing the entire genome, including *Mycoplasma* genome [1], yeast chromosome [2], *Escherichia coli* genome [3] and mouse mitochondrial genome [4]. Large assemblies may be unstably maintained in *E. coli* but can be easily completed in yeast [5]. Viral genomes [6] have also been constructed, including poliovirus, simian immunodeficiency virus, coxsackievirus, adenovirus, tobacco mosaic virus, human endogenous retrovirus, coronavirus, as well as bacteriophage ϕX174, T7, AP205 and G4 [7]. The cloning of many phage genomes in bacteria is stalled by possible gene virulence [8]. Genome synthesis of viruses provides a powerful tool for studying the function of viral genes and *de novo* synthesis allows modification in naturally occurring genomes [9,10].

Cyanophages, the viruses of blue-green algae, are abundant in aquatic ecosystem and play an important role in regulating the cyanobacteria biomass and community structure, which have been considered as a potential biological method for controlling the cyanobacterial bloom [11]. In recent years, a very limited number of freshwater cyanophages, compared to marine cyanophages, have been studied and most of them are highly host-specific while water blooms are usually caused by multiple cyanobacteria [12]. Refactoring and engineering artificial genomes can be obtained using synthetic biology strategies. However, no synthetic genome of cyanophage has been reported to date. The freshwater cyanophage A-4L (GenBank accession no. KF356198) infecting model cyanobacteria *Nostoc* sp. strain PCC 7120 [13] was isolated in 1981 [14] with its complete genome sequenced and analyzed in 2015 [15]. We suggest that the *de novo* synthesis of cyanophage A-4L can be helpful to better understand the interactions of virus–host system and thus to promote biological properties of natural cyanophages by functional characterization and genetic modifications.

We reported here the bottom-up assembly of synthetic cyanophage A-4L genome and surveyed possible toxic ORFs that hindered the assembly of the whole genome in *E. coli*. Low-copy-number plasmid was recommended for the construction of sequences that may be toxic to host, since the difficulties of cloning observed here was shown to correlate with copy number. We proposed a strategy to segment genome sequences according to annotated ORFs and applied it for the synthesis

of the 16 kb fragment of A-4L genome in *E. coli*. In the absence of the promoter and start codon, ORFs can be cloned without considering the adverse effects of toxic gene expression products. Based on the above studies, the completion of the artificial full-length A-4L genome was finally achieved by transformation-associated recombination (TAR) [16] in yeast.

2. Material and methods

2.1. Strains, competent cells, plasmids, and culture conditions

All strains, competent cells and plasmids used in this study are listed in Supplementary Table S1. *E. coli* strains were grown in Luria-Bertani (LB) broth with appropriate antibiotics (Ampicillin: 100 μg/ml; Chloramphenicol: 25 μg/ml) and rotated at 37 °C, 220 rpm. *Saccharomyces cerevisiae* BY4741 was cultured in YPD (Yeast Extract Peptone Dextrose) medium at 30 °C, 200 rpm. Yeast strains carrying plasmids were grown in synthetic complete medium without histidine/leucine/uracil (SC–H/L/U). *Nostoc* sp. PCC 7120 strain was grown in BG11 medium at 30 °C, 150 rpm with a light density of 2000 lux. *E. coli* competent cells were purchased from Biomed, China.

2.2. Deletion strategy for possible toxic ORFs

ORFs between adjacent B-fragments were designed to be deleted to achieve the construction of C-assemblies that could be successfully transformed into *E. coli*. In a general way, a single ORF among all possible toxic ORFs was deleted first, and then two or more ORFs were deleted simultaneously until the recombinant plasmid can be transferred into *E. coli*. Another scheme that deletion started with all ORFs deleted at the same time was put into use.

2.3. Plasmids construction for verification of predicted promoters

The chloramphenicol acetyltransferase (*cat*) gene was used as the report gene by replacing its own promoter with individual predicted promoter sequence. To avoid the initiation of the *cat* gene caused by unannotated promoters, the T7Te terminator was added. The *cat* gene and T7Te terminator were obtained from plasmid pLS0 by PCR amplification. Predicted promoter, with about 30 bp terminal overlaps introduced by primers, was inserted between the *cat* gene and the T7Te terminator *in vitro* assembly. The β-lactamase (*bla*) gene conferring resistance to ampicillin in the backbone of pUC57, played a role in transformant screening. The function of the promoter was further determined according to phenotype of chloramphenicol resistance of the recombinant plasmid. The negative control was plasmid containing the *cat* gene with no promoter sequence. The positive control was consisted of the original promoter together with the *cat* gene.

2.4. Serial dilution assays

E. coli colonies of strains for predicted promoters were grown overnight in 5 mL of LB medium with 25 μg/ml chloramphenicol added at 37 °C and then were serially diluted in 10-fold increments in water and plated. The initial concentration of each sample was limited to OD600 = 0.02. All LB plates were incubated at 37 °C for 12–16 h.

2.5. Syn-A-4L design and assembly

According to the reference sequence, the natural A-4L genome except direct terminal repeats (DTRs) was divided into A1-A81 with 700–800 bp in length, which were all synthesized, verified and provided from company (Tsingke Biological Technology, Beijing, China). *Not I*(5′-GCGGCCGC-3′) restriction sites were added in both sides of each DNA cassette for the convenient release from cloning vector. 2-4 adjacent A-fragments were joined together to generate 2 kb segments of B1–B25 *in*

vitro homologous recombination and transferred to *E. coli*. The assemblies were obtained by gel separation and purification (TIANgel Purification Kit, TIANGEN BIOTECH, Beijing, China). B11–B25 were recombined in sets of 2–4 to produce C4–C8 in *E. coli*, whereas B1–B10 were failed to generate C1–C3. The assembly of the 16 kb fragment was achieved *via* TAR in yeast. The direct terminal repeat named C9, which overlaps with ORF1 by 379 bp, was ligated with pUC57 and cloned in *E. coli*. Two different vectors containing only yeast regulatory elements was used to achieve complete genome assembly. The *URA3* selection marker was PCR amplified using plasmid pRS426 [17] as the template. The yeast centromere sequence and autonomously replicating sequence (*CEN/ARS*) from pRS415 [18] allowed Syn-A-4L2.1 to be stably propagated at a low expression dosage. *2μ*-based high-copy-number origin of replication was used in Syn-A-4L2.2 to increase the concentration of plasmids extracted from yeast cells. The full-length synthetic genome of cyanophage A-4L was assembled in yeast from fragments described above.

2.6. In vitro assembly

2-4 fragments with about 20 bp overlaps could be recombined in 15 min at 50 °C *in vitro* assembly (EASY®-Basic Seamless Cloning and Assembly Kit, TransGen Biotech, Beijing, China). The reaction products were transferred to chemically competent cells. *E. coli* cells from single colony were resuspended into 20 μL of ddH$_2$O and treated with the PCR program in thermocycler to firstly verify the correct connection between fragments.

2.7. Yeast assembly

S. cerevisiae BY4741 competent cells were prepared as described [19]. DNA fragments with about 60 bp overlaps [20], linearized vector and 50 μl competent cells were co-transformed referring to the LiAc/SS carrier DNA/PEG method [21]. The recombined plasmids extracted (TIANprep Mini Plasmid Kit, TIANGEN BIOTECH, Beijing, China; HiPure BAC DNA Mini Kit, Magen Biotechnology, Guangdong, China) from yeast were transferred into electroporation-competent cells. Yeast colony PCR analysis (YCPCR) was performed according to a modified method [22]. PCR primers were designed to produce amplicon across each of the junctions.

3. Results

3.1. The promotors on the 16 kb sequence of cyanophage A-4L are active in E. coli

23 ORFs on the left arm of A-4L genome are transcribed rightward and encode several relatively conserved proteins, which are involved in DNA replication and modification [15]. In previous study, clone of the 16 kb (351 bp~15,930 bp) fragment containing these ORFs was hampered by what was speculated to be genes that toxic for *E. coli* [23]. C1–C3, about 5 kb segments randomly split from the 16 kb sequence, were also found to fail transferring into *E. coli*. However, ten short DNA fragments (B1–B10 of 2 kb in size) subdivided were respectively transformed and all survived in *E. coli*, suggesting that the sequences between adjacent B-fragments may be toxic (Fig. 1A and B).

We speculated that the cyanophage genome might contain genes encoding proteins that are toxic to *E. coli*, so we tested the putative promoter activity of the A-4L genome in *E. coli*. In view of the presence of several consecutive ORFs, 12 predicted promoter sequences were explored (Fig. 1A). The sequence about 200 bp before individual ORF was used to replace the original promoter of the *cat* gene (Fig. 1C). The *cat* gene and all fragments were joined together *in vitro* assembly. Chloramphenicol resistance of resulting plasmid was observed in *E. coli*. Although some phenotypic differences emerged (Fig. 1D), all tested promoters were proved to be functional, which meant that ORFs of
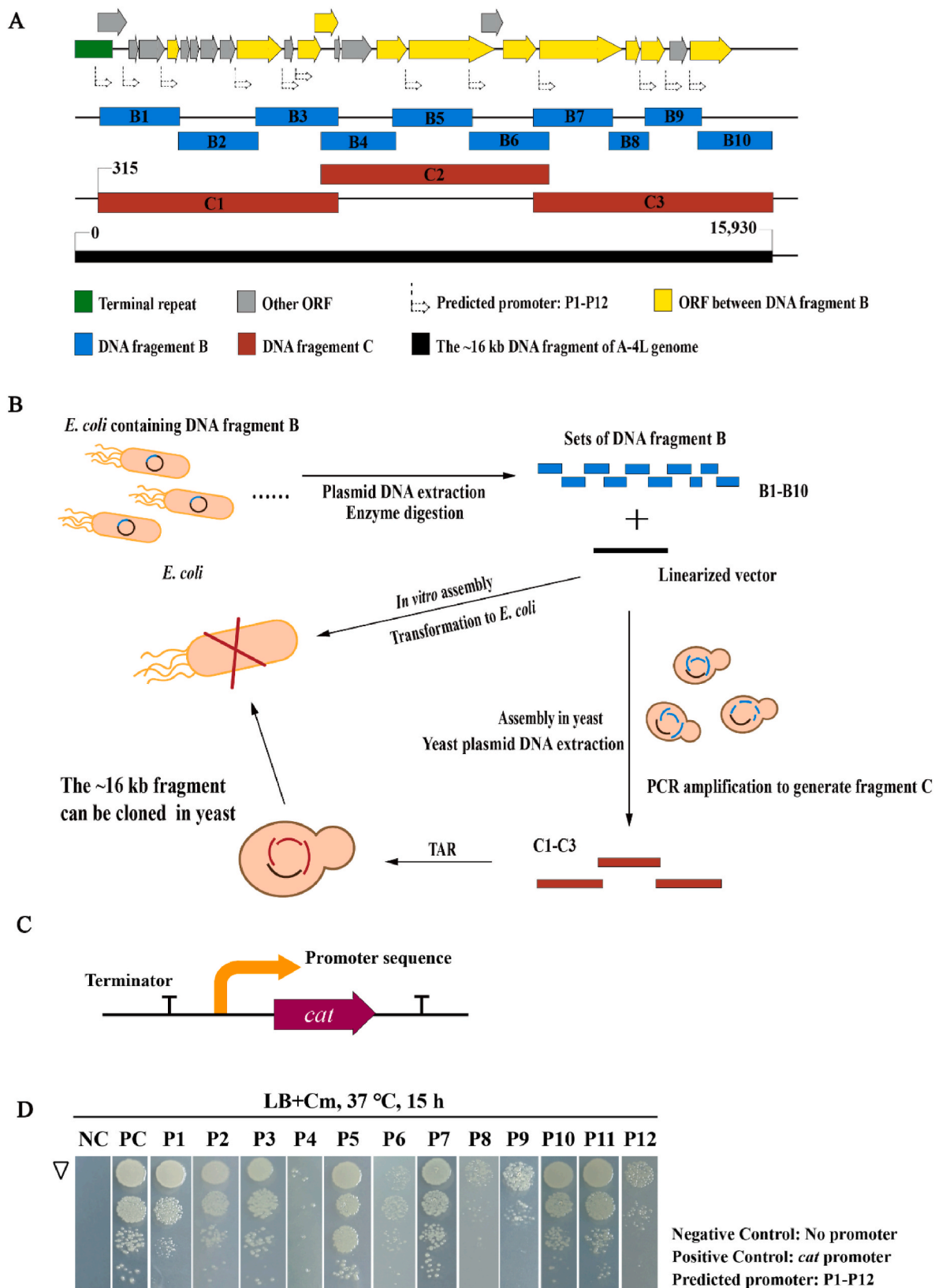
Fig. 1. The 16 kb sequence of A-4L genome. (A) The ~16 kb fragment contains 23 ORFs and can be divided into ~5 kb fragments C1–C3/~2 kb fragments B1–B10. The 12 predicted promoter sequences (P1–P12) are respectively located in front of ORF1, ORF2, ORF4, ORF9, ORF10, ORF11, ORF16, ORF17, ORF19, ORF21, ORF22 and ORF23. (B) The ~16 kb fragment and C-fragments can only be maintained in yeast but B-fragments can exist in *E. coli*. (C) The *cat* gene was used as the report gene to verify the function of predicted promoters. (D) Phenotypic analysis of predicted promoter sequences.

between adjacent B-fragments can be expressed.

3.2. Screening for potential toxic ORFs in E. coli by ORF-deletion and ORF-clone

To identify toxic ORFs, the strategy of ORF deleting in each C-fragment was first used in order to avoid false-negative results due to operational errors. We explored ORFs located between B-fragments (Supplementary Table S2), which were deleted as designed (Table 1). *S. cerevisiae* can stitch together multiple DNA sequences efficiently and with high fidelity [5]. Recombinant plasmids with pRS415 as backbone were assembled *via* TAR in yeast and transferred into *E. coli* (Fig. 2A). There are three ORFs between adjacent B-fragments in C1. Colonies appeared when ORF11 was deleted. Plasmids lacking a single ORF in C2 were failed introducing into *E. coli*. Colonies only appeared when ORF15 and ORF16 were deleted simultaneously. No colonies appeared when all four ORFs to be investigated in C3 were deleted at the same time. We surmised that ORF22, the fifth ORF remaining in the fragment, possibly is toxic. The second round of screening showed that plasmid with ORF19-23 deleted at once could be transformed into *E. coli* but plasmid with only ORF22 deleted could not. Failures of electroporation were also observed in assemblies deleted four ORFs including ORF22. Analysis of the assembled C1ΔORF11, C2ΔORF1516 and C3ΔORF19-23 by digestion with *Eco*R I was shown in Fig. 2B. The 16 kb fragment with eight toxic ORFs deleted was constructed from C1, C2 and C3 as described above *via* TAR. However, no colonies appeared when this plasmid was electroporated into *E. coli*. It seemed that there were other lethal factors in this region which make the large segment unacceptable to *E. coli*.

Next, we cloned 23 ORFs in the 16 kb region of A-4L genome to further verify toxic ORFs (Fig. 2C). Each ORF containing predicted promoter and about 100 bp downstream overlap was assembled with the same *E. coli*-yeast shuttle vector used in ORF-deletion. Plasmid isolated from yeast cells was then electroporated into *E. coli*. The result showed that ORF11, ORF16, ORF19 and ORF20 among toxic ORFs could not be cloned in *E. coli* as expected, while ORF15, ORF21, ORF22 and ORF23 were surprisingly clonable. In addition, ORFs except for ORF8, ORF17 and ORF18 were all transferrable into *E. coli*. However, ORF8 was proved to be cloned when it was included in the C1 fragment where ORF11 was deleted, as well as ORF17 and ORF18 were turned out to be innocuous in C2ΔORF1516, as shown in ORF-deletion. We also constructed a recombinant plasmid containing both ORF17 and ORF18, which was found to be cloned and maintained stably in *E. coli*. Plasmids constructed in ORF-clone, which were capable of introducing into *E. coli*, were verified by enzyme digestion (Supplementary Fig. S1). Combining the screening results of ORF-deletion and ORF-clone (Fig. 2D), we speculated that a portion of the observed lack of transfer is likely due to

detrimental gene expression products.

3.3. Clone strategies for the 16 kb fragment in E. coli

The *de novo* synthesis begins with chemically synthesized oligonucleotides and stepwise large DNA fragments up to full-length genome are assembled. The first few stages of assembly usually were done by *in vitro* recombination and cloned into *E. coli*. However, cloning of randomly divided DNA fragments in *E. coli* had met similar difficulties to that encountered in cloning the 16 kb fragment of A-4L genome [24]. Gene-transfer barriers may be associated with toxicity of the transferred gene to the host [25–27]. Therefore, those segments, which turned out to be problematic, had to get cloned in some other ways.

High-copy-number plasmid often become unstable or kill the hosts due to their high gene dosage when expressing toxic sequences [28]. Therefore, we prepared another shuttle vector from a low-copy-number plasmid pLS0 (Fig. 3A) for ORF cloning. pLS0 behaves as a single copy under the regulation of origin of replication for bacterial F plasmid and contains several genes for plasmid partitioning and a cos site [29] for accommodating large DNA sequence. Expectedly, eight toxic ORFs screened out and three additional ORFs failed cloning using pRS415 could all be transformed into *E. coli* together with pLS0. The fragments used here were verified by sequencing. Successful cloning meant that genes could be innocuous at the low level of expression. A portion of transfer barriers described in this study likely was gene copy–number dependent. We proposed that low-copy-number plasmids should be preferred as cloning vectors in the synthetic assembly of genomes with potentially toxic sequences.

Given the cloning failures of which the gene expression products might be toxic for the host, we proposed a cloning strategy that abandoned the promotor and the start codon of ORFs. As shown in Fig. 3B, DNA fragment of wild-type genome was partitioned referring to the predicted ORFs and the resulting segments were cloned in *E. coli*. In this way, the sequence of each segment started after the start codon of the upstream ORF and contained about 100 bp overlap with the next ORF. The 16 kb fragment of A-4L genome in this study was split into 23 segments to be cloned. DNA fragment was joined to linearized vector of pUC57 *in vitro* recombination and directly transformed into *E. coli* cells. In the absence of the promoter and start codon, bacterial growth was observed with the introduction of plasmids containing every single ORF in the 16 kb region. ORFs can all be propagated in *E. coli* because no harmful gene product was generated. The results of restriction enzyme digestion were shown in Supplementary Fig. S2. This approach can be applied in the bottom-up assembly with predicted ORF-information provided, and toxic ORFs whose expression products probably are lethal to *E. coli* can be cloned as well.

3.4. Design and assembly of full-length cyanophage A-4L genome

The A-4L genome is a 41,750 bp of linear dsDNA molecule and contains 810 bp long direct repeat at both ends [15]. The construction of synthetic A-4L genome began with oligonucleotides. In a general way, the genome except DTRs was divided into 81 overlapping fragments (A1-A81) with an average length of 700–800 bp at first. In the second step, all the A-fragments were assembled *in vitro* homologous recombination into 25 fragments (B1–B25) of approximately 2 kb in size. In the third step, we found that merely B11~B25 were able to produce C4–C8 in *E. coli* but B1–B10 failed to generate C1–C3 because of the toxicity for *E. coli* (Fig. 4A). The 16 kb fragment could be obtained using the strategy proposed above. 23 ORFs were separately cloned without the promoter and start codon. The directed terminal repeat sequence with 379 bp overlap of ORF1 (C9), instead of the original two DTRs at each end of the natural linear genome was retained. Then, two sets of neighboring ORFs could be assembled by OE-PCR (overlap-extension PCR) to form b1~b12 (the start codon of ORF1 herein was recovered from C9 *in vitro* assembly), and the assembled products were then co-transformed into yeast
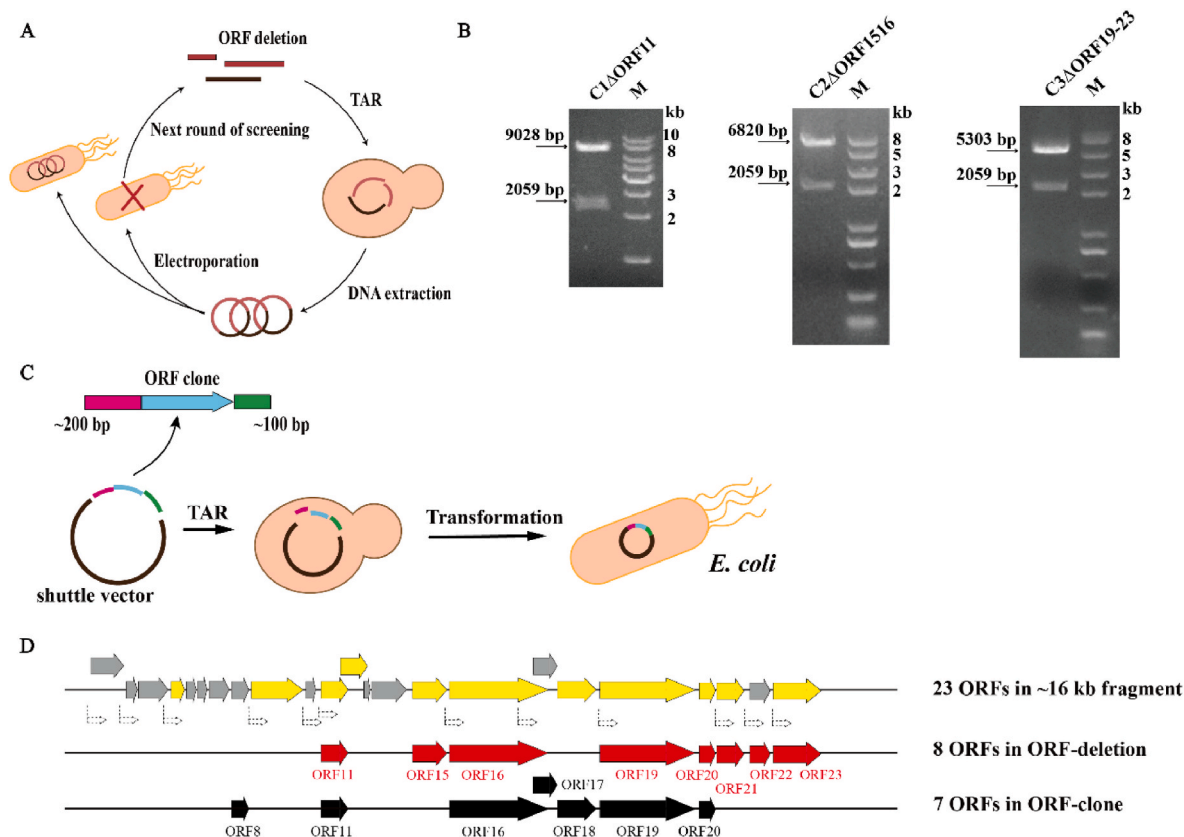
Table 1
Scheme of ORF deletion.

	ORFs deleted in Round 1	ORFs deleted in Round 2	ORFs deleted in Round 3
C1	ORF4/ ORF9/ **ORF11**/	–	–
C2	ORF12/ ORF15/ ORF16/ ORF18/	ORF12, ORF15/ ORF12, ORF15/ ORF12, ORF16/ ORF12, ORF18/ **ORF15, ORF16**/ ORF16, ORF18/	–
C3	ORF19, ORF20, ORF21, ORF23/	**ORF19, ORF20, ORF21,** **ORF22, ORF23**/ ORF22/	ORF19, ORF20, ORF21, ORF22/ ORF19, ORF20, ORF23, ORF22/ ORF19, ORF21, ORF23, ORF22/ ORF20, ORF21, ORF23, ORF22/

Fig. 2. Scheme and result of toxic ORFs screening. (A) ORF-deletion strategy: recombinant plasmids deleted ORF(s) were constructed in yeast and transferred into *E. coli*. (B) C1ΔORF11, C2ΔORF1516 and C3ΔORF19-23 that can be propagated in *E. coli* were digested by *Eco*R I into two segments. (C) ORF-clone strategy: ORF was cloned in yeast along with ~200 bp sequence before and ~100 bp sequence after the ORF and transformed into *E. coli*. (D) The results of toxic ORFs screening based on ORF-deletion and ORF-clone.

cells to generate c1~c3 (Fig. 4A). The synthetic cyanophage A-4L was designed to be a circular genome (Fig. 4B) referring to rolling-circle (RC) mechanism commonly seen in replication of many bacteriophages DNA and plasmids of bacteria [30,31].

Considering that the whole A-4L genome cannot exist in *E. coli*, plasmids for complete genome assembly were designed to contain only yeast regulatory elements and the selection marker (*URA3* used in here). To stably maintain the synthetic cyanophage genome, low-copy-number plasmid with *CEN/ARS* was used to construct Syn-A-4L2.1 (42,956 bp). High-copy-number plasmid with *2μ ori* instead were used in Syn-A-4L2.2 (43,678 bp) to increase the yeast plasmid concentration. Restriction-modification (R-M) systems is an innate immunity evolved by bacteria to prevent phage invasion, which protects hosts from exogenous DNA [32]. Type II restriction endonucleases are part of restriction modification systems [33]. *Nostoc* sp. PCC 7120 has been proved to have R-M systems, possessing *Ava I*, *Ava II*, and *Nsi I*, an isoschizomer of *Ava III* [34] (Supplementary Fig. S3). To avoid the inefficient transformation of artificial cyanophage genome into cyanobacteria, the restriction sites on the vectors were removed by mutation (Supplementary Fig. S4A). The assemblies of the synthesized A-4L genome were verified by PCR amplification of DNA fragments (Fig. 4C) and yeast colony PCR (Supplementary Fig. S4B).

4. Discussion

To our knowledge, this is the first full-length synthetic cyanophage genome has been constructed. Previously, no other than a 31-kb long DNA sequence from the cyanophage PP genome [29] and a truncated cyanophage A-4L genome of 35-kb in size [23] were generated. The *de novo* synthesis genome of A-4L (~40 kb) in this study was obtained from

yeast but could not be maintained in *E. coli*. Assembly of overlapping cassettes in *E. coli* found that there might be toxic ORFs in the 16 kb region of A-4L genome. ORF11, ORF15, ORF16 and ORF19-23 were screened out to be detrimental to *E. coli* by deleting ORFs between B-fragments. A single wrong base in an essential gene can be the reason of inactivation, while changes in nonessential parts of the genome have no observable effect on viability [35]. Therefore, mutations detected in assemblies capable of transformation into *E. coli* (C1ΔORF11, C2ΔORF15, 16, C3ΔORF19, 20, 21, 22, 23), some of which were located in ORFs encoded hypothetical proteins and others were in untranslated regions, could also be noteworthy (Supplementary Table S3).

We cloned every single ORF in *E. coli* to explore whether toxic gene products are responsible for the inability of transfer. As described in ORF-clone, ORF8, ORF11, ORF16-20 were unclonable. The C-fragments respectively containing ORF15 and ORF21-23 were incapable of transferring, while the four potential toxic ORFs were separately cloned. Another interesting finding was that ORF8, ORF17 and ORF18 were solely unclonable, whereas ORF8 in the C1-fragment containing ORF1-10 can be cloned in *E. coli*, and ORF17 together with ORF18 can simultaneously coexist in *E. coli*. The predicted protein of these ORFs is poorly understood, except that ORF17 encodes endonuclease and ORF21 encodes recombinantion protein. The results showed that only part of the cloning difficulties could be explained by toxic expression products and cloning of some ORFs could be affected by other ORFs (single ORF or multiple ORFs).

We suggested assembling DNA sequences that may be toxic to the host with low copy-number vector. It turned out that all problematic ORFs in this study can be cloned with pLS0. However, the 16 kb fragment of A-4L genome with or without toxic ORFs still failed to be existed in *E. coli* despite using pLS0. In addition, we proposed a cloning strategy
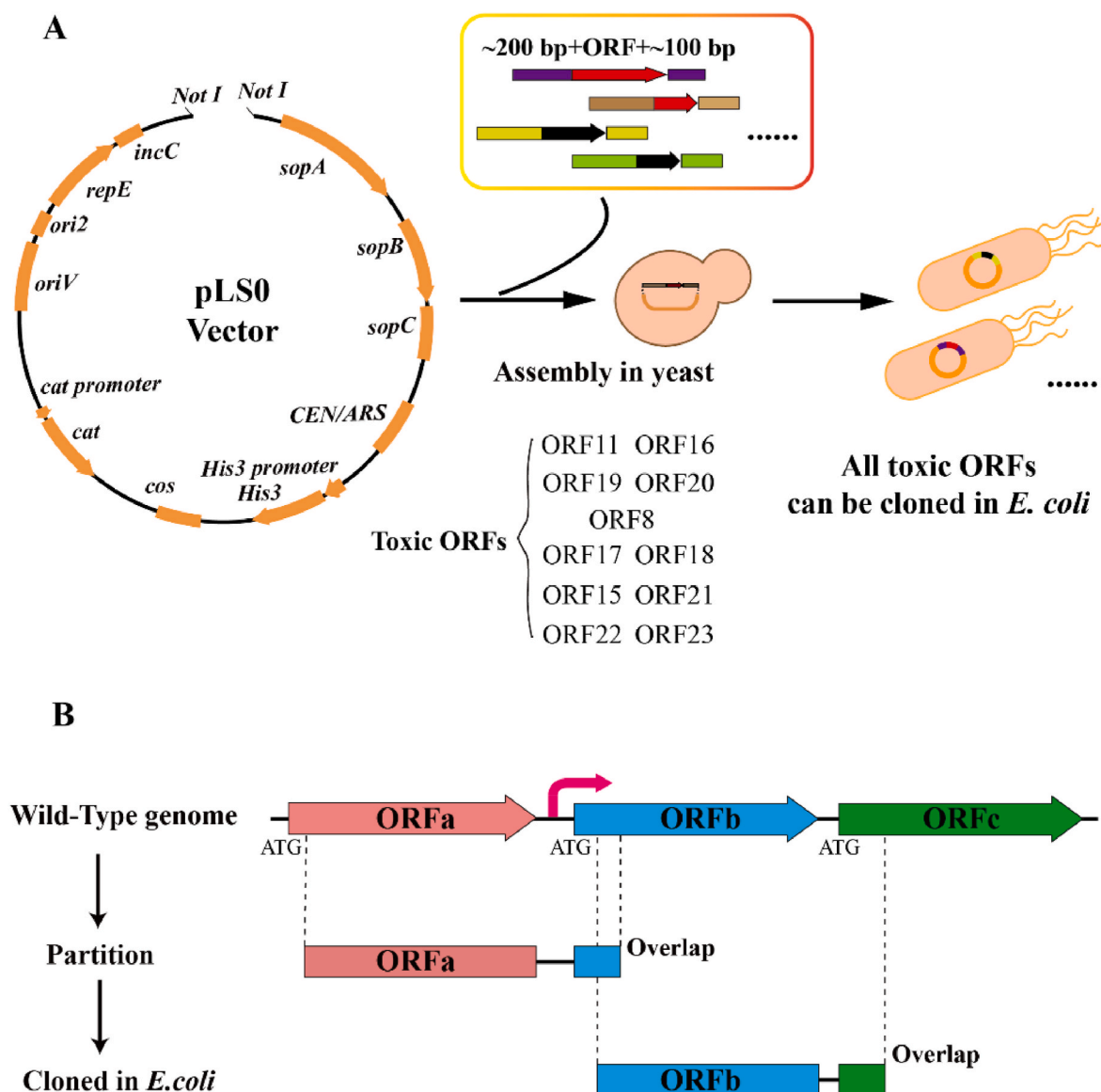
Fig. 3. Two strategies for DNA fragment clone. (A) 11 toxic ORFs were respectively assembled in yeast with the low-copy-number vector prepared from plasmid pLS0 and all can be transformed into *E. coli*. (B) Scheme for the partition of genome sequence. Sequence used in cloning in bacteria abandoned its promoter and the start codon.

and applied it for the generation of c1-c3 fragments. Synthetic A-4L genome assembled in yeast was confirmed by Sanger sequencing. The details were shown in Supplementary Table S3, including several missense mutations, some single-base substitutions and one synonymous mutation. DNA cassettes (A-fragments) were synthesized by the commercial corporation and confirmed to be correct. We speculated that most of the errors occurred during PCR amplification to obtain large fragments and the process of transformation. Conjugation appears to be the most common means to introduce foreign DNA into cyanobacteria [36]. As synthetic A-4L genome cannot be existed in *E. coli*, the resurrection of artificial cyanophage through conjugation is unfeasible. Electroporation [37], protoplast fusion [38] and cell-free system [7] were attempted but all failed in rebooting, which robbed us of verifying the impact of mutations on the synthetic genome.

In conclusion, we reported here the first synthetic entire genome of cyanophage A-4L. Predicted promoters in the 16 kb region of A-4L genome were proved to be capable of recognizing by *E. coli* transcriptional machinery and driving the expression of the report chloramphenicol acetyltransferase gene. In screening for toxic ORFs, ORF11,

ORF15, ORF16 and ORF19-23 were found out in ORF-deletion and ORF8, ORF11 and ORF16-20 were chased down in ORF-clone. Two scenarios that the interaction of expressed genes and the toxicity of gene products were considered to explain the lack of transfer of DNA segments into *E. coli*. Low-copy-number plasmids are good choices for cloning sequences that may be detrimental to the host. The strategy of partitioning genome sequence and cloning ORF without the promoter and start codon can be applied for the bottom-up assembly, providing reference to the construction of synthetic genomes.

CRediT authorship contribution statement

Ting Zhang: Methodology, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Bonan Xu:** Methodology, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Jia Feng:** Methodology, All authors have read and agreed to the final manuscript. **Pingbo Ge:** Methodology, All authors have read and agreed to the final manuscript. **Guorui Li:** Methodology, All authors have read and agreed to the final manuscript. **Jiabao Zhang:**
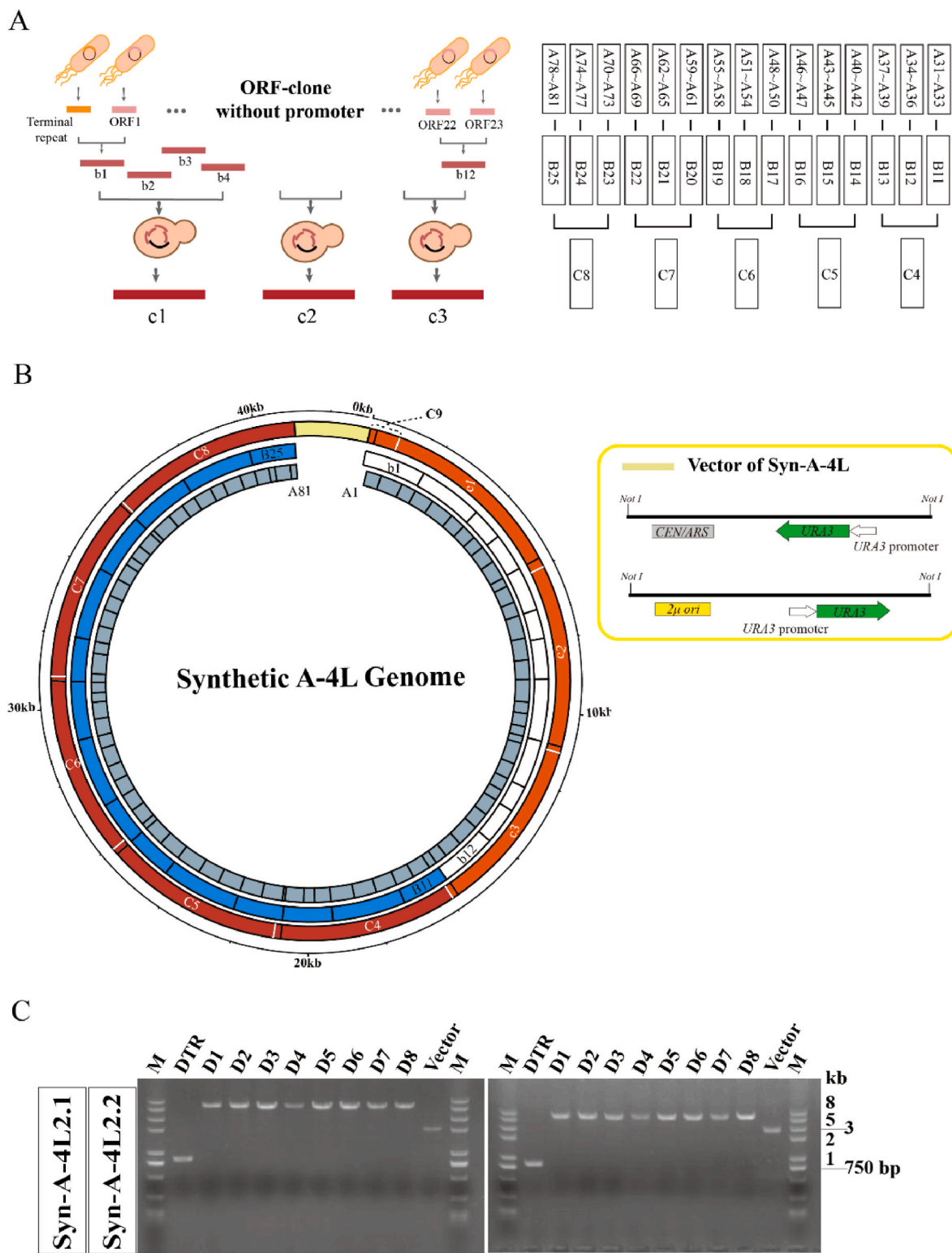
Fig. 4. Design and assembly of full-length synthetic cyanophage A-4L genome. (A) Scheme for the assembly of the synthetic A-4L genome. Assemblies in boxes were obtained in *E. coli*. (B) Circular map of Syn-A-4L. Red parts show C4–C8 fragments with 5 kb in size that existed in *E. coli*. Blue parts show B-fragments of ~2 kb in length. Orange parts show c1-c3 fragments assembled from yeast by the method of partition. White parts show b-fragments assembled by OE-PCR. Grey ring shows 81 A-fragments. Yellow part is the yeast vector of plasmids. (C) Amplicons of all the 9 cyanophage A-4L DNA fragments were present in Syn-A-4L, sized: DTR (Direct terminal repeat): 810 bp; D1: 5680 bp; D2: 5500 bp; D3: 5300 bp; D4: 5400 bp; D5: 5380 bp; D6: 5400 bp; D7: 5578 bp; D8: 5562 bp; Vector of Syn-A-4L2.1: 2016 bp; Vector of Syn-A-4L2.2: 2738 bp.

Methodology, All authors have read and agreed to the final manuscript. **Jianting Zhou:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Jianlan Jiang:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare no competing financial interest.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2022.12.004.

References

[1] Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, et al. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 2008;319(5867):1215–20. https://doi.org/10.1126/science.1151721.

[2] Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, et al. Total synthesis of a functional designer eukaryotic chromosome. Science 2014;344(6179):55–8. https://doi.org/10.1126/science.1249252.

[3] 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. https://doi.org/10.1038/s41586-019-1192-5.

[4] Gibson DG, Smith HO, Hutchison CA, Venter JC, Merryman C. Chemical synthesis of the mouse mitochondrial genome. Nat Methods 2010;7(11):901–3. https://doi.org/10.1038/nmeth.1515.

[5] Koster CC, Postma ED, Knibbe E, Cleij C, Daran-Lapujade P. Synthetic genomics from a yeast perspective. Front Bioeng Biotechnol 2022;10:869486. https://doi.org/10.3389/fbioe.2022.869486.

[6] Coradini ALV, Hull CB, Ehrenreich IM. Building genomes to understand biology. Nat Commun 2020;11(1):6177. https://doi.org/10.1038/s41467-020-19753-2.

[7] Yeom H, Ryu T, Lee AC, Noh J, Lee H, Choi Y, et al. Cell-free bacteriophage genome synthesis using low-cost sequence-verified array-synthesized oligonucleotides. ACS Synth Biol 2020;9(6):1376–84. https://doi.org/10.1021/acssynbio.0c00051.

[8] Mohanraj U, Wan X, Spruit CM, Skurnik M, Pajunen MI. A toxicity screening approach to identify bacteriophage-encoded anti-microbial proteins. Viruses 2019;11(11). https://doi.org/10.3390/v11111057.

[9] Wimmer E, Mueller S, Tumpey TM, Taubenberger JK. Synthetic viruses: a new opportunity to understand and prevent viral disease. Nat Biotechnol 2009;27(12):1163–72. https://doi.org/10.1038/nbt.1593.

[10] Wimmer E, Paul AV. Synthetic poliovirus and other designer viruses: what have we learned from them? Annu Rev Microbiol 2011;65:583–609. https://doi.org/10.1146/annurev-micro-090110-102957.

[11] Zhang Y, Huang L, Wang X, Ding S, Liu Z, Tong Y. Regulation of cyanobacteria population density by cyanophage and its effect on material circulation in water. J Lake Sci 2022;34(2):376–90. https://doi.org/10.18307/2022.0202.

[12] Zhang D, He Y, Gin KY-H. Genomic characterization of a novel freshwater cyanophage reveals a new lineage of cyanopodovirus. Front Microbiol 2021;12:768868. https://doi.org/10.3389/fmicb.2021.768868.

[13] Kaneko T, Nakamura Y, Wolk CP, Kuritz T, Sasamoto S, Watanabe A, et al. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium Anabaena sp. strain PCC 7120. DNA Res 2001;8(5). https://doi.org/10.1093/dnares/8.5.205.

[14] Hu NT, Thiel T, Giddings TH, Wolk CP. New Anabaena and Nostoc cyanophages from sewage settling ponds. Virology 1981;114(1):236–46. https://doi.org/10.1016/0042-6822(81)90269-5.

[15] Ou T, Liao X-Y, Gao X-C, Xu X-D, Zhang Q-Y. Unraveling the genome structure of cyanobacterial podovirus A-4L with long direct terminal repeats. Virus Res 2015;203:4–9. https://doi.org/10.1016/j.virusres.2015.03.012.

[16] Kouprina N, Larionov V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. Nat Protoc 2008;3(3):371–7. https://doi.org/10.1038/nprot.2008.5.

[17] Christianson TW, Sikorski RS, Dante M, Shero JH, Hieter P. Multifunctional yeast high-copy-number shuttle vectors. Gene 1992;110(1):119–22. https://doi.org/10.1016/0378-1119(92)90454-w.

[18] Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics 1989;122(1):19–27. https://doi.org/10.1093/genetics/122.1.19.

[19] Ando H, Lemire S, Pires DP, Lu TK. Engineering modular viral scaffolds for targeted bacterial population editing. Cell Syst 2015;1(3):187–96. https://doi.org/10.1016/j.cels.2015.08.013.

[20] Noskov VN, Koriabine M, Solomon G, Randolph M, Barrett JC, Leem SH, et al. Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning. Nucleic Acids Res 2001;29(6):E32. https://doi.org/10.1093/nar/29.6.e32.

[21] Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc 2007;2(1):31–4. https://doi.org/10.1038/nprot.2007.13.

[22] Xie Z-X, Li B-Z, Mitchell LA, Wu Y, Qi X, Jin Z, et al. "Perfect" designer chromosome V and behavior of a ring derivative. Science 2017;355(6329). https://doi.org/10.1126/science.aaf4704.

[23] Liu S, Feng J, Sun T, Xu B, Zhang J, Li G, et al. The synthesis and assembly of a truncated cyanophage genome and its expression in a heterogenous host. Life 2022;12(8):1234. https://doi.org/10.3390/life12081234.

[24] Thi Nhu Thao T, Labroussaa F, Ebert N, V'Kovski P, Stalder H, Portmann J, et al. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. Nature 2020;582(7813):561–5. https://doi.org/10.1038/s41586-020-2294-9.

[25] Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. Genome-wide experimental determination of barriers to horizontal gene transfer. Science 2007;318(5855):1449–52. https://doi.org/10.1126/science.1147112.

[26] Karas BJ, Tagwerker C, Yonemoto IT, Hutchison CA, Smith HO. Cloning the Acholeplasma laidlawii PG-8A genome in Saccharomyces cerevisiae as a yeast centromeric plasmid. ACS Synth Biol 2012;1(1):22–8. https://doi.org/10.1021/sb200013j.

[27] Itaya M, Tsuge K, Koizumi M, Fujita K. Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. Proc Natl Acad Sci U S A 2005;102(44):15971–6. https://doi.org/10.1073/pnas.0503868102.

[28] Na G, Wolfe A, Ko C, Youn H, Lee Y-M, Byun SJ, et al. A low-copy-number plasmid for retrieval of toxic genes from BACs and generation of conditional targeting constructs. Mol Biotechnol 2013;54(2):504–14. https://doi.org/10.1007/s12033-012-9591-y.

[29] Hou Z, Zhou Z, Wang Z, Xiao G. Assembly of long DNA sequences using a new synthetic Escherichia coli-yeast shuttle vector. Virol Sin 2016;31(2):160–7. https://doi.org/10.1007/s12250-016-3730-8.

[30] Khan SA. Plasmid rolling-circle replication: recent developments. Mol Microbiol 2000;37(3):477–84. https://doi.org/10.1046/j.1365-2958.2000.02001.x.

[31] Khan SA. Rolling-circle replication of bacterial plasmids. Microbiol Mol Biol Rev 1997;61(4):442–55. https://doi.org/10.1128/mmbr.61.4.442-455.1997.

[32] Oliveira PH, Touchon M, Rocha EPC. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res 2014;42(16):10618–31. https://doi.org/10.1093/nar/gku734.

[33] Lyra C, Halme T, Torsti AM, Tenkanen T, Sivonen K. Site-specific restriction endonucleases in cyanobacteria. J Appl Microbiol 2000;89(6):979–91. https://doi.org/10.1046/j.1365-2672.2000.01206.x.

[34] Elhai J, Vepritskiy A, Muro-Pastor AM, Flores E, Wolk CP. Reduction of conjugal transfer efficiency by three restriction activities of Anabaena sp. strain PCC 7120. J Bacteriol 1997;179(6):1998–2005. https://doi.org/10.1128/jb.179.6.1998-2005.1997.

[35] Gibson DG, Glass JI, 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. https://doi.org/10.1126/science.1190719.

[36] Elhai J, Wolk CP. Conjugal transfer of DNA to cyanobacteria. Methods Enzymol 1988;167:747–54. https://doi.org/10.1016/0076-6879(88)67086-8.

[37] Toyomizu M, Suzuki K, Kawata Y, Kojima H, Akiba Y. Effective transformation of the cyanobacterium Spirulina platensis using electroporation. J Appl Phycol 2001;13(3):209–14. https://doi.org/10.1023/a:1011182613761.

[38] Biggins J. Preparation of metabolically active protoplasts and particle preparations from the blue-green alga. Phormidium luridum. Plant Physiol 1967;42(10):1442–6. https://doi.org/10.1104/pp.42.10.1442.