Genome-scale top-down reduction of phages to generate viable minimal phage genomes

Shengjian Yuan
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
https://orcid.org/0000-0003-0410-9591

Juan Shi
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

Jianrong Jiang
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

Yingfei Ma (✉ yingfei.ma@siat.ac.cn)
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
https://orcid.org/0000-0002-2563-5390

Keywords: Tailed-phage Genomes, Phage Synthetic Biology, CRISPR/Cas9, Metagenomic Sequencing, Mutant Phage Populations

DOI: https://doi.org/10.21203/rs.3.rs-633927/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Reduction of tailed-phage genomes to generate viable minimal genome phages is important for expanding our understanding of phage biology, providing insights for phage synthetic biology. Many efforts have been made to minimize living cells, but such work remains a challenge for phages due to the extraordinary genomic diversity and lack of genome-scale editing techniques. Here, we developed a CRISPR/Cas9-based iterative phage genome reduction (CiPGr) approach to detect the nonessential gene set of phages and minimize phage genomes. By CiPGr, inactivated genes accumulated on the phage genome, and mutant progeny with robust growth gradually arose, eventually becoming predominant in the populations. CiPGr was applied to four distinct tailed phages (model phages T7 and T4; wild-type phages seszw and selz), resulting in mutants of these phages with deletion of 8–20% (3.3–33 kbp) sequences, and leading to minimal genomes. Metagenomic sequencing of the mutant phage populations generated showed that 46.7 to 65.4% of genes of these phages were removed. Loss of some genes (39.6%-50%) in the removable gene sets was likely severely detrimental to phage growth. This made the corresponding mutant progenies recede in the populations, leading to the failure of detection of these genes in the genomes of the isolated mutants. In summary, our results for these four distinct tailed phages demonstrated that CiPGr is a generic yet effective approach suitable for use in novel phages without prior knowledge.

Background

Bacteriophages (phages) are the most abundant\textsuperscript{1} and genetically diverse\textsuperscript{2-4} biological entities on earth. For more than a century, phage research has been critical to numerous fundamental biological discoveries\textsuperscript{5,6} and has provided important biotechnical tools in molecular genetics\textsuperscript{7-9}. Recent advances in viral metagenomic sequencing have discovered a great number of phage sequences in the human gut\textsuperscript{2,3} and various other environments\textsuperscript{4}. The majority of these sequences (> 75%) are new\textsuperscript{4,10}, and over 95% belong to tailed phages with a double-stranded (ds) DNA genome\textsuperscript{11,12}. The novelty of tailed phages limits our understanding of phage biology and their roles in the ecosystem\textsuperscript{4}. In addition, phages have been recognized as natural antimicrobial agents for treating bacterial infections\textsuperscript{13,14}. Some major concerns, e.g., extremely high host specificity to a small set of bacterial strains and the rapid emergence of bacterial resistance, remain regarding the use of wild-type (WT) phages in clinical phage therapy\textsuperscript{15}. Therefore, many efforts have been made to genetically engineer phages to overcome these limitations\textsuperscript{16,17}.

Phage synthetic biology can integrate functional genes or gene circuits into phage genomes, bolstering their antibacterial activity and potential in various bioengineering applications\textsuperscript{16,17}. This integration can be challenging because of the limited DNA encapsulation capability of viral particles\textsuperscript{18,19}. The accessory genes of phages may help phages better adapt to a wide range of ecological niches but may be nonessential for phage development under given conditions\textsuperscript{20}. A reduction in these nonessential genes can create some space in phage genomes. This raises questions regarding the extent to which a minimal viable phage genome can be generated for a given phage, especially for novel
phages with plentiful unknown functional genes\textsuperscript{4,10}. The removal of redundant genes from the phage will help advance the understanding of the phage–host interaction pattern and the phage lifestyle. This understanding can further enable the redesign of a phage genome with more functions\textsuperscript{21,22}, paving the way for fundamental biological discoveries.

Phages are among the simplest biological systems. Nontailed phages, e.g., M13\textsuperscript{23} and phi X174\textsuperscript{24,25}, have very compact genomes (<10 kb) encoding a limited number of genes, making it easy to understand the functions of the genes. Tailed phages often have relatively large (14 kbp to 500 kbp)\textsuperscript{13,14} and extraordinarily diverse genomes\textsuperscript{26}, implying that these phages likely employ complex and heterogeneous propagation mechanisms within their host cells. It remains challenging to obtain a viable minimal genome of tailed phages on a large scale. One challenge is the identification of the nonessential genes of a given phage\textsuperscript{27}. There are no efficient approaches to identify the nonessential genes of phages on a genome-wide scale. A summary of the nonessential gene set of phage T7 (25 genes, 41.7\%) was obtained from accumulated knowledge derived from numerous studies conducted in recent decades\textsuperscript{28,29}. These genes are interspersed throughout the genome, hindering efforts for genome reduction of phage T7. Approaches widely used in genome reduction for living cells, i.e., \textit{Escherichia coli}\textsuperscript{30}, yeast\textsuperscript{31}, \textit{Bacillus subtilis}\textsuperscript{32}, and \textit{Mycoplasma mycoides}\textsuperscript{27}, might not be suitable for application in tailed phages on a large scale due to their exclusive characteristics of self-propagation relying strictly on their hosts. Moreover, the bioinformatic method for comparing complete bacterial genomes to identify the minimal gene set for cellular lives\textsuperscript{33} often fails for phage genomes because of their highly divergent nature\textsuperscript{4}. Recently, the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system has been employed for phage genome editing\textsuperscript{34-37}. Nevertheless, the stepwise deletion of many genomic regions is laborious and time consuming, especially if simultaneous screening of mutant phages that are capable of maintaining robust growth is needed. J. Craig Venter and colleagues placed \textit{M. mycoides} JCVI-syn1.0 genes into three categories, namely, essential, quasi-essential, and nonessential genes, by transposon mutagenesis. Quasi-essential genes are not critical for viability but are nevertheless required for robust growth\textsuperscript{27}. This is likely true for phage genes, representing another challenge for screening mutant phages with viable minimal genomes capable of robust growth from numerous heterogeneous mutants\textsuperscript{38}.

To address these challenges, we report a top-down genome reduction approach termed CRISPR/Cas9-based iterative phage genome reduction (CiPGr). In the CiPGr process, host bacterial cells harboring a dual plasmid system consisting of pTarget with heterogeneous spacers\textsuperscript{39} and spCas9 are infected by a specific phage. During phage infection and injection of DNA into the cell, the single-guide RNA (sgRNA, encoded by pTarget) leads the Cas9 nuclease (encoded by spCas9) to cleave the gene at the designed site, and subsequently, the break is repaired via homologous recombination (HR) induced by the HR sequence (pTarget), causing gene deletion or disruption\textsuperscript{34}. If the gene is not essential for phage growth, the mutant phage progeny can multiply without the gene. The mutant phage population is continuously transferred to fresh cells harboring the dual plasmid system, and gene mutations accumulate in the phage genome. The minimal-genome-harboring phage with growth advantages will be predominant in
the population and finally selected. We assessed the use of CiPGr on four diverse tailed phages (T7, T4, seszw, and selz) and obtained removable phage gene sets and phages with viable minimal genomes from the mutant populations, demonstrating the high value of CiPGr in phage synthetic biology research.

Results

Development of the CiPGr approach

The CiPGr approach was performed according to the scheme presented in Fig. 1. First, to engineer the phage genes of interest, the CRISPR spacer and HR templates were designed in a single oligonucleotide (CiPGr cassette). The 200-bp CiPGr cassette harbored a 20 bp spacer sequence, 36 bp promoter sequence, and 100 bp HR template sequence with 50 bp arms (Fig. 1a). We added barcode (20 bp) and primer (24 bp) sequences to the ends. This design ensured that the cassette sequences could be synthesized on a DNA chip. Second, the synthesized CiPGr cassettes for different phages were separated by the primer and the barcodes and integrated into the plasmid pTarget (Fig. 1b). The CiPGr pTarget plasmids were pooled into 8 libraries and transformed into \textit{E. coli} (spCas9) or \textit{Salmonella enterica} (spCas9) cells, generating CiPGr plasmid library cells. Third, the phage genome reduction process was started by adding the 4 WT phages (T7, T4, seszw, and selz) into the culture medium of the corresponding CiPGr plasmid library cells (Fig. 1c). Due to the heterogeneous pTarget plasmids in each library, various mutant phages were generated. These mutant phage populations were continuously transferred to the culture of fresh CiPGr plasmid library cells, and gene deletions or disruptions accumulated in the mutant progenies with a series of transfers. The mutant phages in the populations were expected to have varying fitness in growth, and consequently, mutants with a growth advantage would become gradually predominant in the population after a series of transfers.

We applied CiPGr to four different tailed phages: T7, T4, seszw, and selz (Supplementary Fig. 1). T7 (39.9 kbp) and T4 (168.9 kbp) are model phages that specifically infect \textit{E. coli} strains and belonging to the families \textit{Podoviridae} and \textit{Myoviridae}, respectively. Phages seszw (45.8 kbp) and selz (154.8 kbp) were newly isolated from sewage water samples using \textit{S. enterica} (ST56) as the host (Supplementary Table 1 and Supplementary Table 2). Phage selz was classified into the family \textit{Myoviridae}. According to the phylogenetic tree, genome comparison, and transmission electron microscopy (TEM) features (Supplementary Fig. 2), seszw was classified into the family \textit{Siphoviridae}. We designed two types of CiPGr cassettes per gene to disrupt the gene by deletion of a 100 bp fragment (cassette-100) and deletion of the whole gene (cassette-gene). Cassette-100 and cassette-gene used the same spacer but different HR templates. Our initial attempt to design cassette-100 was to disrupt the phage genes without reducing the genome size too much. We selected eight spacers for each gene due to the likely varying on-target activities of different sgRNAs, and these spacers spanned the entire gene, ensuring successful gene editing. For some genes, fewer spacers were selected owing to the short gene sequences or few predicted spacers. We did not design CiPGr cassettes targeting the annotated structural genes and the essential genes confirmed in previous reports\textsuperscript{17,25} or the likely regulator elements.
In total, we designed and synthesized 5,828 unique CiPGr cassettes targeting 926 genes of the four phages with an average of 5-6 cassettes per gene. Cloning and delivering the pooled CiPGr plasmid libraries into \textit{E. coli} (spCas9) or \textit{S. enterica} (spCas9) yielded $2 \times 10^5$ to $2 \times 10^6$ colonies and achieved 176 to 5865 colonies per individual plasmid in the libraries (Supplementary Table 3). We assessed the CiPGr plasmid libraries by next-generation sequencing (NGS) and observed that 42.38% ($\pm 2.98\%$) of the cloned CiPGr cassette sequences in the libraries shared 100% identity with the designed sequences. These correct sequences showed that each of the eight libraries captured 97.34% ($\pm 3.36\%$) of the designed cassette sequences, covering 99.33% ($\pm 0.98\%$) of the genes of interest, thus enabling the completeness of our genome-wide engineering (Supplementary Table 4 and Supplementary Fig. 3). Some designed CiPGr cassette sequences (218, 3.74%) were not detected in the libraries, likely due to the low sequencing depth or high error rates in DNA synthesis. Next, the four WT phages were used to infect the CiPGr plasmid library cells (Fig. 1c). Serial transfer of the mutant populations of the four phages in the eight libraries was carried out 300 to 600 times, corresponding to over 2000 generations of the phage life cycle (Supplementary Table 5). Taking the genes gp4.7 and gp5.3 of phage T7 as examples, we detected gene deletions by PCR in the mutant phage populations derived by serial transfer and found that the mutant phages became dominant gradually in the populations (Supplementary Fig. 4a), confirming the feasibility of the approach.

**Detection of the removable gene sets of the phages**

We initially assume that the removal of some genes from these phage genomes would not impair the phage life cycle but could likely have different impacts on their growth fitness. To detect the removable gene sets of the phages, we performed metagenomic sequencing of the mutant phage populations generated during serial transfer. In examining the metagenomic sequencing reads, the results obtained from the two types of plasmid libraries (cassette-gene and cassette-100) demonstrated that the number of disrupted and deleted genes as designed by CiPGr was 28 (46.7%) in T7, 52 in seszw (64.2%), 120 in T4 (43.3%), and 96 in selz (47.8%). In addition, the loss of large fragments or point mutations was also detected in the genomes of the mutants of the four phages, making the number of deleted or disrupted genes increase by 0 in T7, 1 in seszw, 19 in T4, and 17 in selz (Fig. 2a). These losses and mutations were not designed in this experiment, and some of them can likely be explained by the CRISPR-escape mutations (CEMs) of phages\textsuperscript{41}. We detected 3 different large fragments (gp0.3-0.6, 4.3-4.7, and 19.5) in phage T7, 3 (gene3-10, 55-59, and 69-71) in seszw, and 7 (genes 20-21, 74-76, 85-90, 95-100, 114-120, 122, and 125-131) in T4 (Fig. 2a and Supplementary Table 7). In particular, the phage selz has 10 different large fragments that were deleted (gene2-7, 23-31, 51-52, 63-66, 78-83, 107-117, 130-133, 176-178, 189-195, and 200), leading to 53 gene deletions, accounting for almost half of the removable gene set of selz (Fig. 2a and Supplementary Table 7). Moreover, 1322 one- or two-base deletions were detected in the metagenomic sequencing reads in the T4 mutant populations, causing the disruption of 54 genes.
For the model phage T7, 25 nonessential genes have been identified in previous studies\textsuperscript{29}. In this study, we detected 80% (20 genes) of these nonessential genes (except five genes, namely, gp0.3, 1.6, 1.8, 6.3, and 6.5), and an additional eight genes were deleted or disrupted (gp1.3, 1.7, 2.5, 3.5, 5.5-5.7, 7.7, 18.5, and 18.7). For another model phage, T4, 16 nonessential genes and 60 auxiliary genes were reported in a previous study\textsuperscript{42}. Here, CiPGr deleted or disrupted 139 genes, including 13 (81.3%) nonessential genes, 32 (53.3%) auxiliary genes, 86 (67.7%) hypothetical genes, and 8 other functional genes (Fig. 2a and Supplementary Table 7). This result demonstrates the feasibility of using CiPGr to detect the phage removable gene set. Overall, the number of genes in the removable gene sets was 28 in T7 (46.7%), 139 in T4 (50%), 53 in seszw (65.4%), and 113 in selz (56.2%) (Supplementary Table 6). Among these detected removable genes, 57.1% for T7, 61.9% for T4, 75.5% for seszw, and 75.2% for selz were annotated as hypothetical genes (Supplementary Table 6). In the phage seszw, the removable gene set accounted for 65.4% of the total genes encoded by the genome, which was the highest among these four phages and indicated that different phages have distinct tolerances to gene deletion.

The removable gene sets of these phages detected by the two types of libraries were mostly consistent (Fig. 2b and Supplementary Table 7). For both the T7 and seszw phages, the removable gene sets exhibited high correlation (r=0.73, T7; r=0.73, seszw) in the frequency of gene deletion or disruption in the populations between the two types of CiPGr plasmid libraries (cassette-100 and cassette-gene) (Fig. 2b). For the T4 and selz phages, the correlations of the removable gene sets between the two types of libraries (r=0.49, T4; r=0.63, selz) were lower than those of T7 and seszw (Fig. 2b). Considering that the phage genome reduction caused by the cassette-100 plasmid library was less than that caused by the cassette-gene plasmid library, these results suggest that the genome size did not affect our experiment. More importantly, we did not observe a correlation between the frequency of gene mutation and the abundance of the corresponding CiPGr cassette in the plasmid library (Supplementary Fig. 5a), implying that the frequency of gene deletion or disruption did not depend on the abundance of the corresponding CiPGr cassette in the plasmid library.

In examining the occurrence of gene deletion or disruption across serial transfers, we observed that the number of removable genes increased by 18.5% for T7, 8.16% for seszw, 57% for selz, and 78% for T4 from the 20\textsuperscript{th} serial transfer to the last. Moreover, in the last ~100 transfers, the number of removable genes increased by only 0% for T7, 6.1% for seszw, 7.6% for selz, and 10.8% for T4 (Fig. 3), implying that the removable gene sets detected by CiPGr were becoming saturated. When taking only the genes that we designed for deletion and disruption into account, we obtained a similar result as above (Supplementary Fig. 4b). The majority of the removable genes of T7 and seszw (81.5% of T7 and 91.84% of seszw) were detected in the early transfers (T20 of T7 and T32 of seszw), while the number of removable genes of T4 and selz gradually increased with the increase in the number of transfers. This may have been caused by the poor cleavage efficiency of the CRISPR system due to DNA modification\textsuperscript{43-45}. Nevertheless, our results demonstrated the versatility of CiPGr in most cases in detecting removable gene sets of different tailed phages.
Phenotype of the mutant phages

To determine the impacts of gene deletions on the phage phenotype, we first determined the infection dynamics of these mutants on a large scale to evaluate their capability to kill the original bacterial cells (E. coli strain MG1655 and S. enterica (ST56) without the dual plasmid system). According to the removable gene sets, here, we mainly focused on the mutant phages of T7 and seszw generated in the cassette-gene plasmid libraries. Eight large- and eight small-plaques (Supplementary Fig. 6a) were picked and purified from double-layer agar cultures of the mutant phages every ten transfers, resulting in 480 mutants for each phage from the populations of over a total of 300 transfers. Infection dynamics indicate that the deletion of different genes can deteriorate the capability of phages to kill their hosts to various degrees. We observed that the performance of the large-plaque mutants (kill-time of T7 mutants: 66.5±10 min; seszw mutants: 270±76 min) was mostly better than that of the small-plaque mutants (T7 mutants: 108±36 min; seszw mutants: 240±76 min) in terms of killing their hosts (Fig. 4a). Moreover, the capability of both large- and small-plaque mutants of T7 and seszw became weaker with the increase in the number of transfers (kill-time of T7 mutants: 40 to 250 min (WT, 50 min); seszw mutants: 340 to 90 min (WT, 320 min) (Fig. 4a). Only a small number of the mutants acquired an increased capability to kill their hosts (17.9% of T7; 1.5% of seszw) compared to their parental phages.

For T4 and selz, we picked and purified eight plaques every 100 transfers (T4, from the 100th transfer to the 600th; selz, from the 100th transfer to the 400th) and performed the infection dynamics assays. T4 mutants showed a similar trend as the T7 mutants (Fig. 4a), in which the capability to kill host cells became weaker (from 80 min to 250 min) with the increase in the number of transfers, but a high ratio (25%) of the mutants showed an increased capability (kill-time less than 115 min) to kill cells compared to WT T4 (Fig. 4a). Interestingly, gene deletion in the phage selz lengthened the kill-time (OD600 < 0.4) from 235 to 345 min (Fig. 4a). These results demonstrate that the impacts of phage genome reduction on phage growth were heterogeneous but mostly detrimental to phage growth.

Correlation of phenotype and genotype of the mutant phages

To link the genotype of the mutant phages with their phenotype, we next sequenced 60 mutant phages selected from the 480 sampled plaques of T7 and seszw. For each sampling point, we selected the mutants with the strongest and weakest host killing capability regarding their killing curves (Supplementary Fig. 6b). When comparing these 60 mutant phage genomes of T7 (Fig. 4b and Supplementary Fig. 7) and seszw (Fig. 4b and Supplementary Fig. 8), we observed that gene deletions accumulated on the genomes of the T7 and seszw mutants with the increase in the number of transfers, as expected. On the genomes of the T7 mutants, 9 genes (2 kbp) were deleted in 1-100 transfers, 1 gene (0.1 kbp) in 100-200 transfers, and 2 genes (1.1 kbp) in 200-300 transfers; on the genomes of the seszw mutants, 13 genes (4.7 kbp) were deleted in 1-100 transfers, 8 genes (1.9 kbp) in 100-200 transfers, and 5 genes (2.5 kbp) in 200-300 transfers. Plotting of the number of serial transfers
versus the numbers of deleted genes (T7: \( R^2 > 0.87 \), seszw: \( R^2 > 0.71 \)) and the mutant phage genome sizes (T7: \( R^2 > 0.75 \), seszw: \( R^2 > 0.66 \)) yielded rarefaction curves (Fig. 4c, d). The T7 mutants with the strongest and weakest host killing capabilities showed a highly similar rate of gene deletion from their genomes with transfers. Thus, we could extrapolate from these observations (Fig. 4c, d) that the number of deleted genes approached a maximum in the T7 genome, likely generating a viable minimal phage genome among the mutants. In the case of seszw, the number of deleted genes continued to grow, implying that more genes can be deleted if more rounds of CiPGr are conducted.

Next, by examining the genomes, we selected five mutant phages with a large plaque of T7 and seszw, and the gene deletions accumulated on the genomes of the mutant phages with the increase in the number of transfers (Supplementary Fig. 9). One-step growth curves of these mutants showed that as more gene deletions occurred in the phage genome, fewer mutant progenies were generated per cell (Supplementary Fig. 10). The number of mutant progenies per infected cell (burst size) decreased from 134 to 3 for T7 mutants and from 44 to 9 for seszw mutants. (Fig. 5a). Moreover, with the deletion of the genes, the relative fitness of T7 mutants decreased from 31 to 11.5 doublings per hour, and that of seszw mutants decreased from 9.3 to 5.5 doublings per hour (Fig. 5b). These results indicate that gene deletions in T7 and seszw can cumulatively inhibit progeny generation.

We also sequenced the genomes of the mutant phages with the strongest host killing capability selected from the population of the transfer of T4 (the 583rd transfer) and selz (the 232nd transfer) based on their host kill curves and determined their one-step growth curves. Compared to the WT parental phages, the burst sizes decreased slightly from 53 to 31 for the T4 mutant and from 35 to 25 for the selz mutant (Fig. 5a). The genomes of the Myoviridae phages T4 and selz are approximately 3-4 times larger than those of the phages T7 and seszw. The T4 and selz genomes likely acquire and carry more accessory genes. The loss of these genes detected in this study may not be sufficient to greatly impact phage growth\(^{31,32}\). Overall, the mutants of these four different phages showed a similar trend: with the increase in the number of transfers, the bacteriostatic ability of the mutants became weaker, and the burst size decreased. In particular, for the phages T7 and seszw, the parental genomes were nearly optimal in killing their hosts. This observation regarding T7 is consistent with a previous result obtained by simulation\(^{46}\).

To detect the effects of phage genome reduction on phage particle size, the particles of the parental T7 and seszw and their mutant progenies (T7 300L and seszw 300L) with minimal genomes and the largest plaque from the population of the 300\(^{th}\) transfer were observed using TEM. The results showed no significant changes in particle size caused by genome reduction for either phage (Supplementary Fig. 11). Next, we integrated 1.5 kbp and 2.5 kbp yeast genome fragments (no known gene functions) into the genome of T7 300L and found that its fitness remained unchanged (Fig. 5b), implying that instead of the alteration in the genome size, the gene loss was likely the important factor affecting phage growth in this study.

In addition, the one-step growth curves of the mutants of the four phages showed that the life cycle duration of the mutants (~17 min for T7, 50 min for seszw, 40 min for T4, and 60 min for selz) remained
unchanged with the decrease in genome size (Fig. 5a). This observation is consistent with the conclusion of a previous study showing that holin controls the length of the infective cycle for lytic phages.\(^{47}\)

In summary, our experiment indicates that the loss of some genes in the removable gene sets of phages can greatly impair phage growth. These genes likely help the phage adapt to host metabolism, transcription, and translation and turn the host into a phage reproduction machine\(^{48,49}\), but most of them encode unknown functions, necessitating further investigation.

**Capability of CiPGr to generate minimal phage genomes**

Here, according to the study of minimal genome cells\(^{21}\), we defined the minimal phage genome for a given tailed phage that must contain only the genes essential for self-reproduction in its host\(^{11}\). We thus can presumably create viable minimal phage genomes by *in silico* deletion of the removable gene sets and the intergenic region between two adjacent removable genes from the phage genomes, resulting in \(T_7\)\(_{\text{min}}\) with 33,816 bp (16.9% genome deleted), \(\text{seszw}\)\(_{\text{min}}\) with 25,679 bp (44% genome deleted), \(T_4\)\(_{\text{min}}\) with 107,544 bp (36.33% genome deleted), and \(\text{selz}\)\(_{\text{min}}\) with 94,575 bp (38.91% genome deleted).

However, examination of the sequenced genomes of the isolated mutant phages showed that the number of deleted or disrupted genes in the mutant phage genomes was approximately half of the number of genes in the removable gene sets (14/28 \(T_7\), 34/53 \(\text{seszw}\), 79/139 \(T_4\), and 64/113 \(\text{selz}\)) (Supplementary Table 6 and Supplementary Table 7). The majority (\(T_7\) 100%, \(T_4\) 92.4%, \(\text{seszw}\) 94.1%, and \(\text{selz}\) 98.4%) of these gene deletions that were detected in the mutant phage genomes showed a relatively high frequency (>5%) across the mutant phage populations in the serial transfers (Fig. 2b and Supplementary Table 7). This suggests that these genes are less important for phage growth than other genes in the removable gene sets. Thus, according to our results, we can tentatively categorize the phage genes into three groups: (i) genes that were absent in the isolated mutant phage genomes and had relatively high frequencies of gene deletion, over 5%, in the mutant phage populations were classified as nonessential. (ii) Other genes in the removable gene sets (frequency <5%) were classified as quasi-essential. The loss of these genes would generate defective phages with fewer growth advantages than mutants lacking nonessential genes. (iii) Genes that were not detected in the removable gene sets were essential for phage growth. Therefore, CiPGr could reduce the phage genome by deleting nonessential genes and, in theory, generate the minimal viable phage genome (Fig. 5b and Supplementary Table 7).

Comparison of the mutant phage genomes, namely, \(T_7\)\(_{300L}\) with 36,937 bp (8.4% of the genome deleted), \(\text{seszw}\)\(_{300L}\) with 36,655 bp (20.1% of the genome deleted), \(T_4\)\(_{583L}\) with 151,790 bp (10.1% of the genome deleted, 41 genes acquired a premature stop codon), and \(\text{selz}\)\(_{232L}\) with 121,813 bp (21.3% of the genome deleted) were determined to be the minimal viable phage genomes in this study (Fig. 5b and Supplementary Table 8), although it can be extrapolated from the above trend that with additional rounds of CiPGr (Fig. 4b,c,d), a smaller viable phage genome with lower fitness could be obtained. This
indicates that in a given condition, a balance needs to be determined between genome reduction and the growth fitness of phages\textsuperscript{18,40}. In conclusion, these minimal genome mutants will nevertheless be helpful in further applications and contribute to the understanding and redesigning of phage genomes.

**Discussion**

To our knowledge, no literature pertaining to the genome reduction of a given tailed phage to generate the viable minimal phage genome has been reported. The results obtained from numerous studies on the minimal genome of free-living cells have expanded the understanding of fundamental biological principles, highlighting the great interest in the minimal genome of tailed phages because their extraordinary genetic diversity implies that tailed phages employ complex and heterogeneous propagation mechanisms within their host cells. By CiPGr, for the first time, we demonstrated the likelihood of compacting the genomes of different tailed phages to generate viable minimal phage genomes. The discrepancy between the removable gene sets detected by metagenomic sequencing of the mutant phage populations and the gene sets absent from the isolated mutant phage genomes suggests that these genes are of different importance for phage growth. The isolated mutant phages showed that genome reduction is mostly detrimental to phage propagation to varying degrees. Further investigations on the removable gene sets and the minimal genome mutant phages capable of maintaining robust growth will likely help us deeply understand the biological principles of tailed phages, paving the way for the redesign of phages for various applications.

The CRISPR/Cas9-assisted HR method has been used to successfully edit several phage genomes\textsuperscript{34-37}. These studies highlighted multiple factors that should be taken into account before attempting to edit phage genomes by this tool, such as efficient spacer selection or extensive base modification\textsuperscript{35,45}. These factors may adversely affect the editing efficiency of the tool. Due to the poor capability of software programs in evaluating and predicting efficient spacers for phage genomes\textsuperscript{50}, we directly designed multiple spacer sequences for each gene of interest to avoid editing failure. Our results illustrate that in the iterative CiPGr process, the majority of the genes of interest are disrupted or deleted from the genomes of the phages T7 (28/36, 77.8%), seszw (52/60, 86.7%), T4 (120/188, 63.8%), and selz (96/179, 53.6%) (Fig. 2.c), showing the feasibility of our design. Nevertheless, editing was not detected on some genes. We conjecture that this may have three different reasons. First, the corresponding designed spacers could be invalid. We calculated the number of designed spacers and cassette abundance in the libraries for the genes that were not edited (Supplementary Fig. 5c). The results showed that these genes had an average of 6, 5, 6, and 6.5 spacers per gene and 589, 154, 597, and 326 cassettes for T7, T4, seszw, and selz, respectively, confirming the occurrence of gene editing (Supplementary Table. 7). Second, base modifications could be present in the designed sites. It has been shown that the Cas9 nuclease can act on the modified T4 genome, although the efficiency is lower than that for the unmodified genome\textsuperscript{35}. We examined the genomes of the other three phages but failed to find any known genes encoding the enzymes related to modification akin to the T4 genes of cytosine hydroxymethylase (g42) and dCTPase (g56). Consistent with previous results\textsuperscript{35}, the gene deletion or disruption efficiency for
T4 was much lower than that for T7 (Fig. 2g and Fig. 3). Nevertheless, the percentage of the removable genes of T4 (50.2%) was not very different from that of the other three tested phages (T7 46.7%, seszw 65.4%, and selz 56.2%). This suggests that the genome modification effect was negligible in the CiPGr process for determining the removable gene sets of phages. Third, and most likely, the genes could be essential to these phages, although we were not able to completely exclude the above two reasons. For example, in this study, we failed to detect the five genes (gp0.3, 1.6, 1.8, 6.3, and 6.5) in the removable gene set of T7, which used to be considered nonessential. This was apparently not caused by the corresponding cassettes in the plasmid libraries (the abundance of the cassettes, log$_2$$>$5, Supplementary Fig. 5c). A likely explanation is that these genes are important for T7 growth in MG1655 (for example, gp0.3 acts to overcome the DNA restriction system of the host), and the loss of these genes can severely impact the growth fitness of T7 within its host, consequently leading to the rapid recession of the corresponding mutants in the populations. Thus, we can extrapolate that the hypothetical genes that are resistant to deletion in our study must play a critical role in phage growth (Supplementary Table 7).

The iterative process ensures that the mutants with a great advantage in robust growth will be predominant among the mutant population and are thus preferentially selected from the numerous heterogeneous mutants for the next round of CiPGr. Therefore, the order of the genes absent from the mutant phage genomes likely suggests that the genes are of different importance to their phage growth, i.e., the earlier a gene was deleted or disrupted, the less important the gene was to phage growth. This was evident in the metagenomic sequencing data of the mutant populations (Fig. 2a) and the genomes of the isolated mutant phages (Supplementary Fig. 7, 8, and 9b), as indicated by the frequencies of 64.86% (±2.92%) of the removable genes of the four phages increasing across the populations as the number of transfers increased (Fig. 2a and Supplementary Table 7).

The process of CiPGr resulted in thousands of generations of mutant phage progeny (Supplementary Table 5). A previous study indicated that Cas9 can cause rapid evolution of phage T4 mutants in a short time. Unexpectedly, a number of point mutations across the whole genomes of the isolated phage mutants were observed. The number of point mutations was counted in the mutant phages of T7 (60) and seszw (58), and we found a linear relationship between the number of transfers and the number of cumulative point mutations in the mutant phage genomes (Supplementary Fig. 12a). These mutations likely cannot be attributed to the ECM only, as 23.5%±8.3% (T7) and 12%±7.7% (seszw) were located in the spacer and PAM regions, and most of them were spontaneous mutations. Moreover, 90%±6% (T7) and 86.7%±6.9% (seszw) of these point mutations were missense mutations, leading to the corresponding amino acid changes. Due to the positive selection exerted by CiPGr on the most fit mutants, these mutations can likely benefit the growth of the mutant phages to some extent, as we showed that the ratio of nonsynonymous to synonymous substitutions (Ka/Ks, including only the genes without designed spacers) of most mutant phages (94.4% of T7 and 76.9% of seszw) was greater than 1 (2.89±1.66 of T7, 1.82±0.85 of seszw) (Supplementary Fig. 12b).
Although a great number of mutant phages with different deleted or disrupted genes were generated in the process of CiPGr, we failed to correlate any mutant genes with the phenotype of the corresponding mutant phages. For example, the genome of seszw 210L had 6 genes more than that of seszw 290L, but their fitness did not differ much (7.12±0.19 for 210L and 7.09±0.40 for 290L). In addition, 300L and 300S of T7 had the same subset of genes deleted. However, 9 unique point mutations were found in the coding region of 300S, and the kill time changed from 80 min (300L) to 190 min (300S). Therefore, the phenotype of a mutant phage is likely the result of cumulative effects of the mutant genes and point mutations on the genome. The exact function of each gene in the removable gene set needs further investigation.

In conclusion, CiPGr is a top-down phage genome reduction approach that successively eliminates nonessential genes to generate a viable minimal phage genome without prior knowledge about the phages. We believe that the minimal genomes generated in this study are not the ultimate minimal genomes for the parental phages, since this approach cannot delete all nonessential regions from the phage genomes, but the information obtained can help us redesign the minimal genome through the bottom-up approach of chemical synthesis. Nevertheless, this study demonstrated the usefulness and convenience of CiPGr as a powerful approach in phage biology studies by determining nonessential gene sets of phages and obtaining viable minimal phage genomes.

**Declarations**

**Materials and Methods** (see Supplementary information)

**Acknowledgments**

This work received support from the Ministry of Science and Technology of China (http://www.most.gov.cn; grant no. 2018YFA0903600). This work was also supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDPB18), the Guangdong Provincial Key Laboratory of Synthetic Genomics (2019B030301006), and the Shenzhen Key Laboratory of Synthetic Genomics (ZDSYS201802061806209). We thank Chenli Liu from Shenzhen Institute of Synthetic Biology for providing *E. coli* strain MG1655, phage T4, and phage T7. We thank J. Feng from the Institute of Microbiology of the Chinese Academy of Sciences for providing *S. enterica* (ST56).

**Author contributions**

S.J.Y. conducted the experiments, performed analyses, and wrote the manuscript. J.S. and J.R.J. conducted the experiments and performed analyses. Y.F.M. supervised the research and wrote the manuscript.
Competing interests

All authors declare no competing interests. S.J.Y., J.S., J.R.J., and Y.F.M. are employees of Shenzhen Institutes of Advanced Technology.

References

1. Potera, C. Phage renaissance: new hope against antibiotic resistance. *Environ Health Perspect* **121**, a48-53, doi:10.1289/ehp.121-a48 (2013).

2. Ma, Y., You, X., Mai, G., Tokuyasu, T. & Liu, C. A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* **6**, 24, doi:10.1186/s40168-018-0410-y (2018).

3. Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. Massive expansion of human gut bacteriophage diversity. *Cell* **184**, 1098-1109.e1099, doi:10.1016/j.cell.2021.01.029 (2021).

4. Paez-Espino, D. *et al.* Uncovering Earth's virome. *Nature* **536**, 425-430, doi:10.1038/nature19094 (2016).

5. Salmond, G. P. C. & Fineran, P. C. A century of the phage: past, present and future. *Nature Reviews Microbiology* **13**, 777-786, doi:10.1038/nrmicro3564 (2015).

6. Hershey, A. D. Inheritance in bacteriophage. *Ann N Y Acad Sci* **54**, 960-962, doi:10.1111/j.1749-6632.1952.tb39971.x (1952).

7. Roberts, J. W. Termination factor for RNA synthesis. *Nature* **224**, 1168-1174, doi:10.1038/2241168a0 (1969).

8. Kaiser, A. D. & Jacob, F. Recombination between related temperate bacteriophages and the genetic control of immunity and prophage localization. *Virology* **4**, 509-521, doi:10.1016/0042-6822(57)90083-1 (1957).

9. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)* **315**, 1709-1712, doi:10.1126/science.1138140 (2007).

10. Gregory, A. C. *et al.* Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **177**, 1109-1123.e1114, doi:https://doi.org/10.1016/j.cell.2019.03.040 (2019).

11. Brussow, H. & Hendrix, R. W. Phage genomics: small is beautiful. *Cell* **108**, 13-16, doi:10.1016/s0092-8674(01)00637-7 (2002).
Roux, S. *et al.* IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Res.* **49**, D764-D775, doi:10.1093/nar/gkaa946 (2021).

Kupferschmidt, K. Resistance fighters. *Science (New York, N.Y.)* **352**, 758-761, doi:10.1126/science.352.6287.758 (2016).

Reardon, S. Phage therapy gets revitalized. *Nature* **510**, 15-16, doi:10.1038/510015a (2014).

Lu, T. K. & Koeris, M. S. The next generation of bacteriophage therapy. *Curr. Opin. Microbiol.* **14**, 524-531, doi:10.1016/j.mib.2011.07.028 (2011).

Lemire, S., Yehl, K. M. & Lu, T. K. Phage-Based Applications in Synthetic Biology. *Annual review of virology* **5**, 453-476, doi:10.1146/annurev-virology-092917-043544 (2018).

Lenneman, B. R., Fernbach, J., Loessner, M. J., Lu, T. K. & Kilcher, S. Enhancing phage therapy through synthetic biology and genome engineering. *Curr. Opin. Biotechnol.* **68**, 151-159, doi:https://doi.org/10.1016/j.copbio.2020.11.003 (2021).

Cui, J., Schlub, T. E. & Holmes, E. C. An allometric relationship between the genome length and virion volume of viruses. *J. Virol.* **88**, 6403-6410, doi:10.1128/jvi.00362-14 (2014).

Hua, J. *et al.* Capsids and Genomes of Jumbo-Sized Bacteriophages Reveal the Evolutionary Reach of the HK97 Fold. *mBio* **8**, e01579-01517, doi:10.1128/mBio.01579-17 (2017).

Edwards, K. F., Steward, G. F. & Schvarcz, C. R. Making sense of virus size and the tradeoffs shaping viral fitness. *24*, 363-373, doi:10.1111/ele.13630 (2021).

Forster, A. C. & Church, G. M. Towards synthesis of a minimal cell. *Mol. Syst. Biol.* **2**, 45, doi:10.1038/msb4100090 (2006).

Zhang, L. Y., Chang, S. H. & Wang, J. How to make a minimal genome for synthetic minimal cell. *Protein & cell* **1**, 427-434, doi:10.1007/s13238-010-0064-4 (2010).

Ghosh, D., Kohli, A. G., Moser, F., Endy, D. & Belcher, A. M. Refactored M13 Bacteriophage as a Platform for Tumor Cell Imaging and Drug Delivery. *ACS synthetic biology* **1**, 576-582, doi:10.1021/sb300052u (2012).

Smith, H. O., Hutchison, C. A., 3rd, Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15440-15445, doi:10.1073/pnas.2237126100 (2003).

Jaschke, P. R., Lieberman, E. K., Rodriguez, J., Sierra, A. & Endy, D. A fully decompressed synthetic bacteriophage oX174 genome assembled and archived in yeast. *Virology* **434**, 278-284, doi:10.1016/j.virol.2012.09.020 (2012).
26 Al-Shayeb, B. et al. Clades of huge phages from across Earth’s ecosystems. *Nature* **578**, 425-431, doi:10.1038/s41586-020-2007-4 (2020).

27 Hutchison, C. A., 3rd et al. Design and synthesis of a minimal bacterial genome. *Science (New York, N.Y.)* **351**, aad6253, doi:10.1126/science.aad6253 (2016).

28 Studier, F. W. Genetic analysis of non-essential bacteriophage T7 genes. *J. Mol. Biol.* **79**, 227-236, doi:10.1016/0022-2836(73)90002-8 (1973).

29 Calendar, R. L. *The Bacteriophages*. 2nd edn, p. 280 (Oxford Univ. Press, 2006).

30 Yu, B. J. et al. Minimization of the Escherichia coli genome using a Tn5-targeted Cre/loxP excision system. *Nat. Biotechnol.* **20**, 1018-1023, doi:10.1038/nbt740 (2002).

31 Luo, Z. et al. Compacting a synthetic yeast chromosome arm. *Genome Biol.* **22**, 5, doi:10.1186/s13059-020-02232-8 (2021).

32 Reuß, D. R. et al. Large-scale reduction of the Bacillus subtilis genome: consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res.* **27**, 289-299, doi:10.1101/gr.215293.116 (2017).

33 Mushegian, A. R. & Koonin, E. V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10268, doi:10.1073/pnas.93.19.10268 (1996).

34 Martel, B. & Moineau, S. CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. *Nucleic Acids Res.* **42**, 9504-9513, doi:10.1093/nar/gku628 (2014).

35 Tao, P., Wu, X., Tang, W. C., Zhu, J. & Rao, V. Engineering of Bacteriophage T4 Genome Using CRISPR-Cas9. *ACS synthetic biology* **6**, 1952-1961, doi:10.1021/acssynbio.7b00179 (2017).

36 Lemay, M. L., Tremblay, D. M. & Moineau, S. Genome Engineering of Virulent Lactococcal Phages Using CRISPR-Cas9. *ACS synthetic biology* **6**, 1351-1358, doi:10.1021/acssynbio.6b00388 (2017).

37 Hatoum-Aslan, A. Phage Genetic Engineering Using CRISPR-Cas Systems. *Viruses* **10**, doi:10.3390/v10060335 (2018).

38 Springman, R., Molineux, I. J., Duong, C., Bull, R. J. & Bull, J. J. Evolutionary stability of a refactored phage genome. *ACS synthetic biology* **1**, 425-430, doi:10.1021/sb300040v (2012).

39 Bao, Z. & HamediRad, M. Genome-scale engineering of Saccharomyces cerevisiae with single-nucleotide precision. *Nat. Biotechnol.* **36**, 505-508, doi:10.1038/nbt.4132 (2018).

40 Chen, L. et al. Characterization and complete genomic analysis of two Salmonella phages, SenALZ1 and SenASZ3, new members of the genus Cba120virus. *Arch. Virol.* **164**, 1475-1478,
41 Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327-336, doi:10.1038/s41586-019-1894-8 (2020).

42 Miller, E. S. *et al.* Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* **67**, 86-156, table of contents, doi:10.1128/mmbr.67.1.86-156.2003 (2003).

43 Carlson, K. & Overvatn, A. Bacteriophage T4 endonucleases II and IV, oppositely affected by dCMP hydroxymethylase activity, have different roles in the degradation and in the RNA polymerase-dependent replication of T4 cytosine-containing DNA. *Genetics* **114**, 669-685 (1986).

44 Bryson, A. L. *et al.* Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9. *mBio* **6**, e00648-00615, doi:10.1128/mBio.00648-15 (2015).

45 Tao, P. & Wu, X. Unexpected evolutionary benefit to phages imparted by bacterial CRISPR-Cas9. *Science advances* **4**, eaar4134, doi:10.1126/sciadv.aar4134 (2018).

46 Endy, D., You, L., Yin, J. & Molineux, I. J. Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5375-5380, doi:10.1073/pnas.090101397 (2000).

47 Wang, I.-N., Smith, D. L. & Young, R. Holins: The Protein Clocks of Bacteriophage Infections. *Annu. Rev. Microbiol.* **54**, 799-825, doi:10.1146/annurev.micro.54.1.799 (2000).

48 Kutter, E. *et al.* From Host to Phage Metabolism: Hot Tales of Phage T4’s Takeover of E. coli. *Viruses* **10**, 387 (2018).

49 Howard-Varona, C. *et al.* Phage-specific metabolic reprogramming of virocells. *The ISME journal* **14**, 881-895, doi:10.1038/s41396-019-0580-z (2020).

50 Shen, J., Zhou, J., Chen, G. Q. & Xiu, Z. L. Efficient Genome Engineering of a Virulent Klebsiella Bacteriophage Using CRISPR-Cas9. *J. Virol.* **92**, doi:10.1128/jvi.00534-18 (2018).

51 Studier, F. W. Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host. *J. Mol. Biol.* **94**, 283-295, doi:https://doi.org/10.1016/0022-2836(75)90083-2 (1975).

52 Tenaillon, O. *et al.* Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* **536**, 165-170, doi:10.1038/nature18959 (2016).

**Figures**
Figure 1

Schematic diagram of the CiPGr approach.
Figure 2

Removable gene sets of the phages.
Figure 3

Number of removable genes in the mutant phage populations.
Figure 4

Phenotype and genotype of the isolated mutant phages.
Figure 5

Characterization of the mutant phages.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.
• supplementary.docx
• supplementarytable2.xlsx
• supplementarytable7.xlsx
• nrreportingsummaryNCOMMS2126086T.pdf