A fully decompressed synthetic bacteriophage øX174 genome assembled and archived in yeast

Paul R. Jaschke, Erica K. Lieberman, Jon Rodriguez, Adrian Sierra, Drew Endy*

Department of Bioengineering, Y2E2-269B, 473 Via Ortega, Stanford, CA 94305-4201, USA

**Abstract**

The 5386 nucleotide bacteriophage øX174 genome has a complicated architecture that encodes 11 gene products via overlapping protein coding sequences spanning multiple reading frames. We designed a 6302 nucleotide synthetic surrogate, øX174.1, that fully separates all primary phage protein coding sequences along with cognate translation control elements. To specify øX174.1.f, a decompressed genome the same length as wild type, we truncated the gene F coding sequence. We synthesized DNA encoding fragments of øX174.1f and used a combination of in vitro- and yeast-based assembly to produce yeast vectors encoding natural or designer bacteriophage genomes. We isolated clonal preparations of yeast plasmid DNA and transfected *E. coli* C strains. We recovered viable øX174 particles containing the øX174.1f genome from *E. coli* C strains that independently express full-length gene F. We expect that yeast can serve as a genomic ‘drydock’ within which to maintain and manipulate clonal lineages of other obligate lytic phage.

© 2012 Elsevier Inc. All rights reserved.

**Introduction**

Small icosahedral virus particles containing single stranded DNA genomes encoding various *Microviridae* have been found in many environments, from seawater to the human gut (Roux *et al.*, 2012). Given its intrinsic natural importance plus accumulated relevance as an experimentally tractable model system, bacteriophage øX174 is now the best-studied member of the *Microviridae* (Hayashi *et al.*, 1988; Fane *et al.*, 2005). øX174 is also widely used to explore topics ranging from adaptive evolution, to molecular self-assembly, molecular systems biology, and virus particle transmission (Wichman and Brown, 2010; Gordon *et al.*, 2012; Markely and Yin, 2007; Julian *et al.*, 2010).

One notable feature of the øX174 genome is the extent to which the coding sequences for seemingly independent gene products overlap (Sanger *et al.*, 1977). For example, gene E, whose encoded product is responsible for host cell lysis, is encoded entirely within an alternate reading frame internal to gene D, which encodes an external scaffold protein (Barrell *et al.*, 1976; Fane *et al.*, 2005). Overall, 16.8% of the øX174 genome encodes for more than one gene, and there are two genomic regions in which three genes overlap (Fig. 1A, top).

For broader context, ~30% of prokaryotic genes overlap with the coding sequences of neighboring genes (Kingsford *et al.*, 2007), and several models have been proposed to explain the existence of overlapping genetic elements. For example, Yokoo and Oshima (1979) hypothesized that the øX174 genome encodes information from an advanced alien civilization, with overlapping genes as the most likely candidate regions in which to find such messages. Researchers have also considered co-regulation of gene expression, genetic coding efficiency as driven by genome replication or packaging constraints, and coupling of evolutionary selection(s) as terrestrial factors leading to overlapping elements (Chirico *et al.*, 2010). However, direct experimental tests demonstrating the spontaneous emergence of overlapping genes and other genetic elements have been difficult to implement given the starting complexity of already evolved natural systems.

Previously, using the bacteriophage T7 genome as a test case we introduced and explored the idea that natural genetic systems might be “refactored” in order to produce engineered surrogates that are easier to study and apply (Chan *et al.*, 2005). Our refactoring process reflected six engineering goals: (1) well-defined boundaries for all genetic elements, (2) no overlapping elements, (3) only one encoded function per element, (4) enable precise and independent manipulation of elements, (5) a genome that is possible to construct, (6) a genome that encodes a viable system. We constructed 12,179 base pairs of the T7.1 genome and performed in vivo- and yeast-based assembly to manipulate clonal lineages of other obligate lytic phage.

*Corresponding author.
E-mail address: endy@stanford.edu (D. Endy).

0042-6822/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2012.09.020
However, given our prior experiences with T7, we first sought to develop a new system and methods that would be better suited for developing and applying ideas related to refactoring to obligate lytic phage. For example, many DNA fragments encoding phage genetic elements cannot be readily propagated using typical *E. coli* cloning strains and vectors. As a second example, we were limited in working on T7.1 by concerns that any individual genetic design change might produce a non-viable genome that could not be physically recovered, propagated, or amplified; since refactoring projects often attempt to change hundreds of nucleotides simultaneously (Temme et al., 2012), non-viable designs that cannot be independently cloned or propagated are difficult if not impossible to work with.

There have been notable advances in the synthesis, assembly, and handling of nucleic acids over the past decade (Carr and Church, 2009). For example, the genomes of several viruses and one microbe have been assembled from chemically synthesized oligonucleotides, with viable systems being recovered in each case (Cello et al., 2002; Smith et al., 2003; Gibson et al., 2010; Yang et al., 2011). Also of note, a bacterial genome was assembled via transformation-associated recombination in yeast, wherein the yeast-based bacterial genome might be maintained and further manipulated without immediate concern for the viability of the microbe resulting from the so-encoded bacterial genome (Gibson et al., 2008).

Thus, we sought to establish methods enabling the synthesis, assembly, and recovery of a bacteriophage genome via yeast. We first choose to work with bacteriophage øX174 given its small circular genome. We were then further motivated to explore if all genetic element overlaps within the øX174 genome might be eliminated and shown to be non-essential, generating the first fully decompressed phage genome.

**Results and discussion**

**Designing a decompressed bacteriophage øX174**

We designed a completely decompressed version of øX174 in order to test the limits of refactoring with a highly overlapping natural genetic system. Our first design, øX174.1, eliminated all gene overlaps while preserving the use of native ribosome binding sites (RBSs) and promoters as much as possible. We referred to the so-changed coding and RBS sequences via a "letter.1" notation; for example, gene A.1 replaces gene A. In total, the process of genome decompression added 909 nucleotides to the wild-type genome.
We next addressed practical constraints arising from the length of DNA that can be physically packaged within a øX174 capsid without impacts to reproductive fitness. Previous work has shown that the length of a øX174 genome, when packaged in vitro, must be kept within a few percent of the 5386 nucleotide wild-type length in order to avoid any significant fitness decrease (Aoyama and Hayashi, 1985). Similar results were shown in vivo (Russell and Muller, 1984). To reduce the decompressed genome length we removed the first 916 nucleotides of gene F, encoding the coat protein (Air et al., 1978). We chose gene F because a plasmid containing a restriction fragment encoding wild-type gene F was able to complement two conditional gene F mutations (Avoort et al., 1983). Additionally, the gene F coding sequence is greater than the total of the combined increases needed to implement the øX174.1 genome design. The truncated gene F version of the decompressed genome was named øX174.1f. To complement øX174.1f when transformed into host cells we designed a medium copy vector expressing gene F under control of a rhamnose-inducible promoter (Fig. S1).

**Building bacteriophage genomes in yeast**

Our workflow to construct bacteriophage genomes in yeast consisted of several steps (Fig. 2). After designing a genome on a computer we obtained all genomic elements to be assembled via DNA synthesis or PCR cloning. We then used PCR amplification to add homologous overhangs to adjacent elements and any short DNA synthesis or PCR cloning. We then used PCR amplification to linearize the yeast-based assembly alone /m use bet pe p e t/ and transformation into suitable host cells. Finally, when our design and assembly process resulted in viable phage, we picked plaques and sequenced genomic DNA from clonal lineages.

**Transplantation of wild-type øX174 to and from yeast**

To test the validity of our assembly workflow and to determine whether bacteriophage genomes propagated in yeast remain viable, we first attempted to capture and then recover a wild-type øX174 genome from yeast. We used PCR to linearize the yeast shuttle vector pRS425 while also adding overhangs homologous to the ends of a linearized wild-type øX174 genome plus AsiSI restriction sites (Fig. 2A). Both DNA fragments were co-digested and ligated and transformed into E. coli K.1. We did not recover plaques from transformations for which the yeast vector backbone was not removed via AsiSI digestion. Overall, of the 16 yeast clones we screened, 11 were positive by PCR for both the phage-vector junctions and 7 of these clones produced phage, giving an overall yeast assembly followed by rededuction process efficiency of ~44%.

**Assembly and recovery of a decompressed bacteriophage øX174**

We next sought to construct and test a øX174.1f genome. We assembled the øX174.1f genome from nine DNA fragments encoding independent phage elements plus the shuttle vector backbone. Because assembly efficiencies decline with increasing numbers of fragments (Liang et al., 2012), we used longer 30 base pair homologous fragment overhangs. Yeast-based assembly alone yielded no functional phage. Therefore, to increase the proportion of complete assemblies, we used an in vitro isothermal assembly process.
were processed as before and transformed into an AAT silent mutation in the 139th codon. Both of these differences 

et al., 1977) we identified two unexpected nucleotide changes: a GAT (Asn) to AAT (Asp) substitution in the 113th codon, and an AAC to AAT silent mutation in the 139th codon. Both of these differences 

and transformed into bacterial host cells. Resultant plaques are sequenced to verify correctness of the assembly. (A) A linearized wild-type øX174 genome assembled with yeast vector pRS425. Flanking AsiSI restriction sites were added to the assembly via PCR primers used to linearize the vector, so that the phage genome can be released from the vector and re-circularized prior to bacterial transfection. (B) Multiplex PCR screening of pRS425(øX174(AsiSI)) assemblies. The left schematic shows amplicon locations for junction testing. Letters (right) note expected amplicon sizes for correct assemblies. (C) Infection of E. coli C with øX174(AsiSI) genomes recovered from yeast.

from our design likely arose from sourcing of the DNA fragment encoding gene G via PCR from our laboratory phage stock (as opposed to de novo DNA synthesis). Both polymorphisms are consistent with observed variation in the øX174 genome, as Smith et al. (2003) also reported these exact two changes from sequencing of laboratory stocks of øX174.

Initial phenotypic characterization of øX174.1f

We checked for first-order phenotypic effects, if any, resulting from the øX174.1f genome architecture. For reference, separation of bacteriophage T7 early genes resulted in a 50% reduction in plaque size and lysis times (Chan et al., 2005). First, we compared wild-type and decompressed phage plaque formation on an E. coli C host expressing various amounts of gene F under the control of a rhamnose inducible promoter (Fig. 3B). Wild-type plaque sizes did not change significantly with varying gene F induction (Table 2). No plaques were observed for øX174.1f when 0 μm rhamnose was added. øX174.1f plaque sizes increased from ~7.5-fold smaller than wild-type at 1 μm rhamnose to virtually identical to wild-type at 1000 μm (Table 2). These results suggest that gene F expression levels are closely coupled to plaque size.

We also assayed the growth properties of øX174.1f in liquid culture. As expected, the decompressed øX174.1f phage was unable to lyse E. coli C lacking the gene F plasmid (Fig. 3C, left panel). We observed that both wild-type øX174 and øX174.1f phage lysed E. coli C (pJ804(Gene F)) with similar efficiency (Fig. 3C, right panel).

Taken together, when sufficient gene F is supplied in trans the first-order phenotype of øX174.1f is indistinguishable from the wild type. In an artificial fashion, our decompression of the øX174 genome has allowed us to travel ‘backwards in time’ regarding Microviridae evolution. For example, previous bioinformatic analysis has suggested

step (Gibson et al., 2009) prior to yeast transformation. We screened resulting yeast clones using two multiplex primer pairs targeting both phage-vector junctions (Fig. 3A). Positive clones were processed as before and transformed into an E. coli C strain expressing gene F under the control of rhamnose induction. Full genome sequencing of phage DNA recovered from a clonal lineage cultured from a plaque confirmed that all designed changes to the wild-type genome were present in øX174.1f. Additionally, in gene G, where our design recapitulated the wild-type sequence (Sanger et al., 1977) we identified two unexpected nucleotide changes: a GAT (Asn) to AAT (Asp) substitution in the 113th codon, and an AAC to AAT silent mutation in the 139th codon. Both of these differences

Table 2

<table>
<thead>
<tr>
<th>Host strain</th>
<th>Phage</th>
<th>Rhamnose (μM)</th>
<th>Plaque size*</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli C</td>
<td>wt øX174</td>
<td>–</td>
<td>6(1)</td>
</tr>
<tr>
<td>E. coli C</td>
<td>øX174(AsiSI)</td>
<td>–</td>
<td>6(1)</td>
</tr>
<tr>
<td>E. coli C</td>
<td>Refactored øX174.1f</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>E. coli C</td>
<td>øX174(p804(Gene F))</td>
<td>0</td>
<td>4.9(0.7)</td>
</tr>
<tr>
<td>E. coli C</td>
<td>øX174</td>
<td>1</td>
<td>4.6(0.6)</td>
</tr>
<tr>
<td>E. coli C</td>
<td>øX174(p804(Gene F))</td>
<td>10</td>
<td>4(1)</td>
</tr>
<tr>
<td>E. coli C</td>
<td>øX174</td>
<td>100</td>
<td>4(1)</td>
</tr>
<tr>
<td>E. coli C</td>
<td>øX174(p804(Gene F))</td>
<td>1000</td>
<td>4.4(0.8)</td>
</tr>
<tr>
<td>E. coli C</td>
<td>øX174(p804(Gene F))</td>
<td>0</td>
<td>n/a</td>
</tr>
<tr>
<td>E. coli C</td>
<td>Refactored øX174.1f</td>
<td>1</td>
<td>0.6(0.1)</td>
</tr>
<tr>
<td>E. coli C</td>
<td>Refactored øX174.1f</td>
<td>10</td>
<td>0.9(0.2)</td>
</tr>
<tr>
<td>E. coli C</td>
<td>Refactored øX174.1f</td>
<td>100</td>
<td>2.2(0.7)</td>
</tr>
<tr>
<td>E. coli C</td>
<td>Refactored øX174.1f</td>
<td>1000</td>
<td>4.1(0.6)</td>
</tr>
</tbody>
</table>

* Plaque sizes in mm ± standard deviation in brackets.
that øX174 genes evolved in a step-wise fashion, whereby gene A first expanded its 3′ end to overlap gene B, followed by overprinting of genes K and E (Pavesi, 2006). Such findings are supported by the non-essential nature of genes A*, K and E (Colasanti and Denhardt, 1987; Gillam et al., 1985; Pollock et al., 1978). Our aggregate finding that no essential information is lost by full øX174 gene decompression may further support the hypothesis that genome compression results from evolutionary pressure to package more genetic information within a limited capsid size (Chirico et al., 2010). Now that the øX174 genetic sequences have been separated from one another further tests of such evolutionary relationships can be pursued via direct experiments.

Materials and methods

Bacterial, yeast and bacteriophage strains and propagation

Wild-type øX174 (ATCC # 13706-B1) and E. coli host (ATCC # 13706) were used throughout. All competent cells were made by the TSS method (Chung et al., 1989). Bacterial cells were propagated with LB and TK broth (Fane and Hayashi, 1991). E. coli (pJ804(Gene F)) was routinely grown with carbenicillin at 50 mg/mL concentration. Yeast strain W303 was used. øX174 was propagated following published methods (Fane and Hayashi, 1991).

DNA design and manipulation and sequencing

øX174 wild-type sequence NC_001422.1 was used throughout as the reference øX174 sequence. All wild-type sequences are numbered as in NC_001422.1 (from the single PstI cleavage site) as originally designated (Sanger et al., 1977). All refactored sequence coordinates are referenced from the first nucleotide of the H/A-intergenic region. A plasmid Editor (ApE, http://biolgylabs.utah.edu/jorgensen/wayned/ape/) was used to visualize and manipulate all sequences. The Primer3Plus web interface (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) (Rozen and Skaltsky, 2000), NCBI Primer-BLAST (Ye et al., 2012), and NetPrimer (Premier Biosoft) were used for designing and analyzing oligonucleotides. Synthetic DNA for genes A.1, B.1, K.1, C.1, D.1, E.1 and the pJ804(Gene F) plasmid were synthesized by DNA2.0, Inc. The decompressed øX174.1f genome was sequenced via Sanger methods by Elim Biopharmaceuticals Inc., and we assembled sequencing reads into a contig using CodonCode Aligner. Average sequencing depth was greater than 3x at all positions.

Building bacteriophage genomes in yeast

To capture the øX174 genome inside yeast we first linearized the yeast/E. coli shuttle vector pRS425 (Christianson et al., 1992) by using PCR amplification with opposite facing primers at

Fig. 3. Assembly and characterization of a the fully decompressed øX174.1f genome. (A) Multiplex PCR screening of pRS425(øX174.1f) assemblies. Schematic on left shows locations of PCR amplicons. Letters on right denote expected amplicon sizes for correct assembly of the first and last phage parts with the yeast vector. (B) Characterization of wild-type øX174 and decompressed øX174.1f plaque size in E. coli C strain expressing gene F under the control of a rhamnose inducible promoter. (C) Characterization of wild-type øX174 and decompressed øX174.1f liquid lysis dynamics. Left panel depicts infection of E. coli C; right panel depict infection of E. coli C(pJ804(Gene F)) in media supplemented with 10 μM rhamnose. Cells were grown in TK broth with 5 mM CaCl2 and 10 mM MgCl2 added just prior to infection. Assays were carried out at 37°C with an initial average MOI of ~5. Uninfected control (red), decompressed øX174.1f (light blue), and wild-type øX174 phage (dark blue). Error bars give standard deviations from four replicates.
coordinates 2957 and 2971 Genbank Accession #U03452.1, leaving out a small part of the plasmid between the primers. The forward primer 1 (Table S1) encodes (5’–3’) 20 nt of øX174 homologous sequence. AsIIS site, and pRS425 homologous sequence. The reverse primer 2 was the same layout as the forward primer, both binding within lacZ. DpnI digestion followed amplification.

Since the native øX174 genome is circular and single-stranded, we purified a dsDNA intermediate form of the genome (called RFI form) from infected cells (Burch et al., 1999). This purified circular dsDNA was linearized by PCR using opposite facing PCR primers (as for vector linearization above) that bind to the Gene H/A-intergenic region boundary. The forward primer 3 binds at the beginning of the H/A-intergenic region while the reverse primer 4 binds at the 3’-end of Gene H. This site of linearization of the øX174 genome was chosen because the region of the genome: (1) does not encode any known genes, (2) is upstream of the Pα-promoter and origin of viral strand replication located in Gene A (Hayashi et al., 1988), and (3) the junction is not known to encode any other function.

Linear øX174 genomes and pRS425 were co-transformed into yeast spheroplasts (strain W303) (Kouprina and Larionov, 2008; Orr-Weaver et al., 1983). Multiplex screening of the assemblies were performed on resuspended yeast colonies directly picked from 3 to 5 days re-grown spheroplast transformations. Yeast colonies were resuspended in 50 μL Zymolase 20T (MP Biometrics) and 0.05 μL beta-mercaptoethanol, then incubated for 1 h at 37°C. Primer sets for multiplex PCR were designed using mulplex (Rachlin et al., 2005). See Supplementary materials for primer sequences 5–8. Qiagen Multiplex PCR Plus kit was used for all screens. For øX174(AsIIS) assembly screening the following PCR conditions were used: 95°C/5:00, 35 cycles of 95°C/0:30; 65°C/1:30; 72°C/1:00, 68°C/10:00, B/hold. 1–2 μL of yeast resuspensions were used for PCR template, while 1 ng of purified øX174 RFI (New England Biolabs) and pRS425 control templates used per 10 μL reaction. We cultured positive clones and purified plasmid DNA using a modified miniprep procedure (Singh and Weil, 2002).

The decompressed øX174.1f genome was built identically to øX174(AsIIS), but with the following exceptions. Yeast vector pRS425 was linearized by SmaI digestion, followed by gel purification. The middle of gene F to the end of the H/A-intergenic region (coordinates 1917–3980) and gene J (coordinates 832–1000) parts of the assembly were amplified from wild-type øX174 RFI DNA generated from ATCC #13706-B1. All other parts were amplified from plasmids synthesized by DNA2.0. All parts were assembled in vitro using the Gibson isothermal assembly method (Gibson et al., 2009), with 10 fmol of each part for 15 min at 50°C, prior to yeast spheroplast transformation. Yeast clones were screened with multiplex primers 9–12. We initially screened yeast clones with denser multiplex PCR amplicons covering all part-part assembly junctions but dropped this expanded testing since we found the results to be difficult to interpret due to the highly repetitive nature of the øX174.1f genome. Out of the 48 yeast clones we screened, 18 were positive by PCR for both phage-vector junctions and only 1 of the prepared and processed vectors resulted in plaque formation, giving an overall success rate of ~2%. We picked plaques, produced liquid culture phage lysates, and sequenced single-stranded genomic DNA from purified øX174.1f phage particles.

øX174 plaque assays

For plaque assays, we added 100 μL phage in BE or 100 μL øX174 transformed E. coli C cells, plus 100 μL untransformed host E. coli C grown in LB broth, to 5 mL 0.7% TK agar at 46°C. This mixture of cells and agar was poured over 15 mL 1.2% TK agar plate and incubated at 37°C for 12 h. For E. coli C(pBJ804(Gene F)), antibiotics and rhamnose were added to the bottom agar in amounts needed for the total volume of bottom plus top agar.

Lysis curves

To obtain lysis curve data, 96-well plate experiments were set up in a Victor3 plate reader at 37°C. A600 measurements were taken every 5 min, and MATLAB was used to process the data. The volume of each 200 μL well was comprised of phage, E. coli C(pBJ804(Gene F)), 5 mM CaCl₂, 10 mM MgCl₂, variable amounts of rhamnose, and TK broth for the remaining volume.

Author contributions

PRJ and DE conceived the study. PRJ, JR, and AS designed øX174.1. PRJ and EKL performed all experiments. PRJ, EKL, and DE wrote the paper.

Acknowledgments

We thank Monica Ortiz, Francois St. Pierre, Diane Liu, and the Endy and Smolke labs. Funding was provided by the NICMS, NSF SynBERC, and Stanford. PRJ was supported by an NSERC Post-doctoral Fellowship.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2012.09.020.

References


Chung, C.T., Niemela, S.L., Miller, R.H., 1989. One-step preparation of competent E. coli cells, plus 100 μL untransformed host E. coli C grown in LB broth, to 5 mL 0.7% TK agar at 46°C. This mixture of cells and agar was poured over 15 mL 1.2% TK agar plate and incubated at 37°C for 12 h. For E. coli C(pBJ804(Gene F)), antibiotics and rhamnose were added to the bottom agar in amounts needed for the total volume of bottom plus top agar.

Lysis curves

To obtain lysis curve data, 96-well plate experiments were set up in a Victor3 plate reader at 37°C. A600 measurements were taken every 5 min, and MATLAB was used to process the data. The volume of each 200 μL well was comprised of phage, E. coli C(pBJ804(Gene F)), 5 mM CaCl₂, 10 mM MgCl₂, variable amounts of rhamnose, and TK broth for the remaining volume.

Author contributions

PRJ and DE conceived the study. PRJ, JR, and AS designed øX174.1. PRJ and EKL performed all experiments. PRJ, EKL, and DE wrote the paper.

Acknowledgments

We thank Monica Ortiz, Francois St. Pierre, Diane Liu, and the Endy and Smolke labs. Funding was provided by the NICMS, NSF SynBERC, and Stanford. PRJ was supported by an NSERC Post-doctoral Fellowship.


