Virology 434 (2012) 278–284

Contents lists available at SciVerse ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

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

ARTICLE INFO

Available online 15 October 2012

Keywords:
Refactoring
øX174
Phage
Yeast
Recombination
Synthetic biology
Genome engineering
Synthetic genomics
DNA assembly
Irreducible complexity
Overprinting
Overlapping genes

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 øX174.1f, 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.

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 recovered viable chimeric phage, albeit with significantly reduced fitness. Subsequent phage evolution experiments by others reported that ~35% of the T7.1 design changes were lost while selecting for mutant phage that recovered to a near wild-type fitness level (Springman et al., 2012). Taken together, these two studies suggest that genome refactoring to remove overlapping elements followed by experiments that attempt to select for "repacking" might be fruitful.

* 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

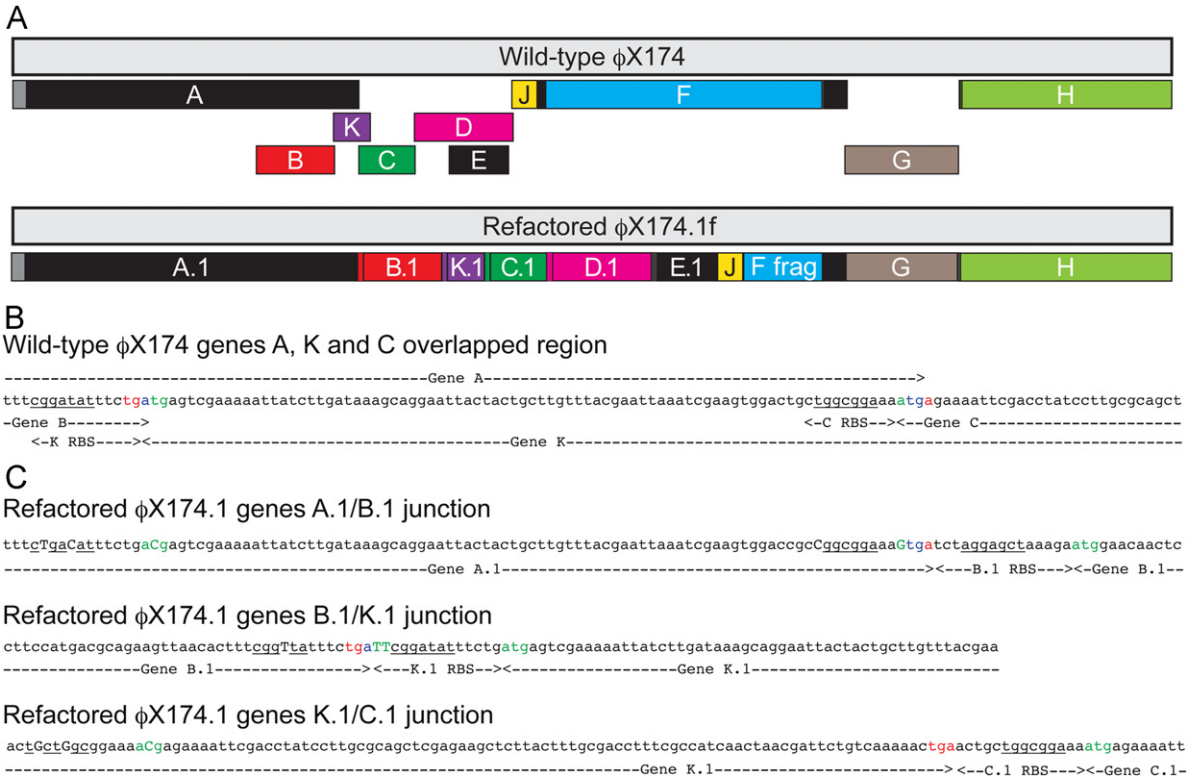

Fig. 1. Design of a fully decompressed bacteriophage øX174 genome. (A) Reading frame comparison of the wild type and decompressed øX174.1f. Wild-type genome reading frames +1 to +3 are given from top to bottom. (B) Expanded view of the wild-type øX174 genes A, K, C junction. (C) Decompression of genes A, K, and C. To separate gene K from gene A (top) the start codon of K was changed from ATG to ACG and its RBS weakened. To decompress gene C from gene A (top) the start codon was changed from ATG to GTG and its RBS weakened. To decompress gene K from gene B (middle) the start codon was changed from ATG to ATT and its RBS weakened. To decompress gene C from gene K (bottom) the start codon was changed from ATG to ACG and its RBS weakened. Start and stop codons are green and red, respectively, while overlapping start/stop codons are blue. RBSs are underlined. Changes made during gene decompression are shown as capital letters.

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 (Fig. 1A, bottom). Our genetic decompression algorithm was similar to that used with T7: (i) overlapping coding sequences were fully separated, with sequence duplication as needed, into physically distinct open reading frames and (ii) vestiges of regulatory elements such as start codons and legacy RBSs were weakened or erased via silent point mutations so as to make translation initiation of legacy reading frames less likely (Fig. 1B and C). We generated synonymous mutations to remove remnant RBSs by reducing the frequency of A and G nucleotides. We used a published RBS strength calculator to confirm that the predicted strengths of so-changed RBSs decreased from wild-type strengths (Salis et al., 2009). Again following Chan et al. (2005), in making synonymous changes we sought to maintain codon usage bias but, when required, choose higher frequency codons.

Next, we used these design rules to separate genes B, K, C from each other and gene A, gene D from gene C, and genes E and J from gene D (Fig. 1, Supplementary Files). We also implemented 22 simultaneous point mutations from the wild-type sequence in order to disrupt legacy start codons and RBSs (Table 1). 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.