

A high-resolution map of bacteriophage ϕ X174 transcription

Dominic Y. Logel, Paul R. Jaschke*

Department of Molecular Sciences, Macquarie University, Sydney, 2109, New South Wales, Australia



ARTICLE INFO

Keywords:

Promoter
Terminator
Overlapped gene
Microviridae
Transcription
Next-generation sequencing
Transcriptomics
Antisense
Escherichia coli
Coliphage
Microbiome
Gut
Rho
RNA-seq

ABSTRACT

Bacteriophage ϕ X174 is a model virus for studies across the fields of structural biology, genetics, gut microbiomes, and synthetic biology, but did not have a high-resolution transcriptome until this work. In this study we used next-generation sequencing to measure the RNA produced from ϕ X174 while infecting its host *E. coli* C. We broadly confirm the past transcriptome model while revealing several interesting deviations from previous knowledge. Additionally, we measure the strength of canonical ϕ X174 promoters and terminators and discover both a putative new promoter that may be activated by heat shock sigma factors, as well as rediscover a controversial Rho-dependent terminator. We also provide evidence for the first antisense transcription observed in the *Microviridae*, identify two promoters that may be involved in generating this transcriptional activity, and discuss possible reasons why this RNA may be produced.

1. Introduction

Bacteriophage ϕ X174 is a member of the *Microviridae* and is an important part of the human gut microbiome (Michel et al., 2010; Reyes et al., 2012). Phage ϕ X174 produces a tailless icosahedral capsid (Fane et al., 2006) and is known to infect the *Escherichia coli* C strain through the capsid attachment to the bacterium's lipopolysaccharide followed by dissociation of the spike protein and conformational changes in the major capsid protein (Hayashi et al., 1988; Sun et al., 2017). The circular single-stranded positive (+) sense DNA genome is transferred across both membranes and the periplasm through an extensible conduit formed by the pilot protein (Sun et al., 2014).

Bacteriophage ϕ X174 has been a useful model for genetics research for over 50 years (Benbow et al., 1971; Hutchison, 1969; Jaschke et al., 2019). The 5386 nucleotide (nt) genome was the first DNA genome sequenced (Sanger et al., 1977) and the number of annotated protein-encoding genes in the ϕ X174 genome has changed over time in response to increasingly sophisticated technologies. Prior to sequencing, ϕ X174 was believed to encode seven genes (Hutchison, 1969), however this was revised upwards several times in response to additional experimental evidence (Benbow et al., 1971; Jeng et al., 1970; Mayol and Sinsheimer, 1970). One main feature of the ϕ X174 genome that puzzled and fascinated phage researchers is the presence of extensive gene

overlaps. Before the genome was sequenced, it was estimated that more sequence length than is covered by the genome was required to encode the observed expressed proteins (Barrell et al., 1976; Benbow et al., 1972). Sequencing solved this mystery by showing that 8 of the 11 genes in ϕ X174 overlap, thereby increasing the available genetic encoding capacity beyond the 5386 nt genome limit (Barrell et al., 1976; Sanger et al., 1977). All ϕ X174 genes were discovered through forward genetic screens with the exception of gene K, which was discovered using amber mutants (Tessman et al., 1980).

The genes in ϕ X174, and other *Bullavirinae* (Godson et al., 1978; Kodaira et al., 1992; Lau and Spencer, 1985), are arranged in functional gene clusters, with four broad types of proteins produced: scaffolding, viral propagation, capsid, and host interaction (Table 1). The functional groups are clustered into two major genomic regions within *Bullavirinae*, the first with no overlapping genes and containing the four capsid proteins: J, F, G, and H. The second genomic region encoding the remaining 7 genes where the genes overlap with at least one neighbouring gene (Fig. 1A).

The ϕ X174 single-stranded positive (+) sense genome that is initially transferred to the *E. coli* cytoplasm upon infection is thought to initiate transcription only after the complementary, or antisense (−) strand is synthesized, producing the double-stranded replicative form genome (Supplementary Fig. S1) (Sinsheimer et al., 1968). ϕ X174

* Corresponding author.

E-mail address: paul.jaschke@mq.edu.au (P.R. Jaschke).

Table 1
Genes of ϕ X174.

Gene	Gene Product Function	Category
A	DNA replication, stages II and III	Viral propagation
A*	Inhibiting host DNA replication and capsid packaging	Host interaction
B	Internal procapsid scaffolding	Scaffolding
C	DNA replication stage II to III switching	Viral propagation
D	External procapsid scaffolding	Scaffolding
E	Host cell lysis	Host interaction
F	Major coat protein	Capsid
G	Major spike protein	Capsid
H	DNA pilot protein, forms a tail for genome transport	Capsid
J	DNA packaging	Capsid
K	Burst size modulation	Viral propagation

transcription is thought to exclusively use the antisense (−) genome strand as template, therefore the mRNA transcript sequence is identical to the viral (+) DNA strand (Supplementary Fig. S1) (Hayashi et al., 1963).

Unlike other coliphages, such as T7, ϕ X174 does not encode an RNA polymerase (RNAP), and as a result, is totally dependent on the *E. coli* host's transcription system. The ϕ X174 RNAP promoter locations were previously identified by a combination of binding assays (Chen et al., 1973) and hybridization mapping (Axelrod, 1976a, b), revealing three major promoters upstream of genes A, B, and D (Fig. 1A and Supplementary Table S1). Additional *in vitro* studies mapped RNAP binding to specific genomic locations using electron microscopy, and determined multiple additional binding sites: the 3' end of gene G, and weak binding at two locations in the 5' end of gene F (Rassart and Spencer, 1978; Rassart et al., 1979). Follow up studies using a super-coiled pKO-1 plasmid and galactose activity to validate the identified promoter sites *in vivo* showed different relationships between *pA*, *pB*, and *pD* depending on the normalization methods employed (Ringuelette and Spencer, 1994; Sorensen et al., 1998) (Supplementary Table S1). Furthermore, sequence context beyond the −35 and −10 regions seemed to have dramatic effects on promoter strengths, indicating that native ϕ X174 transcription was affected by mRNA stability or enhancement sequences which were not transferred to their reporter plasmid (Sorensen et al., 1998). These context-specific effects are also consistent with prior findings (Hayashi and Hayashi, 1985; Hayashi et al., 1989) showing that better *in vivo* measurements of transcription from the ϕ X174 genome itself are needed to address these inconsistencies.

The current understanding of ϕ X174 transcriptional termination is that there are four Rho-independent (intrinsic) terminators located in the intergenic regions between genes J-F, F-G, G-H, and H-A, named T_J , T_F , T_G , and T_H respectively (Fig. 1A) (Hayashi et al., 1981, 1989). However, past work shows a potentially more complex picture where *in vitro* experiments detected a single Rho-dependent terminator in gene A (Axelrod, 1976a, b) and multiple Rho-dependent terminators in gene F (Axelrod, 1976a, b; Kapitza et al., 1979).

Taken together, the contemporary ϕ X174 transcriptome model is that it is driven by three constitutive promoters (*pA*, *pB*, and *pD*) of increasing strength and regulated by a set of four terminators (T_J , T_F , T_G , and T_H) with efficiencies below 100%, resulting in a series of transcripts of various lengths, tiling across the genome and forming six distinct regions with varying total transcript abundance (Fig. 1A). Recently, qPCR measurements of the relative transcript abundance in each of the six regions showed ratios of 1:6:17:11:5:1 for A:B:D:F:G:H transcripts, respectively (Fig. 1B) (Zhao et al., 2012). These results correspond well with existing knowledge of ϕ X174 transcription, however the methods employed were only able to determine broad transcription trends within the phage and did not provide high resolution mapping of the transcriptome.

In this work we use next-generation RNA sequencing to map ϕ X174 transcription for the first time at base-pair resolution. We address a

number of open questions from the ϕ X174 corpus and find evidence for a Rho-dependent terminator, and one new promoter candidate. Surprisingly, we also measure small amounts of antisense RNA that may indicate previously undiscovered transcriptional activity in this phage.

2. Results

To generate a map of the ϕ X174 transcriptome, *Escherichia coli* C122 was infected with wild-type ϕ X174 at a multiplicity of infection (MOI) of 5. After 20 min of growth, corresponding to the point just before lysis, still-intact cells were harvested and RNA isolated from them. The RNA was subsequently used to produce Illumina libraries and sequenced, followed by mapping to the ϕ X174 genome. Read coverage was very deep with every base of the ϕ X174 genome covered by between 900 and 318,502 separate reads. The resulting sequence reads showed a wide range of abundances across the genome that varied in magnitude in a manner that agreed well with previously annotated promoters and terminators (Fig. 1C).

We next compared gene-specific read abundance using the Transcript per Million (TPM) measurement which normalizes against both gene length as well as sequencing depth within each sample (Wagner et al., 2012). The results of this analysis (Fig. 1D) showed strong correspondence overall to previous measurements of ϕ X174 gene expression (Fig. 1B) (Zhao et al., 2012). In contrast to the overall trend, the expression of genes K and E in our measurements were lower than expected by 53% and 55%, while genes J, F, and H were higher than expected by 40%, 19%, and 403%, respectively (Fig. 1D).

2.1. Quantifying promoter and terminator activity

To measure the relative strength of ϕ X174 promoters *pA*, *pB*, and *pD* we compared TPM counts upstream and downstream of each known promoter site. The results of this analysis revealed that *pB* and *pD* increased transcriptional current, the flux of RNA polymerases passing a given point and generating RNA transcripts (Bonnet et al., 2013), by 2.0 and 2.5 fold, respectively (Fig. 2A). Measurement of *pA* activity was complicated by the fact that there is a promoter-terminator overlap within the *pA-T_H* region, where the end of the promoter sequence is immediately followed by the beginning of the terminator stem (Supplementary Fig. S2). We observed a net loss of transcriptional current across the *pA-T_H* feature (Fig. 1C), which we interpret as a result of fewer RNAP initiation events from *pA* than termination events by T_H . Therefore, our calculated strength for *pA* was below 1 (Fig. 2A).

Using the same analysis method, we measured the reduction in transcriptional current across the T_J , T_F , T_G , and T_H terminators. This analysis showed that terminators T_J , T_F , and T_G only blocked 20–40% of transcriptional current flowing through them, whereas T_H was more effective, blocking 70% of transcriptional current (Fig. 2B). The *in silico* RNA folding energies of the terminator sequences do not correlate with measured termination efficiency (Supplementary Fig. S3). Together, our measurements of promoter and terminator strengths update the quantitative description of the cumulative network that controls ϕ X174 transcription (Fig. 2C).

2.2. Determining presence of cryptic promoters and terminators

To determine whether there may be any undiscovered promoter elements within the ϕ X174 genome, we identified areas where read abundances increased without a known promoter upstream. For example, the middle of gene E and the 5'-end of gene F (Fig. 1C). Next we used two computational tools to generate a series of predicted promoter sites and then analysed our data to determine if any of these promoters occurred in locations close to the transcription increases (Supplementary Fig. S4A). These tools were able to predict *pB* and *pD* but not *pA*, likely due to the overlap with T_H (Supplementary Fig. S2). One new promoter candidate (btss49) was identified at the 5' end of

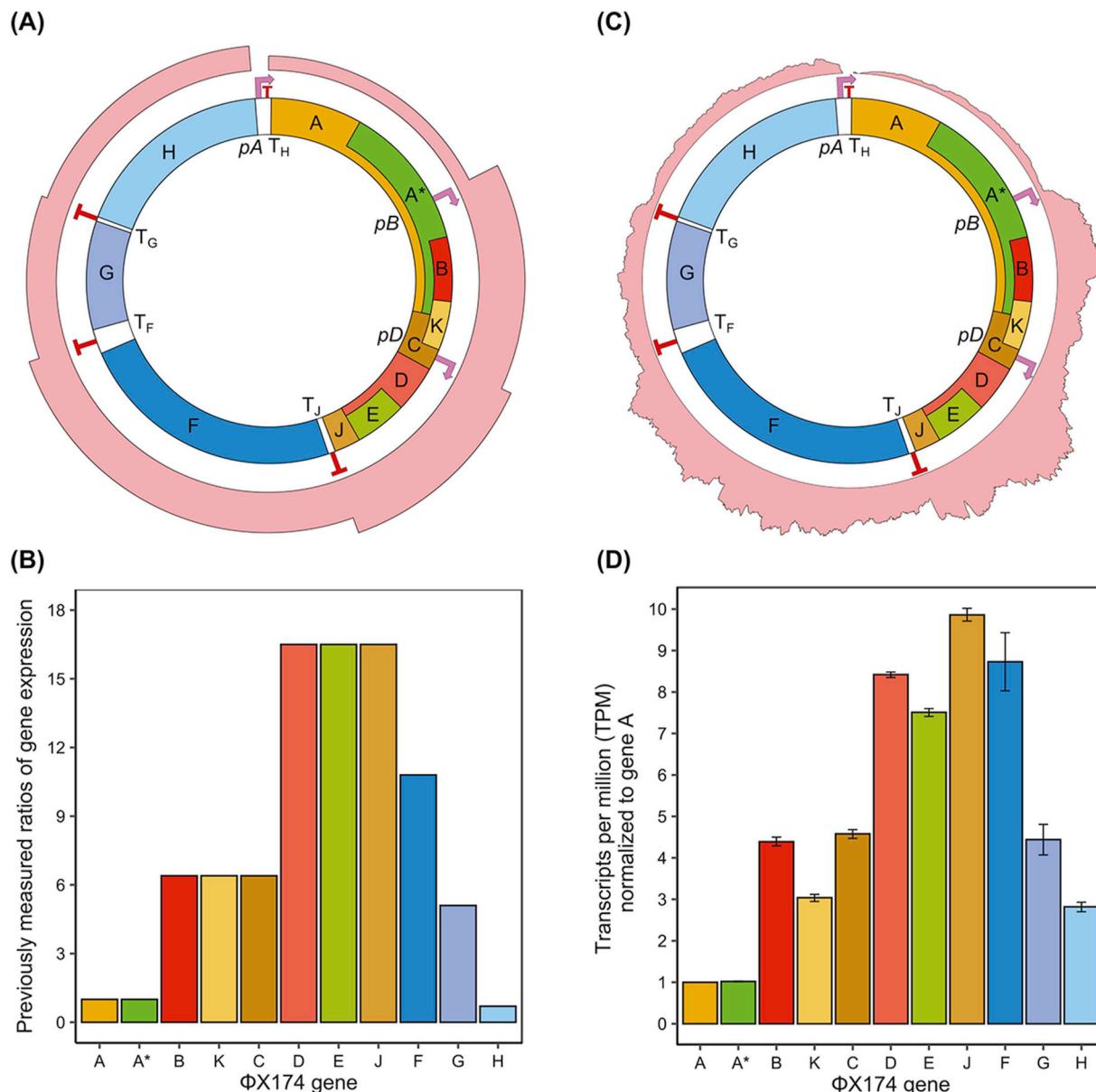


Fig. 1. Models of φX174 transcription. (A) Previous model of φX174 transcription adapted from (Fane et al., 2006). Pink bars represent idealized mRNA quantities with initiation occurring at each known promoter and a step-wise decrease in transcriptional current occurring at each Rho-independent terminator. (B) Previous model of relative transcription of φX174 genes generated from qPCR data (Brown et al., 2010; Zhao et al., 2012). (C) RNA-seq measurements of φX174 infection. Read mapping density across the φX174 genome shown by pink bar. Trimmed reads from biological replicates ($n = 3$) mapped to reference φX174 genome (Genbank no. NC_001422.1) using Geneious Prime software. (D) RNA-seq measurements of relative transcription of φX174 genes. Values are the average of three biological replicates with error bars showing one standard deviation. Transcripts per million calculated from CDS feature annotations using Geneious Prime.

gene C (Fig. 3A) and preceded a noticeable increase in transcription (Fig. 3C), despite a modest measured strength due to local read coverage noise (Fig. 3B). We suggest the name $pD2$ for this potential promoter since it is directly upstream of gene D. The predicted btss49 ($pD2$) promoter was identified as a σ^{24} type, which is generally involved with heat shock, and extracytoplasmic and envelope stress in bacteria (Rhodius et al., 2006). However, a detailed analysis of the potential transcription start site (TSS) revealed that while σ^{24} promoters typically have 5–6 nt between the –10 sequence and the +1 TSS (Rhodius et al., 2006), we see a 23 nt stretch between btss49 and the start of transcriptional increase (Fig. 3C). Several explanations may account for this discrepancy: (1) degradation of the btss49-initiated transcript 5'UTR which shortened the sequence, or (2) the btss49 annotation may be spurious and another non-predicted promoter is at that location. Manual inspection guided by the RNA-seq read increases of the

sequence surrounding btss49 suggest a possible σ^{70} promoter sequence within 25 nucleotides downstream: AAGCTCTTACATTGCGACCTTCG CCATCAACTAACGATT, with potential –35 and –10 sequences conforming to the σ^{70} consensus underlined.

We next looked for cryptic terminators by identifying regions where transcriptional current is reduced without the presence of one of the four canonical terminators. For example, the 5'-end of genes B and E (Fig. 3A). We used three computational tools to predict the presence of Rho-independent terminator locations within the φX174 genome. Despite these tools detecting all known terminators T_J , T_F , T_G , and T_H , they did not find any additional terminators mapping close to a sharp reduction in read abundance (Supplementary Fig. S4B), leading us to conclude that under our experimental conditions there are no Rho-independent terminators within φX174 left to be discovered.

We next looked for Rho-dependent terminators in the φX174

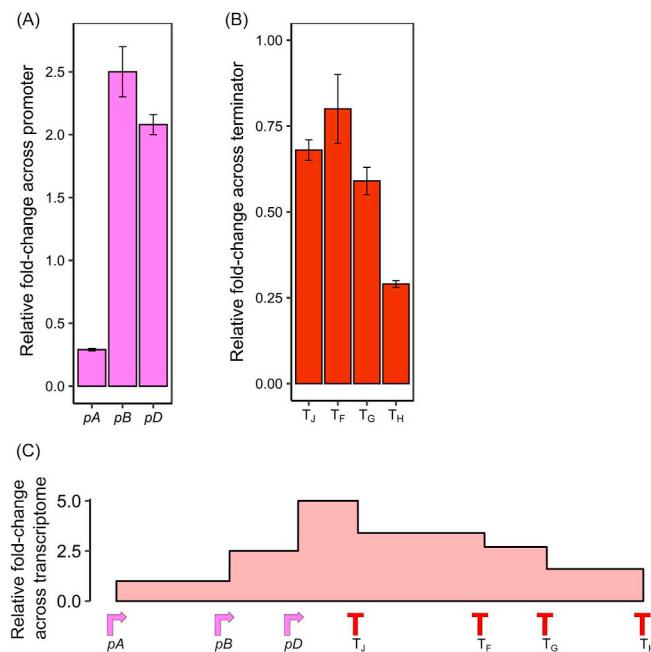


Fig. 2. Measurements of canonical promoter and terminator strengths across φX174 genome using RNA-seq measurements. (A) Relative promoter strengths. (B) Relative terminator strengths. (C) Cumulative transcriptional current from combined promoter and terminator activities. Promoter and terminator strengths were measured by calculating differences in transcripts per million (TPM) read counts 100 bp downstream versus upstream of regulatory feature using Geneious Prime. Values are the average of three biological replicates with error bars showing one standard deviation.

transcriptome. Rho-dependent terminators have been observed before *in vitro* (Axelrod, 1976a; Brendel, 1985; Smith and Sinsheimer, 1976a, b, c) but failed to be substantiated when probed *in vivo* (Hayashi et al., 1981). To identify locations of potential Rho utilisation (RUT) sites, which precede the Rho-terminator region by 10–100 nt (Koslover et al., 2012; Ray-Soni et al., 2016), we used the RhoTermPredict algorithm (Di Salvo et al., 2019). This analysis identified 13 RUT sites across the φX174 genome (Supplementary Fig. S5). Two of the predicted RUT sites, RUT-7 and RUT-8, correspond with previously observed Rho-termination sites within gene F (Kapitza et al., 1979), however, no decrease in transcriptional current across the features was observed under the growth conditions tested (Supplementary Fig. S5).

A third site, RUT-3, corresponded broadly with the previously predicted location of a putative Rho-dependent termination site identified from *in vitro* assays and named T_C (Smith and Sinsheimer, 1976a, b, c). We observed that the transcriptional current decreased significantly within 100 nt downstream of the predicted RUT-3 feature (Fig. 3A and C). We measured the termination strength of RUT-3 and found it is comparable to other terminators within the φX174 genome with a 0.74 fold change occurring across the region containing the RUT site (Fig. 3B) and within 100 nt downstream (Fig. 3C). This analysis leads us to conclude that the RUT-3 Rho-dependent terminator is likely an active transcriptional terminator within φX174, under our laboratory conditions. We have named the terminator T_B because it is closest to the 3'-end of gene B, following current φX174 naming conventions.

2.3. Discovery of antisense (−) RNA transcripts

Previous experiments seeking to identify all φX174 transcripts failed to detect the presence of any antisense RNA messages and predicted that if any messages were to be found in the future, they would constitute < 5% of total reads (Hayashi et al., 1963). The sequencing method we used in this work is designed to enable identification of

which strand each sequencing read originates from. Surprisingly, during our analysis we mapped between 0.21% and 0.38% of the total reads from each biological replicate (Supplementary Table S2) across the entire antisense (−) strand (Fig. 4A).

The antisense reads do not appear to be from replicative form DNA contamination during the sequencing library prep, which would be expected to produce relatively uniform read coverage across the contaminating nucleic acid sequence. Another possible source of contamination could have been the PhiX Control v3 RFI DNA used within Illumina sequencers, which would also have been expected to produce uniform coverage. We specifically did not use this control during our sequencing runs to prevent cross-contamination, but it is possible that small amounts of this control DNA could have been present within the sequencer or reagents. We disproved this idea by comparing the consensus sequence of the mapped antisense (−) sequences to the wild-type φX174 sequence (Genbank No. NC_001422.1) and the PhiX Control v3 sequence, which differ from each other at five positions (587G > A, 833G > A, 2731A > G, 2793C > T, 2811C > T). In all cases our antisense (−) consensus sequence matched NC_001422.1 and not PhiX Control v3, indicating the reads were from our φX174 libraries and not Illumina control contamination. During replication of the complementary (−) strand of the φX174 genome, the primosome synthesizes short 9–14 nt RNA primers using the viral (+) strand as template (Supplementary Fig. S1). These RNA primers would have the same sequences as the detected antisense RNA in this work. We can discount this possibility though because such short sequences would be removed during the read trimming process.

A further possible explanation could be the strand marking protocol allows a small proportion of strand mislabeling and that reads identified as antisense are not. In this case we would expect the antisense reads to perfectly match the abundance pattern of the sense reads, which is the case for the majority of the antisense reads mapping across the φX174 genome in our experiment (Fig. 4A). Interestingly, several areas of large deviation from the sense strand read abundance pattern were also seen within antisense reads mapping to genes G, H, and A (Fig. 4A), which indicated to us that we could also be measuring low abundance authentic antisense reads.

To investigate this antisense RNA phenomena further, we used computational tools to identify any predicted promoter and terminator elements encoded on the φX174 sense (+) strand that could be responsible for driving the production of antisense RNA. We detected 12 potential promoters spread across the entire genome (Supplementary Fig. S6A and Supplementary File S2), but only pR3 and pR12 were located upstream of large increases in transcriptional current (Fig. 4C) that does not mirror the sense read pattern (Fig. 4A).

We also identified 11 potential terminator elements on the antisense (−) strand (Supplementary Fig. S6B), however none were adjacent to transcriptional current declines.

Next, we measured the strength of pR3 and pR12 putative promoters and found transcriptional current to increase across the feature at 2.06 and 1.66-fold, respectively (Fig. 4B). We conclude these newly discovered promoters may be responsible for the small number of antisense reads observed in our experiments that cannot be attributed to sense RNA mislabeled as antisense during library preparation.

3. Discussion

Bacteriophage φX174 has been studied since the dawn of molecular biology, but our understanding of φX174 transcription has not kept pace with recent improvements in sequencing technology. In this work, we sought to update the φX174 transcriptome model by generating a new high-resolution map using next-generation sequencing technology. We used this high-resolution map to identify and quantify the strength of known regulatory elements, identify potentially new regulatory elements, as well as to identify novel antisense transcription and corresponding regulatory elements.

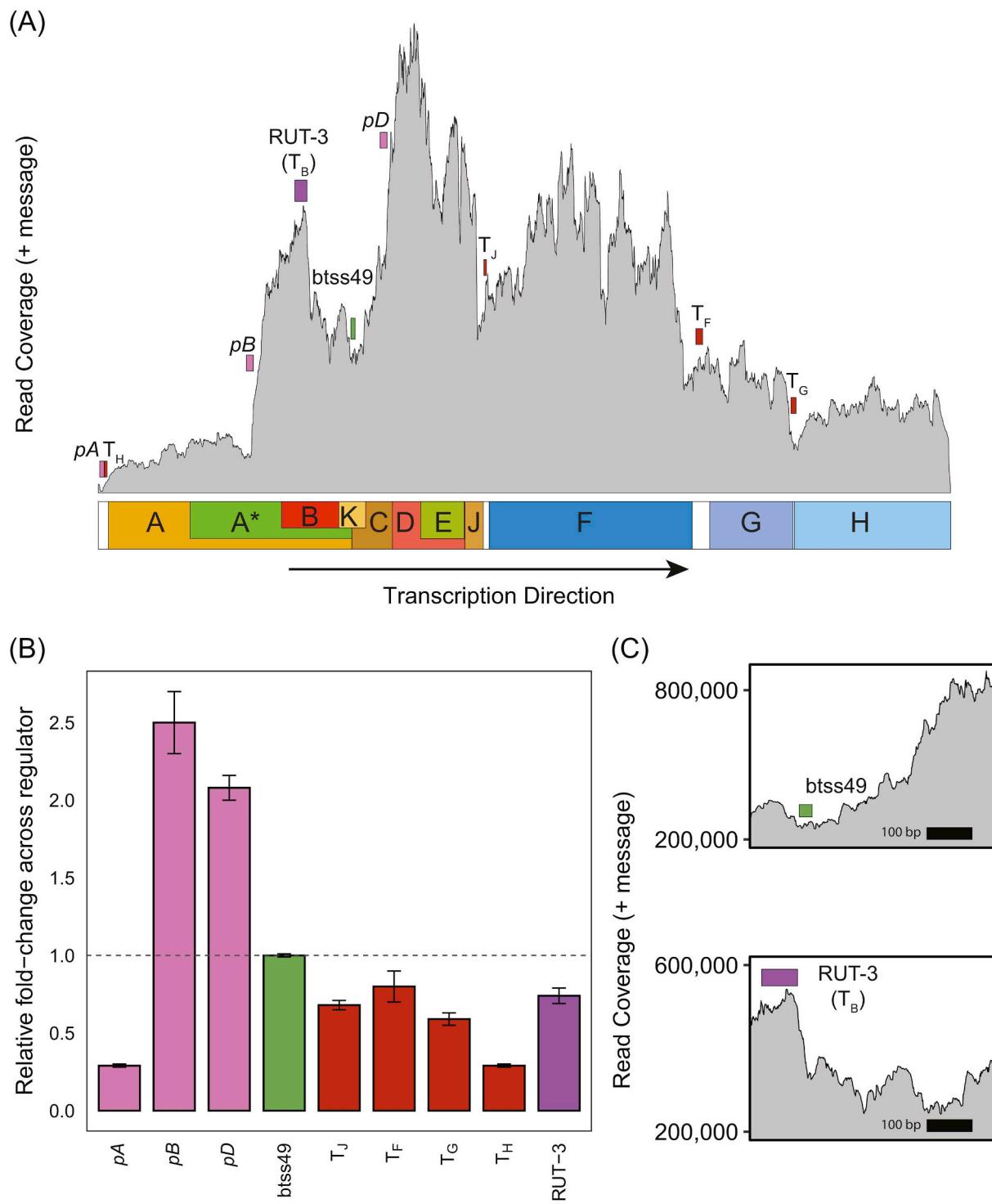


Fig. 3. Computationally identified putative regulatory elements in ϕ X174 genome. (A) Location of canonical and putative regulatory elements overlaid onto average read counts mapped across ϕ X174 transcriptome (grey). Canonical promoters (pink), terminators (red), predicted promoter btss49 ($pD2$) (green), and putative Rho-dependent terminator RUT-3 (T_B) (purple) are annotated to contain total sequence length of feature. (B) Canonical and putative regulator strengths. Strengths were measured by calculating differences in transcripts per million (TPM) read counts 100 bp downstream versus upstream of the feature using Geneious Prime. Values are the average of three biological replicates with error bars showing one standard deviation. (C) Detailed view of predicted btss49 ($pD2$) promoter and RUT-3 (T_B) terminator.

3.1. Quantification of ϕ X174 gene expression

From our analysis we were able to broadly confirm relative ϕ X174 gene expression patterns but observed that RNA abundance decreased significantly within the regions immediately after initiation at pB and pD (Fig. 1C), which was not captured in previous transcription models

(Hayashi et al., 1988; Zhao et al., 2012). Whereas previous models assumed that little to no variation occurred outside of known regulatory events, we were able to detect a distinct decrease in read coverage at the 3' end of gene B, which resulted in a decrease in measured gene K expression compared to literature (Fig. 1B and D). The variation we detected between our TruSeq RNA-seq data and previous TaqMan qPCR

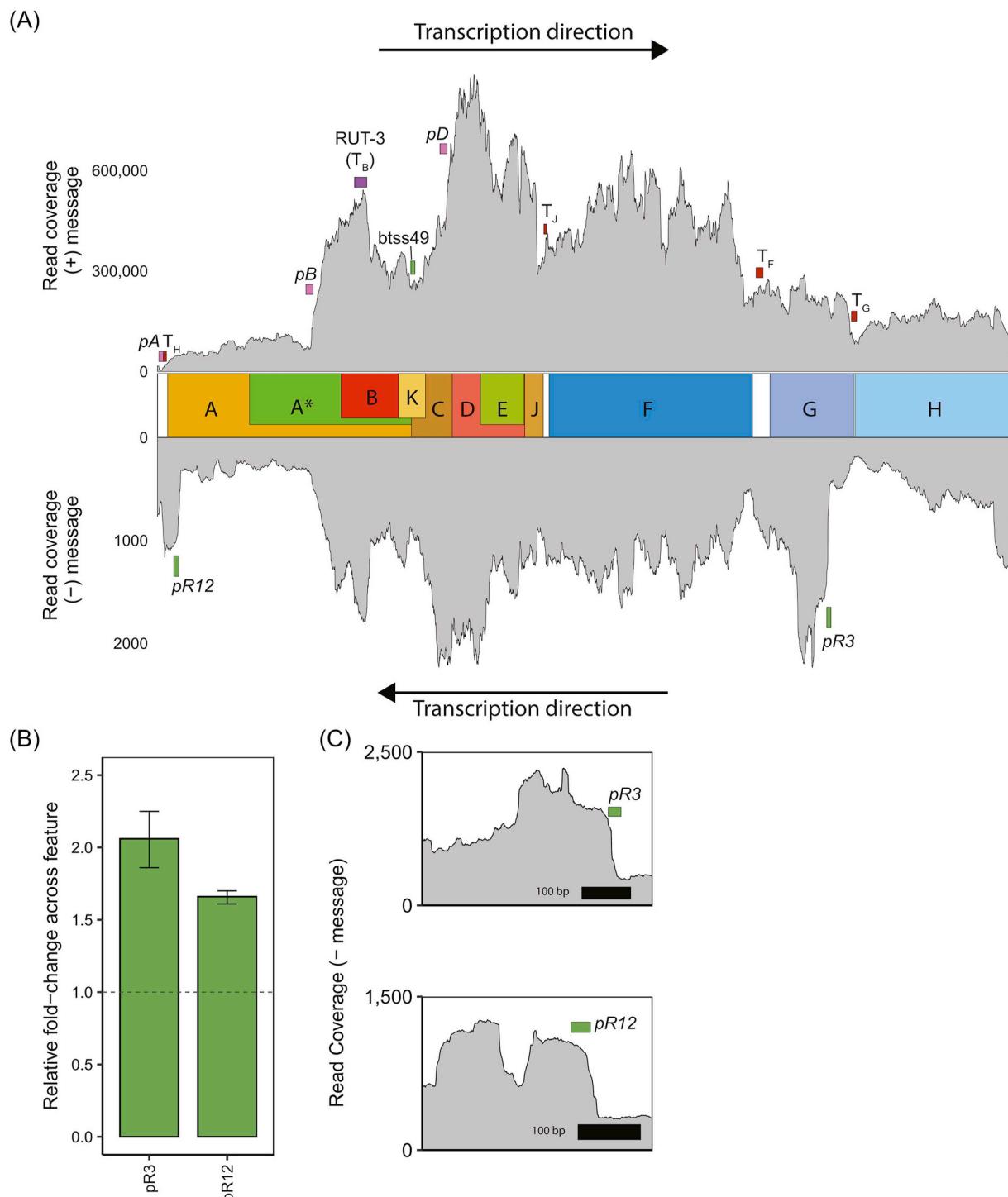


Fig. 4. Comparison of sense (+) and antisense (−) read mapping with putative transcriptional control elements. (A) Sense (top) and antisense (bottom) reads aligned to the ϕ X174 genome. Predicted antisense (−) promoters (green) are shown along with canonical and predicted sense (+) strand regulatory elements. Colored boxes represent the whole annotated sequence predicted from respective software. (B) Putative antisense promoter strengths measured by calculating differences in transcripts per million (TPM) read counts in 100 bp windows downstream versus upstream of each predicted promoter feature. (C) Detailed view of predicted pR3 and pR12 promoters.

data (Zhao et al., 2012) is likely a result of the known differences in sensitivities between the two approaches (Schuierer et al., 2017).

3.2. Measurements of canonical promoter and terminator strengths

We quantified the effect each ϕ X174 promoter has on the transcriptional current and found that pB and pD increases transcriptional

current by 2.5 and 2.0-fold, respectively. The combined pA-T_H element actually has higher termination activity than initiation activity, resulting in a net 0.29-fold transcriptional current change from the 3'-end of gene H to the 5'-end of gene A (Figs. 1C and 2A). The rank order of the promoter strengths agrees with previous work but qPCR measurements showed 6.4, 2.5, and 0.7-fold changes for pB, pD, and pA-T_H, respectively (Zhao et al., 2012).

Compared to other bacterial and phage promoters, ϕ X174 promoters are significantly weaker. For example, multiple common inducible promoters, *pBAD*, *pBAD2*, *pTAC* increase transcriptional current 8, 9, and 10-fold, respectively (Gorochowski et al., 2017), and T7 promoters have been shown to increase 169 fold (Komura et al., 2018), whereas the ϕ X174 promoters are in the range of 2.0–2.5-fold (Fig. 2A).

ϕ X174 transcription is known to generate a range of mRNA species due to terminator read-through (Hayashi et al., 1981). Using our high-resolution sequencing data we measured for the first time, to our knowledge, the termination efficiency of each canonical terminator. This analysis shows that T_J , T_F , and T_G are all relatively poor Rho-independent terminators, reducing the transcriptional current that passes through them by between 0.80 and 0.59-fold (Fig. 2B). These values put these three ϕ X174 terminators in the lowest strength quartile of all *E. coli* terminators, on par with the *endA* and *araBAD* terminators (Chen et al., 2013; Greinert et al., 2012; Jekel and Wackernagel, 1995; Lee et al., 1986). Although measurements of the strength of the T_H terminator were confounded by the overlapping *pA* promoter, the combined element is more than twice as effective at terminating transcriptional current than the next strongest terminator T_G , with the calculated terminator strength placing it in the second lowest strength quartile of *E. coli* terminators (Chen et al., 2013; Greinert et al., 2012; Jekel and Wackernagel, 1995; Lee et al., 1986).

3.3. Antisense (−) transcription in ϕ X174

In this study we observed for the first time antisense (−) RNA within ϕ X174 (Fig. 4 and Supplementary Table S2). Antisense RNA read coverage varied by 25-fold across the genome (Fig. 4A). Within the genes G to A region, putative promoters pR3 and pR12 can be detected upstream of antisense read abundance increases (Fig. 4C). These σ^{70} promoters (Supplementary File S2) appear to be relatively strong as we measure them to increase transcriptional current by 2.0 and 1.7 fold, respectively. This places pR3 at a similar relative strength to *pB*.

Despite our observation of antisense (−) transcripts, the fraction of total reads were less than 1%. Additionally, no past work has ever shown evidence for proteins produced from the antisense (−) transcript (Pollock et al., 1978). Recently, we tried to identify any additional open reading frames ≥ 60 bp initiating with *E. coli*'s most common start codons (Hecht et al., 2017) from both strands of ϕ X174 using mass spectrometry (Jasicke et al., 2019). We could only detect one additional protein produced from the sense (+) mRNA and no proteins from the antisense (−) transcript. While we cannot rule out the possibility of protein produced from the antisense (−) transcript, further investigation using more sensitive mass spectrometry methods such as Parallel Reaction Monitoring (Peterson et al., 2012; Vincent et al., 2019) would be needed to establish the absence of translation from the RNAs detected in the current study. There may be uses, from an evolutionary perspective, for non-translated mRNA production (Yona et al., 2018). For example, many genomes have a spectrum of genic and non-genic transcribed sequences allowing for the development of protein-producing open reading frames over evolutionary timescales from either canonical or non-canonical start codons (Carvunis et al., 2012; Hecht et al., 2017).

3.4. Implications of updated transcriptome model of ϕ X174

The precise step-wise activity of the promoters and terminators of ϕ X174 are important for the phage lifecycle to create different levels of transcripts encoding each gene since the only known control mechanisms are genome concentration, transcription rate, and translation efficiency (Fane et al., 2006; Hayashi et al., 1988).

Our RNA-seq results match well with the previously measured protein abundances with ratios of B:D:J:F:G:H being 60:240:60:60:12 (Fane et al., 2006). In contrast, we see that gene E RNA abundance is lower than expected. Despite falling within a region

of overall high RNA expression, there appears to be some mechanism to reduce transcript levels for gene E (Fig. 3A). This effect may be related to the low concentrations of E protein required for host lysis, aided by a very weak ribosome binding site (Bläsi et al., 1990; Maratea et al., 1985).

The Rho-dependent terminator, T_B , seems to function to reduce gene K and C transcripts (Fig. 3A). While transcript reduction covers both genes K and C, no negative phenotypes have been detected for excess gene K expression, making it likely that gene K reduction is a byproduct and that gene C reduction was the primary selective force at play (Bläsi et al., 1988). Gene C is important in switching from stage II to stage III DNA synthesis which results in increased ssDNA packing into phage capsids as increased protein C concentration outcompetes host single-strand binding protein binding to the ϕ X174 origin of replication (Aoyama and Hayashi, 1986; Doore et al., 2014; Hayashi et al., 1988; Mukai et al., 1979). This balance is apparent in the *Microviridae* α3 where overexpression of protein C results in decreased viral protein synthesis (Doore and Fane, 2016). We suggest that T_B in ϕ X174 functions, in part, to modulate gene C expression.

Early attempts to measure the stability of ϕ X174 transcripts cloned the J-F intergenic region, corresponding to the T_J terminator, directly onto the end of genes B and D and revealed the 3' stabilising effect of the J-F intergenic region (Hayashi and Hayashi, 1985). The *pA*-initiated transcript has long been known to be unstable (Hayashi et al., 1976) and the existence of T_B may finally explain this: the A transcripts terminated by T_B lack the protecting action of the J-F intergenic region.

Subsequent recovery of the transcriptional current after T_B by *pD* is augmented by the putative heat shock promoter btss49 (*pD2*) discovered in this work (Fig. 3A and C). Promoter btss49 (*pD2*) is identified as a σ^{24} type which works in concert with the σ^{32} sigma factor and together are activated under extreme cellular heat shock and capsule and membrane disruption (Raina et al., 1995). Previously, a possible σ^{32} promoter was identified close to the canonical *pB* promoter in experiments subjecting ϕ X174 to heat shock during infection (Zhao et al., 2012). Both σ^{24} and σ^{32} are known to be activated during bacteriophage infection and associated stress response (Bahl et al., 1987; Osterhout et al., 2007). Together, it seems highly likely that ϕ X174 uses some additional promoters that are activated by the stress caused by a phage infection, and corresponding membrane disruption, to drive the formation of phage particles prior to lysis.

This work also shows evidence for the first time of antisense transcription in ϕ X174, which has not been detected in any other member of the *Microviridae*. In fact, the closest related phage that is known to produce antisense RNA are the *Podophage* (Casjens and Grose, 2016; Liao et al., 1987). We have no prior evidence of translation from antisense RNAs in ϕ X174 (Jasicke et al., 2019; Pollock et al., 1978), so it is unlikely that the observed antisense RNA encodes ORFs producing protein. Regulatory antisense RNAs have been observed in lambda phage (Krinke et al., 1991), bacteriophage 186 (Dodd and Egan, 2002), P22 (Liao et al., 1987), and T4 (Belin et al., 1987). Antisense RNA with no known function has been detected in phage φ29 (Mojardín and Salas, 2016) and AR9 (Lavysh et al., 2017).

Although no other ssDNA phages are known to encode antisense RNA, it is relatively common among ssDNA viruses of eukaryotes. For example, Bean golden yellow mosaic virus, Beet curly top virus, Maize streak virus, Tomato pseudo-curly top virus, Porcine circovirus 1, and Bacilladnavirus all produce antisense RNA that contains genes (Briddon et al., 1996; Gilbertson et al., 1991; Lazarowitz, 1988; Niagro et al., 1998; Stanley et al., 1986; Tisza et al., 2020).

A different role for the antisense transcription seen in this work could be attributed to controlling the production of proteins A and A*, which are known to be critically important for ϕ X174 replication and packaging (Roznowski et al., 2020), but also seem to be only needed at very low quantities for successful infections (Jeng et al., 1970). In fact, overexpression of these genes has been seen to be deleterious towards the host, therefore controlling their expression may be crucial to

ensuring early stages of DNA replication are completed successfully before lysis gene E is expressed (Colasanti and Denhardt, 1985). In our work we see several ways that this tightly controlled production of proteins A and A* occurs. In phage known to produce *cis* antisense RNA from transcribing the opposite strand to a gene, one mechanism of gene expression disruption is through transcriptional interference (TI) caused by the head-on collision of the RNAP on each strand (Calen et al., 2004; Shearwin et al., 2005). This could be one possible role for the pR12 promoter located near to *pA* on the opposite strand, where it could play a role in decreasing *pA* transcription initiation, or dislodging any RNAP passing through the T_h site that initiated at *pB* or *pD* (Fig. 4A). Therefore, we speculate that *pA* activity is tightly controlled through several independent mechanisms to suppress the expression of A and A* genes during early infection to ensure host survival until gene E expression and controlled lysis occurs.

4. Materials and methods

4.1. Bacterial strains and ϕ X174 propagation

The wild-type ϕ X174 phage used in this study was created using synthetic DNA to match the original Sanger sequence (Jaschke et al., 2019) (Genbank No. NC_001422.1) and differs from commercially available preparations in at least nine bases (Jaschke et al., 2012; Smith et al., 2003). Overnight cultures of *Escherichia coli* NCTC122 (National Collection of Type Cultures, Public Health England) were grown in 15 mL centrifuge tubes (Westlab, #153–560) in an Infors MT Multitron pro shaker at 250 RCF rotating orbitally at 25 mm diameter at 37 °C containing Lysogeny Broth (LB) Miller media with 2 mM CaCl₂ (phage LB) (Jaschke et al., 2019; Rokytka et al., 2009). Overnight cultures were diluted with fresh media so all had an OD_{600nm} of 1 and then split 1:50 into a 500 mL flat-bottomed Erlenmeyer flask (Duran, #2121644). Cells were grown in triplicate at 250 RCF and 37 °C until a mid-log growth of 0.7 OD_{600nm} was achieved. Cells were centrifuged at 8000 RCF and 4 °C, washed with 5 mL ice-cold HFB-1 starvation buffer (Fane and Hayashi, 1991) (60 mM NH₄Cl, 90 mM NaCl, 100 mM KCl, 1 mM MgSO₄·7H₂O, 1 mM CaCl₂, 100 mM Tris Base, pH 7.4) twice with centrifugation. Cells were finally resuspended with 5 mL ice-cold HFB-2 buffer (60 mM NH₄Cl, 90 mM NaCl, 100 mM KCl, 1 mM MgSO₄·7H₂O, 10 mM MgCl₂·6H₂O, 100 mM Tris base, pH 7.4), and split equally (2.5 mL) into two 15 mL centrifuge tubes on ice. Infected samples of NCTC122 were treated with a ϕ X174 MOI of 5. Infected samples were incubated at 14 °C for 30 min to allow phage attachment without genomic insertion. These were then supplemented with 22.5 mL of 37 °C pre-warmed phage LB media in 250 mL flat bottomed Erlenmeyer flasks (Simax, #1632417106250) and grown at 250 RCF and 37 °C. Bacterial growth parameters have been reported to the best of our knowledge conforming with the MIEO v0.1.0 standard (Hecht et al., 2018).

A 5 mL sample was taken from the triplicates after 20 min and transferred to ice-cold 15 mL falcon tubes and stored on ice. These were then centrifuged at 3500 RCF for 5 min at 4 °C, (Eppendorf, 5430 R), and resuspended in 200 μ L of ice-cold 1x PBS and briefly vortexed to mix the cells. The samples then had a 2:1 ratio of RNAProtect (Qiagen: #76506) added to them and centrifuged at 5000 RCF for 10 min, before the supernatant was removed and the samples stored at –80 °C.

4.2. In silico genetic element prediction

Promoter prediction was performed via web analysis tools BPROM (Salamov and Solovyev, 2011) and bTSSfinder (Shahmuradov et al., 2017) using their default settings and full length sequences for ϕ X174 (Genbank No. NC_001422.1).

Previous Rho-independent terminator sequences from literature (Brendel, 1985; Godson et al., 1978; Hayashi et al., 1981; Otsuka and Kunisawa, 1982) were analysed with RNA folding simulations using the NUPACK web-server with default settings (Zadeh et al., 2011). Lowest-

energy structures were reported.

Potential Rho-independent terminators were identified using the following software packages: FindTerm (Salamov and Solovyev, 2011), iTerm-PseKNC (Feng et al., 2018), and RibEx: Riboswitch Explorer (Abreu-Goodger and Merino, 2005). Data was analysed via 300 nt windows with 50 nt overlaps between windows. The iTerm-PseKNC and RibEx packages were used according to default settings, while FindTerm was modified to set the Energy threshold value to –12.

Potential Rho-dependent terminators were identified by analysing the whole ϕ X174 genome (Genbank No. NC_001422.1) with the RhoTermPredict python script (Di Salvo et al., 2019). All predicted sequences are collected in Supplementary File S2.

4.3. Next-generation sequencing sample preparation

RNA was purified from 5 mL *E. coli* culture following infection using the RNeasy Mini kit (Qiagen: #74106) according to the manufacturer's instructions, with the optional DNase on-column digestion step (Qiagen: #79254). RNA concentrations were measured with the QUBIT RNA assay (Life Technologies: #Q32852). Sequencing library preparation was performed by Macrogen Inc (S. Korea). The rRNA from samples was depleted with a Ribo-Zero Kit (Illumina) and the RNA library generated with a TruSeq Stranded mRNA kit for microbes. Sequencing was carried out with an Illumina HiSeq 2500 instrument in 2x100bp mode. PhiX Control Library was not used during the sequencing run to avoid cross-contamination.

4.4. Sequence analysis

Raw reads were trimmed using Geneious Prime's (Version, 2019-1-3) implementation of the BBDuk trimmer (Bushnell, 2019) using the following command: java -ea -Xmx6000 m -cp ... /current jgi. BBDuk2 ktrimright = t k = 27 hdist = 1 edist = 0 ref = adapters.fa qtrim = rl trimq = 30 minlength = 30 qin = 33 in = input1. fastq out = output1. fastq. Following trimming, sets of paired reads from three biological replicates were mapped to the ϕ X174 reference sequence (GenBank No. NC_001422.1) using the Geneious assembler (Version, 2019-1-3) with the following settings: medium sensitivity, iterate: up to 5 times, minimum mapping quality: 30, and map multiple best matches: to none.

Following assembly, coding sequence (CDS) feature annotation expression levels were calculated using the 'Calculate Expression Levels' function of Geneious Prime with default settings with Ambiguously Mapped Reads: Count as Partial Matches.

Change in transcriptional current across promoter and terminator sequences were calculated by creating 100 bp CDS feature annotations up and downstream of the promoter/terminator followed by applying the 'Calculate Expression Levels' function (Geneious Prime) to those features using the Ambiguously Mapped Reads: Count as Multiple Full Matches setting.

Antisense (–) RNA was discovered by mapping trimmed 'Read 1' and 'Read 2' reads to the ϕ X174 reference sequence. When Read 1 maps in the Forward direction or Read 2 maps in the Reverse direction, they are expected to be the identical sequence to the antisense (–) transcript (Illumina TruSeq Stranded mRNA Reference Guide (1000000040498 v00)). Between 0.21% and 0.38% of Reads 1 and 2 mapped in this way across the three biological replicates (Supplementary Table S2). Further analysis, such as promoter and terminator strength measurements were performed on this subset of the reads. Averaged reads for both strands (Supplementary File S3) were aligned to ϕ X174 genome sequence (Supplementary File S4).

FUNDING

PRJ was supported by the Molecular Sciences Department, Faculty of Science & Engineering, and the Deputy Vice-Chancellor (Research) of

Macquarie University. DYL is a recipient of the Macquarie University Research Excellence PhD scholarship (MQRES) and CSIRO PhD Scholarship Program in Synthetic Biology (Synthetic Biology Future Science Platform).

CRediT authorship contribution statement

Dominic Y. Logel: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Paul R. Jaszche:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare no competing conflicts of interest.

Acknowledgements

We recognize that the intellectual and physical labour of this research was conducted on the traditional lands of the Wattamattagal clan of the Darug nation. We thank Varsha Naidu, Liam Elbourne, and Sasha Tetu for helpful discussions and feedback.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.virol.2020.05.008>.

References

- Abreu-Goodger, C., Merino, E., 2005. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res.* 33, W690–W692.
- Aoyama, A., Hayashi, M., 1986. Synthesis of bacteriophage phi X174 in vitro: mechanism of switch from DNA replication to DNA packaging. *Cell* 47, 99–106.
- Axelrod, N., 1976a. Transcription of bacteriophage phi-X174 in vitro: selective initiation with oligonucleotides. *J. Mol. Biol.* 108, 753–770.
- Axelrod, N., 1976b. Transcription of bacteriophage phiX174 in vitro: analysis with restriction enzymes. *J. Mol. Biol.* 108, 771–779.
- Bahl, H., Echols, H., Straus, D.B., Court, D., Crowl, R., Georgopoulos, C.P., 1987. Induction of the heat shock response of *E. coli* through stabilization of sigma 32 by the phage lambda cII protein. *Genes Dev.* 1, 57–64.
- Barrell, B.G., Air, G.M., Hutchison 3rd, C.A., 1976. Overlapping genes in bacteriophage phiX174. *Nature* 264, 34–41.
- Belin, D., Mudd, E.A., Prentki, P., Yi-Yi, Y., Krisch, H.M., 1987. Sense and antisense transcription of bacteriophage T4 gene 32: processing and stability of the mRNAs. *J. Mol. Biol.* 194, 231–243.
- Benbow, R.M., Hutchison, C.A., Fabricant, J.D., Sinsheimer, R.L., 1971. Genetic map of bacteriophage phiX174. *J. Virol.* 7, 549–558.
- Benbow, R.M., Mayol, R.F., Picchi, J.C., Sinsheimer, R.L., 1972. Direction of translation and size of bacteriophage phiX174 cistrons. *J. Virol.* 10, 99–114.
- Bläsi, U., Nam, K., Lubitz, W., Young, R., 1990. Translational efficiency of phi X174 lysis gene E is unaffected by upstream translation of the overlapping gene D reading frame. *J. Bacteriol.* 172, 5617–5623.
- Bläsi, U., Young, R., Lubitz, W., 1988. Evaluation of the interaction of phi X174 gene products E and K in E-mediated lysis of *Escherichia coli*. *J. Virol.* 62, 4362–4364.
- Bonnet, J., Yin, P., Ortiz, M.E., Subsoontorn, P., Endy, D., 2013. Amplifying genetic logic gates. *Science* 340, 599–603.
- Brendel, V., 1985. Mapping of transcription terminators of bacteriophages phi X174 and G4 by sequence analysis. *J. Virol.* 53, 340–342.
- Briddon, R.W., Bedford, I.D., Tsai, J.H., Markham, P.G., 1996. Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly top virus, suggests a recombinant origin. *Virology* 219, 387–394.
- Brown, C.J., Zhao, L., Evans, K.J., Ally, D., Stancik, A.D., 2010. Positive selection at high temperature reduces gene transcription in the bacteriophage ϕ X174. *BMC Evol. Biol.* 10, 378.
- Bushnell, B., 2019. BBMap.
- Callen, B.P., Shearwin, K.E., Egan, J.B., 2004. Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol. Cell* 14, 647–656.
- Carvunis, A.R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charlotteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B., Brar, G.A., Weissman, J.S., Regev, A., Thierry-Mieg, N., Cusick, M.E., Vidal, M., 2012. Proto-genes and de novo gene birth. *Nature* 487, 370–374.
- Casjens, S.R., Grose, J.H., 2016. Contributions of P2- and P22-like prophages to understanding the enormous diversity and abundance of tailed bacteriophages. *Virology* 496, 255–276.
- Chen, C.-Y., Hutchison, C.A., Edgel, M.H., 1973. Isolation and genetic localization of three ϕ X174 promoter regions. *Nat. New Biol.* 243, 233–236.
- Chen, Y.-J., Liu, P., Nielsen, A.A.K., Brophy, J.A.N., Clancy, K., Peterson, T., Voigt, C.A., 2013. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* 10, 659–664.
- Colasanti, J., Denhardt, D.T., 1985. Expression of the cloned bacteriophage phi X174 A* gene in *Escherichia coli* inhibits DNA replication and cell division. *J. Virol.* 53, 807–813.
- Di Salvo, M., Puccio, S., Peano, C., Lacour, S., Alifano, P., 2019. RhoTermPredict: an algorithm for predicting Rho-dependent transcription terminators based on *Escherichia coli*, *Bacillus subtilis* and *Salmonella enterica* databases. *BMC Bioinf.* 20, 117.
- Dodd, I.B., Egan, J.B., 2002. Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch. *Mol. Microbiol.* 45, 697–710.
- Doore, S.M., Baird, C.D., Roznowski, A.P., Fane, B.A., 2014. The evolution of genes within genes and the control of DNA replication in microviruses. *Mol. Biol. Evol.* 31, 1421–1431.
- Doore, S.M., Fane, B.A., 2016. The microviridae: diversity, assembly, and experimental evolution. *Virology* 491, 45–55.
- Fane, B.A., Brentlinger, K.L., Burch, A.D., Chen, M.I.N., Hafenstein, S., Moore, E., Novak, C., Uchiyama, A., 2006. PhiX174 et al., the Microviridae. In: Abedon, S.T., Calendar, R.L. (Eds.), *The Bacteriophages*, 2 ed. Oxford University Press, USA, pp. 129–145.
- Fane, B.A., Hayashi, M., 1991. Second-site suppressors of a cold-sensitive prohead accessory protein of bacteriophage phi X174. *Genetics* 128, 663–671.
- Feng, C.-Q., Zhang, Z.-Y., Zhu, X.-J., Lin, Y., Chen, W., Tang, H., Lin, H., 2018. iTerm-PseKNC: a sequence-based tool for predicting bacterial transcriptional terminators. *Bioinformatics* 35, 1469–1477.
- Gilbertson, R.L., Faria, J.C., Hanson, S.F., Morales, F.J., Ahlquist, P., Maxwell, D.P., Russell, D.R., 1991. Cloning of the complete DNA genomes of four bean-infecting geminiviruses and determining their infectivity by electric discharge particle acceleration. *Phytopathology* 81, 980–985.
- Godson, G.N., Barrell, B.G., Staden, R., Fiddes, J.C., 1978. Nucleotide sequence of bacteriophage G4 DNA. *Nature* 276, 236–247.
- Gorochowski, T.E., Espah Borujeni, A., Park, Y., Nielsen, A.A., Zhang, J., Der, B.S., Gordon, D.B., Voigt, C.A., 2017. Genetic circuit characterization and debugging using RNA-seq. *Mol. Syst. Biol.* 13, 952.
- Greinert, R., Volkmer, B., Henning, S., Breitbart, E.W., Greulich, K.O., Cardoso, M.C., Rapp, A., 2012. UVA-induced DNA double-strand breaks result from the repair of clustered oxidative DNA damages. *Nucleic Acids Res.* 40, 10263–10273.
- Hayashi, M., Aoyama, A., Richardson Jr., D.L., Hayashi, M.N., 1988. Biology of the bacteriophage ϕ X174. In: Calendar, R.L. (Ed.), *The Bacteriophages*. Plenum Press, New York and London, pp. 1–71.
- Hayashi, M., Fujimura, F.K., Hayashi, M., 1976. Mapping of in vivo messenger RNAs for bacteriophage phiX-174. *Proc. Natl. Acad. Sci. U. S. A.* 73, 3519–3523.
- Hayashi, M., Hayashi, M.N., Spiegelman, S., 1963. Restriction of in vivo genetic transcription to one of the complementary strands of DNA. *Proc. Natl. Acad. Sci. U. S. A.* 50, 664–672.
- Hayashi, M.N., Hayashi, M., 1985. Cloned DNA sequences that determine mRNA stability of bacteriophage phi X174 in vivo are functional. *Nucleic Acids Res.* 13, 5937–5948.
- Hayashi, M.N., Hayashi, M., Imai, M., 1981. Bacteriophage phi X174-specific mRNA synthesis in cells deficient in termination factor rho activity. *J. Virol.* 38, 198–207.
- Hayashi, M.N., Yaghmai, R., McConnell, M., Hayashi, M., 1989. mRNA stabilizing signals encoded in the genome of the bacteriophage phi X174. *Mol. Gen. Genet.* 216, 364–371.
- Hecht, A., Filliben, J., Munro, S.A., Salit, M., 2018. A minimum information standard for reproducing bench-scale bacterial cell growth and productivity. *Commun. Biol.* 1, 219.
- Hecht, A., Glasgow, J., Jaszche, P.R., Bawazer, L.A., Munson, M.S., Cochran, J.R., Endy, D., Salit, M., 2017. Measurements of translation initiation from all 64 codons in *E. coli*. *Nucleic Acids Res.* 45, 3615–3626.
- Hutchison, C.A.I., 1969. Bacteriophage PhiX174: Viral Genes and Functions.
- Jaszche, P.R., Dotson, G.A., Hung, K.S., Liu, D., Endy, D., 2019. Definitive demonstration by synthesis of genome annotation completeness. *Proc. Natl. Acad. Sci. Unit. States Am.* 116, 24206–24213.
- Jaszche, P.R., Lieberman, E.K., Rodriguez, J., Sierra, A., Endy, D., 2012. A fully decompressed synthetic bacteriophage PhiX174 genome assembled and archived in yeast. *Virology* 434, 278–284.
- Jekel, M., Wackernagel, W., 1995. The periplasmic endonuclease I of *Escherichia coli* has amino-acid sequence homology to the extracellular DNases of *Vibrio cholerae* and *Aeromonas hydrophila*. *Gene* 154, 55–59.
- Jeng, Y., Gelfand, D., Hayashi, M., Shleser, R., Tessman, E.S., 1970. The eight genes of bacteriophages phi X174 and S13 and comparison of the phage-specified proteins. *J. Mol. Biol.* 49, 521–526.
- Kapitza, E.L., Stukacheva, E.A., Shemyakin, M.F., 1979. Effect of *Escherichia coli* rho factor and RNase III on the formation of phi X174 RNA in vitro. *FEBS Lett.* 98, 123–127.
- Kodaira, K., Nakano, K., Okada, S., Taketo, A., 1992. Nucleotide sequence of the genome of the bacteriophage alpha 3: interrelationship of the genome structure and the gene products with those of the phages, phi X174, G4 and phi K. *Biochim. Biophys. Acta* 1130, 277–288.
- Komura, R., Aoki, W., Motone, K., Satomura, A., Ueda, M., 2018. High-throughput evaluation of T7 promoter variants using biased randomization and DNA barcoding. *PloS*

- One 13, e0196905 e0196905.
- Koslover, D.J., Fazal, F.M., Mooney, R.A., Landick, R., Block, S.M., 2012. Binding and translocation of termination factor rho studied at the single-molecule level. *J. Mol. Biol.* 423, 664–676.
- Krinke, L., Mahoney, M., Wulff, D.L., 1991. The role of the OOP antisense RNA in coliphage lambda development. *Mol. Microbiol.* 5, 1265–1272.
- Lau, P.C., Spencer, J.H., 1985. Nucleotide sequence and genome organization of bacteriophage S13 DNA. *Gene* 40, 273–284.
- Lavish, D., Sokolova, M., Slaschcheva, M., Förstner, K.U., Severinov, K., 2017. Transcription profiling of *Bacillus subtilis* cells infected with AR9, a giant phage encoding two multisubunit RNA polymerases. *mBio* 8, e02041 02016.
- Lazarowitz, S.G., 1988. Infectivity and complete nucleotide-sequence of the genome of a south-african isolate of maize streak virus. *Nucleic Acids Res.* 16, 229–249.
- Lee, N., Gielow, W., Martin, R., Hamilton, E., Fowler, A., 1986. The organization of the araBAD operon of *Escherichia coli*. *Gene* 47, 231–244.
- Liao, S.M., Wu, T.H., Chiang, C.H., Susskind, M.M., McClure, W.R., 1987. Control of gene expression in bacteriophage P22 by a small antisense RNA. I. Characterization in vitro of the Psar promoter and the sar RNA transcript. *Genes Dev.* 1, 197–203.
- Maratea, D., Young, K., Young, R., 1985. Deletion and fusion analysis of the phage phi X174 lysis gene. *E. Gene* 40, 39–46.
- Mayol, R.F., Sinsheimer, R.L., 1970. Process of infection with bacteriophage phiX174. XXXVI. Measurement of virus-specific proteins during a normal cycle of infection. *J. Virol.* 6, 310–319.
- Michel, A., Clermont, O., Denamur, E., Tenaillon, O., 2010. Bacteriophage PhiX174's ecological niche and the flexibility of its *Escherichia coli* lipopolysaccharide receptor. *Appl. Environ. Microbiol.* 76, 7310–7313.
- Mojardín, L., Salas, M., 2016. Global transcriptional analysis of virus-host interactions between phage q29 and *Bacillus subtilis*. *J. Virol.* 90, 9293.
- Mukai, R., Hamatake, R.K., Hayashi, M., 1979. Isolation and identification of bacteriophage phi X174 prohead. *Proc. Natl. Acad. Sci. U.S.A.* 76, 4877–4881.
- Niagro, F., Forsthöfel, A., Lawther, R., Kamalanathan, L., Ritchie, B., Latimer, K., Lukert, P., 1998. Beak and feather disease virus and porcine circovirus genomes: intermediates between the geminiviruses and plant circoviruses. *Arch. Virol.* 143, 1723–1744.
- Osterhout, R.E., Figueira, I.A., Keasling, J.D., Arkin, A.P., 2007. Global analysis of host response to induction of a latent bacteriophage. *BMC Microbiol.* 7, 82.
- Osuka, J., Kunisawa, T., 1982. Characteristic base sequence patterns of promoter and terminator sites in ϕ X174 and fd phage DNAs. *J. Theor. Biol.* 97, 415–436.
- Peterson, A.C., Russell, J.D., Bailey, D.J., Westphall, M.S., Coon, J.J., 2012. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics: MCP* 11, 1475–1488.
- Pollock, T.J., Tessman, I., Tessman, E.S., 1978. Potential for variability through multiple gene products of bacteriophage Φ X174. *Nature* 274, 34–37.
- Raina, S., Missikas, D., Georgopoulos, C., 1995. The rpoE gene encoding the sigma E (sigma 24) heat shock sigma factor of *Escherichia coli*. *EMBO J.* 14, 1043–1055.
- Rassart, E., Spencer, J.H., 1978. Localization of *Escherichia coli* RNA polymerase binding sites on bacteriophage S13 and ϕ X174 DNA: alignment with restriction enzyme maps. *J. Virol.* 27, 677–687.
- Rassart, E., Spencer, J.H., Zollinger, M., 1979. Localization of *Escherichia coli* RNA polymerase binding sites on bacteriophage S13 and ϕ X174 DNAs by electron microscopy. *J. Virol.* 29, 179–184.
- Ray-Soni, A., Bellcourt, M.J., Landick, R., 2016. Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.* 85, 319–347.
- Reyes, A., Semenkovich, N.P., Whiteson, K., Rohwer, F., Gordon, J.I., 2012. Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat. Rev. Microbiol.* 10, 607–617.
- Rhodius, V.A., Suh, W.C., Nonaka, G., West, J., Gross, C.A., 2006. Conserved and variable functions of the sigmaE stress response in related genomes. *PLoS Biol.* 4 e2-e2.
- Ringuelette, M.J., Spencer, J.H., 1994. Mapping the initiation sites of in vitro transcripts of bacteriophage S13. *Biochim. Biophys. Acta* 1218, 331–338.
- Rokyta, D.R., Abdo, Z., Wichman, H.A., 2009. The genetics of adaptation for eight microvirid bacteriophages. *J. Mol. Evol.* 69, 229–239.
- Roznowski, A.P., Doore, S.M., Kemp, S.Z., Fane, B.A., 2020. Finally, a role befitting α^{star} : strongly conserved, unessential microvirus α^* proteins ensure the product fidelity of packaging reactions. *J. Virol.* 94, e01593 01519.
- Salamov, V.S.A., Solovyevand, A., 2011. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. *Metagenomics and its Applications in Agriculture*. Nova Science Publishers, Hauppauge, NY, USA, pp. 61–78.
- Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocombe, P.M., Smith, M., 1977. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265, 687–695.
- Schuerer, S., Carbone, W., Knehr, J., Petitjean, V., Fernandez, A., Sultan, M., Roma, G., 2017. A comprehensive assessment of RNA-seq protocols for degraded and low-quality samples. *BMC Genom.* 18, 442.
- Shahmradov, I.A., Mohamad Razali, R., Bougouffa, S., Radovanovic, A., Bajic, V.B., 2017. tBSSfinder: a novel tool for the prediction of promoters in cyanobacteria and *Escherichia coli*. *Bioinformatics* 33, 334–340.
- Shearwin, K.E., Callen, B.P., Egan, J.B., 2005. Transcriptional interference—a crash course. *Trends Genet.* 21, 339–345.
- Sinsheimer, R.L., Knippers, R., Komano, T., 1968. Stages in the replication of bacteriophage phi X174 DNA in vivo. *Cold Spring Harbor Symp. Quant. Biol.* 33, 443–447.
- Smith, H.O., Hutchison 3rd, C.A., Pfannkoch, C., Venter, J.C., 2003. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* 100, 15440–15445.
- Smith, L.H., Sinsheimer, R.L., 1976a. The in vitro transcription units of bacteriophage phiX174. I. Characterization of synthetic parameters and measurement of transcript molecular weights. *J. Mol. Biol.* 103, 681–697.
- Smith, L.H., Sinsheimer, R.L., 1976b. The in vitro transcription units of bacteriophage phiX174. II. In vitro initiation sites of phiX174 transcription. *J. Mol. Biol.* 103, 699–710.
- Smith, L.H., Sinsheimer, R.L., 1976c. The in vitro transcription units of bacteriophage phiX174. III. Initiation with specific 5' end oligonucleotides of in vitro phiX174 RNA. *J. Mol. Biol.* 103, 711–735.
- Sorensen, S.E., Barrett, J.M., Wong, A.K., Spencer, J.H., 1998. Identification of the in vivo promoters of bacteriophages S13 and phi X174 and measurement of their relative activities. *Biochem. Cell. Biol.* 76, 625–636.
- Stanley, J., Markham, P.G., Callis, R.J., Pinner, M.S., 1986. The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. *EMBO J.* 5, 1761–1767.
- Sun, L., Young, L.N., Zhang, X., Boudko, S.P., Fokine, A., Zbornik, E., Roznowski, A.P., Molineux, I.J., Rossmann, M.G., Fane, B.A., 2014. Icosahedral bacteriophage PhiX174 forms a tail for DNA transport during infection. *Nature* 505, 432–435.
- Sun, Y., Roznowski, A.P., Tokuda, J.M., Klose, T., Mauney, A., Pollack, L., Fane, B.A., Rossmann, M.G., 2017. Structural changes of tailless bacteriophage PhiX174 during penetration of bacterial cell walls. *Proc. Natl. Acad. Sci. U. S. A.* 114, 13708–13713.
- Tessman, E.S., Tessman, I., Pollock, T.J., 1980. Gene K of bacteriophage phi X 174 codes for a nonessential protein. *J. Virol.* 33, 557–560.
- Tisza, M.J., Pastrana, D.V., Welch, N.L., Stewart, B., Peretti, A., Starrett, G.J., Pang, Y.-Y.S., Krishnamurthy, S.R., Pesavento, P.A., McDermott, D.H., Murphy, P.M., Whited, J.L., Miller, B., Brenchley, J., Rosshart, S.P., Rehermann, B., Doorbar, J., Ta'alá, B.A., Pletnikova, O., Troncoso, J.C., Resnick, S.M., Bolduc, B., Sullivan, M.B., Varsani, A., Segall, A.M., Buck, C.B., 2020. Discovery of several thousand highly diverse circular DNA viruses. *Elife* 9, e51971.
- Vincent, R.M., Wright, B.W., Jasicke, P.R., 2019. Measuring amber initiator tRNA orthogonality in a genetically recoded organism. *ACS Synth. Biol.* 8, 675–685.
- Wagner, G.P., Kin, K., Lynch, V.J., 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theor. Biosci.* 131, 281–285.
- Yona, A.H., Alm, E.J., Gore, J., 2018. Random sequences rapidly evolve into de novo promoters. *Nat. Commun.* 9, 1530.
- Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M., Pierce, N.A., 2011. NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.* 32, 170–173.
- Zhao, L., Stancik, A.D., Brown, C.J., 2012. Differential transcription of bacteriophage phiX174 genes at 37 degrees C and 42 degrees C. *PLoS One* 7, e35909.