Modeling the Fitness Consequences of a Cyanophage-Encoded Photosynthesis Gene

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1371/journal.pone.0003550">http://dx.doi.org/10.1371/journal.pone.0003550</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Public Library of Science</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sun Apr 09 17:57:26 EDT 2017</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/55289">http://hdl.handle.net/1721.1/55289</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Modeling the Fitness Consequences of a Cyanophage-Encoded Photosynthesis Gene

Jason G. Bragg1*, Sallie W. Chisholm1,2

1 Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, 2 Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

Abstract

**Background:** Phages infecting marine picocyanobacteria often carry a *psbA* gene, which encodes a homolog to the photosynthetic reaction center protein, D1. Host encoded D1 decays during phage infection in the light. Phage encoded D1 may help to maintain photosynthesis during the lytic cycle, which in turn could bolster the production of deoxynucleoside triphosphates (dNTPs) for phage genome replication.

**Methodology / Principal Findings:** To explore the consequences to a phage of encoding and expressing *psbA*, we derive a simple model of infection for a cyanophage/host pair — cyanophage P-SSP7 and *Prochlorococcus* MED4 — for which pertinent laboratory data are available. We first use the model to describe phage genome replication and the kinetics of *psbA* expression by host and phage. We then examine the contribution of phage *psbA* expression to phage genome replication under constant low irradiance (25 μE m⁻² s⁻¹). We predict that while phage *psbA* expression could lead to an increase in the number of phage genomes produced during a lytic cycle of between 2.5 and 4.5% (depending on parameter values), this advantage can be nearly negated by the cost of *psbA* in elongating the phage genome. Under higher irradiance conditions that promote D1 degradation, however, phage *psbA* confers a greater advantage to phage genome replication.

**Conclusions / Significance:** These analyses illustrate how *psbA* may benefit phage in the dynamic ocean surface mixed layer.

Introduction

The marine picocyanobacteria *Prochlorococcus* and *Synechococcus* are numerically dominant phytoplankton in nutrient-poor open ocean ecosystems, and are an important contributor to photosynthesis in the oceans [1–3]. They are infected by cyanophages including members of the families Podoviridae, Myoviridae and Siphoviridae [4], which can be abundant in regions where these cells dominate (e.g. [5–8]). Several genomes of these marine cyanophages have been sequenced, revealing gene content and organization broadly similar to confamilial phages [9–11]. For example, cyanophage P-SSP7, which infects *Prochlorococcus* MED4, has many genomic similarities to the T7 phage that infects *Escherichia coli* [11].

The genomes of marine *Synechococcus* and *Prochlorococcus* cyanophages often contain genes that are absent from the genomes of morphologically related phages that do not infect marine cyanobacteria [11]. A striking example of this is the *psbA* photosynthesis gene [12,13]. This gene is found in the genomes of a large proportion of cyanophages known to infect marine picocyanobacteria [14,15], suggesting that it confers a fitness advantage. The product of the *psbA* gene in the host cell, the D1 protein, forms part of the photosystem II reaction center, and turns over relatively rapidly during photosynthesis [16]. Over the course of phage infection, host-encoded D1 proteins decline following the inhibition of host transcription and the decay of host *psbA* transcripts [17], while phage-encoded D1 proteins increase [17]. It is hypothesized that the latter replace damaged host D1 proteins, and help to maintain photosynthesis throughout the lytic cycle. This, in turn, could increase the relative fitness of phage that carry the *psbA* gene [12,18]. In some cyanophages, reproduction (e.g. [19]) and genome replication [17] are severely limited in the dark, indicating that photosynthesis can be important for phage genome replication, which potentially limits the production of phage progeny. During the cyanophage P-SSP7 lytic cycle, *psbA* is transcribed contemporaneously with several metabolism genes that have probable roles in dNTP synthesis (e.g. ribonucleotide reductase), as well as genome replication enzymes [20]. This adds weight to the suggestion that the *psbA* gene helps phage P-SSP7 to acquire resources to make dNTPs during infection.

Models of phage infection in well established phage/host systems, such as T7/*E. coli* [21], have provided significant insights into factors affecting phage reproduction and fitness [22–24]. Inspired by these works, we have developed an intracellular model of infection of *Prochlorococcus* MED4 by the Podovirus P-SSP7. The model concentrates on processes of phage genome replication, the
production of dNTPs, and the expression of psbA by host and phage, and can find good agreement with experimental measurements collected over the cyanophage P-SSP7 lytic cycle [17,20]. We use the model to ask basic questions about the advantages to the phage of carrying this gene that are not yet tractable experimentally: How much can phage psbA expression benefit phage genome replication? To what degree is this contingent on environmental conditions, particularly the ambient light environment?

Methods
Model Development

(a) Approach. After the genome of cyanophage P-SSP7 enters a host cell, phage genes are expressed, the phage genome is replicated, new phage particles are assembled, and the host cell is lysed — all over a period of about 8 hours [20]. Many of these processes are carried out using products of phage-encoded genes, which are expressed at different times during the cycle of infection [20]. Our model links phage genome replication to the production of deoxynucleoside triphosphates (dNTPs), which in turn is linked to photosynthesis and the kinetics of host and phage psbA expression (Fig. 1). It incorporates elements of previous models of phage genome replication [21,25], and D1 protein kinetics [26]. More specifically, we model phage genome replication within a host cell as a function of the availability of dNTPs, which are supplied by (i) scavenging from the degraded host genome, and (ii) a pathway for the synthesis of new deoxynucleotides. We assume that the supply of dNTPs from each of these sources can depend on photosynthesis. In turn, photosynthesis is modeled as a function of the number of functional photosystem II (PSII) subunits, which become non-functional when their D1 core proteins are damaged, and regain their function when the damaged D1 is excised from the photosystem and replaced with the protein product of either a host or phage psbA gene (Fig. 1).

Our model incorporates processes that are carried out by phage genes that begin to be expressed at different times following infection [20]. We therefore need a way to represent relatively abrupt increases in the velocity of processes that are carried out by different proteins (generically, Pr, at different, specific times following infection. We do this using Hill functions [27], or

\[ P_r = \frac{1}{1 + \left(\frac{t}{t_r}\right)^n} \]

where \( t_r \) is the time since infection, \( t \) is the time at which process \( P_r \) reaches half its maximum rate (the ‘timing parameter’), and \( n \) is a parameter that controls the abruptness of the increase. The Hill function is a sigmoidal curve that increases from zero to one with increasing values of \( t \), and can provide a reasonable description of the expression of some relevant phage genes, at least in terms of mRNA abundances (see Text S1). Below, we represent Hill functions in our equations using ‘parameter’, and

\[ G_H(t) = \frac{V_{Gdeg}G_H}{G_H + K_{Gdeg}} P_{Gdeg}, \]

where \( P_{Gdeg} \) is a Hill function representing the expression of genes that degrade the host genome (with time parameter \( t_{Gdeg} \)), and \( V_{Gdeg} \) is the maximum rate of degradation of the host genome. The term \( G_H(t) \) with small \( K_{Gdeg} \) is approximately equal to \( V_{Gdeg} \) until the host genome is almost entirely degraded. This formulation is based on the observation that the decline in host genomes is approximately linear, following a delay of 4–5 h [20], as well as the necessity for degradation to cease as \( G_H(t) \) approaches 0. We set \( G_H(t) = 1 \) to reflect a single host genome at the time of infection.

We assume the phage can then make dNTPs from the degraded host genome (\( G_{Hdeg} \)). We model the rate of production of dNTPs from this source (\( \xi_G \) dNTPs cell\(^{-1}\) h\(^{-1}\)) as \( \xi_G = 2L_H \frac{1}{t_{Gdeg} + K_{Gdeg}} P_{Gdeg} \).

Here \( L_H \) is the number of deoxynucleotides in the host genome (2 per base pair, \( t_{Gdeg} = t_H \) base pairs per genome), and \( V_z \) is the maximum rate at which the degraded genome can be converted to dNTPs. We set the parameter \( K_{Gdeg} \) to a small value so that when genes for degrading the host genome are expressed and degraded host genome is available, the term \( \frac{1}{t_{Gdeg} + K_{Gdeg}} \) is approximately equal to 1, and dNTPs are produced from degraded genomes at a rate of approximately \( 2L_HV_z \).

We then consider the possibility that the production of dNTPs from degraded genomes (\( 2L_HV_z \)) is limited by photosynthesis. We represent this in the model by letting \( 2L_HV_z = \varepsilon \gamma \hat{P} \), where \( \varepsilon \) represents the rate at which dNTPs can be made from degraded genomes during infection in the dark, and \( \gamma \) represents the photosynthesis-dependent production of dNTPs from host genomes. In turn, we assume that \( \gamma \) is limited by the abundance of functional photosystem II subunits or \( \gamma = \gamma_7 \hat{P}_{PSII} \), where \( \hat{P}_{PSII} \) is the number of functional photosystem II subunits per cell, \( \gamma \) is the rate of photosynthesis per functional PSII, and \( \gamma \) is the efficiency with which photosynthesis is used in converting degraded genomes to dNTPs. Below, we develop a model for the proportion of PSII subunits that are functional (\( \hat{P}_{PSII} \)) during infection. We therefore set \( \hat{P}_{PSII} = \gamma \frac{V_z}{\gamma_7} \), where \( \gamma \) is the total number of PSII subunits per cell. This means we have \( \gamma = \gamma_7 \hat{P}_{PSII} \) or if we represent \( \gamma_7 \) by the parameter \( \xi \) (in dNTPs cell\(^{-1}\) h\(^{-1}\)) \( \gamma = \xi \hat{P}_{PSII} \).

We note that the change over time in the proportion of the host genome in a degraded state is then given by

\[ \frac{d(G_H)}{dt} = \frac{V_{Gdeg}G_H}{G_H + K_{Gdeg}} P_{Gdeg} - \frac{V_NG_HG_{Hdeg}}{G_{Hdeg} + K_N} - P_{Gdeg}. \]
functions in converting ribonucleotides to deoxynucleotides. We assume this source of dNTPs is dependent on photosynthesis, as well as the activity of genes that are encoded by the phage, and whose expression is described by a Hill function, $F_p$. Once these phage genes are expressed, the rate of supply of dNTPs from this source is assumed to be proportional to the rate of photosynthesis, which is limited by the abundance of functional PSII subunits, or $s_p = v F_{psbA} P_{s}$, where $v$ is the efficiency with which products of photosynthesis are converted to dNTPs. We then represent $v U$ using a single parameter, $l$ (in dNTPs cell$^{-1}$ h$^{-1}$), such that $s_p = l f_{PSII} P_{s}$. Presently we lack detailed mechanistic information about this potential extra source of dNTPs, which would be useful for refining the model. For example, if photosynthesis powers the conversion of a finite cellular resource to dNTPs, the depletion of this resource ought to be modeled.

After accounting for the incorporation of free dNTPs into genomes, we have a rate equation for dNTPs per cell ($N$):

$$\frac{d(N)}{dt} = s_G + s_p - 2 \frac{V_r N}{N + K_{nr}} P_r.$$  \hspace{1cm} (4)

We next model the proportion of PSII subunits that are functional ($f_{PSII}$), (i) contain damaged D1 proteins ($d_{PSII}$), and (iii) have had damaged D1 proteins excised (‘empty’ PSII subunits, or $x_{PSII}$) (Fig. 1). For functional and damaged subunits, we track PSII subunits containing host- versus phage-encoded D1 proteins separately. For example, for functional PSII subunits, $f_{PSII} = f_{PSIIH} + f_{PSIIP}$, where $f_{PSIIH}$ and $f_{PSIIP}$ contain host D1 and phage D1, respectively. We also assume that during the course of infection, the total number of PSII subunits ($U$) in a cell is constant, and that $f_{PSIIH} + f_{PSIIP} + d_{PSIIH} + d_{PSIIP} + x_{PSII} = 1$. This yields the following system of equations:

$$\frac{df_{PSIIH}}{dt} = -k_{D1, dam} f_{PSIIH} + k_{D1} R_{psbA} x_{PSII}$$  \hspace{1cm} (5)

$$\frac{df_{PSIIP}}{dt} = -k_{D1, dam} f_{PSIIP} + k_{D1} R_{psbA} x_{PSII}$$  \hspace{1cm} (6)

$$\frac{dd_{PSIIH}}{dt} = k_{D1, dam} f_{PSIIH} - k_{exc} d_{PSIIH}$$  \hspace{1cm} (7)

$$\frac{dd_{PSIIP}}{dt} = k_{D1, dam} f_{PSIIP} - k_{exc} d_{PSIIP}.$$  \hspace{1cm} (8)

where $x_{PSII} = 1 - (f_{PSIIH} + f_{PSIIP} + d_{PSIIH} + d_{PSIIP})$. Here $k_{D1, dam}$ is the rate
at which D1 proteins in functional PSII subunits are damaged by irradiance, \( k_{\text{dam}} \), is the rate at which damaged D1 proteins are excised from PSII subunits, and \( k_{\text{rec}} \) is the rate at which damaged PSII subunits are repaired using \( p\text{psb}A \) mRNA transcripts. \( R_{\text{HpsbA}} \) and \( R_{\text{PpsbA}} \) are the abundances of host and phage \( p\text{psb}A \) transcripts, respectively. This formulation assumes that D1 proteins are represented only in functional and damaged PSII subunits, and that \( p\text{psb}A \) transcripts are limiting to repair.

The expression of host and phage \( p\text{psb}A \) transcripts are modeled as follows:

\[
\frac{d(R_{\text{HpsbA}})}{dt} = k_{\text{PpsbA}} G_{\text{H}}(1 - P_{\text{psbA}}) - d_{\text{PpsbA}} R_{\text{HpsbA}} \tag{9}
\]

\[
\frac{d(R_{\text{PpsbA}})}{dt} = k_{\text{PpsbA}} P_{\text{psbA}} - d_{\text{PpsbA}} R_{\text{PpsbA}}. \tag{10}
\]

Here \( d_{\text{HpsbA}} \) is the decay rate of \( p\text{psb}A \) mRNA transcripts. \( k_{\text{HpsbA}} \) and \( k_{\text{PpsbA}} \) are the maximum rates of transcription of host and phage \( p\text{psb}A \) mRNAs, respectively. Host \( p\text{psb}A \) is transcribed until either the host genome is gone, or until host RNA polymerase is inhibited. The inhibition of host RNA polymerase by a phage protein is represented using the term \((1 - P_{\text{psbA}})\), where \( P_{\text{psbA}} \) is a Hill function. \( P_{\text{psbA}} \) is a Hill function representing the commencement of transcription of phage \( p\text{psb}A \) at a time of approximately \( t_{0}\).

The above formulation includes assumptions that can be tested experimentally, and improved in future versions of the model. We assume, for example, that host and phage \( p\text{psb}A \) transcripts have identical rates of decay \( (d_{\text{HpsbA}}) \). We also assume that empty PSII subunits can be repaired at identical rates using products of host and phage \( p\text{psb}A \) genes, that PSII subunits containing host and phage D1 proteins have similar rates of damage \( (k_{\text{D1,dam}}) \) and excision \( (k_{\text{rec}}) \), and that functional PSII subunits containing host and phage D1 have similar rates of photosynthesis. In reality, these properties of host and phage \( p\text{psb}A \) transcripts or D1 proteins could be different. For example, it has been suggested that phage D1 might be more resistant to photodamage than host D1 [28]. Furthermore, we assume that the total number of photosystem II subunits and the maximum rates of excision and repair are constant over the course of infection, while in reality, these values may decay as a function of time. It would be useful to measure these properties of infected cells experimentally, and revise the model if necessary. More broadly, our model clearly uses a highly simplified representation of photosynthesis, an extremely complex process influenced by a large number of factors [29]. Our goal was to abstract this complexity with a focus on the potential advantage to phage of supplementing the supply of D1 during infection.

## Results

### Model validation

**a) Approach.** The parameterization of the model is described in detail below, and parameter values are listed in Table 1. Our general approach was as follows: We began by considering parameters that govern the abundance of host and phage \( p\text{psb}A \) transcripts, and then estimated parameters for the abundance of host and phage D1 proteins. In the experiments that are the basis for the model, cells were grown under continuous light [17]. We therefore assumed the abundances of host \( p\text{psb}A \) mRNAs and the proportions of functional, damaged and empty PSII subunits were in steady state prior to infection, and set equations (5), (6) and (9) equal to 0. This imposed relationships between parameters and initial conditions of some variables (Table 1), reducing the number of free parameters. Finally, we used data for genome replication in the light and dark [17] to estimate parameters for the dependence of dNTP acquisition on photosynthesis. Data from Lindell et al. [20] were used to estimate parameters for the degradation of host genomes and for the timing of expression of phage genes involved in genome replication and dNTP production (see Text S1).

Lindell et al. [17,20] studied populations of cells that were infected with phage, while our model is based on infection of a single host cell. In comparing model predictions to these experimental data, we assume that our model represents infection of an average cell. To estimate the number of phage genomes per host cell at different times after infection, we normalized by the number of phages measured at 1 h post-infection. Given that our estimates of phage genome replication depend on this normalization, we place our emphasis on the proportional advantage or disadvantage conferred by phage \( p\text{psb}A \), rather than the absolute number of genomes. Lindell et al. [17] used a low multiplicity of infection (0.1 phage for every host cell) for the experiment in which phage genome replication was measured. Under these conditions, most infected cells would have been infected by a single virion. When using data for intracellular levels of host \( p\text{psb}A \) transcripts, D1 proteins, and genomes, we assumed that 50% of cells were infected (see Text S1 for analyses that consider the implications of varying this assumption). A higher multiplicity of infection (3 phage per host cell) was used in the experiments from which these data were collected, and 50% represents the maximum level of infection that has been observed for this phage [17]. We also assumed that measurements of D1 protein abundance made by [17] detected both functional and damaged D1 proteins.

We integrated equations (1)–(10) using ode45, a MATLAB® (The MathWorks, Natick, MA) variable step numerical ODE solver, which implements a medium order Runge-Kutta scheme.

### b) Expression of photosynthesis genes.** Experimental evidence shows that following infection by the cyanophage P-SSP7, the abundance of \( p\text{psb}A \) transcripts in the host cell declines [17]. We assumed that host transcription was largely inhibited \((1 - P_{\text{psbA}})\) by 1 hour after infection (Fig. 2A, Table 1), and calculated the decay constant \( (d_{\text{HpsbA}}) \) using experimental observations [17]. We set the initial value of host \( p\text{psb}A \) mRNA to \( R_{\text{HpsbA}}(0) = 1 \), and normalized the abundance of host \( p\text{psb}A \) transcripts to this initial (maximum) value.

Following the decay of host \( p\text{psb}A \) transcripts, Lindell et al. [17] observed a drop in the level of host D1 proteins, such that host D1 abundance had decreased to approximately 45% of its maximum value (measured 1 hour after infection) after 8 hours of infection (Fig. 2B). In parameterizing the dynamics of host D1 proteins, we first assumed cells were in steady state prior to infection, which constrained parameters according to \( R_{\text{HpsbA}}(0) = k_{\text{psbA}}(1 + V_{\text{psbA}}(0)) = k_{\text{psbA}}(1 + V_{\text{psbA}}(0)) = k_{\text{psbA}}(1 + V_{\text{psbA}}(0)) = 1 \). We did not have independent estimates of the parameters \( k_{\text{D1,dam}} \), \( k_{\text{rec}} \), and \( D_{\text{psbA}} \) for *Prochlorococcus* under the conditions of the experiment [17], and values of \( k_{\text{D1,dam}} \) and \( k_{\text{rec}} \) may vary substantially among organisms and growth conditions [26]. We therefore used measurements from a study of *Prochlorococcus* PSII function and D1 protein abundance under transient exposure to high irradiance [30] to estimate possible ranges of parameters \( k_{\text{D1,dam}} \), \( k_{\text{rec}} \), and \( D_{\text{psbA}} \). We then solved our model of D1 dynamics 3060 times, comprising all combinations of 17 values of \( k_{\text{D1,dam}} \), 15 values of \( k_{\text{rec}} \), and 12 values of \( D_{\text{psbA}} \) (see Table 1). Out of these 3060 simulations, 126 resulted in a drop in...
Table 1. Model parameters and initial conditions.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Units</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_p$</td>
<td>phage genomes</td>
<td>genomes cell$^{-1}$</td>
<td>1*</td>
</tr>
<tr>
<td>$G_h$</td>
<td>host genomes</td>
<td>genomes cell$^{-1}$</td>
<td>1*</td>
</tr>
<tr>
<td>$G_{deg}$</td>
<td>degraded host genomes</td>
<td>genomes cell$^{-1}$</td>
<td>0</td>
</tr>
<tr>
<td>$N$</td>
<td>dNTPs</td>
<td>dNTPs cell$^{-1}$</td>
<td>0</td>
</tr>
<tr>
<td>$\beta_{PSIIH}$, $\beta_{PSII}$</td>
<td>proportion of PSII subunits that are functional and contain host, phage D1</td>
<td>dimensionless</td>
<td>$\beta_{PSIIH}$ = 0.46, $\beta_{PSII}$ = 0*</td>
</tr>
<tr>
<td>$\alpha_{PSIIH}$, $\alpha_{PSII}$</td>
<td>proportion of PSII subunits that are damaged and contain host, phage D1</td>
<td>dimensionless</td>
<td>$\alpha_{PSIIH}$ = 0.04, $\alpha_{PSII}$ = 0*</td>
</tr>
<tr>
<td>$k_{exc}$</td>
<td>proportion of PSI subunits that are empty</td>
<td>dimensionless</td>
<td>0.5*</td>
</tr>
<tr>
<td>$R_{psbA}$, $R_{psbA}$</td>
<td>psbA transcripts</td>
<td>dimensionless</td>
<td>1, 0*</td>
</tr>
<tr>
<td>$L_p$, $L_P$</td>
<td>genome length of host, phage</td>
<td>bp genome$^{-1}$</td>
<td>1657990, 44970</td>
</tr>
<tr>
<td>$V_r$</td>
<td>max velocity of phage DNA elongation</td>
<td>dNTP h$^{-1}$ cell$^{-1}$</td>
<td>1332000</td>
</tr>
<tr>
<td>$K_s$</td>
<td>half-saturation for DNA replication</td>
<td>dNTP cell$^{-1}$</td>
<td>1224</td>
</tr>
<tr>
<td>$\tau$</td>
<td>timing parameter for phage genome replication</td>
<td>h</td>
<td>2</td>
</tr>
<tr>
<td>$V_{Gdeg}$</td>
<td>max velocity of host genome degradation</td>
<td>genomes cell$^{-1}$ h$^{-1}$</td>
<td>0.35</td>
</tr>
<tr>
<td>$K_{Ggen}$</td>
<td>half-saturation for host genome degradation</td>
<td>genomes cell$^{-1}$</td>
<td>0.000001</td>
</tr>
<tr>
<td>$t_{Gdeg}$</td>
<td>timing parameter for host genome degradation</td>
<td>h</td>
<td>5</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>production of dNTPs from degraded host genome in the dark</td>
<td>dNTP h$^{-1}$ cell$^{-1}$</td>
<td>127665</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>production of dNTPs from degraded host genome in the light</td>
<td>dNTP h$^{-1}$ cell$^{-1}$</td>
<td>0</td>
</tr>
<tr>
<td>$G_s$</td>
<td>half saturation for dNTP production from degraded host genome</td>
<td>genomes cell$^{-1}$</td>
<td>0.000001</td>
</tr>
<tr>
<td>$\delta$</td>
<td>production of dNTPs in the light</td>
<td>dNTP h$^{-1}$ cell$^{-1}$</td>
<td>1027800</td>
</tr>
<tr>
<td>$t_s$</td>
<td>timing of dNTP synthesis from source $s_p$</td>
<td>h</td>
<td>4</td>
</tr>
<tr>
<td>$k_{D1dam}$</td>
<td>damage to functional D1 proteins</td>
<td>h$^{-1}$</td>
<td>0.35*</td>
</tr>
<tr>
<td>$k_{Exc}$</td>
<td>excision of damaged D1 proteins</td>
<td>h$^{-1}$</td>
<td>4*</td>
</tr>
<tr>
<td>$k_{D1}$</td>
<td>repair of empty PSII subunits</td>
<td>h$^{-1}$</td>
<td>$k_{D1} = k_{D1dam} = 0.32$</td>
</tr>
<tr>
<td>$d_{psbA}$</td>
<td>psbA transcript decay</td>
<td>h$^{-1}$</td>
<td>0.27*</td>
</tr>
<tr>
<td>$R_{psbA}$</td>
<td>host psbA transcription</td>
<td>h$^{-1}$ (genomes cell$^{-1}$)</td>
<td>0.27*</td>
</tr>
<tr>
<td>$k_{psbA}$</td>
<td>phage psbA transcription</td>
<td>h$^{-1}$</td>
<td>0.016</td>
</tr>
<tr>
<td>$\tau_{psbA}$</td>
<td>timing parameter for inhibition of host RNA polymerase</td>
<td>h</td>
<td>1</td>
</tr>
<tr>
<td>$\tau_{psbA}$</td>
<td>timing parameter for transcription of phage psbA</td>
<td>h</td>
<td>1.3</td>
</tr>
<tr>
<td>$n$</td>
<td>Hill parameter</td>
<td>dimensionless</td>
<td>5</td>
</tr>
</tbody>
</table>

*Initial condition.

Values were systematically varied in exploring the kinetics of D1 protein degradation, excision and repair. All combinations of the following values were used: $F_{psbA} = (0.01, 0.025, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.3, 0.35, 0.4, 0.5)$. 

This estimate is based on microarray measurements of host psbA mRNA expression. Measurements of host psbA transcript abundances that were made using RT-PCR ([17] suggested a greater value of $\alpha_{psbA} = 0.72$ in Text S1.

We next modeled the abundance of phage psbA transcripts using the decay constant ($k_{D1dam}$) calculated above for host psbA transcripts. Our model describes the shape of the experimentally derived curve of phage psbA mRNA abundance reasonably well (Fig. 2A), though modeled levels of mRNAs approached an asymptotic level more slowly than observed [17].

The empirical observation that phage D1 proteins accumulated to approximately 10% of all D1 after 8 hours of infection [17] was predicted by the model when phage psbA transcription was set to 5.9% of the rate at which host psbA was transcribed prior to infection ($\beta_{psbA} = 0.016$; Fig. 2B) and with the same values of $k_{D1dam}$ and $k_{Exc}$ that were used for host D1. The model predicted the increase in the level of phage D1 slightly sooner than it was observed experimentally. This could be due to a time delay for the translation of D1, or may simply reflect experimental variability.
is assumed to be a sphere with diameter 0.6 μm. We assumed that phage genome replication enzymes were expressed approximately 2 hours post-infection (t = 2), based on observations of phage DNA polymerase transcript abundance in [20] (see Text S1). We then found that e = 127,665 could provide a reasonable description of genome replication in the dark, if all dNTPs used in phage genome replication in the dark were derived from the host genome (Fig. 3).

Our model includes the possibility that photosynthesis increases the production of dNTPs from degraded host genomes, and the possibility that photosynthesis promotes the synthesis of new deoxynucleotides. However, since we do not know the relative importance of these possible sources of dNTPs, we analyze their potential contribution to dNTP production during infection in the light separately. Here we present analyses that assume extra dNTPs made in the light were derived from the synthesis of new deoxynucleotides (i.e., e > 0 and μ = 0). However, we confirmed that similar results are obtained if we assume instead that the extra dNTPs made in the light were derived from the degraded host genome (see Text S1).

For infection in the light, we use the same values of parameters V_r, K_m, t_e, and e, and found that e = 1,027,800 (dNTP h⁻¹ cell⁻¹) gave a reasonable description of phage genome replication (Fig. 3). With this parameterization, dNTP availability is strongly limiting to genome replication: a 10% increase in dNTP production by photosynthesis (e) results in a 7.3% increase in genome replication, whereas a 10% increase in the maximum velocity of genome replication (V_r) results in almost no increase in genome replication.

**In silico knockout of phage psbA**

The major goal of this study is to consider the fitness consequences to a phage of encoding and expressing the psbA gene. Having described the kinetics of infection reasonably well with our model (Figs 2 and 3), we can now turn off transcription of phage psbA (k_PpsbA = 0), and study how this affects the predicted number of phage genomes in infected cells after 8 h of infection. Using the parameter values presented in Table 1, we predict that a phage unable to express psbA would produce 2.81% fewer genomes after 8 h of infection. However, if a phage did not encode psbA, its genome would be shorter, by approximately

![Figure 2](image_url). Measured (data points) and modeled (lines) levels of host and phage psbA transcripts (A) and D1 protein product (B) during the lytic cycle of infection of Prochlorococcus MED4 by cyanophage P-SSP7. For modeled levels of D1 protein, solid lines represent the sum of functional and damaged D1, and dotted lines represent functional D1 only. Data are from [17]. Data for host expression levels were transformed assuming that 50% of cells were infected [17].

**Figure 2.** Measured (data points) and modeled (lines) levels of host and phage psbA transcripts (A) and D1 protein product (B) during the lytic cycle of infection of Prochlorococcus MED4 by cyanophage P-SSP7. For modeled levels of D1 protein, solid lines represent the sum of functional and damaged D1, and dotted lines represent functional D1 only. Data are from [17]. Data for host expression levels were transformed assuming that 50% of cells were infected [17].

doi:10.1371/journal.pone.0003550.g002

**Figure 3.** Measured (data points) and modeled (lines) genome copies of cyanophage P-SSP7 during the lytic cycle under light (25 μE m⁻² s⁻¹) and dark conditions. Genome copies were measured as genomes per ml of culture [17], and were transformed to a per cell basis for comparison to the model.

![Figure 3](image_url)
1080 bp. Taking this into account, a phage that did not encode phd4 would produce only 0.53% fewer genomes after 8 h of infection than a phage that encodes and expresses phd4. In the dark, where there is presumably no advantage to expressing phd4, a phage without this gene is predicted to produce 1.97% more genomes than a phage with it.

Ideally, we would like to consider the consequences of phd4 to phage genome replication under different and changing levels of irradiance. However, many of the parameters used in our model are likely to change as a function of irradiance, in ways that can be difficult to predict (see [29]). Therefore we limit ourselves to one specific case, where cells are moved from 25 μE m⁻² s⁻¹ to 50 μE m⁻² s⁻¹ one hour after infection has begun, when the capacity of the cells to respond to the changing light may be largely compromised by infection. This means we can use the same initial conditions and parameter values as in our previous simulations (Table 1), except for two parameters that will be affected directly by the increased irradiance (δ₂[dim] and δ). We assume that the rate of damage to functional PSII subunits (κD1dam) increases proportionally with irradiance (i.e., δ₂[dim] is doubled; [26]), and that the rate of photosynthesis of Prochlorococcus MED4 is greater at 50 μE m⁻² s⁻¹ than at 25 μE m⁻² s⁻¹ by a factor of approximately 1.75 [33,34].

We found that in the case where irradiance increases from 25 μE m⁻² s⁻¹ to 50 μE m⁻² s⁻¹ one hour after infection, a phage that does not express phd4 is predicted to produce 4.31% fewer genomes than a phage that does. Here, a phage that does not express or encode phd4 is predicted to produce 21.0% fewer genomes than a phage that does encode and express phd4. We therefore predict that phd4 will have a greater impact on phage genome replication under this switch to a higher level of irradiance, such as could occur in the surface mixed layer of the oceans. We note that this prediction also holds if photosynthesis (and δ) increases by a factor of either 1.5 or 2 under the switch to higher irradiance, rather than by a factor of 1.75 (see Text S1). Further, we performed analyses similar to the above using a range of different values of δ₂[dim], λ, and ψ[phd4(0)]. Across these simulations, expressing phd4 usually led to a modest increase in genome replication in continuous light (of between 2.5 and 4.5%), though this increase was typically smaller (between 0.3 and 2.3%) when the cost of encoding phd4 was considered. The predicted advantage of expressing and encoding phd4 was typically greater when a switch to higher light was simulated during infection, though the precise size of the advantage conferred by phd4 varied (see Text S1).

Cyanophage dNTP diets

In addition to phd4, marine cyanophages encode a variety of genes that potentially help them acquire dNTPs (e.g., ribonucleotide reductase, transaldolase; see [35]). It is therefore interesting to ask more broadly: Under what set of circumstances can a phage increase its total genome replication by encoding an additional gene or module of genes that help it acquire extra dNTPs?

Consider a phage that encodes genes that allow it to access dNTPs from a single source, s₁ (e.g. scavenging from the host genome). Now, a mutant acquires an extra gene or module of genes that allow it to access an extra source of dNTPs, s₂. If genes needed to access this second source of dNTPs elongate the wild type genome, we want to know when the mutant will make more genomes than the wild type by some time post-infection, or when

\[ G_M(t) > G_W(t) \]  (10)

where \( G_M(t) \) and \( G_W(t) \) are the numbers of genomes in cells infected by mutant and wild type phages (respectively) at time \( t \).

We explore this question using our original model as a starting point, but with modifications that allow it to be studied analytically. We assume that the supply of dNTPs is highly limiting to genome replication (\( N < K_m \)), such that we can rewrite equation (1) as

\[ \frac{dN}{dt} \approx \frac{1}{L} \sum_{i} \nu_i \cdot N \]

where \( V_R \) is the rate of DNA elongation (here in bp dNTP⁻¹ h⁻¹ cell⁻¹). We also assume that the dNTP revenue from each phage-encoded source can be expressed as a function of time, such that a phage using \( s_1 \) and \( s_2 \) has \( \frac{dN}{dt} = s_1(t) + s_2(t) - \frac{2V_R}{N} \). In reality, processes of phage-encoded dNTP acquisition (\( s_1 \) and \( s_2 \)) and genome replication may begin at different times post-infection. For simplicity, we assume these processes all begin at the same time (\( t_0 \) hours after infection), and let time \( t = 0 \) in this model refer to this time \( t_0 \) when these processes begin. Solving for \( N(t) \), and then \( G(t) \) yields

\[ G_P(t) = 1 + \frac{1}{2L_P} \left[ N(t_0) \cdot (1 - \exp(-2V_R t)) + s_1(t) \right] \\
+ \frac{1}{2L_P} \left[ (1 - \exp(-2V_R t)) + s_2(t) \cdot (1 - \exp(-2V_R t)) \right] \]  (11)

where \( N(t) \) is the number of dNTPs in the cell at time \( t \); \( s_1 \) is the amount of dNTPs contributed / cost in genome length, and \( s_2 \) represents a covolution product. If we assume \( N(t_0) \) is very small (\( N(t_0) = 0 \)) and sub (11) into (10), we get

\[ \frac{1}{2(L_R + L_1 + L_2)} \left[ \Psi_1 + \Psi_2 > \frac{1}{2(L_R + L_1)} \right] \]  (12)

where \( \Psi_1 \) and \( \Psi_2 \) represent dNTPs derived from \( s_1 \) and \( s_2 \) (respectively). \( L_2 \) and \( L_1 \) represent the length of genes needed to encode \( s_1 \) and \( s_2 \) (respectively) and \( L_R \) represents the length of the rest of the genome. This can be expressed as \( \frac{\Psi_1}{\Psi_2} > \frac{L_2}{L_2 + L_1} \)

meaning that a new module of genes will increase phage genome replication if it leads to a proportional increase in dNTP production that is greater than the proportional increase in genome length that it causes. Alternatively, we could say that for a new module of genes to increase phage genome replication, it must have a ratio of dNTPs contributed / cost in genome length, \( \frac{\Psi_2}{\Psi_1} \)

that exceeds a threshold, \( \frac{\Psi_1}{\Psi_2} \).

While this model is oversimplified, and has required assumptions that limit its applicability, it nevertheless may help us to understand some of the variability among cyanophages in methods they use to acquire dNTPs. For example, it suggests that if two similar phages acquire dNTPs by scavenging from the genomes of their hosts (i.e., their \( s_1 \)), but one phage infects a host with a smaller genome from which fewer dNTPs can be produced (smaller \( \Psi_1 \)), this phage might be more likely to exploit an additional source of dNTPs, if given the opportunity. This may be one factor that helps to explain why cyanophages infecting Prochlorococcus, which has a very small genome, might encode genes that help acquire dNTPs from other sources (see [36] for discussion of related issues).

Further, it can be shown that if the new source of dNTPs, \( s_2 \), is highly profitable, the phage may no longer be advantaged by encoding \( s_1 \). We would expect this to be the case when

\[ \frac{1}{2(L_R + L_1)} \left[ \Psi_2 > \frac{1}{2(L_R + L_1)} \right] \]  (13)

or when \( \frac{\Psi_1}{\Psi_2} > \frac{L_2}{L_2 + L_1} \). This illustrates a way in which one source of dNTPs could replace another in the genome of a phage, over evolutionary time.

This analysis has strong parallels with diet theory models that predict when a foraging animal should incorporate an encountered prey item into its diet, based on the energetic gain from the prey item, balanced against the cost in terms of time of pursuing it [37,38]. It thus adds to an impressive list of circumstances in which
phage strategies can be understood using analogies to theory developed for foraging animals (e.g. [39,40]).

**Discussion**

The goal of this simple modeling exercise was to predict the advantage conferred to a cyanophage of carrying and expressing the psbA gene. More specifically, we consider the hypothesis that phage psbA expression augments the photosynthetic apparatus of the host during infection, following the decay of host psbA transcripts, and we do not consider possible alternative or additional advantages of phage psbA. We have intentionally oversimplified the complex processes of infection, photosynthesis and dNTP synthesis in an effort to match the model to the scope and resolution of the available data. The modeled predictions serve as hypotheses to be tested when the means to knock out specific genes in these cyanophage genomes are eventually developed.

First, we predict that under low continuous irradiance, phage psbA expression increases phage genome replication, and potentially phage fitness, relative to a ‘mutant’ that does not contain this gene. This advantage is substantially reduced, however, if one accounts for the cost to the phage of elongation of the cyanophage P-SSP7 genome by psbA. Second, we predict that the slight advantage conferred by phage-encoded psbA may be greater under conditions of light stress, such as an increase in irradiance during infection. This is due to the more rapid decay of host D1 proteins at higher irradiance, and could contribute substantially to the advantage conferred by psbA to cyanophage P-SSP7 in the dominant habitat of this particular *Prochlorococcus* host — the surface mixed layer of the ocean. Finally, the model predicts that during infection in the dark, where there is presumably no advantage to expressing psbA, encoding psbA would result in a net decrease in genome replication of approximately 2%. Taken together, these results illustrate how the benefits of psbA to cyanophage genome replication may vary substantially among infections that occur at different times over the diel cycle, or for cells that are subject to different conditions of irradiance due to mixing [18]. These are all testable hypotheses.

It is clear that the selective advantage of psbA to phage will be determined by the benefit it confers during all conditions under which infection occurs, weighted by their frequency of occurrence. Therefore to fully understand the fitness consequences of carrying the psbA gene to phage, we will need to better understand how psbA influences genome replication over a much broader range of conditions, including at different times over the diel cycle where properties of host photosynthesis will change dynamically [29] and hosts will contain different numbers of genomes [41] and free dNTPs. To connect these predictions for genome replication to fitness, we will also need a better understanding of when genome replication limits phage burst size (e.g. see [24,42,43]) and of the interactions between burst size and other factors, such as the timing of cell lysis, and the availability and quality of hosts (e.g. [39,40,44–47]).

We have learned recently that marine cyanophage encode a number of genes that are absent in the genomes of non-marine phages and share homology with genes involved in microbial metabolism [11]. As we attempt to understand both the evolutionary significance of these genes and the distribution of phage genes in the ocean [35], it will be useful to have theoretical tools. To begin building such tools, here we have developed a model exploring the selective advantage of one specific gene of host origin that is commonly encoded by marine cyanophages, as well as more general tradeoffs between acquiring dNTPs and elongating the genome. We hope these models will form the basis for a more powerful and predictive modeling framework, and contribute substantially to our understanding of phage dynamics in marine microbial communities.

**Supporting Information**

**Text S1** Description of simulations using a range of different parameter values. Found at: doi:10.1371/journal.pone.0003550.s001 (0.34 MB PDF)

**Acknowledgments**

We thank S. Abedon, D. Campbell, D. Lindell, M. Follows and S. Dutkiewicz for valuable comments on earlier versions of this manuscript. We thank M. Sullivan, D. Lindell, D. Endy and members of the Chisholm lab for helpful discussions of this work.

**Author Contributions**

Conceived and designed the experiments: JGB. Analyzed the data: JGB. Wrote the paper: JGB SWG.

**References**


