Characterization of Phage Fitness through Single-Phage Assays

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science with Honors in the Computational Biology concentration of Brown University

Jeffrey Yuan

First Reader: Professor Daniel Weinreich

Second Reader: Suzanne Sindi
Characterization of Phage Fitness through Single-Phage Assays

Abstract

Bacteriophage (phage) life history traits include lag time, and burst size. To characterize the distributions of these traits, individual $\phi X174$ virions were isolated in each sample and combined with E. coli at the start time. The number of phage present in each sample was assayed at various time-points—samples in which many phage were observed were denoted bursts. No bursts were observed before a certain time-point, and the proportion of phage that burst was seen to increase over time, which was modeled with a cumulative exponential distribution. The size of bursts was seen to fit a normal distribution, and that distribution was shown to depend on time. The distributions that model burst size and lag time can be used to determine the variance in these two traits, which has significant implications for modeling phage evolution and the characterization of phage adaptation.
Acknowledgements

I would like to thank the Undergraduates in Biology and Mathematics (UBM) Program which gave me the opportunity to start working in Professor Weinreich’s lab, with the wonderful assistance of Meghan Baker and Diane Proctor. Working with the two of them along with Professor Weinreich helped me immensely in adjusting to working in a lab, becoming familiar with all the standard procedures, and not feeling completely on my own when I was utterly lost in what I was supposed to be doing. The other members of the lab were also immensely helpful in my transition into the lab: Fei Cai, Nicole Demari, Ayoosh Pareek—thanks for making overnights and taking things out of the incubator at odd hours, and especially Jen Knies who I pestered with questions all the time. I would like to thank Paul Joyce and Craig Miller for performing the analyses on the data, coming up with the statistical models and testing them, and coming through for us whenever we needed more analysis done. Without them, this thesis would basically be lacking a results or discussion section. Thanks also to Suzanne Sindi for agreeing to read through this thesis and being so eager to hear about the progress we’ve made in the course of the year. I cannot express how grateful I am to Chris Baker, who always took the initiative in starting chores, arranging times to meet in the lab, setting up experiments, and working with me side-by-side through all of the titers and all of the experiments this year. Finally, I would like to thank Professor Weinreich immensely for teaching me all about the lab, patiently answering all of my questions, being so willing to set aside time to meet with us and make sure that we’re on the right track, accepting the mistakes that we make every once in a while, and most of all, for always inspiring to look at these biological phenomena as fascinating problems to figure out and infusing so much encouragement and energy into this project. Thank you for helping me learn how to be a real scientist, push through all of the bumps in the road, carefully assess the reliability of our data, and showing me what a real passion for science looks like.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>2</td>
</tr>
<tr>
<td>Background/Introduction</td>
<td>4</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>7</td>
</tr>
<tr>
<td>Results/Analysis</td>
<td>15</td>
</tr>
<tr>
<td>Discussion/Conclusion</td>
<td>27</td>
</tr>
<tr>
<td>References</td>
<td>31</td>
</tr>
</tbody>
</table>
**Background/Introduction:**

In recent years, bacteriophage (phage) has been used as a model organism to study the mechanisms of evolution (Abedon 2009; Bull *et al.* 1997, 2000, 2004; Bull 2006; Burch and Chao 1999, 2000; Heineman and Bull 2007; Patwa and Wahl 2008; Wang *et al.* 1996; Wang 2006). Its short generation time, easily manipulated simple genome, and relatively simple life history make it an excellent candidate for the study of adaptation and the fixation of mutations. One of the approaches to understanding evolution is through the life history of an organism (Abedon 2009; Bull *et al.* 2004; Bull 2006; Heineman and Bull 2007; Patwa and Wahl 2008; Wang 2006). Life history characteristics refer to key factors in the processes involved in evolutionary fitness such as age of reproduction, length of reproductive fertility and number of offspring. Commonly, there are tradeoffs implicit between the optimization of different characteristics, and the environment determines the allotment of each characteristic that is most adaptive (Burton *et al.* 2010). Life history characteristics relevant to the fitness of the phage include adsorption rate, lag time, and burst size (Abedon 2009; Hyman and Abedon 2009; Patwa and Wahl 2008; Wang *et al.* 1996; Wang 2006). In the life cycle of a phage, the adsorption rate refers to the rate at which the phage is able to find and adsorb to the surface of a bacteria cell; the lag time refers to the amount of time between the injection of phage DNA into the cell and the lysis of the cell after the replication and assembly of phage progeny within the cell; and the burst size refers to the number of phage progeny released from a lysed host cell for an individual phage. For the purposes of this study, we restrict our attention to lag time and burst size.

Many prior studies have attempted to model the effects of burst size and lag time on phage fitness (Bull 2006; Heineman and Bull 2007; Patwa and Wahl 2008; Wang *et al.* 1996; Wang 2006). It has often been assumed that there is a tradeoff between an early lag time and a high burst size (which has been shown for a few phage species, see Heineman and Bull 2007). Thus, one or the other will be favored by the phage at different times, depending on the environment. The marginal value theorem of optimal foraging theory has been applied to these characteristics to determine how the phage will evolve to achieve the lag time and burst

However, the distributions for burst size and lag time has not been well established in these models—the models tend to assume either a constant burst size and lag time (Patwa and Wahl 2008) or a simple linear relationship between burst size and lag time (Heineman and Bull 2007; Wang et al. 1996, Wang 2006). None of the models develop a statistical model for burst size and lag time, which includes an account of the variance expected in those factors. However, Patwa and Wahl (2008) have shown that, in the context of population bottlenecks, mutations in the (assumed constant) lag time are most likely to reach fixation. Based on these results, we would expect that the variance in lag time would have a large effect on the fitness of the phage, and mutations that affect the variance in lag time would also affect the adaptiveness of the phage. Before these effects can be tested, however, the variance in lag time would have to be experimentally determined, which can occur through the characterization of burst size and lag time.

The burst size and lag time of various phage has been measured in a large host of previous studies (Delbruck 1940, 1945; Ellis and Delbruck 1939; Gutierrez et al. 2010; Heineman and Bull 2007; L. Lin et al. 2011; N. Lin et al. 2010; Wang 2005, 2006). However, in most of these studies, these measurements were done for large populations of phage (Gutierrez et al. 2010; Heineman and Bull 2007; L. Lin et al. 2011; N. Lin et al. 2010; Wang 2005, 2006). These bulk assays only reveal the mean burst size and lag time, not the variance of these factors. Moreover, the calculations for mean burst size do not take into account the question of whether or not all the phage have burst. Simply dividing the number of phage in the sample after bursts have occurred by the number of phage in the initial sample disregards the phage that have not yet burst at the time of sampling—it results in the mean burst size per total phage rather than the mean burst size per phage conditioned on bursting, which is the true number of progeny resulting from the infection of a single phage. In order to obtain the burst size per phage conditioned on bursting, we measured the size of bursts resulting from individual phage through single-phage assays of φX174. By performing single-phage assays rather than bulk assays, we were also able to
characterize precise distributions for the burst size and lag time of φX174, which reveals both the mean and the variance of these two characteristics.¹

¹ Single-phage assays have been performed in previous studies (Delbruck 1940, 1945; Ellis and Delbruck 1939). However, those studies did not attempt to determine the variance or the distributions of burst size and lag time.
Materials and Methods:

Experimental Procedures:

Virus and Bacteria stocks:
A stock of genetically identical bacteriophage ϕX174, strain 300aI, was kindly provided by Art Poon and Lin Chao, and maintained at −20°C (Poon and Chao 2005). A titer of this stock showed it to contain on the order of $10^8$ phage per mL. Initially, $10^{-6}$ dilutions were maintained as stocks in the fridge at 2°C. However, by performing daily titers on four different phage stocks, it was found that phage stocks with 40% glycerol kept in the freezer at −20°C had a lower variance in titer size over time (Fig. 1).

E. coli C were used as the host cells of ϕX174. E. coli were streaked out onto LB plates and incubated at 37°C overnight, before being stored in the refrigerator at 2°C. Each plate of E. coli colonies was used for about a month before new plates were streaked out.

Figure 1 – Phage Titer Data
The change in titer of four different phage stocks over time. In the legend, CG refers to a stock of phage with normal LB (see below) kept in the fridge (2°C); GG refers to a stock of phage with 40% glycerol in the LB kept in the fridge (2°C); CZ refers to a stock with normal LB kept in the freezer (-20°C); GZ refers to a stock with 40% glycerol in the LB kept in the freezer (-20°C). The GZ stock is the most stable over time.
**LB broth and plates:**

*Liquid LB broth (LB)* – 20 g/L Difco™ LB Broth, Lennox. 1 part per thousand of 2 M CaCl₂ was added before use.

*Soft agar (TA)* – LB with an additional 7 g/L Bacto™ Agar. Before use, it is microwaved until all the agar is dissolved, supplemented with 1 part per thousand of 2 M CaCl₂, and allowed to cool. After use, leftover TA is kept at 42°C.

*LB plates* – LB with an additional 15 g/L Bacto™ Agar, poured into Petri dishes.

**Liquid cultures of E. coli used:**

*Stationary phase E. coli* were made the day prior to any titer or experiment. A single colony from the current E. coli plate was put into 10 mL of LB and incubated while shaking at 37°C overnight. It was found that stationary phase E. coli do not burst in the given conditions so exponential phase were used.

*Exponential phase E. coli* were made from a dilution (either 100X, 50X, or 30X) of the stationary phase E. coli—usually 100 µL of the stationary phase E. coli with 10 mL of LB or 250 µL of E. coli with 25 mL of LB for the 100X dilution—and incubated while shaking at 37°C until it reached an optical density (OD) of 0.6 when it was used. If the OD had reached 0.6 prematurely, then a further dilution was done, reducing the OD to 0.4, and the liquid culture was allowed to incubate at 37°C until the OD again reached 0.6.

**Counting Plaques:**

Counting plaques refers to the standard method of determining the number of phage in a sample (Ellis and Delbruck 1939; Hyman and Abedon 2009; Krueger 1930). Plates that contain the sample of phage, E. coli, and TA and that have been incubated for 3 hours are observed to have transparent, circular holes or “plaques” in the surface of the plate or “lawn” of E. coli, and each plaque correspond to an individual phage.

**Making a titer of a bacteriophage stock:**

100 µL of exponential phase E. coli was combined with 10 µL of phage stock in one well of a 96-well plate in a heat block at 44°C and allowed to sit for 1-2 minutes. The phage-bacteria mixture was pipetted into a test tube
containing 3 mL of TA heated to 42°C, vortexed, and then plated onto an LB plate. Triplicate titers of each stock were done. The plates were then incubated at 37°C for three hours before being stored in the fridge at 2°C. The titer of each replicate was determined by counting the number of plaques in each plate.

**Spot-Titering Procedure:**
The phage stock was diluted to an expected average of one phage per spot (5 µL) based on a recent titer.

200 µL of exponential phase E. coli were added to 3 mL of TA kept at 42°C and poured onto each LB plate used.

5 µL of the diluted phage stock were pipetted on top of the lawn of E. coli and TA for each spot. About 25 spots were made in a grid-like fashion for each plate. Plates were then incubated for 3 hours at 37°C and then the numbers of plaques per spot were counted.

**Time-series Sampling (Raging Tiger) Procedure:**
For each trial, the dilution of the phage stock required to achieve the desired average number of phage per sample (50 µL) was calculated based on the previous day’s titer of the stock, and enough of that dilution was made to run the trial. A 96-well plate was kept on a heat block at 44°C, and a rack of test tubes containing 3 mL of TA each was kept on a heat block at 42°C. Exponential phase E. coli were added to each well (100 µL), with a multi-channel pipet. Next, 50 µL of phage were also added into each well; the time was started as soon as the phage were added to the first column with the multi-channel pipet. The time lapse of when the phage were added between adjacent columns is about 15 seconds, so each successive column is 15 seconds earlier than the previous one. Although the start and end time of each trial varied, one well was sampled (see below) every 30 seconds (correcting for the 15-second lapse between columns) to generate data for each time point. After sampling, the plates are incubated at 37°C for 3 hours before being stored in the fridge at 2°C. The number of plaques in each plate were then counted and recorded.
Single Time-point Sampling (Black Cobra) Procedure:
For each trial, a well-plate is prepared as in the time series sampling with 100 µL of exponential phase E. coli along with 50 µL of phage in each well of a 96-well plate (the samples of phage diluted the appropriate amount to achieve the desired average). The time is started as soon as phage is added to the first column of wells with E. coli with a multi-channel pipet. The E. coli-phage mixtures are allowed to sit until the point in time of interest. At that point, the entire well-plate is moved from the hot plate into an ice bath, after which each well is sampled (see below). The resulting plates are incubated for 3 hours at 37°C, and then stored in the fridge at 2°C. A modification was made in certain trials to correct for the 15-second intervals between columns: rather than transferring the whole well-plate to the ice bath, columns of samples were pipetted from the well-plate in the heat block to another well-plate in the ice bath, thus staggering the transfer of each column by 15 seconds and correcting for the time difference. The number of plaques in each plate were then counted and recorded.

Sampling of wells:
Each well contained 100 µL of E. coli and 50 µL of phage. When a well was ready to be sampled, the whole 150 µL was extracted, added to a test tube with 3 mL of TA kept at 42°C, vortexed briefly, and plated onto an LB plate.

Statistical Procedures:
These procedures were developed by our collaborators, Paul Joyce and Craig Miller, specifically to analyze the burst size and lag time of bacteriophage.

Initial Comprehensive Maximum Likelihood Model:
Initially, it was assumed that the burst size of each phage follows a normal distribution with mean $\mu$ and standard deviation $\sigma$ and the lag time follows a Weibull distribution with parameters $\lambda$ and $\alpha$. The phage input for potential bursts follows a Poisson distribution with parameter $\beta$. Using this model, we attempted to find the
distribution for $C_t$, which is the number of phage at a certain time $t$. The probability density function of $C_t$ was found to be:

$$f(c_t|\alpha, \lambda, \mu, \sigma^2) = P(N_t = 0) + \sum_{j=1}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(c_t-j\mu)^2}{2j\sigma^2}} P(N_t = j)$$

where $N_t$ is the number of bursts at time $t$. $N_t$ is a latent variable which means that it is unobserved but found in the likelihood of the data. The probability that $N_t = k$, given the model, was found to be:

$$P(N_t = k) = \begin{cases} e^{-\beta w_t} (w_t \beta)^k / k! & k > 0 \\ 1 - e^{-\beta} & k = 0 \end{cases}$$

where $w_t = P(T \leq t) = 1 - e^{-\left(\frac{t}{\lambda}\right)}$. Thus, given $r = 1,2,\ldots, m$ as the replicate number and $c_{t_i,r}$ as the number of phage at time $t_i$ for replicate experiment $r$, the likelihood of the data can be calculated as:

$$L(\alpha, \lambda, \mu, \sigma^2) = \prod_{r=1}^{m} \prod_{i=1}^{k} f(c_{t_i,r}|\alpha, \lambda, \mu, \sigma^2)$$

Thus, these parameters would be found from the data either through a direct maximum likelihood calculation (which would be very difficult given the complexity of the expressions involved) or through an expectation-maximization method (EM). When applied to real data, however, it was found that this model is unstable for high variance in burst size. Thus, a new model was made in which estimators for burst size (mean and variance) and lag time are determined separately.

**Heuristic Mode-based Estimator of Burst Size:**

The mode-based model considers that the mode of all of the phage counts of the bursts can be an estimator for the mean of the burst size ($\mu$) of a single phage (Fig. 2).
The dotted line in each plot represents the mode. Each successive peak in a given plot represents an increasing number of bursts, which occurs with an increasing initial number of phage in each sample. The Poisson character of the number of phage per sample results in the decreasing densities of successive peaks—the density of a certain peak corresponds to the frequency of having that many initial phage in the sample given a Poisson distribution with mean 1.0 ($\beta = 1.0$). Therefore as long as $\sigma$ is smaller than $\mu$ and the Poisson input has a higher frequency of outputting 1's than 2's ($\beta > 2$), the mode of the bursts is a good estimator of the mean of the burst size. From this estimated mean, all of the bursts to the left of the mean can be seen as the left half of a normal distribution, which thus provides as estimate of the variance of burst size. To test how well this estimator predicts $\mu$ and $\sigma$, both of these parameters were estimated from simulations with varying variance (Fig. 3).
Maximum Likelihood Estimator for Determining the Lag-Time Parameters:

It was found that there was no significant difference between applying a Weibull distribution for the lag time variable and applying a simpler, exponential distribution with input parameter $\lambda$. Thus, rather than using a Weibull distribution, the lag time was assumed to follow an exponential distribution after an initial time when bursts started occurring, $t_0$. In this new model, $t_0$ was estimated as 30 seconds prior to the first time-point when a burst is observed. Furthermore, bursts are considered only to be a binary event, with each burst event assumed only to involve a single phage. Following these assumptions (including the assumption that the input variable is Poisson), the likelihood of the data is found to be

$$
\ln(L) = \sum_{d=1}^{n\text{days}} \sum_{t=1}^{n\text{time}} \sum_{r=1}^{n\text{reps}} -\beta_d w_t (1 - Y_{dtr}) + Y_{dtr} \ln(1 - e^{-\beta_d w_t})
$$

$$
+ \sum_{d=1}^{n\text{days}} \sum_{t=1}^{n\text{time}} \sum_{r=1}^{n\text{reps}} I[N_t = 0](-\beta_d (1 - w_t) + m_t \ln(\beta_d (1 - w_t)))
$$
where $\beta_d$ is the Poisson parameter that dictates how many phage are in a well for all the assays of a given day; $Y_{dtr}$ is a dummy variable that takes on value 1 when a well has a burst event (a count > 10 in our analysis), and 0 otherwise; $m_t$ is the number of unburst phage observed in a well at time $t$; and $w_t$ is the probability that a phage burst by time $t$, given by:

$$w_t = 1 - e^{-\frac{(t-t_0)}{\lambda}}$$

$\lambda$ is estimated by finding the likelihood for all $\lambda$ values from 1 to 10 with an interval of 0.1 and determining which $\lambda$ value maximizes the likelihood. In applying this estimator to all of the data combined, rather than estimating $t_0$ from the earliest burst time, both $\lambda$ and $t_0$ were jointly estimated by maximizing the likelihood of both values together. Thus, estimators for the parameters characterizing lag time, including $t_0$ and $\lambda$ were determined.

**Results/Analysis:**

We sought to determine the distribution of burst size and lag time in $\phi$X174 through single phage assays. To accomplish this task, we tried to isolate an individual phage in each sample. Due to the stochastic nature of dilutions, we expect the actual number of phage in each sample to follow a Poisson distribution. Thus, given a titer of the phage stock, a dilution can be made to attain an expected number of phage per sample—the value of the $\beta$ input to the Poisson distribution—and the Poisson distribution can then be used to calculate the probability that there are $k$ phage in each sample. The Poisson distribution is given as:

$$P(Y = k) = \frac{e^{\beta} \beta^k}{k!}$$

$E(Y) = \beta$

$V(Y) = \beta$

**Spot-Titering:**

First, we tested whether the number of phage in each sample did indeed follow a Poisson distribution. One round of spot-titering was done (with 96 spots), and the number of plaques was counted for each spot. Although the expected number of phage per spot was 1.0, the mean calculated from the number of plaques was
2.02. This simply means that either the dilution was not done precisely or the titer on which the calculation was based was not accurate; nevertheless the results should still be Poisson. The frequency of observing \( k \) plaques per spots was calculated and compared with the theoretical frequency of a Poisson distribution with \( \beta = 2.02 \) (Fig. 4). The number of plaques per spot was not found to be statistically significantly different from the Poisson distribution using either a Chi-squared test \( (p = 0.98) \) or a G-test \( (p = 0.80) \). Although the experiment did not have enough power to reject the hypothesis that the number of phage in each sample obeys a Poisson distribution, the close fit of the observed frequencies to the expected frequencies in this trial as well as in all other trials run provides an adequate basis for assuming that the number of phage is indeed Poisson.

**Raging Tiger Results:**

Raging Tiger trials indicate the time when bursts begin to appear (initial burst time) and the size of bursts (Fig. 5). In this trial, bursts begin at 19.5 minutes. Prior to that time, the number of phage observed is expected to follow the Poisson distribution. Thus, those trials can be used to obtain the average number of phage inputted per well in this trial, which is the Poisson \( \beta \) for this trial.
After the time-point when bursts start appearing, observations fall in three different categories. Bursts are observations that are greater than would be expected to appear based on the Poisson $\beta$ for this trial (observations of more than ten phage are generally considered bursts). Observations of zero occur in samples that did not contain any phage and so would not be expected to ever contain bursts—these observations can also be used to determine the Poisson $\beta$, assuming the distribution is Poisson. Other observations within the range of the Poisson distribution for this trial, however, are expected to burst eventually. Thus, observations of one plaque (such as at 20.5 minutes) are phage that were sampled before their lag time. Such observations indicate that not all phage burst at the same time. Multiple Raging Tiger trials were done to better characterize the mean and variance of burst size and lag time (Fig. 6). It was found that initial burst times range from 16.5 minutes to 20 minutes and that burst sizes range from 22 to 338. However, initial burst times are skewed by the prevalence of samples with no phage, and high burst sizes may represent multiple bursts when more than one phage is inputted into a well.
Figure 6 – Compiled Raging Tiger Data
Compilation of all the data collected using the Time-Series Sampling procedure. Bottom portion of the graph represents observations of phage that have not burst while the upper portion represents bursts. The legend indicates the Poisson mean of the number of phage inputted into each well.
This data was analyzed with both the mode-based estimator for the burst size and the maximum-likelihood estimator for the lag-time parameters (the burst size variance was found to be too large to apply the initial maximum likelihood estimator for all parameters at once). Directly applying the mode-based estimator to all of the bursts observed in the data (and disregarding the time at which the burst occurred), the mean ($\mu$) and standard deviation ($\sigma$) in burst size were found (Fig. 7, $\hat{\mu} = 160.4$ and $\hat{\sigma} = 67.4$). The values of the parameters closely matched up to those found from the Late Black Cobra trials (see below), but the shape of the graph is different from expected based on simulations: there is a higher density of values lower than the mode than higher than the mode, which is the opposite of what is expected from the model.

**Figure 7 – Density Plot of Raging Tiger Data**

This plot contains the bursts of all the Raging Tiger trials (from Fig. 3). The mode-based estimator was used to determine the mean ($\mu$) and standard deviation ($\sigma$) of burst size. The dotted line indicates the mode—i.e. the estimate for $\mu$; all data points to the left of $\hat{\mu}$ were used to estimate $\sigma$ (values indicated on the graph). The data-points coded by time-sampled are plotted along the bottom of the graph.
The maximum-likelihood estimator was used to determine the time when phage begin to burst \((t_0)\) and the input parameter of the exponential distribution for lag time \((\lambda)\). Note that the time when bursts are observed is not the time when the bursts occur, but an upper bound on when the bursts occur. The actual lag time is thus unobservable (termed a latent variable) and must be estimated from the data using maximum-likelihood. Following the procedure outlined in the methods, values for the start time of bursts and the exponential distribution parameter for lag time were estimated (Fig. 8, \(\hat{t}_0 = 16\) min, and \(\lambda_{\text{unbiased}} = 5.33\) with 95% confidence interval between 3.72 and 7.27).

Since the Raging Tiger data include many time-points at which no bursts are expected, only a small fraction of the samples contained bursts—the high chance of observing no phage at these time-points further depressed the number of bursts observed. This low number may be responsible for the irregular shape of the density plot (Fig. 7). The low number of bursts observed in Raging Tiger trials makes them inefficient for determining the burst size distribution and detecting whether burst size depends on time of bursting. Thus, the Black Cobra trials were performed to accumulate more observations of bursts.

![Figure 8 – Maximum Likelihood Estimator of Lag Time Parameters for Raging Tiger Data](image)

The time at which bursts occurred for the Raging Tiger trials (as seen in Fig. 3) were used to determine the time at which bursts started \((t_0)\) and the input to the lag-time exponential distribution \((\lambda)\). Parentheticals following \(\lambda\) and dotted lines indicate a 95% confidence interval. \(t_0\) was estimated from the earliest time a burst was observed and \(\lambda\) was estimated with a maximum-likelihood model (both bias-corrected, values indicated on the graph).
Black Cobra Results:

Black Cobra trials were run to better characterize the distribution of burst size and to determine whether that distribution depends on time (Fig. 9). Two sets of trials were performed: Late trials were sampled at 26 minutes to accumulate more data on burst size, while Early trials were sampled from 16 to 20 minutes to determine the start time of bursts. These two trials will also be compared to test whether the model is accurate in assuming that the burst size distribution is independent of time (in which case later bursts will not have larger burst sizes than earlier bursts on average) and that the probability of observing bursts (as opposed to observing phage that have not yet burst) increases over time according to an exponential distribution.

Using the mode-based estimator, the mean ($\mu$) and standard deviation ($\sigma$) in burst size was found for both the Early set of trials (Fig. 10, $\hat{\mu} = 97.59$ and $\hat{\sigma} = 48.81$) and the Late set of trials (Fig. 11, $\hat{\mu} = 174.8$ and $\hat{\sigma} = 67.4$). Compared to the Raging Tiger trials, the shape of the density plots for both of these trials generally corresponds to the shape expected from the model (See Fig. 2, $\sigma = 60$). Although the Late Black Cobra

Figure 9 – Compiled Black Cobra Data
Single Time-Point Sampling procedures were followed as outlined in the methods. Early time-points were sampled between 16-20 minutes. Late time-points were sampled at 26 minutes. If samples contained too many plaques to count, those samples were counted as containing 300 plaques.

Using the mode-based estimator, the mean ($\mu$) and standard deviation ($\sigma$) in burst size was found for both the Early set of trials (Fig. 10, $\hat{\mu} = 97.59$ and $\hat{\sigma} = 48.81$) and the Late set of trials (Fig. 11, $\hat{\mu} = 174.8$ and $\hat{\sigma} = 67.4$). Compared to the Raging Tiger trials, the shape of the density plots for both of these trials generally corresponds to the shape expected from the model (See Fig. 2, $\sigma = 60$). Although the Late Black Cobra
parameter values line up very closely with those found for the Raging Tiger data, the parameter values for the Early Black Cobra data are considerably lower for both the mean and standard deviation, which suggests that burst size depends on time. All of the data was compiled to test this hypothesis as well as the normality of burst size (whether burst size actually fits a normal distribution).

![Figure 10 – Density Plot of Early Black Cobra Data](image)

This plot contains the bursts of the Early Black Cobra trials (light blue portion of Fig. 6). The mode-based estimator was used to determine the mean ($\mu$) and standard deviation ($\sigma$) of burst size. The dotted line indicates the mode—i.e. the estimate for $\mu$. All data points to the left of $\hat{\mu}$ were used to estimate $\sigma$ (values indicated on the graph). The data-points coded by time-sampled are plotted along the bottom of the graph.

![Figure 11 – Density Plot of Late Black Cobra Data](image)

This plot contains the bursts of the Late Black Cobra trials (dark blue portion of Fig. 6). The mode-based estimator was used to determine the mean ($\mu$) and standard deviation ($\sigma$) of burst size. The dotted line indicates the mode—i.e. the estimate for $\mu$. All data points to the left of $\hat{\mu}$ were used to estimate $\sigma$ (values indicated on the graph). The data-points are plotted along the bottom of the graph.
**Compilation of All Data:**

All of the data was combined and then analyzed as one dataset to determine overall estimates of the parameters and to test the assumptions of the models. First, the parameters for lag time—start-time of bursts ($t_0$) and exponential distribution parameter ($\lambda$)—were estimated using the maximum-likelihood estimator on all of the data (Fig. 12, $\hat{t}_0 = 14$ and $\hat{\lambda} = 5.81$). The bias of the estimator was tested by simulating 100 datasets with these values for $t_0$ and $\lambda$; by estimating these parameters for the simulations, the estimator was found to be approximately unbiased, with a slight overestimate for $t_0$ ($\hat{\lambda} = 5.97$ and $\hat{t}_0 = 14.51$, data not shown).

![Figure 12](image)

**Figure 12 – Maximum Likelihood Estimator of Lag-Time Parameters for All Data Combined**
The maximum likelihood estimator were used to calculate both the time at which bursts started ($t_0$) and the input to the lag-time exponential distribution ($\lambda$); the calculated values were $\hat{t}_0 = 14.0$ minutes and $\hat{\lambda} = 5.81$. The points represent the proportion of wells that contained bursts (among wells that contained phage), with the numbers indicating the number of wells with phage at each time point.

Next, the mean ($\mu$) and standard deviation ($\sigma$) of burst size were found by applying the mode-based estimator to all of the data (Fig. 13, $\hat{\mu} = 116.76$ and $\hat{\sigma} = 53.61$). It is important to note that by combining all of the data in calculating these parameters, we are assuming that all of the observations come from the same normal distribution—i.e. that burst size does not depend on time. However, this assumption, as shown below, may very well be erroneous. The shape of this density plot roughly corresponds to the shape of simulations of the mode-based estimator, which supports the model (See Fig. 2).
To statistically test the shape of the plot and verify its normality, we performed a qq-plot on all of the data combined (Fig. 14). We find that the data are roughly normal but with a larger number of smaller bursts than expected from the normal distribution.

**Figure 13 – Density Plot of All Data Combined**
The mode-based estimator was used to determine the mean ($\mu$) and standard deviation ($\sigma$) of burst size. The dotted line indicates the mode—i.e. the estimate for $\mu$. All data points to the left of $\hat{\mu}$ were used to estimate $\sigma$ (values indicated on the graph). The data-points coded by time-sampled and by dataset are plotted along the bottom of the graph.

**Figure 14 – Assessment of Burst-Size Normality**
On the left, a qq-plot is used to test the normality of the data. A straight line indicates that the data is completely normal. On the right, a density histogram is plotted (z-scores against frequency). The shape of the qq-plot and the high bars for z-scores of -0.5 and -0.75 indicate that there are more smaller bursts than expected from a purely normal distribution.
Finally, we tested whether the burst size distribution is a function of time, which was suggested by the different values for $\mu$ and $\sigma$ that were obtained from the Early and Late Black Cobra trials. Three different tests were run to determine whether burst size depends on time. In the first test, we consider that wells with two or more phage in them are more likely to have higher burst counts at later times due to the increased probability of more phage bursting. This bias is corrected for by transforming the data into $z$-scores, which accounts for the chance that more phage were in the well. Since the $z$-scores of all the data are shown to increase with time ($p < 0.0001$, data not shown), this test suggests that burst size increases as time increases.

In the second test, only the wells with observations less than the estimated mean $\hat{\mu}$ are plotted against time to eliminate the confounding effect of wells initially containing more than one phage. Again, the test suggests that burst size increases with time ($p < 0.05$, Fig. 15). The slope of the line in the second test suggests that the burst size increases by 2.2 phage per minute.

![Figure 15](image)

**Figure 15 – Regression of burst size against time for wells with smaller bursts**

The graph plots burst size against time for the bursts of all the datasets smaller than $\hat{\mu}$, the estimate for the mean ($n = 102$). The regression was found to have a positive slope of 2.23 with a $p$-value of $p = 0.011$, suggesting that burst size does increase with time.
In the final test, all of the observed bursts were divided into early bursts and late bursts based on the mean sample time, and an F-test is employed to determine whether these two datasets come from the same population. The F-test rejects the null hypothesis that the early and late bursts come from the same population, thus supporting the notion that the burst size distribution is a function of time ($p < 0.0005$, data not shown).

**Discussion/Conclusions:**

By sampling one phage at a time, we were able to characterize the distributions of burst size and lag time in our stock of φX174. From both the Raging Tiger and the Black Cobra data, we developed statistical models for characterizing both the burst size and the lag time. The burst size of φX174 was found to generally follow a normal distribution, the parameters of which can be estimated with the mode-based estimator. The probability of bursting at a certain time $t$ was found to initially be zero for $t < t_0$ and then gradually increase according to an exponential distribution—both $t_0$ and $\lambda$ of the exponential distribution can be estimated with a maximum likelihood estimator. Values for all the parameters involved were estimated from all the data combined, but not enough data was accumulated to be assured of the accuracy of these values.

The value obtained for mean burst size, ($\hat{\mu} = 116.76$, Fig. 13) is somewhat higher than that achieved for previous bulk assays of burst size for this strain ($\hat{\mu} \approx 100$, data not shown). A possible explanation for this observation is that the estimate for mean burst size achieved through single phage assays is the mean per phage conditioned on bursting whereas the estimate achieved through bulk assays is the mean per all phage in the sample, not conditioned on bursting. Since the bulk assays do not take into account the phage that do not burst by the time of sampling, the mean burst size observed in those assays will always be an overestimate of the burst size for the infection of an individual phage. This bias can be corrected for by dividing the result by the proportion of phage that are expected to burst by the time of sampling:

$$\mu_{\text{per burst phage}} = \frac{\mu_{\text{per total phage}} \cdot \# \text{burst phage}}{\# \text{total phage}}$$
This distinction has not been made clear in previous studies of phage burst size, but can be clearly observed when bursts of only single phage are counted. Nevertheless the previous bulk assays were not performed in the same conditions as these experiments, and so more bulk assays under these conditions have to be performed to test how well this distinction holds. Moreover, if burst size varies with lag time (discussed below), then only the means from the same time point can be compared with the equation above.

The observation that burst size was not entirely normal in the density plot for all of the data (Fig. 14) raises some questions about the validity of the model. Possible explanations for this observation include: (1) that burst size does not truly follow a normal distribution but some other similar distribution, (2) that there is a different biological phenomenon responsible for some of the smaller bursts, and (3) that burst size depends on time and thus it is not valid to assume that different quantities of data from different time-points can be plotted together. Three different statistical tests support the notion that burst size does depend on time, which makes (3) a likely candidate for explaining the observed departure from normality.

It is interesting to consider the implications of the observed 2.2 phage per minute increase in burst size over time (Fig. 15). Clearly, it cannot be the case that the rate of phage replication and assembly within the host is 2.2 phage per minute from the time of infection (approximated by \( t = 0 \)) because then there would only be 40 phage observed in bursts that occur at 18 minutes when there are many observations of more than 100 phage at that time. However, a 2.2 phage per minute increase is only the average rate of increase between the time \( t = 16 \) minutes and \( t = 26 \) minutes, so the rate of phage replication and assembly could be much higher at some time prior to \( t_0 \) and then decrease potentially due to the increased concentration of phage particles in the cell or the decreased rate of viral gene expression. The small size of this increase may also be due to the fact that only bursts smaller than the mean were plotted in the graph. Alternately, we must consider that bursts sampled at later time points may actually have burst already at earlier time points. The failure to account for this fact would also decrease the observed difference in burst size between bursts sampled at early time-points and
late time-points. Thus, the small rate of increase in burst size is likely not reflective of the overall rate of phage replication and assembly in the cell.

As previously mentioned, if burst size depends on lag time, then the tradeoff between the two factors has significant implications for phage fitness and evolution. Here we find that burst size does change with lag time, implying that under selection, the balance between short lag time and high burst size that optimizes fitness would be achieved. If the relation between burst size and lag time could be precisely defined and their relationship to growth rate known, then the optimal values for $\mu$, $\sigma$, $t_0$, and $\lambda$ could theoretically be determined. Determining these optimal values could lead to a model to describe the evolutionary trajectory of phage undergoing selection. Thus, characterizing the parameters involved in the fitness of phage is the first step towards predicting the evolutionary trajectory of bacteriophage. If these values are known, then further work elucidating the molecular and genetic mechanisms that produce certain values for these parameters would be required to be able to model the evolutionary pathway the phage would take.

It was also mentioned that variance in lag time can have a substantial impact on the fixation probability of beneficial mutations in populations that experience periodical bottlenecks. This result under conditions often encountered by phage shows that variance in lag time may have a greater effect on phage evolution than variance in burst size. Under the model that lag time follows an exponential distribution after an initial period of time $t < t_0$, the variance in the probability of bursting is given by $1/\lambda^2$. An interesting property of the exponential distribution is that the mean is equal to the standard deviation (both being $1/\lambda$). This implies that for a larger mean, the standard deviation is larger as well. Intuitively, one would expect that a mutation with a larger variance in lag time have a better chance of reaching fixation—it is more likely that at least one of the phage in the population will have an early burst time and beat out the wildtype phage. However, a larger mean also implies that the lag time would occur later on average, which would lower the fitness of the phage. Thus, the effect of the larger mean lag time would at least partially counter the effect of the larger variance. Further
studies regarding the tradeoff between later lag times and higher variances in lag time would have to be done to determine which one would be more beneficial to the phage.

Many aspects of the models described here must be tested further. Although the model for lag time seems to closely match the data, the actual values for the parameters must be confirmed with more data. More analysis and data collection must be done to determine whether burst size actually does vary with time—taking into account that lag time is a latent variable and does not correspond to time of sampling. Only after it is determined whether burst size depends on time can the distribution for burst size be accurately characterized (unless large amounts of data were collected for a single time-point, in which case the resulting distribution would only be valid for that time-point). If it is found that burst size does not depend on time, then the deviation from normality in the density plot for all the data must be explained (Fig. 14). If, as is more likely, it is found that burst size does change with time, the burst size distribution for individual time points must first be characterized with greater power (through more Black Cobra trials) and then the relation between the distributions could be used to determine how burst size changes with time.

Future studies that would extend the results of this paper include performing a bulk-assay of the phage stock with similar methods to verify the value for mean burst size estimated here, performing Raging Tiger and Black Cobra trials on mutant strains of φX174 to determine how the mutation affects the parameters of burst size and lag time, determining the effect of variance in lag time versus earlier lag time on the fitness of a phage strain, and determining the molecular and genetic mechanisms behind the burst size and lag time parameters described here.

Overall, statistical distributions for burst size and lag time in φX174 were characterized. Parameters relevant to these distributions were identified and estimated with empirically collected data and statistical methods. Although these results only represent preliminary steps towards a rigorous characterization of burst size and lag time, the parameters described here play an integral role in determining the fitness of
bacteriophage. The verification and elaboration of these distributions would lead to a clearer understanding of the factors involved in phage evolution.
References


