

Selection Affects Genes Involved in Replication during Long-Term Evolution in Experimental Populations of the Bacteriophage ϕ X174

Celeste J. Brown^{1,2*}, Jack Millstein^{1,2}, Christopher J. Williams^{2,3}, Holly A. Wichman^{1,2}

1 Department of Biological Sciences, University of Idaho, Moscow, Idaho, United States of America, **2** Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, Idaho, United States of America, **3** Department of Statistical Science, University of Idaho, Moscow, Idaho, United States of America

Abstract

Observing organisms that evolve in response to strong selection over very short time scales allows the determination of the molecular mechanisms underlying adaptation. Although dissecting these molecular mechanisms is expensive and time-consuming, general patterns can be detected from repeated experiments, illuminating the biological processes involved in evolutionary adaptation. The bacteriophage ϕ X174 was grown for 50 days in replicate chemostats under two culture conditions: *Escherichia coli* C as host growing at 37°C and *Salmonella typhimurium* as host growing at 43.5°C. After 50 days, greater than 20 substitutions per chemostat had risen to detectable levels. Of the 97 substitutions, four occurred in all four chemostats, five arose in both culture conditions, eight arose in only the high temperature *S. typhimurium* chemostats, and seven arose only in the *E. coli* chemostats. The remaining substitutions were detected only in a single chemostat, however, almost half of these have been seen in other similar experiments. Our findings support previous studies that host recognition and capsid stability are two biological processes that are modified during adaptation to novel hosts and high temperature. Based upon the substitutions shared across both environments, it is apparent that genome replication and packaging are also affected during adaptation to the chemostat environment, rather than to temperature or host per se. This environment is characterized by a large number of phage and very few hosts, leading to competition among phage within the host. We conclude from these results that adaptation to a high density environment selects for changes in genome replication at both protein and DNA sequence levels.

Citation: Brown CJ, Millstein J, Williams CJ, Wichman HA (2013) Selection Affects Genes Involved in Replication during Long-Term Evolution in Experimental Populations of the Bacteriophage ϕ X174. PLoS ONE 8(3): e60401. doi:10.1371/journal.pone.0060401

Editor: Anthony R. Poteete, University of Massachusetts Medical School, United States of America

Received: December 11, 2012; **Accepted:** February 26, 2013; **Published:** March 22, 2013

Copyright: © 2013 Brown et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funding for this research was provided by grants NCCR P20RR16448, 8P20GM103397 from the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: celesteb@uidaho.edu

Introduction

Determining the biological processes that are affected when organisms adapt to their environment is an important step in understanding the evolutionary process. The molecular changes that arise in DNA or protein sequences during adaptation affect phenotypic traits and fitness via these biological processes. Following the connections from genome sequence to biological processes and fitness in complex organisms is difficult and often relies upon finding a preponderance of adaptive substitutions in genes affecting the same biological process [1–3]. Another way to make connections between sequence, process and fitness is by experimental evolution [4]. For example, experimental adaptations of *Escherichia coli* have related increased fitness to global changes in genome sequence and transcriptional networks [5–7]. However, even the *E. coli* genome is so large and complex that the results of fine-scale dissection of the biological processes involved in adaptation are often ambiguous unless the selection is very specific [8,9].

Bacteriophages and their hosts are a common model for experimental evolution studies [10–18]. The size of phage genomes, ease of culturing and rapid generation time make them ideal for studying the molecular basis of adaptation in a selective

environment. The small microvirid phage, ϕ X174, is particularly suitable because its genome is exceptionally small, its genome structure is simple, and its molecular and structural biology are well-characterized [19–22]. Table 1 shows the gene names and protein functions for each of the 11 genes of the ϕ X174 genome, as well as the biological processes in which they participate. Almost all molecular functions related to replication, transcription and translation are performed by the host cellular machinery. In environments in which host evolution is minimized, adaptive changes accumulate mainly in the phage genome [23]. Thus, bacteriophage grown under such conditions provide an experimental system in which biological processes involved in adaptation can be studied easily.

Previous evolution experiments with ϕ X174 illustrate many consequences of short-term selection under varied environmental conditions [14,15,17,24–27]. Some of these experiments used a two-stage chemostat to minimize adaptation in the host. A two-stage chemostat grows the host cells in the first stage, and then transfers growing cells at a constant rate into the second stage in which there are phage. The results of these studies have shown consistently that phage adapt to new hosts, new temperatures or the high multiplicity of infection environment of a chemostat

Table 1. Biological processes in which ϕ X174 proteins participate.

| Gene | Protein Function | Biological Process | Obs/Exp † |
|------|------------------------|---|-----------|
| A | Replication Initiation | Replication, DNA packaging | 44/29 |
| A* | Endonuclease | Unknown | 16/19 |
| B | Internal Scaffold | Capsid assembly | 2/7 |
| C | DNA/Protein Binding | DNA packaging | 3/5 |
| D | External Scaffold | Capsid assembly | 2/9 |
| E | Lysis | Host-phage interaction | 3/5 |
| F | Major Capsid | Host recognition, Capsid assembly, Capsid structure | 24/24 |
| G | Major Spike | Host recognition, Capsid assembly, Capsid structure | 1/10 |
| H | Pilot or Minor Spike | Host recognition, Capsid assembly, Capsid structure, DNA transport, Replication | 16/18 |
| J | ssDNA Binding | DNA packaging, Capsid assembly, Capsid structure | 1/2 |
| K | Non-essential | Unknown | 2/3 |

†Percent of sites with observed (Obs) substitutions in each gene compared with the percent expected (Exp) due to the length of the gene. A* is a subset of A. Values do not add to 100% due to overlapping genes.

doi:10.1371/journal.pone.0060401.t001

through the accumulation of many common substitutions and several unique substitutions [28]. The common substitutions reveal much about the biological processes involved in adaptation. For example, during growth on *Salmonella typhimurium*, adaptive changes occur in the major capsid protein, which directly interacts with components of the bacterial cell wall. These adaptations can be reversed when the phage are grown on *E. coli* [25]. There are also adaptive changes that occur in the major capsid protein during growth at high temperature [29]. These results suggest that host recognition and capsid assembly are important biological processes involved in adaptation to these environmental challenges.

For the data presented here, replicate two-stage chemostats for each of two environments, 37° with *E. coli* C as the host and 43.5° with *S. typhimurium* LT2 as the host, were grown for 50 days. These two environments were meant to bestow weak and strong selection, respectively. Over 50 years ago ϕ X174 was domesticated by growing on *E. coli* at 37° [30] and doubles its population size 20 times per hour when grown in batch culture under these conditions [31]. Hence this combination of host and temperature should be benign and selection due to these conditions should be weak. On the other hand, *S. typhimurium* at 43.5° is both a novel host and a non-permissive temperature for the lab-adapted strain of ϕ X174, whose population only doubles nine times per hour under these conditions [31]. Hence selection in this environment is quite strong [24,26]. Additionally, the experiments presented here were continued over a longer time period than most previous experiments, which lasted 10 or 11 days up to 22 days, and minimally covered 3600 phage generations [27,28]. The longer time period was specifically chosen to observe whether selection persisted within the chemostat environment once the initial adaptation to host and temperature was attained. The rate of molecular evolution was compared among the chemostats presented here, and the comparison indicated that, although initially the strong selection environment had a greater evolutionary rate, both environments continued to evolve at approximately the same rate for the full 50 day period [23]. Indeed, one of the chemostats growing on *E. coli* at 37° was continued for another 130 days, and adaptation to this environment persisted over the entire time [27]. The persistence of evolution over time was attributed to competition for host cells in the chemostat

environment, where phage greatly outnumber the available hosts [23].

Here we present the complete sequence data for these four chemostats and compare them to genetic variation found in wild phage. Counts of sequence changes in these chemostats were previously summarized in [23] and [28], and an analysis of the 180 day chemostat was presented in [27]. We compare the results from these long term chemostats to previously described experiments at a level of detail that allows us to provide new insights into the biological processes that may be affected by these adaptive changes and the specific selection pressures, temperature, host or high multiplicity of infection, driving evolution in these experiments.

Materials and Methods

Strains

A lab-adapted strain of ϕ X174 was used as the ancestor for these studies (GenBank Accession: AF176034). The bacteria used for phage propagation were *Escherichia coli* C and the Type I restrictionless (*hsd*), CpX 174^s *Salmonella typhimurium* LT2 strain IJ750 [*xyl-404 met A 2 2 metE551 galE71 9 trpD2 ilv-452 hsdLT6 hsdSA29 hsdSB121 fla-66 rpsLI20 H1-b H2-e nix*]. These are the same phage and bacterial hosts used in previous similar experiments [24,25].

Two-stage chemostat

A chemostat consisting of two 100×15 mm glass test tubes was used to select phage populations as previously described [24] except that the media (LB broth: 10 g NaCl, 10 g Bacto Tryptone, and 5 g yeast extract/liter, 0.1% antifoam B, Sigma A5757) contained 2 mM CaCl₂. Briefly, host cells were maintained at a density suitable for phage growth in the first tube, and media containing cells from this tube was drawn continuously into the second tube, which contained phage. The volume of the phage tube was replaced on the order of 100 times per day, which is equivalent to a growth of 144 phage population doublings per day given the continuous nature of replacement [27]. The chemostat was reassembled every two days and inoculated with 0.2 mL of filtered phage lysate, thus discarding any phage-resistant cells.

Sequencing

Whole genome sequencing was carried out as previously described [27]. In most cases, the sequence was determined from a PCR of the heterogeneous lysate from the chemostat culture (referred to as “population consensus” sequences). Data for *E. coli* C, 37° chemostats came from both a population consensus and zero to four isolates. One of the *E. coli* C, 37° chemostats evolved an interfering particle [23], and sequence changes were determined only from two isolates at each time point while the particle was present. These data are reported simply as changes that were detected at a particular time point, and combining the isolate data this way makes them roughly equivalent to the population-derived consensus sequence. A consensus sequence does not identify changes present at low or moderate frequency in the culture and can exhibit genuinely ambiguous bases where polymorphisms occur at intermediate frequency, but it also provides an expedient means of identifying changes that are common in the population.

Statistical analysis

Substitutions occurring in the four 50-day chemostats were compared to those observed in previous evolution experiments. Chemostat experiments previously described in Wichman and Brown [28] can be divided into types based upon temperature and host (Table 2). These experiments were conducted in chemostats that lasted from 10 to 22 days under four conditions: the host was *S. typhimurium* and, the temperature was 42° to 43.5° (S42); the host was *E. coli*, and the temperature was 37° (E37); the host was *E. coli*, and the temperature was 42° to 43.5° (E42); and the host was *S. typhimurium*, and the temperature was 37° (S37). Hypotheses developed based upon the chemostats in the present study, i.e., the index chemostats grown on *E. coli* at 37° or *S. typhimurium* at 43.5°, were tested using the data from these other chemostats, i.e. the test chemostats, thus providing an independent test of our hypotheses (Table 2). The hypotheses of interest were whether or not a set of substitutions occurred more frequently under certain chemostat conditions. The sets of substitutions were determined by their distribution among the index chemostats: found in both chemostat environments, in only the *S. typhimurium* at 43.5° chemostats or in only the *E. coli* at 37° chemostats. Thus the average number of substitutions from each set that were found in each experiment were compared among chemostat conditions. Since the distributions of the numbers of substitutions were nonnormal, Kruskal-Wallis (KW) tests were used to test each hypothesis and permutation-based P-values were computed [32]. When follow-up comparisons between pairs of conditions were performed, permutation-based Wilcoxon tests were used [32]. To test the hypotheses of 1) proportional distribution of substitutions among genes, or 2) homogeneity of gene substitution proportions between experimental and wild phage, exact chi-squared tests were conducted [32]. Other hypotheses concerned the distribution of individual substitutions across experimental conditions and were tested using Fisher's exact test [32]. For situations in which a difference was found between conditions in the index experiments, we might confirm this difference in the test experiments. However, for situations where there were no differences between conditions in the index experiments, if we fail to detect a difference in the test experiments then we typically will not have enough power to definitively conclude that no difference exists. We can, however, note when the test experiment result is consistent with the index experiment result.

Table 2. Number of chemostat experiments for each environmental condition in the index and test sets and the publications in which sequence information was presented.

| Set | Reference | Environment* | | | |
|-------|-----------------------------|--------------|-----|-----|-----|
| | | S42 | E37 | E42 | S37 |
| Index | This study, [27] | 2 | 2 | 0 | 0 |
| Test | [24] | 4 | 0 | 5 | 0 |
| | [26] | 2 | 0 | 0 | 0 |
| | [17] | 0 | 0 | 1 | 0 |
| | [25] | 3 | 0 | 2 | 0 |
| | GenBank AF299300 – AF299314 | 5 | 5 | 0 | 1 |
| | Unpublished data | 1 | 3 | 0 | 2 |

*S = *S. typhimurium*, E = *E. coli*, 42°/42°, 37 = 37°.

doi:10.1371/journal.pone.0060401.t002

Results and Discussion

Four experimental lineages were started from independent isolates of the same ancestral strain of ϕ X174 and grown in two-stage chemostats for 50 days under constant conditions: two lineages were grown on *Escherichia coli* C at 37° (designated E37a and E37b), and two on *Salmonella typhimurium* LT2 at 43.5° (S43a and S43b). The populations of these phage lineages progressively accumulated changes throughout their 50-day history, assayed every 10 days. Frequencies of some substitutions that arose in these chemostats were previously summarized in [23] and [28]. E37b developed an interfering particle as reported in [23]. E37a was run for a total of 180 days, and some of the sequence data for E37a were presented previously [27]. We first provide a summary of the genetic changes that arose in each chemostat, and then we discuss the biological processes that might have been affected by these changes.

The substitutions that arose to detectable levels in the four 50-day chemostats are presented in Figure 1. (For the reader's convenience, data reported in [27] are also presented in columns marked by a superscript a.) Missense substitutions are indicated by red, silent substitutions by blue and intergenic substitutions by green. To highlight parallel changes that occurred in this study, the order of presentation is based upon number of chemostats in which the substitution arose, then on the earliest sample in which it was detected, and then by position in the genome based upon the sequence. Thus, rows 1–9 are substitutions that arose in both environments, rows 10–17 arose in both S43 chemostats, rows 18–24 arose in both E37 chemostats, and rows 25–97 arose only in a single chemostat in these experiments. For easy comparison to other similar experiments, the far right columns indicate the number of times that a substitution has been detected in a particular chemostat environment in other similar experiments. These are the test set in Table 2. Below the labels designating environment are the number of other experiments that were conducted under the listed conditions. So, for instance, S42 indicates chemostats at 42° or higher with *S. typhimurium* as host, and the number of other experiments with these conditions was 15.

In the 50-day chemostats, the substitutions that were found in both environments were in genes involved in genome replication and packaging – the replication initiation protein A (4), the pilot protein H (2), the DNA maturation protein C (1) – and in two intergenic regions (Figure 1, rows 1–9). Most of these substitutions have been seen in other experiments including ones that arose

Figure 1. Substitutions detected in four chemostats over 50 days. S43a and S43b were grown on *S. typhimurium* as host at 43.5°; E37a and E37b were grown on *E. coli* at 37°. Order by row is based upon the number of chemostats in which the substitution arose, then on the earliest 10 day sample in which it was detected and then by position in the genome. The column labeled '180' indicates the presence of a substitution in E37a after 180 days of evolution [27]. The last four columns indicate the number of other chemostat experiments ([17,24–26], unpublished data) in which the substitution arose at each environmental condition (*S* = *S. typhimurium*, *E* = *E. coli*, 42 = 42°, 37 = 37°). Red indicates missense substitutions, blue silent substitutions and green substitutions in intergenic regions. ^aPreviously presented in [27]. doi:10.1371/journal.pone.0060401.g001

under a different environmental condition (*E. coli* host at >42°). The wide distribution of these substitutions under different environmental conditions led us to test the hypothesis that the average number of these substitutions per experiment are equal among environments. Using only the data from previous experiments (test set) and excluding the information from the 50-day chemostats, we failed to reject this hypothesis (KW = 2.55, df = 3, P = .51). As mentioned above, we do not have enough power to conclude that no difference exists, but these results are consistent in direction with the index experiment results. The similar distribution of these substitutions across multiple experimental conditions in both the index set and the test set suggests that a feature that is common to all environments, such as the high multiplicity of infection that exists within the chemostats [23,27], was the selective pressure to which these populations were responding.

The substitutions that arose in both S43 50-day chemostats are in the major capsid protein F, the replication initiation protein A, the pilot protein H, and the intergenic region between *F* and *G* (Figure 1, rows 10–17). All of these substitutions have been seen in other experiments, but predominantly in *S. typhimurium* at 43°. The test of equality of the average number of these eight substitutions among the four environments in the test experiments was rejected (KW = 9.75, df = 3, P = .014), supporting the alternative hypothesis that these substitutions are specifically adaptive in the S43 environment. We infer, as have others, that these substitutions are responding to either the novel host, the high temperature or both [17,29].

Substitutions that were detected in both E37 50-day chemostats were also in the F and A proteins, and three substitutions were silent in A or H (Figure 1, rows 18–24). Only one of these, T5365C, has been seen in more than one other experiment. Using the test set, the test for homogeneity of proportion across the four conditions for this single site is not significant (Fisher's exact $\chi^2 = 4.5$, df = 3, p = 0.20), a result which is NOT consistent with the index experiments. The three parallel silent substitutions are intriguing. Although silent sites are generally considered to evolve neutrally, there are several hypotheses for why this may not necessarily be true. These sites may be important in the folding of the ssDNA genome within the capsid or during ejection. The relationship between silent substitutions and codon bias in gene A of ϕ X174 has been discussed previously [27]. We present a further possibility below. Each of these hypotheses require further investigation.

A large fraction of substitutions arose in only a single 50-day chemostat (Figure 1, rows 25–97). These unique substitutions include all of those that arose in genes *D* or *E*, *J*, *G*, and *K*, as well as almost half (20 of 43) of the substitutions that arose in gene *A*. Of the 73 substitutions that were seen in only one chemostat in this study, 35 were found in previous studies including eight that arose in E37a after day 50. Note, however, that the experiments presented here are the first that were run for more than 22 days, and the vast majority of earlier experiments were run for only 10 or 11 days. Almost all of the truly unique substitutions were detected after day 10, so it is possible that the lack of parallel

evolution at some of these sites is an artifact of the shorter length of these earlier experiments.

Substitutions in sequences involved in transcription. There are three promoters and four transcription terminators in ϕ X174. In the 50-day chemostats, substitutions arose in the –35 sigma factor binding site of one of the promoters: G319T, which was seen in S43b and E37a (Fig. 1, row 7), and A323G, which was seen in S43a (Fig. 1, row 26). These substitutions are functionally equivalent in that they down-regulate transcription from this promoter and are adaptive at high temperature [31]. These substitutions also affect the amino acid sequence of protein C (see below), however, adaptive substitutions at adjacent silent sites have similar effects on transcription and a slightly smaller effect on fitness.

There are two adjacent substitutions immediately downstream of the first transcription termination signal, A986G, which arose in S43a (Fig. 1, row 29), and C987T, which arose in the other three 50-day chemostats (Fig. 1, row 5). Predictions of the strength of this rho-independent transcription terminator are unaffected by these substitutions, however, the C987T substitution increases the number of Ts in the T-tail, which affects readthrough of the RNA polymerase to the downstream genes [33,34]. This site is also an extra-origin cleavage site for the endolytic activity of A/A* and both of these substitutions alter conserved nucleotides in this site [33,35]. Determining whether these intergenic substitutions are functionally equivalent and whether they affect mRNA or genome stability awaits further experimentation.

Substitutions in genes involved in phage genome replication and packaging. The proteins involved in genome replication and packaging are A, H and C. The replication initiation protein A accrues the most substitutions in these experiments. Fully 44% of all substitutions in the 50-day chemostats were in the A gene, and two thirds of these were in the N-terminal third of the protein (Fig. 2A, Table 1). The distribution of substitutions among genes is significantly different than expected given their lengths in nucleotides (Fisher's exact $\chi^2 = 21.6$, df = 7, P = 0.005). The greatest contribution to this χ^2 value is from the excess number of substitutions in gene A ($\chi^2 = 7.9$, df = 1, P < 0.005).

Gene *A* contains the coding sequences for proteins A, A* and B and the N-terminus of protein K, as well as the origin of replication (*ori*). Protein A* shares all but the first 172 residues with protein A and also shares many of the functions of A except the ability to initiate rolling circle replication and package viral DNA into the capsid [36]. The N-terminal third of A interacts with the host helicase, which unwinds the DNA during viral strand synthesis [37,38], and probably with host single-stranded DNA binding protein [35]. The replication initiation protein A acts *in cis* to initiate rolling circle replication, which is accomplished by concomitant transcription, translation and binding to the dsDNA from which transcription was initiated [39,40]. Because protein A only aids in the replication of its own DNA, directional selection may act on this protein to increase replication without danger of being hijacked by coinfecting phage, and this may account for the large number of substitutions in A relative to the other genes.

Three substitutions in protein A have been found in 12 (G4122A), 14 (A4168G) and 9 (C5185T) experiments (Fig. 1, rows

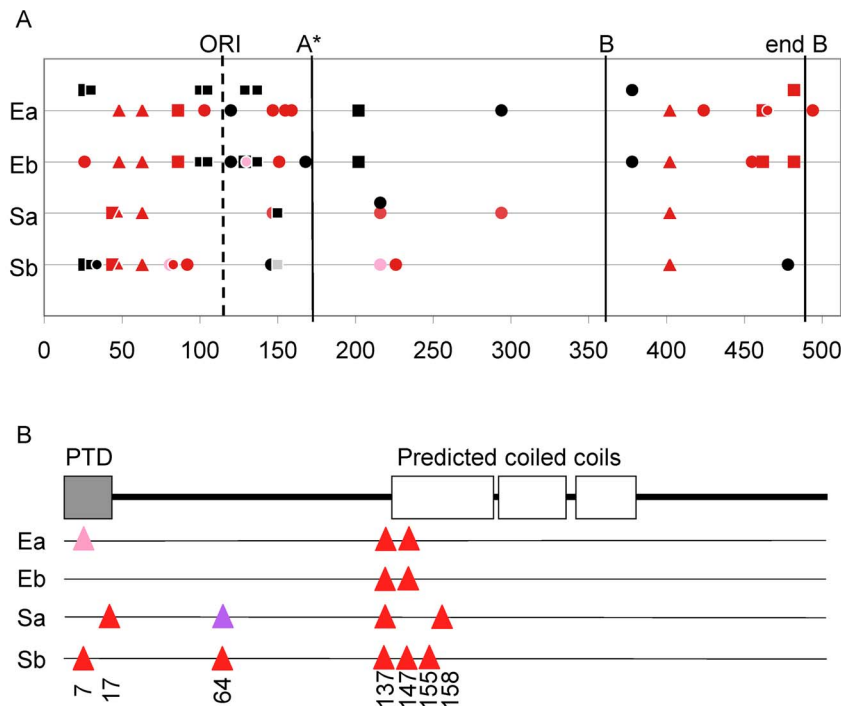


Figure 2. Distribution of substitutions in proteins A and H are not uniform. A) Substitutions in the 512 codons of gene A. The dashed vertical line indicates the origin of replication, the other vertical lines indicate the starts of A* and B, and the end of B (residues: A = 1–512; A* = 173–512; B = 364–484). Red indicates missense substitutions and black indicates silent substitutions in genes A or A*, not B. Fainter colors indicate the substitution was not detected at day 50. Triangles mark substitutions found in all four chemostats, squares mark substitutions found in more than one chemostat and circles are unique to each chemostat. Sa = S43a, Sb = S43b, Ea = E37a, Eb = E37b. The symbols above the line for Ea indicate parallel substitutions that arose in Ea by 180 days [27]. The black circle above the Sa line indicates a silent substitution that affected the same amino acid as the red circle on the line. B) Substitutions in protein H. Graphic of predicted domains from [64]. Red triangles indicate missense substitutions, pink indicates the substitution was not detected at day 50, and purple indicates a second missense substitution at that residue. The vertical numbers indicate position in the protein sequence. PTD = predicted transmembrane domain. Not to scale. doi:10.1371/journal.pone.0060401.g002

2–4). These substitutions are also found in or, in the case of 5185T, are fixed in wild ϕ X-like phage [28]. The proportion of each of these substitutions is not significantly different among the various environments (S42, E37, E42, S37) when the test set is used (Fisher's exact test, P-values: 0.4–0.95), a result which is consistent in direction with the index experiments. One constant among these environments is their high MOI, a condition that may well occur in the waste water treatment plants from which the wild phage were collected. Thus, competition with conspecifics may select for adaptation in the process of genome replication. The elucidation of the exact mechanism by which these adaptive substitutions affect replication awaits further study.

An advantage of the nucleotide substitutions in gene A, both silent and missense, may take the form of improved DNA binding affinity for the helicase or faster DNA unwinding, such that new viral strands are produced more quickly [41]. The accumulation of substitutions around the *ori* in E37b may allow faster DNA unwinding (Fig. 2A). Of the 12 substitutions in this region, both missense and silent, seven were G to T transversions and two were C to T transitions. Overall this led to a decrease in the GC content around the *ori* and especially downstream. A similar clustering of substitutions in this region occurred in the E37a chemostat after day 50 [27]. In this case, the predominance of silent substitutions led to a pronounced increase in GC content upstream of the *ori*. Of the 17 sites upstream of the *ori* that underwent substitutions, five were G or C and 12 were A or T in the ancestral phage. By day 180, 13 of these sites were G or C and only four were A or T. This was primarily due to transitions from A or T to G or C.

Both of these results are consistent with the mechanism of rolling circle replication and suggests a selective advantage for the silent as well as missense substitutions in gene A. A reduction in GC content in the direction in which the helicase unwinds the dsDNA would potentially reduce the energy cost and increase the speed of unwinding. On the other hand, the DNA replication machinery requires a double-stranded template from which to start replication. Increasing the GC content upstream of the *ori* may strengthen the double-stranded nature of the template DNA. That these increases in silent substitutions occur very late suggests that they are mutations of small benefit with relatively weak selective advantage and can only accrue when the population size is quite large and when they are not competing with mutations of larger effect. This could explain the dearth of such mutations in S43a (Figure 2A), where the titer was significantly lower than the other chemostats by day 50 [23].

Protein H is a multi-functional protein that has been associated with many biological processes, including ejection and replication (Table 1). Protein H, also called the pilot protein, brings the viral strand into the cell (ejection) and transports it to sites for DNA replication of ssDNA to dsDNA [42–44]. It has been proposed that the N-terminal transmembrane helix mediates the transition across the bacterial cell wall [45]. Thus, substitutions in this region, especially C2980T (Fig. 1, row 55), which has been found in seven experiments, may affect ejection. The number of sites for second-strand synthesis are limited [46], and other substitutions in protein H may be due to competition for the host replication machinery. The locations of multiple and parallel substitutions in

the 35 amino acids from 136 to 170 suggest that this region, immediately adjacent to and within the first coiled coil domain, is important in the chemostat environment. Over multiple evolution experiments there have been 31 substitutions at sites including five sites that have undergone parallel substitutions in multiple environments [28]; and unpublished data. These environments include three bacterial host species (*E. coli*, *S. typhimurium*, *Shigella sonnei*) and are evenly divided between chemostats run at either 37°C or >42°C. Removing the index set (17 sites), the test for homogeneity of proportions across environments is not significant (Fisher's exact $\chi^2 = 6.4$, $df = 3$, $P = 0.1$), a result which is consistent in direction with the index experiments.

Two non-coding regions appear to be important for replication from ssDNA to dsDNA. One is in the intergenic region between genes H and A. This sequence may interact with a component of the host replication machinery that is present in limiting amounts, possibly the sites for second-strand synthesis mentioned above [47,48]. Competition for this host factor may drive selection in this intergenic region. There are 4 substitutions in this region, one that was only present at detectable levels on day 40 (Fig. 1, line 66), and one that occurred in several experiments under a number of environmental conditions (Fig. 1, line 8). This number of substitutions, let alone parallel substitutions, is much greater than would be expected by chance given the short length (64 nt) of this region. However, these two substitutions are also in the predicted transcription termination signal for H. Determining whether they affect replication or transcription awaits further experimentation. The other non-coding region that is important for ssDNA to dsDNA replication is the intergenic region between F and G (F inter). One substitution that arose in this region in the S43 chemostats and in a third experiment (T2305G; Fig. 1, row 15) is within the primosome binding site [49], and substitutions in this site may improve its interaction with the *S. typhimurium* primosome.

Protein C is involved in packaging ssDNA into the capsid (Table 1). This protein must bind to the dsDNA prior to binding by host replication factors in order to start packaging ssDNA into the procapsid. As well as down-regulating transcription from the D promoter, substitutions at G319T and A323G alter adjacent amino acids in protein C, V63F and D64G, respectively [50]. These two sites have arisen in 11 experiments, but never in the same phage. In addition, during host switching experiments in which a number of substitutions in the capsid proteins reverted to the original amino acid, V63F also reverted, suggesting that it may be interacting with either the capsid or the host helicase. The number of times that these substitutions arose in chemostats that had *E. coli* C as the host (4), however, suggests that it is the capsid with which they are interacting.

These substitutions also affect transcription, as mentioned above, and they may also be affecting genome replication. The transcription and replication machinery move in the same direction around the phage genome. Studies in *Bacillus subtilis* show that these two machineries may interfere with each other [51]. Down regulating expression from the D promoter may decrease the amount of transcript for the capsid proteins, however, the B promoter still produces transcripts that include these proteins. It is an intriguing possibility that the loss in transcript due to down regulating the D promoter results in a sufficiently faster replication rate that increases fitness of phage carrying these regulatory substitutions.

Substitutions in genes involved in host recognition. Another process in which protein H is involved is binding to the cell receptor [42,43,52]. The substitutions at amino acid 64 that have only been found in S43-like chemostats

may well be involved in binding to the cell receptor, which differs in structure between *E. coli* and *S. typhimurium* [53].

Protein F, the major capsid protein, is involved in recognizing and attaching to the host lipopolysaccharide. The substitution in F that persisted throughout the duration of the two E37 chemostats, A1301G, alters the amino acid sequence at a site that is adjacent to two sites that are known to be important for host recognition [25,54]. The three substitutions in F that arose in both S43 chemostats (Fig. 1, rows 10–12) have arisen in multiple experiments and have been attributed to adaptations to either novel host (C2085T), high temperature (C1727T) or both (A1613T) [17,29]. Thus, some of the substitutions in F can be attributed to its role in host recognition and attachment and some to capsid assembly. Interestingly, two of these three substitutions in F eventually arose in E37a after this chemostat was continued for another 130 days (Fig. 1, rows 11,12). It seems likely that these substitutions have a smaller but positive selective advantage in that environment as well.

Substitutions in genes involved in capsid assembly. The major capsid protein F interacts with proteins B, D, G, J and H during capsid assembly and is part of the capsid [19,55]. The role of adaptive substitutions in protein F has been discussed extensively and recently reviewed in [28], therefore we will not discuss them here. It is interesting to note, however, that there were very few substitutions in genes *B*, *D*, *G* and *J*. The possibility that substitutions in F are interacting with substitutions in H has not yet been explored.

Distribution of substitutions compared to wild phage. The large number of substitutions in gene *A* could indicate simply that *A* is hypermutable, or the fitness effect of changes in these genes is nearly neutral. Similarly, the dearth of substitutions in genes *D* or *E*, *J*, *G*, and *K* may be because they are highly conserved genes. If these explanations hold then the amount of variation in natural populations should reflect the variation seen in the chemostats. We compared our results to 16 natural microvirid phage isolates in the ϕ X174 clade [56]. In these 16 isolates, there are substitutions at 706 sites relative to the wild phage consensus sequence; 144 (20%) are missense, 545 (77%) are silent and 18 (3%) are intergenic. Substitutions shared among wild phages are assumed to be identical by descent, although this is a simplifying assumption. The distribution of substitutions among genes is significantly different between the 50-day chemostats and the wild phages ($\chi^2 = 35.3$, $df = 6$, $P < 0.001$).

Figure 3 shows the frequency of sites with missense and silent substitutions for each gene in the wild phages and in each environment. *A** is not included as it is an in-frame variant of *A*. For seven of the remaining 10 genes, the wild phages have the typical pattern of substitutions for genes undergoing purifying selection; there are far fewer missense substitutions (20%) than would be expected under a neutral model of evolution, which for this genome is about 77% of sites. The remaining three genes overlap with the coding sequence of other genes, *B* with *A/A**, *E* with *D*, and *K* with *A/A** and *C*. In the wild phage, about 75% of the substitutions in *B*, *E* and *K* are missense substitutions, and these are generally silent in their overlapping genes. The apparent neutral evolution of these three genes has been noted previously [56]. Thus, the frequency of substitutions in the chemostat experiments does not reflect the frequency of substitutions in natural populations of phages, and we can reject the notion that the patterns seen in the chemostat experiments reflect the degree of conservation of the individual genes.

Although the overall distribution of missense and silent substitutions suggests that purifying selection is the predominant

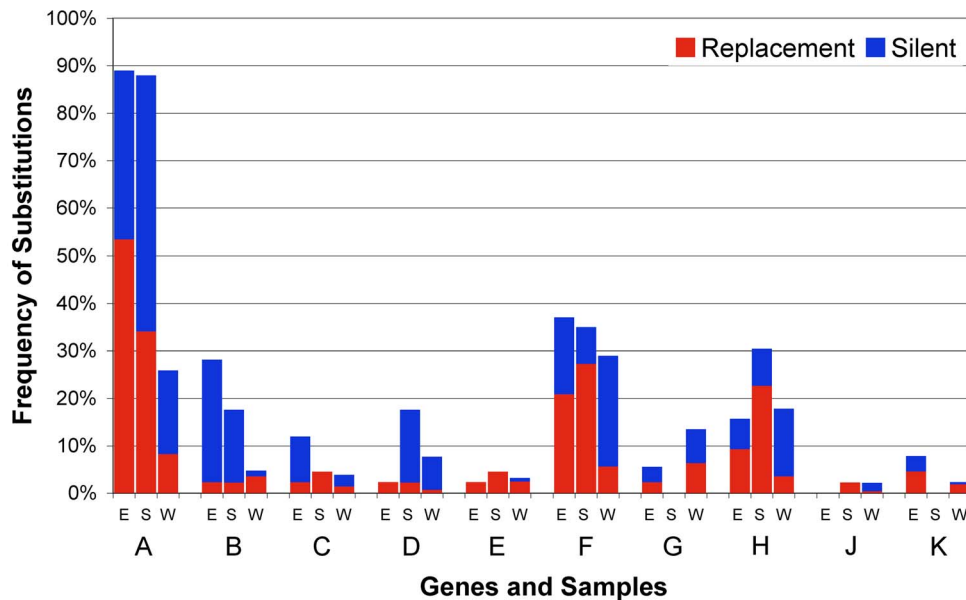


Figure 3. Distribution of substitutions is different between chemostats and wild phage. Percent of all missense (red) and silent (blue) substitutions for each environment (E = E37, S = S43) and among wild phage (W) [56] by gene (A–K). doi:10.1371/journal.pone.0060401.g003

force in natural populations of phage, there are indications of similar adaptive mechanisms in natural and experimental phage populations [28]. Twenty seven of the substitutions in the 50-day chemostats are also found in the wild phage populations, and this is twice what would be expected by chance (13). In four cases, the adaptive substitution is a reversion to the monomorphic wild phage nucleotide: C5185T and A3340G occurred in all four chemostats; G384A, which is a silent substitution in gene C, arose and persisted in only a single chemostat; C1727T arose in the S43 chemostats and eventually in E37a. This last substitution is a well known response to growth at high temperatures that produces a leucine to phenylalanine substitution in the capsid protein F [29]. Its presence in wild populations of phage suggests that they experience high temperatures in nature. If C5185T and A3340G are indeed arising in response to high MOI, then such conditions may also exist in nature.

At other sites, the wild phage are polymorphic at the same sites that mutated in the 50 day chemostats. At three sites (T4628C, C1460G, C2191T), the ancestral sequence is the same as the less common sequence in the wild phage, and adaptive substitutions are convergent to the consensus sequence for the wild phage. At 18 sites (G4256A, T4289C, C4391T, A1809G, G1304C, A4168G, C2085T, T4058C, C39T, G4122A, C4110T, A986G, T4082C, G4418A, C2980T, C1720T, T2014C, G2275A), the ancestral sequence is the same as the wild phage consensus, and the substitution is to a nucleotide that is found in at least one wild phage. Two sites (G4367T, C2711T) had two low frequency substitutions in the wild phage: one is the same as the experimental substitution and one is different. Four sites (G575C, T3122G, A1308T, G4238T), are polymorphic in the wild phage, but did not have the same substitution as the experimental phage. The remaining 66 sites are monomorphic in the wild phage. Thus about a third of the adaptive substitutions that were seen in these experiments are also found in natural populations of phage, suggesting that evolutionary pressures found in these experiments are also found in nature [28].

Conclusions

This study provides valuable insights into the biological processes that are affected when ϕ X174 adapts to a chemostat environment. Although the primary force leading to adaptation in the S43 chemostats appears to be related to host recognition or temperature, the secondary force in these chemostats and the primary force in the E37 chemostats appears to be competition amongst the phage. Two steps are obvious from these and other experiments [26,27]. In chemostats with novel hosts or temperatures, substitutions arise first in genes involved in host recognition and capsid assembly. Thus the greatest initial increase in fitness comes from these processes, leading to increases in population numbers within the chemostat [23]. The second evolutionary step, therefore, is adapting to high viral density where there is intense within-host competition. One way to outcompete coinfecting conspecifics is to have a higher growth rate. Consistent with this, the genes that appear to have accumulated substitutions in both environments are involved in genome replication and packaging.

Genome replication is an obvious target of selection for higher growth rate. Indeed, drugs that target the RNA polymerases of HIV and other eukaryotic viruses select for escape mutations that increase viral replication rates [57,58]. Growing arboviruses only in an arthropod cell line, rather than switching between their arthropod and vertebrate hosts, selects for substitutions that increase the production of the reverse transcriptase needed for genome replication [59]. Multiple studies on eukaryotic viruses indicate that growth in novel hosts leads to evolution in the viral-encoded polymerases to improve their interactions with host proteins such as RNA helicase and importin α [60,61]. Another example of evolution of a foreign replication protein that interacts with a host helicase is found in bacterial host switching experiments with plasmids [62,63]. These experiments lead to evolution of the plasmid replication initiation protein, especially in the region that interacts with the host helicase, leading to increased plasmid copy number, which is analogous to evolution in the replication initiation protein of ϕ X174. Although molecular mechanisms by which evolution increases replication rates may

be very different, especially in eukaryotic viruses, the process of genome replication is a major target for adaptation.

Acknowledgments

We thank Dr. James Bull and the University of Idaho phage group for helpful discussions.

References

- Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat Genet* 42: 260–263.
- Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, et al. (2008) Patterns of positive selection in six mammalian genomes. *PLoS Genet* 4: e1000144.
- Montoya-Burgos JI (2011) Patterns of positive selection and neutral evolution in the protein-coding genes of Tetraodon and Takifugu. *PLoS One* 6: e24800.
- Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* 4: 457–469.
- Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, et al. (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461: 1243–1247.
- Stoebel DM, Dean AM, Dykhuizen DE (2008) The cost of expression of *Escherichia coli lac* operon proteins is in the process, not in the products. *Genetics* 178: 1653–1660.
- Zhong S, Khodursky A, Dykhuizen DE, Dean AM (2004) Evolutionary genomics of ecological specialization. *Proc Natl Acad Sci U S A* 101: 11719–11724.
- Lee DH, Palsson BO (2010) Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a Nonnative carbon source, L-1,2-propanediol. *Appl Environ Microbiol* 76: 4158–4168.
- Fong SS, Joyce AR, Palsson BO (2005) Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res* 15: 1365–1372.
- Betancourt AJ (2009) Genomewide patterns of substitution in adaptively evolving populations of the RNA bacteriophage MS2. *Genetics* 181: 1535–1544.
- Bollback JP, Huelsenbeck JP (2009) Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. *Genetics* 181: 225–234.
- Bull JJ, Molineux IJ (2008) Predicting evolution from genomics: experimental evolution of bacteriophage T7. *Heredity* 100: 453–463.
- Denchey JJ, Abedon ST, Turner PE (2007) Host density impacts relative fitness of bacteriophage $\phi 6$ genotypes in structured habitats. *Evolution* 61: 2516–2527.
- Dickins B, Nekrutenko A (2009) High-resolution mapping of evolutionary trajectories in a phage. *Genome Biology and Evolution* 2009: 294–307.
- Rokyta DR, Abdo Z, Wichman HA (2009) The genetics of adaptation for eight microvirid bacteriophages. *J Mol Evol* 69: 229–239.
- Sachs JL, Bull JJ (2005) Experimental evolution of conflict mediation between genomes. *Proc Natl Acad Sci U S A* 102: 390–395.
- Wichman HA, Scott LA, Yarber CD, Bull JJ (2000) Experimental evolution recapitulates natural evolution. *Philos Trans R Soc Lond B Biol Sci* 355: 1677–1684.
- Knies JL, Izem R, Supler KL, Kingsolver JG, Burch CL (2006) The genetic basis of thermal reaction norm evolution in lab and natural phage populations. *PLoS Biol* 4: e201.
- Fane BA, Brentlinger KL, Burch AD, Chen M, Hafenstein S, et al. (2006) $\phi X174$, et al. *The Microviridae*. In: Calendar R, editor. *The Bacteriophages*. 2nd ed: Oxford Press. 129–145.
- Sanger F, Coulson AR, Friedmann T, Air GM, Barrell BG, et al. (1978) The nucleotide sequence of bacteriophage $\phi X174$. *J Mol Biol* 125: 225–246.
- McKenna R, Xia D, Willingmann P, Ilag LL, Rossmann MG (1992) Structure determination of the bacteriophage $\phi X174$. *Acta Crystallogr B* 48 (Pt 4): 499–511.
- Hayashi M, Aoyama A, Richardson DL, Hayashi MN (1988) *Biology of the Bacteriophage $\phi X174$* . In: Calendar R, editor. *The Bacteriophages*. New York: Plenum Press. 1–71.
- Bull J, Millstein J, Orcut J, Wichman H (2006) Evolutionary feedback mediated through population density, illustrated with viruses in chemostats. *American Naturalist* 167: E39–E51.
- Bull JJ, Badgett MR, Wichman HA, Huelsenbeck JP, Hillis DM, et al. (1997) Exceptional convergent evolution in a virus. *Genetics* 147: 1497–1507.
- Crill WD, Wichman HA, Bull JJ (2000) Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* 154: 27–37.
- Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ (1999) Different trajectories of parallel evolution during viral adaptation. *Science* 285: 422–424.
- Wichman HA, Millstein J, Bull JJ (2005) Adaptive molecular evolution for 13,000 phage generations: a possible arms race. *Genetics* 170: 19–31.
- Wichman HA, Brown CJ (2010) Experimental evolution of viruses: Microviridae as a model system. *Philos Trans R Soc Lond B Biol Sci* 365: 2495–2501.
- Bull JJ, Badgett MR, Wichman HA (2000) Big-benefit mutations in a bacteriophage inhibited with heat. *Mol Biol Evol* 17: 942–950.
- Sinsheimer RL (1991) The discovery of a single-stranded, circular DNA genome. *Bioessays* 13: 89–91.
- Brown CJ, Zhao L, Evans KJ, Ally D, Stancik AD (2010) Positive selection at high temperature reduces gene transcription in the bacteriophage $\phi X174$. *BMC Evol Biol* 10: 378.
- Higgins JJ (2004) *An Introduction to Modern Nonparametric Statistics*. Belmont, CA: Brooks/Cole. 366 p.
- Hayashi MN, Hayashi M, Muller UR (1983) Role for the J-F intercistronic region of bacteriophages $\phi X174$ and G4 in stability of mRNA. *J Virol* 48: 186–196.
- Kingsford CL, Ayanbule K, Salzberg SL (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* 8: R22.
- van Mansfeld AD, van Teeffelen HA, Fluit AC, Baas PD, Jansz HS (1986) Effect of SSB protein on cleavage of single-stranded DNA by ϕX gene A protein and A* protein. *Nucleic Acids Res* 14: 1845–1861.
- van der Ende A, Langeveld SA, Van Arkel GA, Weisbeck PJ (1982) The interaction of the A and A* proteins of bacteriophage $\phi X174$ with single-stranded and double-stranded phi X DNA in vitro. *Eur J Biochem* 124: 245–252.
- Tessman ES, Peterson PK (1976) Bacterial rep- mutations that block development of small DNA bacteriophages late in infection. *J Virol* 20: 400–412.
- Ekechukwu MC, Oberste DJ, Fane BA (1995) Host and $\phi X 174$ mutations affecting the morphogenesis or stabilization of the 50S complex, a single-stranded DNA synthesizing intermediate. *Genetics* 140: 1167–1174.
- Francke B, Ray DS (1972) Cis-limited action of the gene-A product of bacteriophage $\phi X174$ and the essential bacterial site. *Proc Natl Acad Sci U S A* 69: 475–479.
- Henry TJ, Knippers R (1974) Isolation and function of the gene A initiator of bacteriophage phi-chi 174, a highly specific DNA endonuclease. *Proc Natl Acad Sci U S A* 71: 1549–1553.
- Eisenberg S (1980) The role of gene A protein and E. coli rep protein in the replication of $\phi X174$ replicative form DNA. *Proc R Soc Lond B Biol Sci* 210: 337–349.
- Jazwinski SM, Lindberg AA, Kornberg A (1975) The gene H spike protein of bacteriophages $\phi X174$ and S13. I. Functions in phage-receptor recognition and in transfection. *Virology* 66: 283–293.
- Jazwinski SM, Marco R, Kornberg A (1975) The gene H spike protein of bacteriophages $\phi X174$ and S13. II. Relation to synthesis of the parenteral replicative form. *Virology* 66: 294–305.
- Azuma J, Morita J, Komano T (1980) Process of attachment of $\phi X174$ parental DNA to the host cell membrane. *J Biochem* 88: 525–532.
- Cherwa JE Jr., Young LN, Fane BA (2011) Uncoupling the functions of a multifunctional protein: the isolation of a DNA pilot protein mutant that affects particle morphogenesis. *Virology* 411: 9–14.
- Bravo A, Serrano-Heras G, Salas M (2005) Compartmentalization of prokaryotic DNA replication. *FEMS Microbiol Rev* 29: 25–47.
- van der Avoort HG, van Arkel GA, Weisbeck PJ (1982) Cloned bacteriophage $\phi X174$ DNA sequence interferes with synthesis of the complementary strand of infecting bacteriophage $\phi X174$. *J Virol* 42: 1–11.
- van der Avoort HG, van der Ende A, van Arkel GA, Weisbeck PJ (1984) Regions of incompatibility in single-stranded DNA bacteriophages $\phi X174$ and G4. *J Virol* 50: 533–540.
- Arai K, Kornberg A (1981) Unique primed start of phage $\phi X174$ DNA replication and mobility of the primosome in a direction opposite chain synthesis. *Proc Natl Acad Sci U S A* 78: 69–73.
- Fujisawa H, Hayashi M (1977) Functions of gene C and gene D products of bacteriophage $\phi X174$. *J Virol* 21: 506–515.
- Merrikh H, Machon C, Grainger WH, Grossman AD, Soutanas P (2011) Co-directional replication-transcription conflicts lead to replication restart. *Nature* 470: 554–557.
- Jazwinski SM, Lindberg AA, Kornberg A (1975) The lipopolysaccharide receptor for bacteriophage $\phi X174$ and S13. *Virology* 66: 268–282.
- Inagaki M, Kawaura T, Wakashima H, Kato M, Nishikawa S, et al. (2003) Different contributions of the outer and inner R-core residues of lipopolysaccharide to the recognition by spike H and G proteins of bacteriophage $\phi X174$. *FEMS Microbiol Lett* 226: 221–227.
- Pepin KM, Wichman HA (2007) Variable epistatic effects between mutations at host recognition sites in $\phi X174$ bacteriophage. *Evolution* 61: 1710–1724.
- Cherwa JE Jr., Organtini IJ, Ashley RE, Hafenstein SL, Fane BA (2011) In vitro assembly of the $\phi X174$ procapsid from external scaffolding protein oligomers and early pentameric assembly intermediates. *J Mol Biol* 412: 387–396.

Author Contributions

Conceived and designed the experiments: HAW. Performed the experiments: JM. Analyzed the data: CJB CJW HAW. Wrote the paper: CJB CJW HAW.

56. Rokyta DR, Burch CL, Caudle SB, Wichman HA (2006) Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J Bacteriol* 188: 1134–1142.
57. Ordonez P, Hamasaki T, Isono Y, Sakakibara N, Ikejiri M, et al. (2012) Anti-human immunodeficiency virus type 1 activity of novel 6-substituted 1-benzyl-3-(3,5-dimethylbenzyl)uracil derivatives. *Antimicrob Agents Chemother* 56: 2581–2589.
58. Forbes NE, Ping J, Dankar SK, Jia JJ, Selman M, et al. (2012) Multifunctional adaptive NS1 mutations are selected upon human influenza virus evolution in the mouse. *PLoS One* 7: e31839.
59. Weaver SC, Brault AC, Kang W, Holland JJ (1999) Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. *J Virol* 73: 4316–4326.
60. Gabriel G, Herwig A, Klenk HD (2008) Interaction of polymerase subunit PB2 and NP with importin $\alpha 1$ is a determinant of host range of influenza A virus. *PLoS Pathog* 4: e11.
61. Bortz E, Westera L, Maamary J, Steel J, Albrecht RA, et al. (2011) Host- and strain-specific regulation of influenza virus polymerase activity by interacting cellular proteins. *MBio* 2.
62. Sota M, Yano H, J MH, Daughdrill GW, Abdo Z, et al. (2010) Shifts in the host range of a promiscuous plasmid through parallel evolution of its replication initiation protein. *ISME J* 4: 1568–1580.
63. Yano H, Deckert GE, Rogers LM, Top EM (2012) Roles of long and short replication initiation proteins in the fate of IncP-1 plasmids. *J Bacteriol* 194: 1533–1543.
64. Ruboyanes MV, Chen M, Dubrava MS, Cherwa JE Jr., Fane BA (2009) The expression of N-terminal deletion DNA pilot proteins inhibits the early stages of $\phi X174$ replication. *J Virol* 83: 9952–9956.